NEURAL NETWORK BASED METHODS IN THE EXTRACTION OF KNOWLEDGE FROM ACCOUNTING AND FINANCIAL DATA

By
Duarte Trigueiros
September the 23^{th}, 1991

University of East Anglia
School of Information Systems

Preface

This study is about knowledge acquisition in Accountancy and Finance. It develops new concepts and techniques for the extraction of meaningful information from data like returns on assets and accounting reports.

Traditional areas of concern such as the homogeneity of industrial groups and the relation between market perception of risk and accounting numbers are explored in the light of such concepts. The result is a unified view and methodology, leading to an easier formulation and modelling of problems.

The implementation of this formulation is carried out using flexible tools, known as Neural Networks. The aim is to show how these tools are able to apportion qualitative, very significant improvements in areas where there is already a well developed set of results.

We organized our research in two main bodies. The first one comprises five chapters and its goal is the building of a framework allowing for the statistical modelling of accounting information. The second body includes an introduction to Neural Networks and three applications.

Both parts contain a detailed introduction. We defer the usual comments on subjects to such introductions. Here, we would like to highlight just a few achievements of this study along with suggestions for readers.

Describing the cross-sectional characteristics of accounting data: Empirical evidence gathered in chapter 1 and 2 make it possible to develop appropriate models for the description of the cross-sectional behaviour of accounting information. Such models are extensions of financial ratios contemplating deviations from linearity and proportionality (chapter 3).

The most intriguing aspects of the cross-sectional behaviour of accounting data are explained in chapters 2 and 4. We focus on the existence of outliers, the heteroscedasticity found in models, the distribution of ratios and the nature of their information content.

Extracting appropriate ratios from examples: We show that Neural Network-like algorithms are capable of implementing the developed models. Using them, they learn an accounting relation from sets of examples. As a result, this technique will build optimal structures interpretable in terms of financial ratios (chapter 7).

This approach effectively avoids the search of appropriate ratios by the analyst and some other major drawbacks of the multivariate statistical modelling techniques used in accountancy. It is also self-explanatory, yielding not just a model but also an interpretable set of ratio-like structures.

Improving diagnosis specificity with Neural Networks: Neural Networks are also used to automate financial diagnosis (chapter 8). Firstly, graphical tools are developed so as to give a joint answer to two complementary questions often found in financial diagnosis. Then, Neural Networks automatically interpret them. The set of rules generated in this fashion can be seen as a pre-processing for symbol-based expert systems.

Relating firm features to expected return: Our study also typifies financial risk according to the features of large U. K. industrial firms as perceived by the market (chapter 9). It uses arbitrage considerations to establish a relation between sensitivities to factors underlying expected returns and stable features of these firms. The APT offers the possibility of looking into the expected returns of firms from several points of view. APT indices are able to capture with more detail covariance of risk components with some characteristics of the firm. Neural Networks are especially good at modeling information-flow relations involving cross-effects such as those which link accounting reports with market expectations.

Complementary remarks: Appendices contain matters which, if placed in the main text, would break the flow of the reading.

Figures and graphics were used generously throughout the text. The goal is to facilitate the understanding of the subject. Figures are not intended to provide evidence. In most cases they simply show examples or characteristic features. In a few occasions it was impossible to ensure that the presentation of tables or figures would be placed near and after their reference in the main text. This stems from their abundance.

Suggestions for reading: The reader interested in knowledge acquisition in Accountancy and Finance — the subject of this work — should go through the original sequence of chapters. Such reader needs to have a basic background in linear algebra and statistics.

Accountants mainly interested in ratio analysis or in the building of simple statistical models could omit the reading of chapter 3 and the last section of chapter 5. From the second part of the study they would find the first two sections of chapter 8 eventually interesting. The contributions of our research for the problem of the distribution of ratios can be found mainly in chapters 1 and 4.

We would like to make this study accessible to a broad range of potential readers, especially those engaged in research requiring the use of numerical methods in Accountancy and Finance. But it would be outside of the scope of this work to provide detailed explanations for concepts easily found in text-books. Anyway, we tried to avoid all unnecessary algebra.

Acknowledgements

We owe our thanks first and foremost to Professor Robert H. Berry, who supervised our research. The influence of his writings can be traced in many parts of our research. His valuable advice, criticism and very interesting suggestions are also present throughout this study.

Next, we would like to mention with gratitude Professor Augusto de Albuquerque whose qualities of leadership and enthusiasm were in the origin of this work.

The author also wants to express his affection and high regard for all those who funded his period of research in Britain. Firstly, the ISCTE, *Instituto de Ciencias do Trabalho e da Empresa* in Lisbon. But also the AESE, *Associacao de Estudos Superiores da Empresa*.

We are indebted in a very particular way to Professor Luis B. Almeida of INESC. He introduced us to the world of Neural Networks and later, during this study, we have benefitted from many invaluable suggestions and criticisms from him. We are also grateful to Graham Tattersall and his research team at UEA for his suggestions, especially in many practicalities of the implementation of Kohonen Maps of Patterns and other Neural Network algorithms.

Many friends have been generous in giving us the benefit of their experience, in commenting on the manuscript or in general encouragement. We particularly wish to thank Professor Antonio Beltran Almeida, Dr Jose Maria Andre and Dr Pedro Regojo for many helpful comments. We also have benefitted from comments from Professor Elisabete Reis.

We also take this opportunity to express our thanks to the people who have, in various ways, helped us towards this study. Namely, the Accounting and Finance Sector at SYS and the advisory staff at the CPC.

Finally, we wish to express our deep gratitude and esteem to Peter Kemp and his family. They were truly a family for us during our staying in Britain.

Contents

Part I

The Cross-Sectional Characterization of Accounting Data

Introduction To This Part

The subject of the following chapters is the description of the cross-sectional characteristics of a set of accounting items. The final goal is the establishment of an appropriate framework for the extraction of knowledge from accounting reports and related outcomes.

Accounting reports are an important source of information for managers, investors and financial analysts. Statistical techniques have often been used to extract such information from databases where accounting reports and related outcomes are gathered. The goal is to construct models suitable for prediction or for isolating the main features of the firm.

An early model is that of Beaver [7] who used accounting ratios to predict financial distress. Many other researchers followed him, mainly using more sophisticated techniques (see [4] or [125]). Other examples of accounting statistical models are the prediction of bond ratings [63], the relationship between market and accounting risk [9], and the structures of costs and output in various industries [51].

The procedures used to obtain these models are quite similar. The first stage consists of forming a set of ratios from selected items on an accounting report. This selection is typically made in accordance with the beliefs and expectations of researchers. Next, the normality of these variables is discussed. Outliers are discarded. Transformations are applied. Finally some linear modelling technique is used to find optimal parameters in the Least-Squares sense. Linear Regressions and Fisher's Multiple Discriminant Analysis are the most popular algorithms. However Logistic Regression can also be found in some studies. Foster [44] offers a review of accounting modelling practice.

Ratios as Input Variables for Statistical Models: All such models use ratios as predictors. The use of ratios as input variables in accounting statistical models seems to be an extrapolation of their normal use in Accountancy. Ratios are supposed to capture in a simple and standard way some noteworthy feature of the firm. However, there are difficulties involved in using ratios as input variables. As M meaningful accounting variables can generate up to $M^2 - M$ ratios, some research seems to get lost in a prolific use of all sorts of combinations of variables. It is easy to find in the accounting literature models with forty and more predictors.

The problem of selecting adequate ratios as input variables soon became a source of empiricism. It is very difficult to justify the selection of one particular set of ratios instead of any other. Some research clearly opts by accepting them all.

In a survey study dating back to 1981, Chen *et al.* [22] report the use of 65 different financial ratios in 26 studies. From these,

> ... "forty one are considered useful and/or are used in the final analysis by one or more researchers. Given such a heterogeneous set of useful financial ratios, the decision-maker has to be at a loss in selecting which ratios to use for the task at hand. Conceivably, 41 ratios can not all be significant or equally important in a multi-ratio model. The decision-maker may hesitate to omit a ratio if it has been useful in one or more of the empirical studies."

The empirical nature of existing procedures: Accounting statistical modeling practice rely heavily on general-purpose recipes.

- Improvements in normality are sought by pruning out tails and empirically trying different transformations, not always the most appropriate ones. It is a common practice to mix up in the same model square root and log transformations.
- Heteroscedasticity is treated as a separate phenomenon and requires further manipulation, typically a weighting of cases with another variable.
- Multicolinearity is also viewed as an accident in its own right so that the measures recommended for general cases are applied — A few Principal Components are extracted and used instead of the original variables.

The model parameters, after this pruning, scaling and rotating of the input variables become difficult to interpret. The entire routine tends to a broad empiricism.

The Method of this Study: Financial ratios are a simple and intuitive way of capturing features of the firm. Those who use accounting information consider ratios as a starting point for further elaboration. However, when the goal is the description of the behaviour of accounting data it seems clear that ratios are the end of a process rather than the beginning.

We argue that any attempt to build a basis for the description of accounting statistical procedures by looking into the world of financial ratios is not likely to succeed. Ratios are two-variate relations. Their behaviour will be determined by the one of their components plus the interaction between them. This interaction can be itself generated by internal mechanisms of the firm or by external ones, like accounting identities.

It is clear that the final characteristics of any ratio will be determined by complex factors. In this study we begin by the opposite end. We show that the statistical behaviour of accounting data becomes greatly clarified after explaining the statistical behaviour of individual items. And when doing so, also the problems quoted above — normality, transformations versus outliers, heteroscedasticity, multicolinearity — are solved.

Contents: Chapter 1 introduces the data. Then, it empirically assesses the statistical nature of the selected accounting items. Both individual groupings and overall samples are studied.

Chapter 2 examines the regularities devised in chapter 1 and extract the most immediate consequences for the modelling of accounting relations. The existence of outliers and heteroscedasticity in accounting data is explained. The use of regressions is discussed. Then, the joint behaviour of more than one item is approached. The regularities found in the multi-variate behaviour of items are described. Finally, this chapter discusses the statistical nature of items which are a subtraction of two other items.

Chapter 3 builds up the statistical framework to be used throughout this study. New models, which are extensions of financial ratios able to account for non-linearity and non-proportionality, are introduced as a consequence of the empirical observations gathered in previous chapters.

Chapter 4 studies the influence of accounting identities and other external forces in the statistical nature of ratios. The main rules governing their distribution are described. Consequences for the statistical modelling of accounting data are extracted.

Chapter 5 is about size and industrial grouping as the main sources of variability present in our data. A proxy for size is developed and discussed. The homogeneity of some industrial groups and the complexity of its random effects are assessed. We conclude that non-linear modelling tools are required in some occasions when modelling accounting relations.

Chapter 1

Empirical Evidence on the Distribution of Accounting Items

Accounting items, as found in databases containing collections of annual reports of firms, can be viewed as statistical variables. Each firm is a case. For a given item, say, Fixed Assets or Sales, a particular collection of firms form a cross-sectional sample.

In this chapter we show that the lognormal distribution cannot be rejected as a parameterization of the probability density function governing a set of accounting items. We also explore several other possibilities and we show that they are not tenable. In order to assess the importance of the results of the performed tests we compare them with simulated ones. We also examine individually the few cases of departures from lognormality.

We use data from British firms belonging to 14 industrial groups in four broad areas: Engineering, Processing, Textiles and Food Manufacturers. Both individual groups and overall samples are examined.

Our study contemplates a period of five years (1983 - 1987) in order to check the importance of regularities by tracing them during more than one period. But it is a cross-sectional study. Each year is studied individually.

PROCESSING:	14	Building Materials	32	Metallurgy
	54	Paper and Packing	68	Chemicals
ENGINEERING:	19	Electrical	22	Industrial Plants
	28	Machine Tools	35	Electronics
	41	Motor Components		
TEXTILES:	59	Clothing	61	Wool
	62	Miscellaneous Textiles	64	Leather
FOOD:	49	Food Manufacturers		

Table 1: List of the industrial groups examined by this study and their SEIC number.

TA	Total Assets	NW	Net Worth
FA	Fixed Assets	DEBT	Long Term Debt
D	Debtors	C	Creditors
CA	Current Assets	CL	Current Liabilities
I	Inventory	TC	Total Capital Employed
WC	Working Capital		
EX	Operating Expenses less Wages	S	Sales
EBIT	Earnings Before Interest and Tax	W	Wages
OPP	Operating Profit		
FL	Gross Funds From Operations	N	Number of Employees

Table 2: List of accounting items examined by this study and their abbreviations.

1.1 A Review of Previous Research

Some neglect in considering accounting items as statistical variables stems from the fact that their practical importance is not evident. Unlike well-behaved variables like blood pressure or the rate of telephone calls, accounting figures can vary between almost unlimited magnitudes. Reports from companies like ICI or BP, containing values with many digits can be found in the same database along with others which hardly reach four digits.

In order to deal with the problem of such differences in size, accountants use ratios, not the items themselves, to extract useful information. Therefore the statistical properties of basic accounting items received little attention in the literature.

Ratios, of course, have been the object of a much bigger effort. It is worth summarizing briefly this research. As a secondary product some evidence can be collected on the items themselves.

Empirical Research on the distributions of ratios: Horrigan [62] (1965) is an early work on this subject. He analyzed 17 ratios for 50 companies over the period 1948-57 reporting positive skewness. Horrigan explained it as a result of effective lower limits of zero for these variables.

Other studies followed. O'Connor [91] (1973) discovered that for all his 10 ratios in a set of 127 companies during the period 1950-66, skewness was once again prevalent.

Also Bird and McHugh [14] (1976) analyzed 5 ratios for 118 firms over the period 1967-71 in Australia finding skewness. But they considered it as an accident and implicitly suggested the pruning or winsorizing of distributions until achieving normality.

The Deakin study [27] (1976) shows that the positive skewness could not be ignored in his sample of 11 ratios for the period 1955-73. He concluded that

> …"as a result of this analysis it would appear that assumptions of normality for financial accounting ratios would not be tenable except in the case of TD/TA (Total Debt/Total Assets) ratio. Even for TD/TA the assumption would not hold for the most recent data observations."

Deakin also points out that studies suggesting that ratio distributions could be approximated to normality seem to do so for reasons of convenience:

> "With absence of knowledge about these distributions, there is a tendency to rely upon the normal distribution as an approximation due to the availability of statistical techniques designed to analyze relationships among normal variates."

The Bougen and Drury study [17] (1980) was based on U.K. firms. It collected data on 700 industrial firms for 1975 and analyzed 7 ratios, concluding that skewness could not be ignored. Also Buijink [21] (1984) reported the persistency of skewness over a large period. Barnes [5] (1982) argued that skewness on ratios could be the result of deviations from strict proportionality between the numerator and the denominator. This idea that ratio behaviour should be understood by examining the behaviour of the component accounting variables is basic to the present research.

Frecka and Hopwood [45] (1983) extended the Deakin's 1976 study for a longer period and reported similar findings. They also tried to achieve normality by applying square root transformations and pruning the remaining outliers, proposing such procedure as the standard way of dealing with the problem of deviations from normality.

Ezzamel and Mar-Molinero wrote two recent reports (1987 and 1990) on the distribution of ratios using U.K. data [39] [38]. Both are extensive and detailed. The authors also investigate the effects of a family of transformations in the distribution of ratios.

Two studies by McLeay [86] [87] (1986) are somehow out of the previous line of research. They firstly refer to items, not to ratios. McLeay distinguishes two broad classes of items:

Acounting Sums (Σ): Those which are sums of similar transactions, which sign remains the same. This applies both to accounting stocks such as Fixed Assets as well as to flows such as Sales. In cross-section, such items should be bounded at zero with a skewed distribution. Size proxies would fit into this class.

Accounting Differences (Δ): Net items which could be of either sign (or zero), such as Earnings and Working Capital.

McLeay argues that the Σ variables ought to be lognormal since they are directly related to the size of the firm which can be seen as a stochastic process adhering to Gibrat's Law of Proportional Effect [48]. Therefore, ratios formed with Σ variables should also be lognormally distributed.

McLeay's notation is useful and we shall adopt it. However, it somehow seems to induce a qualitative difference between the Σ and the Δ variables. The last ones could wrongly be interpreted as not related to the size of the firm.

Many other studies on the distribution of ratios are not referred to here. There is a wide diversity of observed distributions none of them deserving general agreement. The only feature of accounting ratios which received some credit is the positive skewness.

1.2 The Method and the Scope of this Study

In order to test the lognormality of accounting items we use a logarithmic transformation. We then apply to the transformed data the Shapiro-Wilk test of normality in an improved version due to Royston [105]. This test can cope with large or small sample sizes and is generally recommended as a superior omnibus test. It has been used by some authors for testing the normality of ratios [39] [38]. Berry and Nix (1991) [11] discusses it in more detail.

The Shapiro-Wilk test yields a statistic, W, ranging from zero to one. Values of W approaching 1 mean increasing normality. The significance, P, of W is dependent on the size of the sample. In this study a value of $P < 0.05$ leads to the rejection of the null hypothesis that the tested sample could have been drawn from a normal population.

P should be considered as the probability of obtaining such a W value, or such a sample as the tested one, when the population is normal. It is important to notice that when many tests are performed, the likelihood of obtaining a few cases in which $P < 0.05$ becomes very high.

The Transformation: The logarithmic transformation applied in this study to any item called X is in some cases not just the log of X. The lognormal distribution can have three parameters, not just the two natural extensions of the Gaussian distribution [1]. Therefore, the test of lognormality must make allowance for this third parameter.

In all performed tests when a simple log transformation leads to $0.001 \leq P < 0.05$ — when the probability of getting such a W for a lognormal population is small but not below the used precision — we repeat the test introducing an estimated value for the third parameter, δ so that the transformation of X leads to a new variable T:

$$T = \log(X + \delta) \tag{1}$$

δ is a constant. It accounts for the existence of overall displacements. We call δ a base-line. This designation stems from its role in generative processes leading to lognormal variables.

The above procedure is required if we want to model lognormality with all generality. Lognormal distributions often are three-parametric. The δ are in general small.

An introduction to the lognormal distribution can be found in a book by Aitchison and Brown [1]. When no δ is needed for achieving normality these authors refer to the parameterization as a Two-Parametric Lognormal. For significant δ they refer to it as a Three-Parameters one. This terminology is generally accepted.

Estimating the Base-Line: Some procedures available for estimating δ are also described by Aitchison and Brown. In our case, δ is estimated by trial following a suggestion of Royston [105]. It is a method basically similar to the one originally used by Gibrat and about which Aitchison and Brown say that "it is much more an art than a science". Royston seems to rely on a new factor,

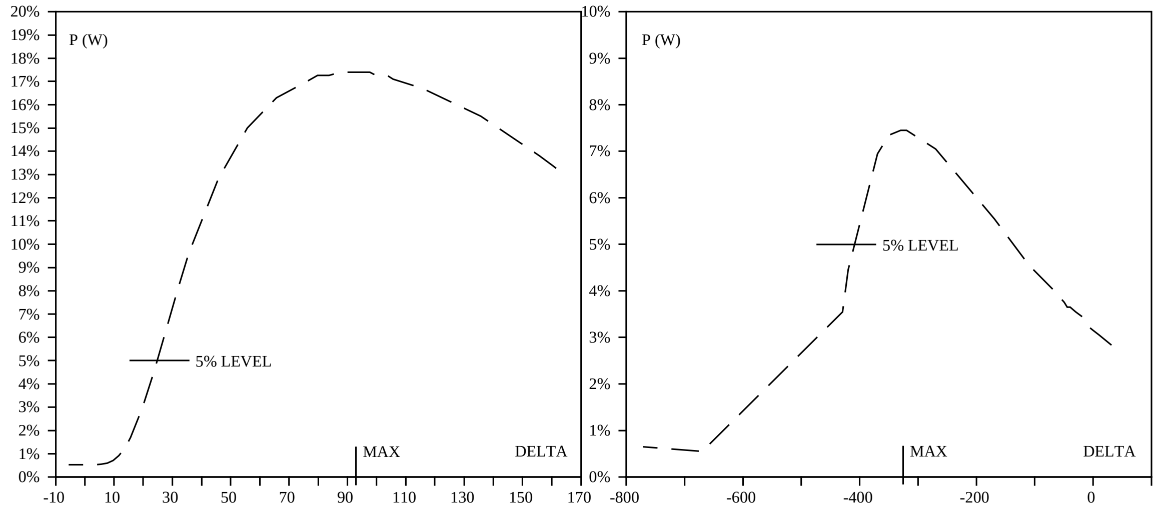

Figure 1: The significance, (P), of the Shapiro-Wilk's W for varying deltas. In these two cases we would accept a three-parametric lognormal distribution with $\delta = +90$ (left) and $\delta = -300$ (right).

the precision of the Shapiro-Wilk test, to overcome this criticism. In fact, the Shapiro-Wilk test is much more reliable than the graphical methods Gibrat used for assessing normality.

Basically, this method consists of discovering by trial which δ maximizes W or $P(W)$. It is indifferent to use the statistic or its significance. Figure 1 shows two examples. The significance, P, of W improves when we add to all observations in the sample a constant small δ before transforming the data. By trying increasing δ we find an optimal W or $P(W)$.

In the case displayed on the right (Total Capital Employed, Electrical industry in 1983), $\delta = -300$ is the value beyond which $P(W)$ no longer improves. Therefore we take this $\delta = -300$ as the correct third parameter to introduce in the log transformation. On the left we see how a small δ of $+90$ enhances the lognormality of Working Capital (Paper & Packing, 1985).

We found sharp, well contrasted optimal values for W, making it possible to estimate δ in this way. In our study the estimated δ were often related to the absolute smallest value in the sample.

Obviously, the values allowed for δ were always smaller than the smallest case in the sample having opposite sign. Therefore the use of δ cannot be considered as similar to the practice of adding a value to all the cases in a sample to avoid negative cases. We never let δ change the sign of any value in the sample.

The Variables: We examined 18 different accounting items. They are listed in table 2 (page 6). There are 11 items from the Balance Sheet, 5 from the Profit and Loss Account, 1 from the Sources and Applications of Funds statement, and one which is not standard. These items are widely used in the second part of this study. They are also frequent as components of ratios used in statistical modelling. Of course, the number of selected items could grow to much higher figures.

The selection of EX, (Operating Expenses less Wages) is a way of getting a picture of the cost structure of firms using disclosed data. The inclusion of N (Number of Employees) concerns findings discussed later on (chapters 5 and 7). This variable convey important information regarding the classification of industrial groups. It is also useful for comparing accounting items with non-accounting variables exhibiting similar statistical behaviour.

In our study this set of selected items can be divided in two groups.

1. Those having only positive cases, like Sales or Inventory. They are broadly the items McLeay calls Σ variables.

2. Those items which can have both positive and negative cases in the same sample like Earnings and Working Capital. They are referred by McLeay as Δ variables. In this second group we perform, when possible, two tests of lognormality.

   (a) Using only the positive-valued cases in each sample, and

   (b) using only the absolute value of the negative cases. For small samples this test could not be carried out because the number of negative cases was non-significant or non-existent.

In section 2.3 (page 48) we discuss the potential applicability and usefulness of these split cross-sectional samples.

The abbreviations, when used, are usual in the literature (see table 2 on page 6).

The Samples: All companies quoted on the London Stock Exchange are classified into different industry groups according to the Stock Exchange Industrial Classification (SEIC) The SEIC aims to group together companies which results are likely to be affected by the same economic, political and trade influences [95]. Although the declared criteria are ambitious, the practice seems to be more trivial, consisting of classifying firms mainly on a end-product basis. The SEIC classifies firms according to a perception of groupings of firms.

The tested samples were drawn from the Micro-EXSTAT database of company financial information provided by EXTEL Statistical Services Ltd, which covers the top 70% of UK industrial companies. We selected 14 manufacturing groups according to the SEIC criteria (see table 1 on page 5). Two kinds of samples were examined.

All Groups Together, in which the 14 industrial groups are gathered in one unique sample.

Industry Groups, for samples of only one industrial group at a time.

Each described test is performed five times for reports from 1983 to 1987. None of the companies present in the original groups were excluded from the tests.

Item	Significant $P(W)$				
	1983	1984	1985	1986	1987
Sales					
Operating Expenses less Wages					
Wages	0.015		0.001	0.01	0.02
EBIT					
Operating Profit					
Gross Funds From Operations					0.03
Debtors					
Current Assets				0.03	0.02
Inventory					
Fixed Assets					
Total Assets		0.01	0.005	0.01	0.005
Working Capital					
Creditors	0.01			0.006	
Current Liabilities	0.03			0.04	
Net Worth					
Long Term Debt					
Total Capital Employed					
Number of employees			0.02	0.04	

Table 3: Two-parametric lognormal items for all groups together. Cases of three-parametric lognormality are those for which $P < 0.05$. The number of cases in each sample ranges from 550 to 700.

1.3 Results

In this section we display the results of testing the lognormal hypothesis for two kinds of samples. Firstly, the large ones containing all the 14 industries together. Secondly, the small ones drawn from one industry at a time. These two groups of tests represent two possible levels of homogeneity worth exploring.

In fact, one single group, if homogeneous, yields homogeneous samples. Two or three groups, each of them homogeneous, can yield samples which are severely non-homogeneous. But 14 groups in the same sample, all of them sharing a common attribute, are likely to apportion random effects rather than fixed ones. In that case, a second level of homogeneity could be attained. The examination of such samples becomes interesting since they represent the common attribute they share.

1.3.1 All Groups Together

The large samples mentioned above are now examined. The common attribute to consider here as a possible source of homogeneity is the industrial character of all the gathered firms.

Positive cases: In appendix A we display the number of cases in each sample and the statistics obtained when applying the Shapiro-Wilk test along with some usual measures of normality (kurtosis and skewness) to the 13 positive-valued accounting items and to the positive values of the 4 items having both positive and negative cases. We also included Long Term Debt for which only the non-zero cases were selected.

Item	Statistic	1983	1984	1985	1986	1987
EBIT	W	0.9572	0.9838	0.988	0.951	0.946
	sig W				0.042	0.048
	N. Cases	40	42	48	57	47
Operating Profit	W	0.9537	0.9851	0.9925	0.9686	0.9596
	sig W					
	N. Cases	57	58	62	74	69
Gross Funds From Operations	W	0.9594	0.9649	0.9792	0.9066	0.9644
	sig W				0.005	
	N. Cases	27	24	30	36	38
Working Capital	W	0.9640	0.9743	0.9647	0.9678	0.9609
	sig W					
	N. Cases	50	61	67	61	62

Table 4: Results of applying the Shapiro-Wilk test to the absolute values of negative cases. All groups together. Departures from the two-parametric assumption have $P < 0.05$.

Table 3 on page 11 shows, by year, a short summary of the results of applying the Shapiro-Wilk test to such items. When nothing is said items yielded non significant departures from two-parametric lognormality. When there is a significant difference, the significance is displayed. In all the significant departures observed, the introduction of a small δ made it vanish.

The results show that 11 on the 18 items are two-parametric lognormal in the whole period of 1983 to 1987.

Sales and Operating Expenses less Wages, Net Worth, Debtors, Fixed Assets, Inventory and Total Capital Employed, along with the positive cases of Earnings, Operating Profit, Long Term Debt and Working Capital, are persistently two-parametric lognormal.

The remaining 7 variables are either two-parametric or three-parametric lognormal depending on the year. None is persistently three-parametric for the five years. In at least one year all variables achieved lognormality with just a simple log transformation.

Total Assets and Wages require a three-parametric transformation in four of the observed years. These samples have their smallest values far away from zero. In general the positive values of McLeay's Δ variables are more near two-parametric lognormality than the Σ ones. This is because their smallest cases approach zero. Only Gross Funds From Operations exhibit one departure from a two-parameters distribution, in 1987.

The values optimal δ assumes whenever a three-parametric transformation is required often follow the smallest case in the sample. It is worth noticing that the effect of adding such a δ completely vanishes for larger cases. The base-line affects only the smallest cases in the sample.

Negative cases: We also checked the negative values of items having both positive and negative cases. We selected first the set of negative cases and then we applied logs to their absolute values.

The results are displayed in table 4 (page 12). Operating Profit and Working Capital are two-parametric lognormal for the whole period. Earnings and Gross Funds From Operations are, in one or two years, three-parametric. Therefore losses are also lognormal. We shall see later on that these

Ind. / Items	S	CL	TA	N	TC	CA	EX	W	C	NW	D	I	FA	T.Ind.
Leather														
Metallurgy														
Motor c.														
Textiles m.														
Wool														
Clothing	1						1							2
Building m.										1			2	3
M. tools				1		1					1	1		4
Chemicals		3							1			1	1	6
Paper	1					1	3				2			7
Electrical	1		3	2	4	2		1		2		3		18
Food	4	3	3		2	2	4	2	4	1				25
I. plants	4	2	2	3	2	1	1	1	2	2	2	1	2	25
Electronics	3	3	3	4	2	2		5	1	2	2			27
T. item	14	11	11	10	10	9	9	9	8	8	7	6	5	117

Table 5: Items having only positive cases. Number of cases yielding a significant departure from two-parametric lognormality by industry and item.

negative values have a distinct behaviour.

1.3.2 Within Industrial Groups

In this section we show that the lognormality of our set of accounting items cannot be rejected also when each sample is drawn from the same industrial group. Since the number of samples involved is very large we shall not display here the detailed results. Instead, we present proportions of departures from the two-parametric lognormality by industry and by item.

The detailed results of applying the Shapiro-Wilk test to all samples — the eighteen observed items for fourteen industries during five years — can be found in appendix A.

How to interpret the displayed proportions: All the presented proportions are calculated using marginal totals: When we refer to a value of 12% of departures by industrial group the total used to calculate such proportion is not the overall total of different samples tested (say, 910) but the 65 different possibilities of sampling by industry. That is, since we are working in this case with five years and 13 variables there are $5 \times 13 = 65$ different ways in which a particular industry can be sampled.

For example, the total used for displaying the important issue of persistency of lognormality for the whole period of five years was 182 because there are 13 items and 14 industrial groups which makes 182 different samples to be tested involving this period. The tables displayed in appendix A allow anyone interested in to calculate any particular proportion.

Positive-valued items by industrial group: For the selected 14 industrial groups, 13 positive-valued items were checked for lognormality with the Shapiro-Wilk test. This procedure was repeated for five years (1983-1987). Therefore the total of different samples tested under this item was 910.

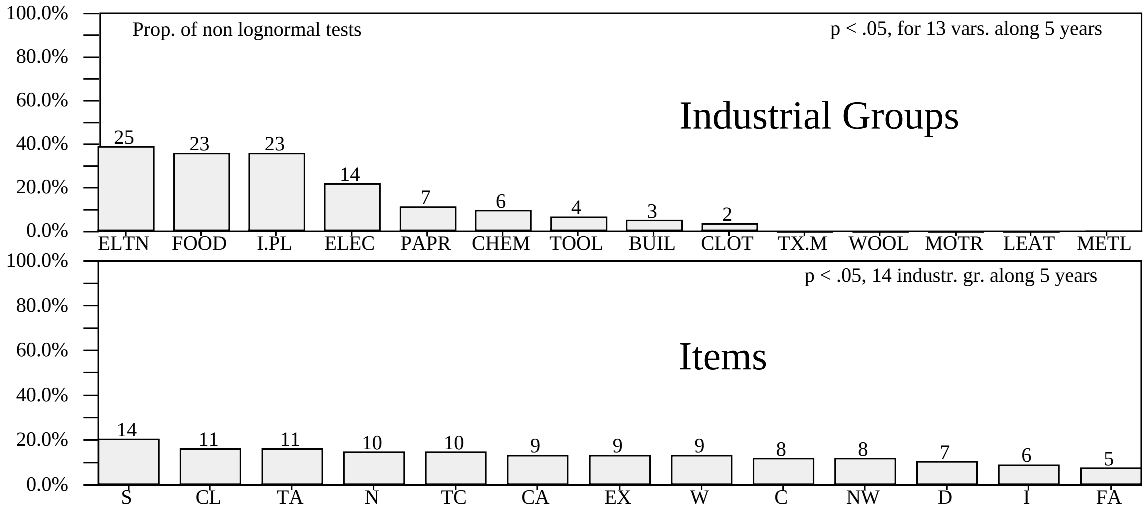

Figure 2: The incidence of departures from two-parametric lognormality for positive-valued accounting items and industrial groups.

Lognormality was observed for the generality of cases. A total of 793 samples (87.1%) yielded two-parametric lognormality.

Table 5 on page 13 shows by industry and by item the number of departures from two-parametric lognormality in the considered period of five years. The table has been sorted by number of departures. As an example, the value 2 in column FA and raw BUILD means that departures from a two-parametric lognormality were observed twice in the five years period for Fixed Assets (FA), in the Building Materials group.

Only in one case (Wages in the Electronics industry) we obtained a persistent $P < 0.05$ for all the five years. Electronics is also the group having more such cases (almost 40%). Next comes Food Manufacturers and Industrial Plants with more than 30% followed by Electricity, about 20%.

Groups like Paper and Packing, Chemicals, Machine Tools, Building Materials and Clothing have less than 10% of departures. Miscellaneous Textiles, Wool, Motor Components, Leather and Metallurgy have no departures at all.

Industries are particular regarding lognormality. Some are two-parametric for all the observed items and some others have a significant number of items which are three-parametric. There is not such a clear separation for items as for industries. The worst item, Sales, had 20% of non two-parametric lognormality for the five years in all groups. The best one, Fixed Assets, had less than 8%. Three-parametric lognormality seems to be related to industrial groups, not to items. Figure 2 on page 14 shows such a contrast in more detail.

Positive values of items having both positive and negative cases by industrial group: In this paragraph we observe the behaviour of the positive values of four items having both positive and negative cases like Working Capital or Earnings. We also include one item, Long Term Debt,

Ind. / Items	EBIT	OP.P.	W.C.	DEBT	FL	T.Ind.
M. tools						
Textiles m.						
Metallurgy						
Leather						
Electrical					1	1
Paper			1			1
Building m.	1				1	2
Wool			1	1		2
Motor c.				1	1	2
Clothing		1			2	3
Chemicals				2	1	3
Electronics	2	2	2		1	7
I. plants	1	2	2	1	1	7
Food	2	1	1	4	3	11
T. item	6	6	7	9	11	39

Table 6: Items having positive and negative cases. Number of cases yielding a significant departure from two-parametric lognormality by industry and item.

for which only the non-zero cases were selected. That makes a total of 350 samples to be tested: Five variables for the same fourteen industries during a period of five years.

A total of 311 samples (88.9%) yielded no significant departure from the two-parametric lognormal hypothesis.

The results by industrial group and by variable are similar to those of the previous paragraph. Food, the second worst for positive variables, is now the worst. The industries exhibiting no departures at all are Leather, Metallurgy and Miscellaneous Textiles as before, along with Machine Tools (see table 6 on page 15).

Again, the five observed items had a more homogeneous behaviour than the industrial groups. Departures were observed in 16% to 8% of the samples, Funds Flow being the worst. Figure 3 on page 16 shows the contrast between the behaviour of industries and items regarding the incidence of three-parametric lognormality.

Negative values of items having both positive and negative cases by industrial group: We could not test the lognormality of all industries when considering only absolute values of negative cases. The size of the resulting samples would be too small or non-existent.

Table 7 on page 17 contains the list of the samples we were able to select and the values of W for each one of them. Only two groups were large enough to provide a few negative cases. All the examined samples were two-parametric lognormal.

Simulation of Negative Working Capital: Since the empirical data was not conclusive we simulated a difference between two lognormal variates having the same features encountered in the items themselves.

We selected a sample, Food 1987, and measured the mean values and correlation coefficient of $\log CA$ and $\log CL$. Then we generated a two-variate sample with 2000 cases obeying the observed

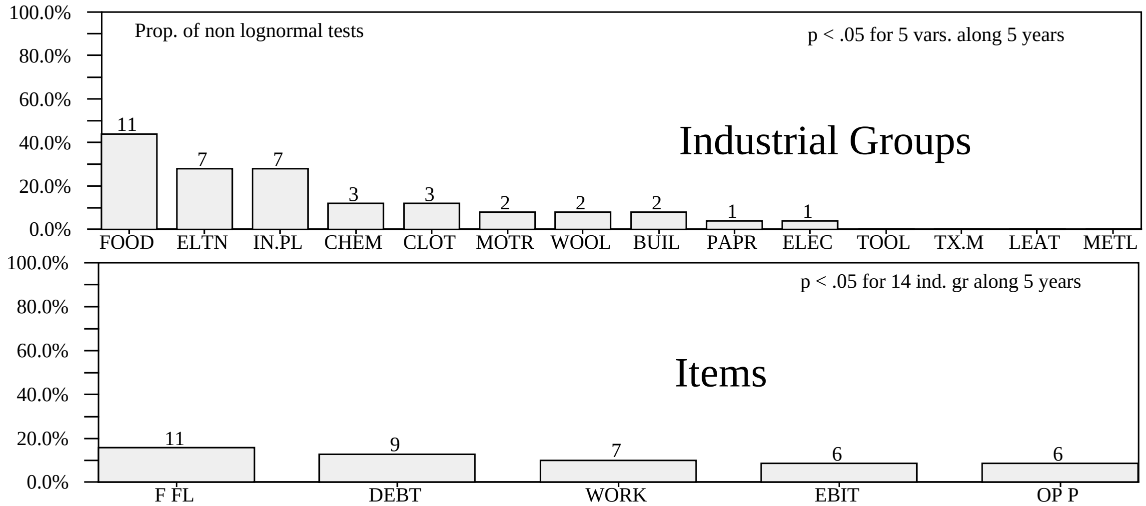

Figure 3: The incidence of departures from two-parametric lognormality for positive cases of accounting items and for industrial groups.

parameters. Finally we found the anti-logarithms and subtracted them. The simulated Working Capital had 463 negative cases. The absolute values of this set were examined and they were lognormal. The results are displayed in appendix A along with details of the used procedure.

The simulated values of CA and CL for negative variates of Working Capital were also lognormal. This result is quite the expected one since $\log CA$ is strongly correlated with $\log CL$. In a two-variate distribution, the line separating the pairs $\{\log CA, \log CL\}$ corresponding to positive WC from the negative ones is parallel to the principal axis (the one along which there is a maximum of variability). The two distributions resulting from cutting the original one along such axis have marginal distributions which are not very different from the original. Should the slicing be done along an oblique axis, then the distortion would be noticeable.

A similar reasoning explains the lognormality of Working Capital. In section 2.3 we further explore the problems posed by the statistical modelling of negative-valued lognormal items.

The lognormality of absolute values of negative cases is not enough to make us consider them as having the same statistical characteristics as the positive ones. Their correlation with other items and in general with the size of the firm is smaller than the usual for positive deviates. Losses are indeed correlated with size but less than profits. In section 2.2.2 we examine correlations between items and more simulations similar to this one are carried out.

Persistency of departures from the two-parametric hypothesis: We measured the number of times a significant departure from a two-parametric lognormal distribution was observed in a given sample for the period of five years. Such a measure can give us an idea of the persistency of simple lognormality. For example, when considering items we noticed that Working Capital in three industrial groups departed once in five years. Two other groups departed twice. And when

Year	Industry	N. Cases	Variable	W	$P(W)$
1983	Food	15	W. Capital	0.929	0.28
1984	Electronics	13	W. Capital	0.946	0.51
	Food	19	W. Capital	0.927	0.16
1985	Electronics	17	EBIT	0.971	0.82
		25	W. Capital	0.950	0.27
		12	F. Flow	0.983	0.97
	Food	14	W. Capital	0.917	0.19
1986	Electronics	16	W. Capital	0.957	0.59
		15	EBIT	0.929	0.26
		24	EBIT	0.918	0.05
	Food	16	W. Capital	0.933	0.30
1987	Electronics	16	W. Capital	0.923	0.19
		13	F. Flow	0.953	0.61
		15	EBIT	0.962	0.69
	Food	16	Op. Profit	0.898	0.09
		17	W. Capital	0.961	0.63

Table 7: Negative cases. Observed samples. No departures from two-parametric lognormality.

considering groups instead of items, Metallurgy had no departures at all, Building Materials had three items which departed once in five years and one item which departed twice. And so on.

In appendix A we display these results in detail. Next table is a summary.

Departures from two-parametric lognormality in five years	Σ items	Δ items
Departures are never observed	69.2%	62.9%
A departure is observed once	11.0%	22.9%
A departure is observed twice	10.4%	11.4%
A departure is observed three times	5.5%	1.4%
A departure is observed four times	3.3%	1.4%
A departure is observed five times	0.5%	

In 69% of the tests performed in samples having only positive values there is no departure at all during the five-years period. One departure or two can be observed in 21% of the cases. Three departures only occur in 5% and four in 3%. Only once (Wages in Electronics) persistence of departure from the two-parametric assumption is observed for the whole period of five years.

The study of persistency for positive values of items having both positive and negative cases yielded similar results. We consider that a non-persistent departure from two-parametric lognormality is interesting. It shows that deviations from the two-parametric model are sporadic.

1.4 Other Possible Parameterization

The logarithmic transformation can be viewed as a way of reducing an excessive amount of skewness in a distribution. Therefore it makes sense to ask if the reduction achieved with logs is the appropriate one in the case of accounting items.

In case less reduction is required we should use a square root or other appropriate root. In case more reduction is required we should use the Pareto distribution (log-ranks) or other of its class.

Item		Square Root	Cubic Root	Fourth Root	Sixth Root	Eighth Root	Log
SALES,	1985, Electronics						
	Skewness	3.59	2.43	1.86	1.32	1.07	0.412
	Kurtosis	15.77	7.97	4.93	2.62	1.76	0.091
	Shapiro-Wilk's W	0.627	0.7783	0.8472	0.9041	0.9266	0.968
WAGES,	1985, Electronics						
	Skewness	4	2.84	2.26	1.7	1.45	0.777
	Kurtosis	19.43	10.2	6.61	3.82	2.72	0.488
	Shapiro-Wilk's W	0.56	0.711	0.785	0.851	0.879	0.94

Table 8: Comparing several root transformations with the log transformation using two samples.

1.4.1 Root Transformations

We tested first a scale of possible roots progressively approaching the effect of a log transformation. For each of them we measured the skewness, kurtosis and W.

From the samples described above we checked only those yielding a poor figure for the significance of W. In all such cases the log transformation was the one which maximized W.

Table 8 shows the results for two particularly badly-behaved groups. Clearly the signs of normality increase with increasing roots, achieving much more acceptable values with logs.

Part of this table is replicated graphically in figure 5 (page 20) on the right. On the left of the same figure a de-trended normal probability plot reveals the fact that only a log transformation seems to account for the skewness present in the data.

In figure 4 (page 19) we display the graphical evolution of a frequency distribution when progressively larger fractional exponents were used for transforming the data. The last case is a logarithmic transformation.

Characteristic behaviour of items: When assessing the distribution of ratios it is usual to find cases in which no transformation seems to improve the normality of the data. However, when we look at single accounting items the situation is very different. Not only the log transformation emerges as the most appropriate. Also there is a clear progress towards normality for fractional exponentiation of increasing degree.

Accounting items show a statistical behaviour much simpler than the one found in ratios. In the particular case of transformations, it is frequent to find in the literature references to an unpredictable outcome in the distribution of ratios after applying transformations such those we use here. We explain this behaviour of ratios in chapter 4. At the moment it is important to notice that there is no reason to expect in items the same kind of interactions present in ratios.

1.4.2 The Pareto Hypothesis

After discarding transformations less effective than logs in neutralizing skewness we tried one which is more powerful.

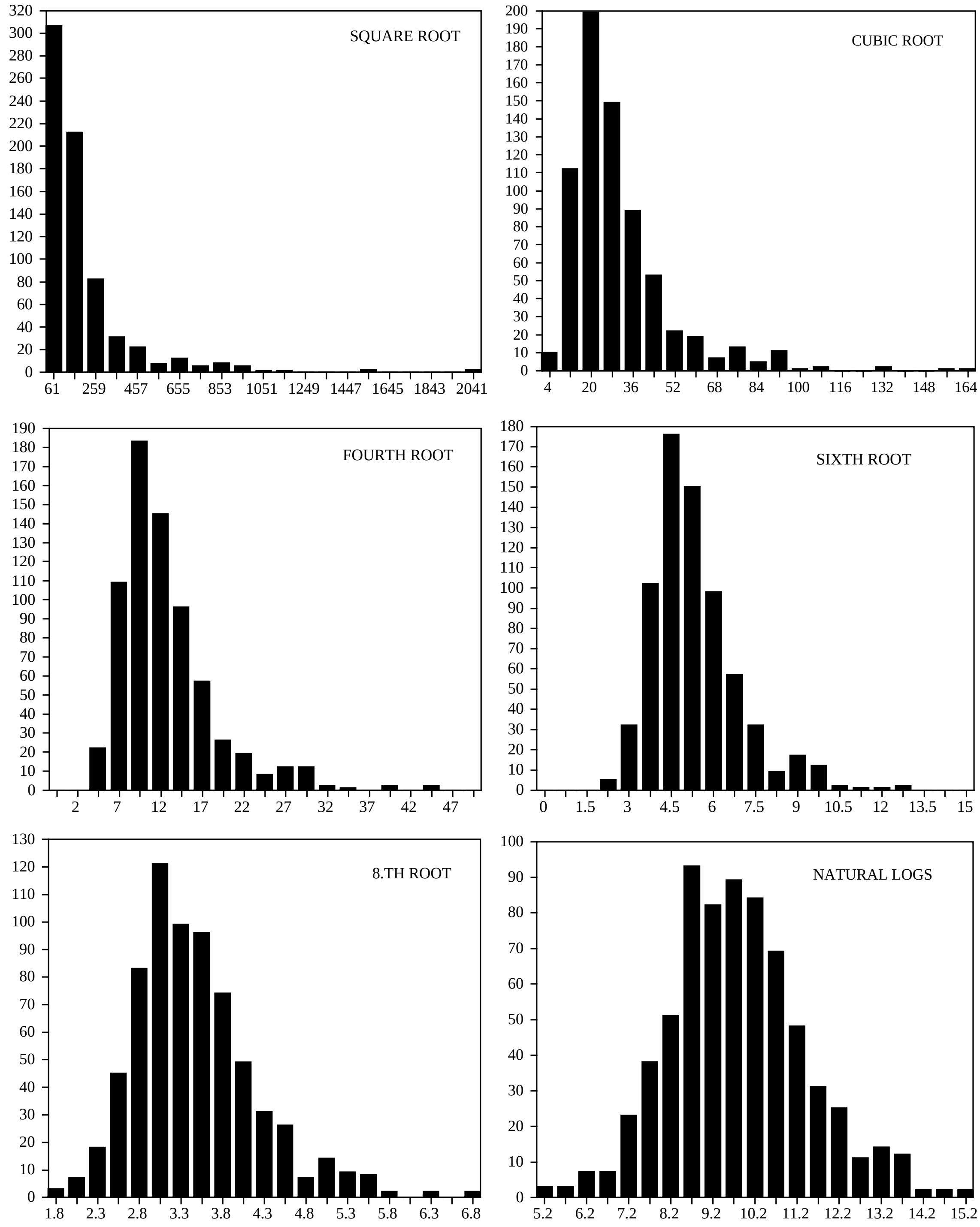

Figure 4: Evolution of the frequency distribution of Current Assets, 1986, all groups, when several root and a log transformation were applied.

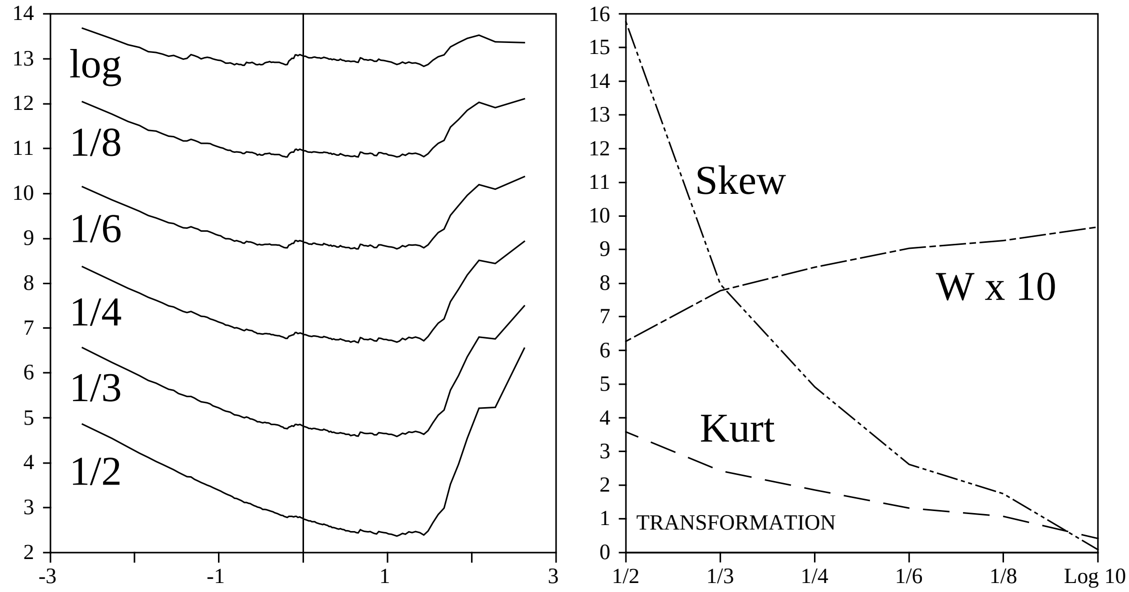

Figure 5: On the left, de-trended normal probability plot for several root transformations and logs. Absolute normality would mean a horizontal straight line. On the right, the evolution towards normality of three known statistics when increasingly high fractional exponents are applied.

Pareto processes have cumulative distributions for which the relation between the observed values and the rank is logarithmic. If x is a random variable governed by a Pareto process we should observe

$$\log x = \log M - \beta \times \log r$$

in which r is the rank of x (the largest x is assigned the rank 1 and so on) and M, β are constants.

Therefore, if we rank accounting items in a sample from large to small the log of the item and the log of the rank should be linearly related (with a slope of β) for the Pareto hypothesis to be acceptable. It is not the case.

The Pareto process and our data: A clear downward concavity of the distribution was observed for all items. This departure from the Pareto hypothesis is very significant. The actual size of firms occupying the middle of the rank is more than twice as large as that predicted by the Pareto distribution.

Figure 6 (page 21) shows on the left an example of the shape accounting items assume when their logs are compared with the logs of their rank in the sample. No signs of linearity can be observed.

In the same graphic we show the shape of an equivalent lognormal deviate (solid line). Clearly, accounting values follow much more closely the lognormal deviate than any straight line. The hypothesis of a Pareto or similar process governing accounting items does not seem tenable. Ijiri and Simon [65] also report the same kind of concavity for Sales and the number of employees in samples of U.S. firms.

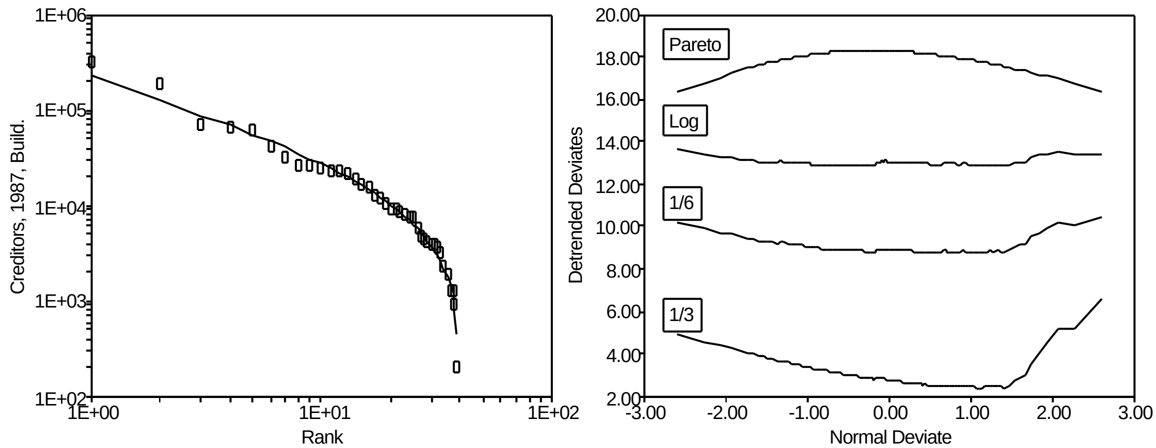

Figure 6: On the left: Pareto processes would exhibit a linear relation between log ranks and log values. The solid line is the corresponding lognormal deviate. On the right, a de-trended normal probability plot with two root transformations and a log one, along with a Pareto deviate.

On the right of figure 6 we can see the shape a Pareto deviate assumes in the de-trended lognormal probability plot. It illustrates the meaning of transformations accounting for progressively larger skewness. According to this scale the lognormal process appears in between the root and the Pareto ones.

The Pareto process and the growth of firms: In some literature it is usual to link the growth of firms with a Pareto process (see for example Steindl [121]). If the growth of firms is Pareto-like, we should observe cross-sections of accounting items like Sales, Total Assets and Net Worth drawing straight lines on a log-ranks vs. log-items scatter. Since this is not the case, it seems as if our results contradict this belief.

The models of growth used for justifying a Pareto distribution of firm sizes are inspired by the Gibrat Law [48]. The Gibrat Law leads to a whole class of skewed distributions depending on the conditions imposed on the growth process. As we recall in chapter 3, the most immediate outcome of the Gibrat Law is lognormality. Lognormality, however, is too simple and general an outcome. It requires a random walk as the growth rate of firms.

The literature concerning the growth of firms considers Pareto processes instead of the lognormal ones because of this scarcity of assumptions the lognormal hypothesis allows. In fact, when known influences like the serial correlation during growth, the disinvestment or the effect of mergers and acquisitions — and especially the birth and death of firms, see [65] — are accounted for, the resulting cross-sectional distribution should be a Pareto or a Yule one, not the lognormal.

We think that the lognormality observed in items cannot be ignored. The models based on the Pareto distribution are attractive but they don't seem able to explain the actual distribution observed in the data. Lognormality therefore should be taken into account in future theoretical developments.

More reasonable is the discussion of whether or not the Pareto process influences large firms. The testing of such a hypothesis is difficult since lognormal processes and the Pareto ones lie very near one another in the upper part of their distributions. If we rank a perfectly lognormal deviate by size and then discard the smallest 2/3 of cases we get a set of values which are not far from a Pareto process. This can also be observed in figure 6. The differences between a straight line and a lognormal variate are small in the portion of the sample containing the largest cases.

1.5 Assessing the Importance of Multiple Tests

In this section we discuss the meaning and importance of W and corresponding P values obtained from applying the Shapiro-Wilk test to a large number of different samples.

We simulated many samples drawn from a strictly Gaussian population. All simulated samples had the same number of cases as found in each performed test. Then we compared the distribution of the P obtained when applying the Shapiro-Wilk test to this set of simulated samples, with the distribution of P from the real tests.

We also examined the possible existence of correlations between such values and the number of cases in every sample.

The Logit transformation: The Gaussian distribution emerges as the result of many independent random causes. It would be interesting to compare our set of P values with the Gaussian distribution in order to measure in what extent their spread can be considered as caused by many independent random events as those influencing any mechanism of sampling.

A set of P values such as the one obtained from the repetition of a significance test for many samples cannot be directly compared with the Gaussian distribution because probabilities or any relative frequencies are bounded by 0 and 1. However, it is possible to transform probabilities so that the resulting variable is normal for Gaussian generative processes. We used the simple logit transformation as an acceptable approximation of the relation linking relative frequencies with Gaussian deviates.

For a given P, $[0 \leq P \leq 1]$, we computed

$$Logit\ P = \log \frac{P}{1-P}$$

$Logit\ P$ (also known as log-odds) now ranges from $-\infty$ to $+\infty$ and is approximately normal for random normal events. A value of $Logit\ P = 0$ is the expected or central one. Negative logits mean $P < 0.5$ while the positive ones are obtained for $P > 0.5$. $Logit\ P$ can boldly be taken as the number of standard deviations for normal distributions.

The logistic function

$$P = \frac{1}{1 + 10^{-logit\ P}}$$

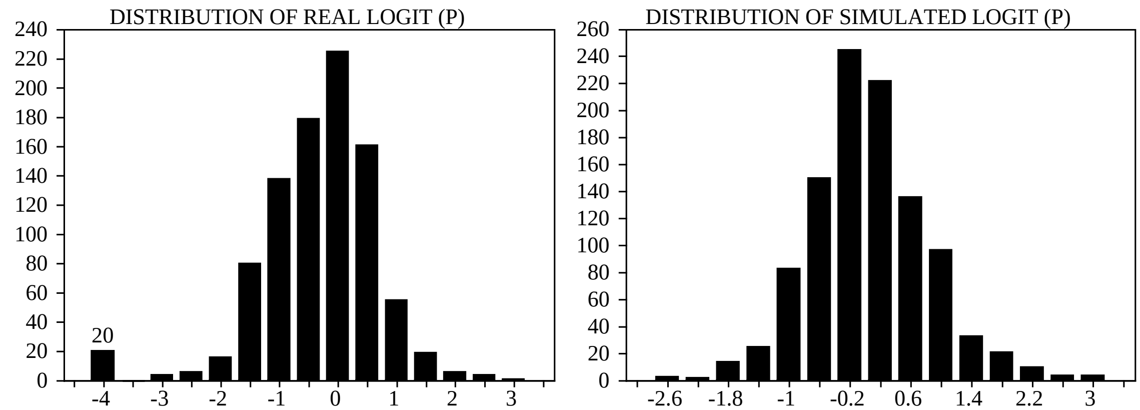

Figure 7: The frequency distributions for real (left) and simulated (right) *Logit P*. The cluster of bad cases has been added to the distribution of real *Logit P* with an arbitrary, very small, value.

transforms log-odds back to probabilities.

Comparing distributions of real and simulated Logits: First we examined the distribution of the log-odds obtained after performing the $910 + 350$ tests described above. In order to do this we had to exclude $16 + 4$ tests having $P < 0.001$ because they would yield infinitely small log-odds for the number of significant digits we were working with. The resulting distribution had a mean value of -0.25 (equivalent $P = 0.49$) and a standard deviation of 0.85. Both skewness and kurtosis were very small. The aspect of the frequency distribution was very much the one of a normal process.

Simulated normal deviates yielded a mean value of 0.04 (equivalent $P = 0.53$) and a standard deviation of 0.8. The skewness was acceptable but the kurtosis was around 1. 3.6% of the simulated tests yielded $P < 0.05$. Notice that supposing normality, 95% of the tests would have to fall inside an interval of $\{-1.6, +1.6\}$ logits which is $P = \{0.025, 0.975\}$ as they did.

The fact that when performing multiple tests some of them are expected to exhibit $P < .05$ even when the samples were drawn from a normal population means nothing wrong with the Shapiro-Wilk test itself. When many samples are drawn from a perfectly normal population it is likely that some of them will be far from normality in some degree due to the random nature of sampling.

Figure 7 shows the frequency distributions of both real and simulated tests. Some interesting conclusions arise from comparing them.

- When many tests are performed we can expect at least 3.6% of them to show significant departures from normality even when the population from which the samples were drawn is strictly normal. Such result is the expected one. Samples don't have to be normal even when the generative mechanism is. In our case we put aside 16 tests as special cases. The proportion of cases having $0.001 < P < 0.05$ is now 11.3%. It seems as if only something like 7.7% of those $(11.3 - 3.6)$ should be considered as real, unexpected, departures from lognormality.

- The second conclusion is induced by the similarity of spreads and the normality observed in the real distribution. Normality means random, independent causes. Now, if some hazardous sampling can introduce a spread of 0.8 in an otherwise perfectly normal collection of samples, the real spread, also 0.8, should be assigned to an hazardous mechanism of sampling and not to any particular cause.

Therefore we can take the difference between estimated expected values as the sole factor affecting lognormality of accounting items. This difference can be expressed by saying that while the expected probability associated with the Shapiro-Wilk W is, for a normal population, 0.53 it becomes 0.49 in the case of accounting items. A question of four to one hundred odds. This is what seems to separate accounting items from a strict lognormal generative mechanism.

Of course, we are interested in accessing the plausibility of assumptions which concern the population, not particular samples. That's why the above simulations are important. They provide a firm ground for interpreting the results.

Relation between the size of samples and the significance of W: The Shapiro-Wilk test is very robust concerning the size of the sample. In the considered interval no correlation was observed between the size of the sample and the significance of W.

1.6 Examination of Bad Cases

We noticed in previous sections that a few tests of lognormality yielded values of P (Shapiro-Wilk's W significance) which were so small that it would not be possible to apply logits. A small P means a departure from the lognormal assumption, that is, from the null hypothesis that the examined log sample could have been drawn from a Gaussian population.

Such set of bad cases behave differently from the other cases. The significance (P values) obtained from all other tests form in logit space a clearly normal distribution. Bad cases do not fit well in such distribution. They are more numerous than expected and they form a cluster sticking out well below the lower normal values of *Logit P*. See figure 7, page 23, on the left.

They are also insensitive to a three-parametric transformation. No δ exists able to turn them lognormal. This feature is a very particular one since in the other cases yielding significant departures a δ exists able to bring W to non-significant values.

We examined each one of such bad cases in order to find out the reasons for their erratic behaviour. In appendix A we describe them. The causes for the existence of the described bad cases seem to be twofold:

Anomalous Cases. Errors, extreme outliers. Values which are orders of magnitude away from the rest of the sample. Most of them are clearly erratic.

Non-Homogeneous Groups. The existence of clusters of firms well detached inside the same industrial group is perhaps a result of a temporary expansion of the sector. Or it can be a consequence of an intrinsic non-homogeneity of an industrial group. It happens from 1984 to 1987 in the Electronics and Food industries. But it only affects Sales, the number of employees and Wages.

Lognormality, either two or three-parametric, is linked with the homogeneity of the sample in logarithmic space. When an industry is not homogeneous — for example, when there are two groups of firms instead of one — the tests cannot classify the sample as lognormal, even if the underlying mechanism governing the behaviour of accounting items is lognormal.

We conclude that there seems to be an explanation or cause for each one of the observed strong departures from the lognormal hypothesis. These causes should be considered as external to the generative mechanism governing the cross-sectional characteristics of accounting items. Their frequency, 20 cases in 1260 different samples, makes them exceptional.

1.7 Discussion and Conclusions

For the observed cases lognormality emerges as a very general and stable feature of cross-sectional samples of accounting items. Not only the large samples formed with several groups are lognormal. Each individual group is lognormal too. Not only "sums of similar transactions which sign remains the same" [86] are lognormal. Items with very different origins are lognormal as well. On the whole, we found only 20 non-lognormal samples in 1260. And there seems to be a good explanation for such departures.

Three-parametric lognormality: Most of the observed samples are two-parametric. Some are three-parametric. In lognormal distributions the third parameter accounts for displacements of the entire sample. We call these displacements base-lines.

In a three-parameters distribution the smallest cases are not approaching zero as they should. They approach a given base-line instead. The two-parameters lognormality only contemplates situations in which the smallest variates approach zero. In accounting numbers, as in many other lognormal variables, base-line effects are expected.

Small base-lines only affect the smallest cases in the sample. The base-lines we estimated are near the magnitude of the smallest cases in the sample. For a sample with smallest values of 10, 15 and so on, the base-line will be non-important. The distribution would be two-parametric. For samples with smallest values of 800, 900 and so on, the introduction of a δ is required in the transformation for achieving normality. This agrees with the three-parametric mechanism.

Three-parametric lognormality is sporadic. It emerges in some years but not in others. And it relates to industries rather than to items.

Applicability of the results: Our study clearly excludes non-industrial firms. Also those which are too small for being collected into the Micro-EXSTAT data-base. We believe that purely commercial firms or mixed ones will not depart from the general lognormal pattern. In the case of small or very small firms, we think that base-line effects and the existence of samples with extremely similar cases (groups with very small variance) can change the results substantially.

In general, our results seem to suggest that the cross-sectional behaviour of accounting items is best analysed in relative or scale-independent space.

Lognormality and the existing research: The finding of a regularity in distributions of items contrasts with the existing scenario. Hitherto no strong regularity or consistency could be expected when examining accounting data. This is because only ratios have been explored. Ratio outputs are the residuals of two-variate relations. They are exposed to the effect of external constraints like accounting identities. And they also reflect internal mechanisms of the firm. This makes ratios different from one another.

There is no reason to expect such a variety when looking into items. The variability in items seems to be mainly determined by the different sizes of the firms gathered in the same sample. Items are mainly a given proportion of the size of the firm.

Lognormality and homogeneous groups: Our results also suggest that for a given industry strict lognormality depends on the homogeneity of such industry. Most of the observed severe departures from lognormality are due to the existence of clusters inside an industrial group. It is interesting to study one by one the 20 "bad cases" described in appendix A. Apart from a few samples with errors or very strange cases, most of these samples are not one unique group of firms. There are two or three distinct groups, at least when considering Sales, Wages or the number of employees.

We also examined the lognormality of samples when gathering two or three industrial groups in the same sample. The proportion of strong departures from lognormality ($P < 0.001$) increases very significantly. We conclude that, for small samples, lognormality is conditional on the homogeneity of the industry.

This conclusion is trivial. It arises in all statistical phenomena. Whenever we hypothesize a generative mechanism for explaining a given distribution, we must isolate homogeneous sets for testing such a hypothesis. Two factors should be considered as conditioning an observed distribution: The underlying generative mechanism, which determines the population's distribution. And the existence of groups or other sources of non-homogeneity in the defined sample.

Our results seem to suggest that the underlying generative mechanism determining the distribution of items has a trend towards lognormality, even when the particular sampling yields non-homogeneous groups. For statistical modelling purposes the real important feature to consider is the underlying distribution. The particular grouping is an accident and can be accounted for.

Another interesting result is the lognormality of samples obtained by gathering all groups together. Since these samples contain 14 different groups, we conclude that a second level of homogeneity is reached when the number of groups is high. One single group, if homogeneous, yields homogeneous samples. Two or three groups, each of them homogeneous, can yield samples which are severely non-homogeneous. But 14 groups in the same sample approach a random effects model rather than a fixed effects one. The grouping is near continuity, yielding a new, overall, homogeneity. This overall homogeneity is the one expected when sampling firms at random according to a new single attribute — the selected firms being industries.

Lognormality and accumulative phenomena: We now focus on lognormality as a quality, that is, as a trend determined by a generative mechanism, regardless of the actual distribution observed in a particular sample.

Lognormality should not be considered as a strange or unexpected quality. It can be found in many accumulation processes, like growth. The weight and height of children with different ages is lognormal. Income distribution and many other processes in Economics also belong to the same class [65].

Any stochastic accumulation, that is, the growth proportional to the size already attained, leads to cross-sectional samples which are lognormal or belonging to the same class of skewed distributions.

If a particular item grows, for several firms, at different, Gaussian paces the final aspect of such a sample is the one of a lognormal distribution. We shall explore this mechanism in chapter 3. At the moment let us retain this simple, known, fact. Gaussian accumulations lead to lognormal final realizations.

Lognormality and cross-section: The last remarks are useful. But they can lead to a misleading interpretation of this study and its method.

A major difficulty in understanding lognormality and other regularities found in this study stems from picturing the problems discussed here as concerning one unique firm and its internal mechanisms instead of a cross-sectional sample. For example, a typical reasoning would be: Sales is likely to be lognormal since it is an accumulation. But Dividends is very unlikely to be lognormal. Dividends are dictated by considerations which have nothing to do with accumulations.

All the above reasoning is about the internal behaviour of the firm as perceived by using ratios. There is nothing on it which can be surely related to a cross-sectional sampling of items. Instead of the item Dividends for many firms of very different sizes, this reasoning is picturing a different thing — Dividends per share. The item Dividends is itself proportional to the item Net Worth — lognormal in our samples. Firms with a large Net Worth will have a proportionally large Dividends item. Small firms having a small Net Worth will have a proportionally small Dividends item — even if they pay large dividends. Dividends, the item, is likely to reflect — in a given proportion — the size of the firm.

The item Sales is not particularly more lognormal than any other item because of being an accumulation. Sales is lognormal because the growth of the firm as a whole is itself an accumulation — a trivial finding —. Sales just reflects, on average, a given proportion of the size of the firm.

Cross-sectional samples gather firms of very different sizes at the same moment in time (ideally). Cross-sectional studies examine the joint behaviour of many firms. Common features are taken as statistical regularities. A cross-sectional study is not about any particular firm. The internal behaviour of firms is not contemplated as such — it can emerge in the residuals though. For example, firms which pay large dividends will have large residuals —.

Only a joint behaviour creates regularities. And cross-sectional regularities mean something common to many different objects.

Lognormality and the developments presented in next chapters: The next three chapters explore lognormality and other related findings. We expect to provide a coherent view enabling the statistical modelling of our data. Lognormality in items emerges as a trivial result. And as a very useful result too. It is trivial because it simply means that items reflect mainly size. And it is useful because it means that cross-sectional samples of accounting items tend to be Gaussian in a scale-independent space.

The size of firms is not bounded by any central trend. Firms are free to be as large as they manage to, inside an economy. A cross-section of many firms is expected to exhibit the same kind of pattern governing other well known unbounded variables like income, wealth or the size of cities.

Our accounting items seem to reflect a given proportion of size. If Sales is taken as a measure of size, then Inventory is expected to reflect, say, 1/3 of it. And Dividends, in average, 1/20 of it. Lognormality generates constant ratios.

But lognormality is a much broader condition. It expresses the statistical or expected proportionality of random effects, not just a strict proportionality like ratios do. This topic is worth exploring. It leads to models which are beyond ratios.

Chapter 2

The Multi-Variate Characterization of Accounting Items

What are the immediate conclusions to extract from the lognormal nature of the observed items? Can lognormality be generalised to a multi-variate context?

In this chapter we first point out that the sole consideration of lognormality is enough to account for the persistent emergence of outliers referred to in the literature. The also mentioned heteroscedasticity of residuals is then discussed. We show that the direct Least-Squares modelling of lognormal data is, in general, not the most adequate procedure. Other models ought to be developed. Specifically, we study the consequences of using ordinary or weighted regressions and the usefulness of trimming lognormal tails.

This chapter then turns to the multi-variate behaviour of items. First, their Gaussian parameters in log space are examined for regularities. The mean values and standard deviations for different industries display some easy to interpret characteristics. Next we observe the variance and co-variance matrices of log items by industry. All the observed items have in common a strong portion of their variability. They are well described as a unique process with some amount of particular randomness superimposed.

Finally, we discuss items obtained by subtraction of other items. We explain why positive differences of two items maintain lognormality. Then we suggest some procedures to apply when such subtraction yields negative cases.

The meaning of the log transformation: We often use the term "space" with a qualifier. For example, we refer to the rotated space or to the log space. Our goal is to emphasize the fact that a given set of variables have been jointly transformed in a well known and consistent way, thus defining a formal system characterized by a set of entities and the corresponding axioms relating to them. For example, the log space is the set of log transformed items.

The accounting literature is cautious about transformations. They are seen as a means of massaging data. We should be alive to the fact that a log transformation cannot be considered just as a manipulation of values to make them more tractable. The log function has a precise meaning. By using logs we select a particular way of looking into the data. We switch to a proportional or scale-independent space. The next quotation shows a pitfall resulting precisely from considering logs as a simple resource to render some sample more tractable. It has been frequently quoted in accounting papers to reinforce warnings about the dangers of using transformations.

> In the case of the log transformation there is also an implicit assumption being accepted where such a transformation is employed. That is, the transformed variables give less weight to equal percentage changes in a variable where the values are large than when they are smaller. If, for example, the variable being transformed was firm size the implications would be that one does not believe that there is as much difference between a \$1 billion and a \$2 billion size firms as there is between a \$1 million and a \$2 million size firms. The percentage difference in the log will be greater in the latter than in the former case. [34]

This comment allows us to illustrate the meaning of switching to a proportional space. First we look into the figures.

Log of one million	=	6	Log of one billion	=	9
Log of two millions	=	6.301	Log of two billions	=	9.301
The difference is		0.301	The difference is		0.301

Logs yield similar differences whenever the ranges are proportional, that is, when they are similar except for scale. The variations of $\log X$ are the relative or proportional variations of X.

By working with log variates we are precisely giving the same weight to proportionally similar changes. We accept that a difference or growth from one million to two millions is exactly as impressive as a growth or difference from one billion to two billions.

In log space we no longer compare firms in terms of real (absolute) size but in terms of relative size. We make differences independent of scale or measure.

Of course, we should avoid applying proportions to log distances. It would be as if we were building proportions of proportions. This is the pitfall, a very reasonable one if we consider the log transformation just as a bit of massaging of the data.

In appendix B we briefly discuss the possible distortions resulting from mixing-up transformations in input variables for statistical models. For consistently transformed input data we don't see what kind of unexpected distortions could arise by working in the proportional space.

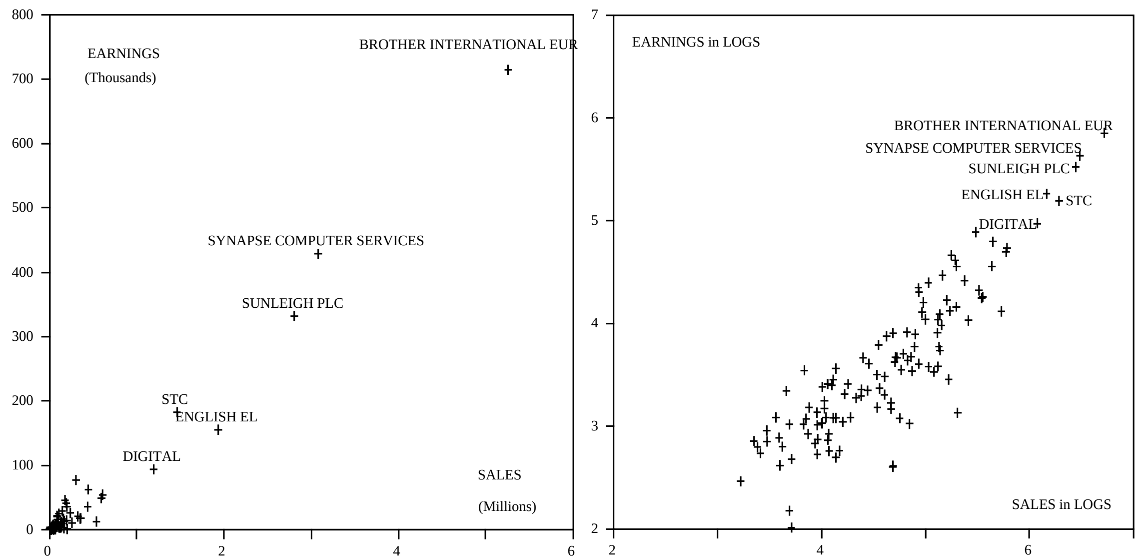

Figure 8: The awkward aspect of a lognormal spread (left) and the homogeneous one of the same data in log space (right). Electronics, 1986.

Before finishing this introduction to logs, let us recall that the decimal log basis allows us to interpret any value in log space as the number of digits of the non-transformed data. In log space, a value of 4 means a number with four digits. A value of 6 means a figure larger than one million and smaller than ten millions.

2.1 The Variability of Accounting Data

This section comments on two problems widely discussed in the literature: The existence of outliers and the heteroscedasticity emerging when using additive Least-Squares algorithms for modelling accounting relations.

2.1.1 Lognormality and Outliers

The presence of outliers in the residuals of the ratio model is consistent with the lognormality of items. In a first approximation, if items are lognormal, ratios should be positively skewed (in chapter 4 we explain why some ratios are Gaussian or even negatively skewed). Lognormal variates are likely to exhibit severe skewness of the kind easily taken as outliers. For increasing values of the variate we observe a strong increase in variance and an also strong decrease in the density of cases. The variance spreads proportionally to the square of the variate. The density of cases decreases in proportion too. The coefficient of variation (the standard deviation expressed as a proportion of the expected value of the deviate) remains constant.

As a consequence, lognormal data draw shapes in which many cases concentrate in a small region and very few of them spread out along a large range — as in figure 8 on the left (page 31). It is easy to take these few extreme values as outliers: they lie out. However, they are not real outliers. Their behaviour is homogeneous providing the adequate distribution is assumed.

Outliers require an assumption about the underlying distribution: The notion of outlier is entirely dependent on a previous assumption. In order for a case to be an outlier there must be a previous acceptance of a particular distribution as the one of the population.

When the literature refers to a case saying that it is an outlier, the underlying distribution is the Gaussian. But since accounting items are not Gaussian, it turns out that most of the cases referred to as outliers are not outliers.

The proper space for checking the existence of outliers is indeed the logarithmic or relative one. Figure 8 (page 31) shows an example of the adequacy of the logarithmic space for observing accounting data. On the left we can see a scatter-plot of Earnings versus Sales for the Electronics industry in 1986. Hardly anything can be sorted out. Apart from the biggest companies, the remaining ones (about 140) are concentrated in a small region at the bottom left of the plot.

When drawing the same plot in log space each case becomes distinctly separable and the two-variate distribution emerges as homogeneous. Even more interesting, a small non-linear relation between the two components is now visible. This non-linear relation turns out to be important for the understanding of non-proportionality in ratios. It became visible because of the adequacy between log space and accounting features.

Along this study we shall see that many other pieces of evidence hitherto hidden from direct observation become visible in log space.

Lognormality and a very small scatter: Notice that the above description of the characteristics of the lognormal distribution should not be taken as a general rule. Perfectly lognormal samples can have small skewness when its standard deviation is very small too. This is the case for some residuals of ratios when the variances and expected values of their log components are similar.

It is generally accepted that coefficients of variation smaller than 0.25 denote distributions which can be approached by the Gaussian one. Such exceptional cases cannot explain why a few ratios are near normality. The literature on the distribution of ratios uses the ratio output, not the ratio residual, to assess their distributions. When using ratio outputs the spread will seldom be small enough: In lognormal variates the spread grows with the mean.

2.1.2 The Heteroscedasticity in Models With Additive Residuals

There is a well known claim for regressions to be used instead of ratios. Non-proportional relations between the ratio components are the basis of such a suggestion. On the whole the discussion resulting from this claim was very revealing. It drew the attention of researchers for accounting

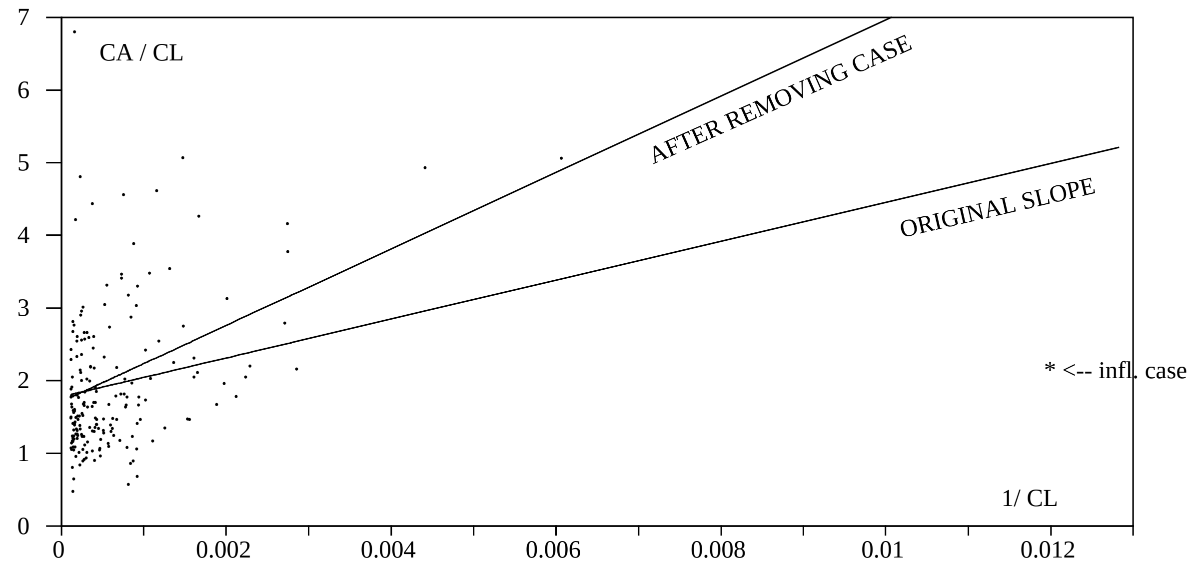

Figure 9: One single influential point in a weighted Least-Squares regression makes the resulting slope change in a significant way. The original data is lognormal (CA vs. CL, 1984, all groups).

items instead of just ratios. And it introduced a strong aspiration towards more accurate models. Berry and Nix (1991) [11] review this research. However, given the lognormal nature of accounting items, the use of regressions is inadequate.

In fact, lognormality implies input vectors which are high-leverage cases. The reason is the same as for the emergence of outliers. Notice that high-leverage cases need not to be influential. But they are likely to be. There is a tendency for high-leverage cases to become influential.

Leverage and Influence: The notion of leverage concerns an assumed model, not an assumed distribution. Leverage regards inputs, not outcomes. Leverage cases are those which are far away from the rest of the input vectors.

Models suffer distortions if one or two input vectors are influential. An influential case monopolizes the fit of a model. When it is excluded the parameters are significantly different from when it is present. This happens because influential cases manage to have small residuals when they should have large ones, at expenses of the whole model. The quadratic nature of Least-Squares algorithms avoids any large residual by modifying the fitted model. Figure 9 shows an example.

In regressions using accounting items as input variables, a large firm will be a leverage case, likely to become influential, just by being large. This is a consequence of the lognormality of items. In fact, when a lognormal distribution is taken as Gaussian, the outliers are always in its tail — the largest firms. In the literature this problem is generally referred to as the non-acceptable heteroscedasticity of accounting data. It is correct to address the lognormal scatter as a case of non-homogeneous variance. But this is a too general way of putting it. There are many kinds of

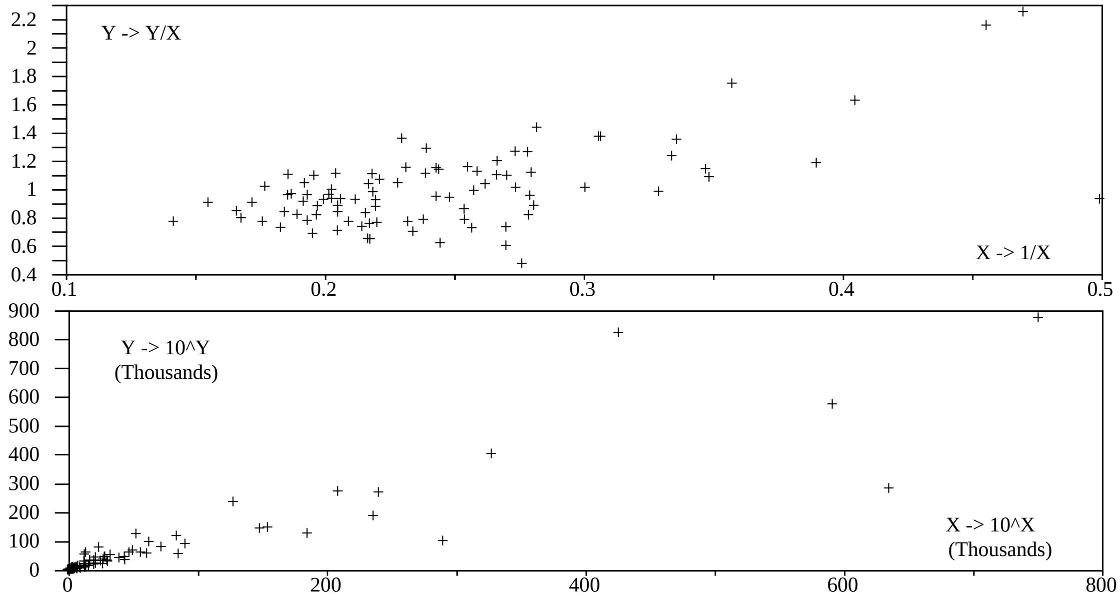

Figure 10: Typical two-variate distributions requiring weighted Least-Squares (above) and Log transformation (below).

heteroscedasticity. In order to cope with each one, some knowledge about its nature is required. Recipes adequate for one particular form will not work in different situations. Next we comment on one of such recipes widely used in accounting research.

The use of weighted Least-Squares: Discussion. Since the direct use of accounting items as inputs for statistical models yields unacceptable heteroscedasticity, weighted Least-Squares has been called upon. If items are lognormal this seems to be the wrong recipe to apply.

A weighted regression uses ratios to control for increasing variance on the predicted variable. Instead of the usual $y_j = A + B \times x_j + \varepsilon_j$, a weighted least squares model is

$$\frac{y_j}{x_j} = \frac{A}{x_j} + B + \frac{\varepsilon_j}{x_j}$$

For variances increasing with x this procedure should allow a Least-Squares modelling. But in lognormal deviates it is the standard deviation, not the variance, which increases with x in average. The variance grows with the square of x. We can easily understand the meaning of this distinction by observing figure 10. The scatter-plot above shows a two-variate relation ideal for a weighted Least-Squares transformation. In fact, this sample has been obtained from a perfectly Gaussian set by applying the inverse of a weighting transformation. Below, the same Gaussian set but after applying the anti-logarithmic transformation instead of the inverse of a weighting. The aspect of both sets is typical of data requiring weighting (above) or logs (below).

Giving lognormality, the best thing to do is recognizing its presence. But in case this is not

Ordinary Least-Squares			Weighted Least-Squares			OLS in log space		
Name	Rank	Cook D.	Name	Rank	Cook D.	Name	Rank	Cook D.
G. E.	1	12.185	TELFORD	129	7.8906	NATIONAL	17	0.077
STC	4	1.4857	MISYS	142	0.1665	G. E.	1	0.0734
IBM UK	2	0.9855	M.M.T.	145	0.1522	POLYTECH.	126	0.0577
ENG. EL.	3	0.2508	FORWARD	139	0.0796	IBM UK	2	0.053
STC C.	6	0.2138	KLARK-TEK	137	0.0577	BELL & H.	59	0.0398
DIGITAL	7	0.0626	HEADLAND	135	0.0531	CASIO	55	0.0391
AMSTRAD	14	0.0331	AMS IND.	136	0.0487	AMS IND.	136	0.036
UNISYS	27	0.0135	SOUNDTRA.	138	0.0442	ZYGAL DY.	125	0.035
FERRANTI	10	0.0128	AVESCO	131	0.0395	ENG. EL.	3	0.034

Table 9: The largest Cook Distances for OLS, WLS and LOG. Electronics, 1986.

achieved, the next best thing to do is to use tests of heteroscedasticity enabling not just the identification of its presence but also the assessment of its degree. Berry and Nix [11] recommend one such test since it "crucially gives an indication not only of the presence of heteroscedasticity but also of the power of the transformation needed to remove it".

An example: The Cook Distance. The Cook Distance [25] can be used to assess how influential cases are. This statistic measures influence: The effective degree in which an input commands the whole fit, for a particular sample.

As an example of the correlation between influential cases and the size of the firm, we selected the Electronics industry, 1986. Three different regressions — OLS, WLS and LOG (a regression in log space) were compared. Sales was the input and EBIT the outcome.

Tolerable Cook Distances should not exceed 1. Values larger than the unit mean a fit monopolized by the case in which it occurs. Figure 9 on page 33 shows an example of one simple case distorting the whole fit in a weighted regression using accounting items. Table 9 shows the firms which were traced as most influential in each regression. The column labeled "rank" signals the ranking of each firm according to size. In this column, 1 means the largest firm, and so on. This ranking was obtained from a size proxy developed in chapter 5.

OLS has two firms which are influential. WLS has one. When using OLS the most influential firms tend to be the largest. In the case of WLS they tend to be the smallest. The largest Cook Distance in log space is far below the maximum value emerging when using OLS or WLS. Figure 11 on page 36 shows the Cook Distances of the entire sample when compared with the size of the firm. There is a clear correlation between the Cook Distance and the extreme sizes, both in OLS regressions and in WLS. In the first case, the largest firms tend to be the most influential. In the second, the smallest are candidates for becoming influential. There is no such a correlation when regressing in log space.

Notice that, when using WLS instead of OLS, some improvements are observed. WLS makes the spread of the resulting variates smaller. And the skewness, in lognormal variates, is very dependent on the amount of the spread. Also, a fortunate coincidence can make some accounting identity

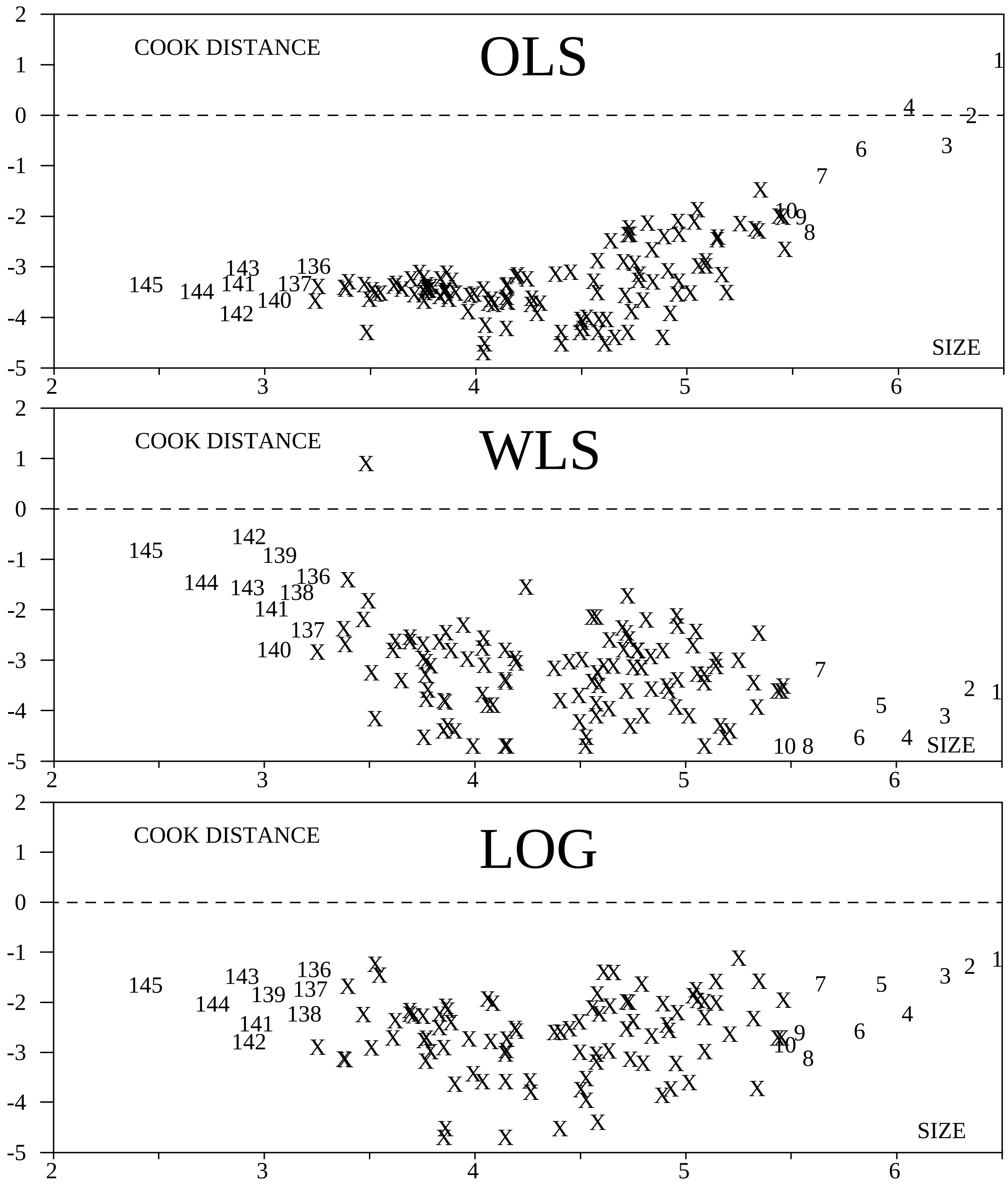

Figure 11: The log of Cook Distances against log s. Values obtained when $Sales$ was used to explain $EBIT$ in three different regressions. Electronics, 1986. The numbers used as marks denote ranking according to size.

constrain the new variables so that their quotient would not be allowed to have long tails: In the case of industries, it is unlikely that firms could exhibit Earnings larger than sales and weighting would work.

2.1.3 Scale-Independence and Trimming

Another point the lognormal trend of items elucidates is why it seems so unfruitful to trim outliers.

Lognormal multi-variate distributions exhibit self-similarity of features across scales. This directly stems from scale-independence. Any shape which holds for billions also holds for thousands.

The shape like a "<" — typical of a lognormal two-variate scatter of very correlated deviates — will never change across scales. Such a shape, along with the correspondent gradient in the density of cases, is continuously generating influential cases across scales.

As a consequence, there is little point in excluding large firms from the sample for obtaining a more homogeneous set. If we exclude the largest cases from the sample new cases will emerge as outliers. A different way of viewing this is noticing that a trimming would be equivalent to a reduction in the scale in which we are observing the data. And, as the phenomenon commanding the emergence of influential cases holds in different scales, new outliers will appear again and again — until the overall variance becomes so small that normality can be taken as a good approximation. Figure 12 on page 38 illustrates this mechanism.

In the above example (table 9 on page 35), if we measure the Cook Distance associated with each case after excluding the two influential firms (G.E. and STC), we get three new cases with a non-acceptable weight in the regression: SUNLEIGH PLC (with a Cook Distance of 1.6), ENGLISH ELECTRIC (1.9) and BROTHER INTERNATIONAL (19.8). The new situation is worse than before trimming.

If we exclude also these three firms, SYNAPSE COMPUTER SERVICES emerges with a new Cook Distance of 80.5. And excluding also this firm, MISYS PLC suddenly appears as having a new Cook Distance of 0.8. Not too bad. But much higher than the highest values observed in LOG.

Notice that DIGITAL, despite being amongst the largest firms and therefore a leverage case, didn't become influential. This is because leverage and influence are different concepts. An influential case must be a leverage case. But it is possible for a leverage case not to become influential.

Conclusion: The heteroscedasticity typical of proportional phenomena is well known and documented. For example, Snedecor and Cochran [119] observe: "Logarithms are used to stabilize the standard deviation if it varies directly with the mean, that is, if the coefficient of variation is constant. When the effects are proportional rather than additive, the log transformation brings about both additivity of effects and equality of variance." The log transformation solves at once two problems accountants were used to consider separately: The non-homogeneous variance and the emergence of outliers.

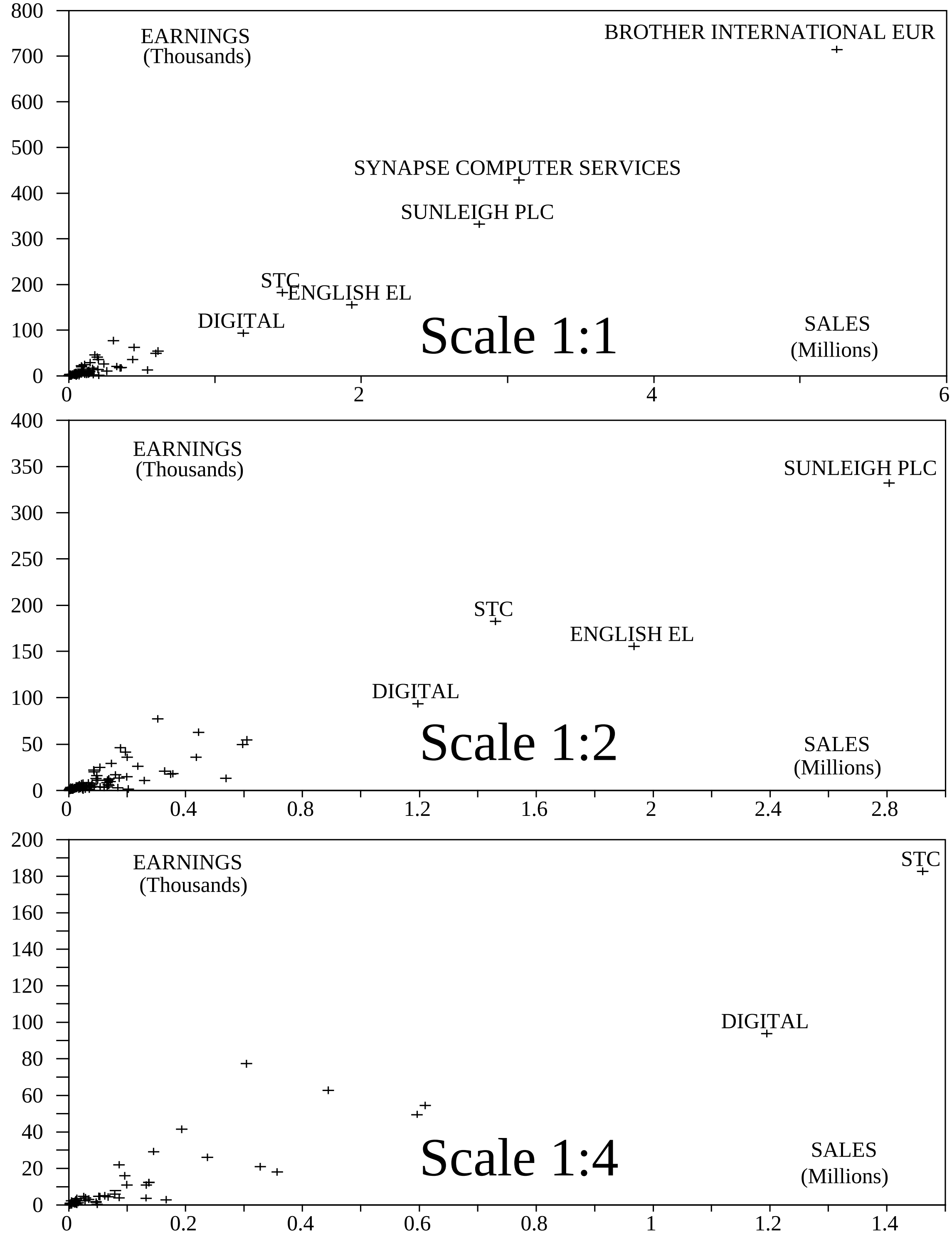

Figure 12: When trimming outliers caused by a lognormal behaviour, new outliers emerge across scales. Electronics, 1986.

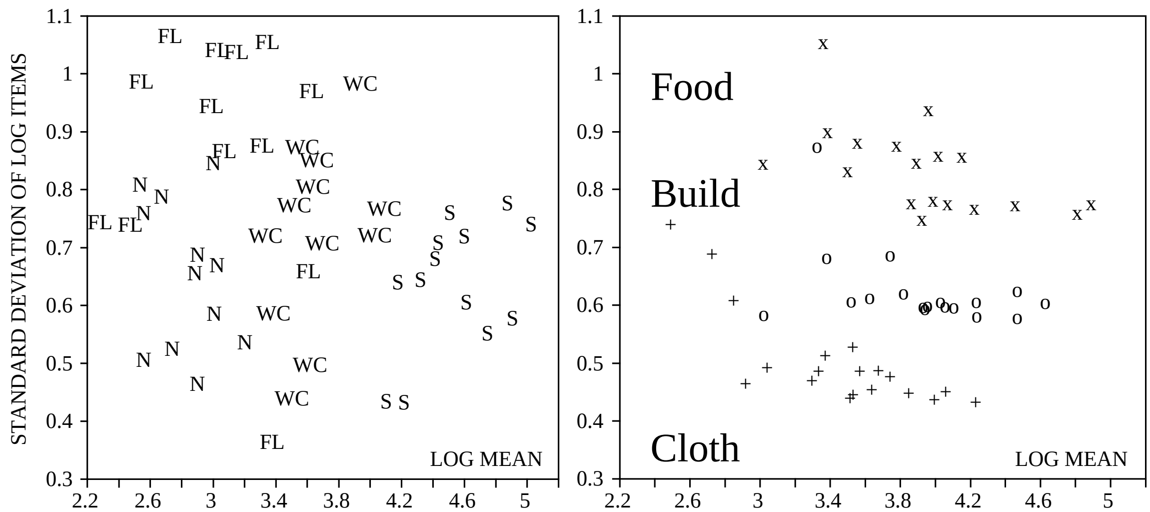

Figure 13: Variables (left) and Groups (right) in parametric space. 1983 data.

2.2 Regularities in Parametric Space

Once accepted the lognormal quality of the observed items it is possible to use the Gaussian parameterization for describing samples. In the simplest case a mean value and standard deviation of one log item will uniquely describe a set of observations. For samples containing more than one observation per case, a vector of means and a matrix of variances and co-variances is required.

In this section we describe the behaviour of the Gaussian parameters of the observed items for different groups. We show empirical evidence on the existence of a common source of variability.

Because this section is all about log space we shall not refer to it explicitly in all occasions. Whenever a statistic, a value or a set of cases are referred to, it is assumed that the observations have been transformed and that the used log base is the decimal one.

The sample is the same as the one used in chapter 1. It contains 14 industrial groups. For each one, 18 items have been observed. This sample can be replicated for a period of five years (1983-1987). Tables 1 on page 5 and 2 on page 6 describe their contents in detail.

2.2.1 Relative Spread and Industrial Groups

In order to observe the mean values and standard deviation of all such cases it is practical to build scatter-plots with means on one axis and standard deviations on the other one. By doing so we visualize the parametric space. Each sample is thus represented by a point.

The described plot allows the searching of common patterns in sets of samples. For example, we can view several different items belonging to the same industrial group and year. Or gather only one particular item and compare three or four industries.

In figure 13 (page 39) we show — on the left — the positions occupied in parametric space by four items. On the right, the position occupied by items belonging to three industries. They form characteristic clusters. All cases are from the 1983 sample.

In figure 14 (page 41) we compare the position of clusters formed in the parametric space by a set of items for six different industrial groups. The most apparent feature is the relation between the smallest standard deviations observed and every industrial group.

The ground value for the relative spread: Groups tend to exhibit similar ground or minimum standard deviations. If we form a set with standard deviations of several items all belonging to the same industry, we notice that these values are seldom found below a ground value which is typical of its group. For example, the standard deviation of items from the Food industry is never smaller than 0.77. For Clothing this value is 0.4 and for Building Materials is 0.6 (see figure 14 on page 41).

The meaning of the log standard deviation: In log space, distances from any case to the mean are no longer real distances. For example, a log displacement of 0.4, — a fairly common value for the log standard deviation observed in some industries — means that the displaced case has been multiplied or divided by 2.5, the anti-logarithm of 0.4. Hence, we may say that a log standard deviation of 0.4 is equivalent to a multiplicative spread of 2.5. Or that each unit of spread in log space measures a scatter of 150%. The log standard deviation is a dimensionless statistic. It is the log of the coefficient of variation of the sample. The spread is assessed as a proportion of the mean.

The fact that distances to the mean are relative will not eliminate the effect of the size of firms. Large standard deviations denote a group for which both large and small firms are possible. Small standard deviations denote a group with firms of uniform size.

2.2.2 Variance and Co-Variance: The Strong Common Variability

We now turn our attention to the multi-variate behaviour of items in log space. The variance and co-variance matrices of our set of items also exhibit some clear features:

- Inside an industrial group variances and co-variances of different items are very similar.
- Strong differences can be found between the variance and co-variance matrices of different industries.
- Even when avoiding linear combinations the variance and co-variance matrices of accounting items are almost singular.
- For samples of significant size no negative co-variances can be found. Nor zero co-variances. As in the case of standard deviations, there is a clear ground value characteristic of the group and both variance and co-variances lie above it.

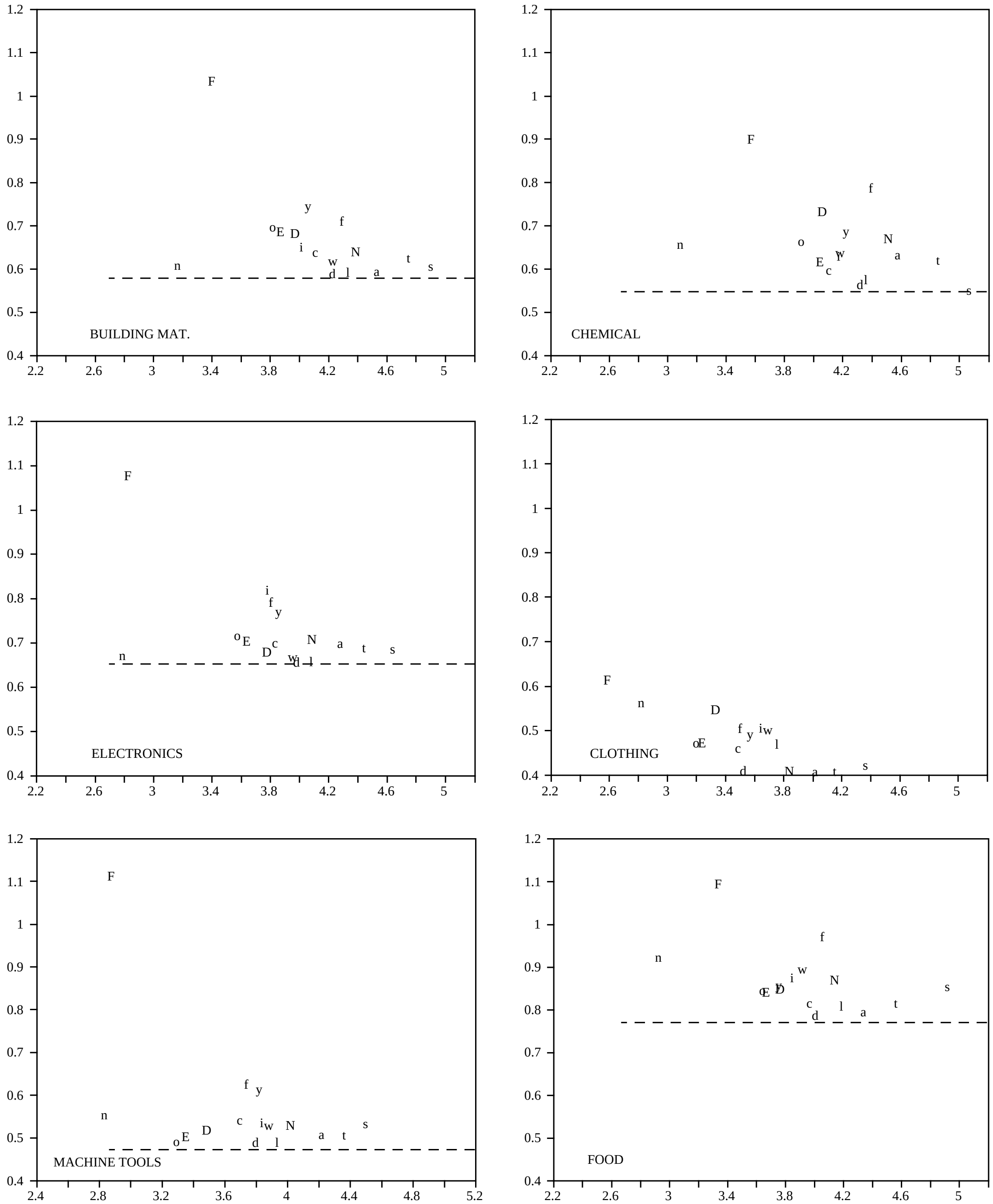

Figure 14: Characteristic shapes of 6 industrial groups in the parametric space. X-axis, the mean. Y-axis, the standard deviation. Marks identify similar items across different plots. For example, “d” is Debtors, “D” is Long Term Debt, “E” is EBIT, “f” is Fixed Assets, “F” is Gross Funds From Operations, “i” is Inventory, “n” is the Number of Employees, “N” is Net Worth, “o” is Operating Profit, “s” is Sales, “t” is Total Assets, “w” is Wages and “y” is Working Capital.

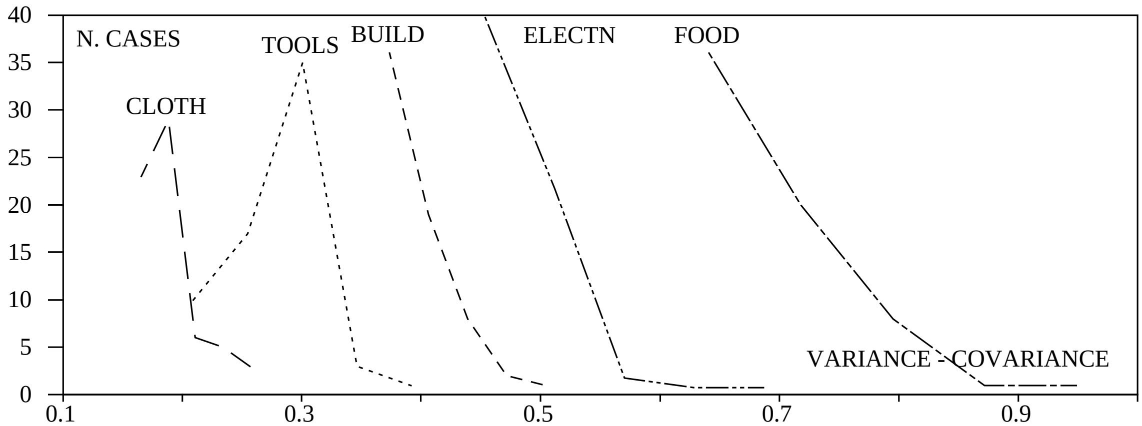

Figure 15: Frequency distributions of variances and co-variances for five groups.

	S	NW	W	I	CA	FA	EBI	S	NW	W	I	CA	FA	EBI
S	0.35							0.21						
NW	0.32	0.39				PAP		0.22	0.26				LEA	
W	0.35	0.36	0.41					0.19	0.21	0.18				
I	0.33	0.34	0.38	0.40				0.19	0.21	0.18	0.19			
CA	0.33	0.32	0.35	0.35	0.34			0.21	0.24	0.20	0.20	0.23		
FA	0.32	0.36	0.37	0.34	0.33	0.40		0.19	0.21	0.18	0.17	0.19	0.20	
EBIT	0.35	0.35	0.36	0.35	0.34	0.35	0.41	0.26	0.29	0.24	0.23	0.27	0.24	0.28
S	0.17							0.55						
NW	0.17	0.20				CLO		0.57	0.73				FOO	
W	0.18	0.18	0.23					0.55	0.64	0.64				
I	0.19	0.20	0.20	0.23				0.59	0.68	0.63	0.73			
CA	0.17	0.18	0.18	0.20	0.19			0.56	0.63	0.58	0.65	0.61		
FA	0.17	0.19	0.21	0.20	0.17	0.26		0.57	0.69	0.65	0.65	0.61	0.71	
EBIT	0.17	0.17	0.18	0.18	0.17	0.19	0.31	0.59	0.71	0.64	0.70	0.66	0.69	0.73

Table 10: Four Σ matrices denoting the homogeneity of variance and co-variance inside industrial groups. Data from 1983.

A graphical description of the spread and co-variance of the observed items is obtained with a frequency distribution of the components of the variance and co-variance matrix. Figure 15 shows a few of them for different groups. They don't overlap and they are J-shaped denoting strong differences between industries and the existence of ground values.

Table 10 (page 42) shows four industry's variance and co-variance matrices (usually referred to as Σ). They describe the features we highlight. In appendix A several other matrices are displayed.

We also display in appendix A the minimum values for elements of the variance and co-variance matrices belonging to each one of the 14 industrial groups observed during a five-years period.

On variance and co-variance grounds the less homogeneous items are Long Term Debt, Inventory and Fixed Assets. In general, such items have higher variances than the usual for the industry. Co-variances of Inventory and Fixed Assets with other log items are not especially different from others. Debt shows differences also in co-variances.

	DB	WC	EB	FA	FL	NW	I	D	C	CL	S	W	CA	mean
METAL	0.40	0.81	0.80	0.59	0.78	0.77	0.81	0.83	0.79	0.78	0.65	0.81	0.90	0.75
CLOTH	0.32	0.56	0.54	0.76	0.67	0.84	0.89	0.79	0.86	0.85	0.92	0.87	0.92	0.75
TOOLS	0.31	0.73	0.70	0.83	0.69	0.78	0.88	0.82	0.92	0.88	0.94	0.95	0.95	0.80
WOOL	0.41	0.78	0.80	0.81	0.86	0.85	0.77	0.88	0.85	0.82	0.91	0.86	0.91	0.81
PLANT	0.57	0.67	0.67	0.82	0.77	0.80	0.88	0.93	0.90	0.93	0.96	0.96	0.95	0.83
PAPER	0.46	0.71	0.84	0.77	0.86	0.83	0.88	0.86	0.91	0.94	0.91	0.92	0.96	0.83
ELECT	0.49	0.71	0.79	0.74	0.84	0.85	0.90	0.92	0.95	0.95	0.93	0.93	0.97	0.84
CHEM	0.62	0.76	0.84	0.84	0.84	0.89	0.80	0.88	0.90	0.92	0.87	0.95	0.92	0.85
ELTN	0.53	0.84	0.80	0.81	0.86	0.92	0.79	0.93	0.90	0.93	0.93	0.90	0.93	0.85
BUILD	0.62	0.74	0.82	0.87	0.86	0.88	0.92	0.96	0.90	0.95	0.97	0.97	0.97	0.88
FOOD	0.73	0.77	0.87	0.86	0.90	0.89	0.90	0.93	0.93	0.94	0.91	0.94	0.94	0.88
MOTOR	0.76	0.80	0.82	0.92	0.88	0.92	0.94	0.96	0.91	0.95	0.96	0.96	0.90	0.90
MISC	0.77	0.78	0.90	0.81	0.92	0.93	0.94	0.94	0.94	0.93	0.96	0.90	0.95	0.90
LEATH	0.68	0.90	0.93	0.87	0.94	0.95	0.97	0.85	0.94	0.96	0.98	0.93	0.97	0.91
mean	0.55	0.75	0.79	0.81	0.83	0.86	0.88	0.89	0.90	0.91	0.91	0.92	0.94	

Table 11: The R^2 (mean of the five years observed) of regressions in which s, a proxy for the size effect, explains 13 items. Rows are industries and columns are items.

Discussion: The outlined features are not usual. They denote a strong regularity.

Whenever similarity between variance and co-variance is observed we know that there is a common source of variability influencing several variables. And the non-existence of negative or null co-variances, along with the ground value, make us realize that we are not really looking into several sources of variability but into a unique one with some smaller variability superimposed. Matrices such those of table 10 are only possible when there is a preeminent source of variability common to all the displayed items.

Any component of these Σ matrices could explain more than 90% of the observed multivariate spread. Many items share up to 94%. of their variability with others. It seems as if accounting items were, in log space, just the same unique process with a bit of particular variability superimposed.

Such a feature of log accounting items is expected. The log transformation does not control for size. Logs make differences in size relative but the effect of size remains. For example, if the Fixed Assets of three firms are worth one thousand, ten thousand and one million pounds, then in log space they will be denoted by the values 3, 4, and 6.

The observed items clearly reflect a common source of variability we identify as the effect of the relative size of each firm in samples gathering many, differently-sized firms. In cross-section, the first and most important source of variability impinging upon different items is the effect of size.

Only after accounting for the size effect is it possible to assess the variability unique to items. The ground value also suggests that such unique variability will show positive or zero correlations with the unique variability of other items. The negative correlations will not be frequent.

Empirical evidence on a strong common effect: A more systematic way of looking into the same phenomenon consists of examining slopes instead of co-variances. Co-variances are non-standardized slopes. When divided by the variance of one of the components they yield a slope.

In order to provide solid evidence on the existence of a strong source of variability common to the observed items, we first created a variate, s, supposed to reflect the common variability present in all the observed items. This can be done in several ways. Ours will be explained later on (section 5.1). Using this variate as the predictor we formed regressions in log space. Each log item was explained in terms of s ($\log x_j = a + b \times \log s_j + \varepsilon_j$). The results show two regularities:

- The common effect explains most of the variability of items.
- All the slopes scatter around and near 1.

Appendix A contains the detailed results of this experiment by industry and year. Also section 5.1, and specifically table 19 on page 115 provides evidence on this subject for all groups together. Here, we present two condensed tables (11 and 12) in which the obtained slopes and proportion of explained variability are displayed by industry and by item. Each value is a mean of five years. Both tables are sorted in ascending order of their marginal contents. Therefore, by reading the headings it is possible to become aware of which items or groups approach the largest values.

Large R^2: First we noticed that the obtained R^2 are very large and very similar. They range from 94% to 74% for items like Sales, Inventory, Debtors, Creditors, Current Assets and Liabilities, Wages, Net Worth, Gross Funds From Operations or Earnings. The results are consistent for the usual period of five years. On the whole, after examining 920 different samples, 90% of them have a R^2 larger than 0.78.

Current Assets is the most explained by s and Debt the least. Fixed Assets, Earnings and Working Capital are below the usual. Long Term Debt is clearly different from the other items. It shows a proportion of explained variability ranging from 30% to 80%, much smaller than any other. However, even in the case of Debt it is not possible to accept the hypothesis of independence from size. The minimum R^2 of 30% represents a correlation of 0.55 with size. This correlation is not negligible. For the examined samples it is significant.

The conclusion is that it is possible to create a unique variate, s, able to account for most of the variability observed in our set of log items.

Industries are similar in terms of explained variability. Only Metallurgy shows consistently values below the usual. Items, on the contrary, display a scale of different explained variability.

It is the contrast with the other items, exhibiting in most of the cases correlations with size larger than 0.9, which makes Debt a particular case. It would be interesting to know if this less strong correlation with size can be found in any of the items we didn't examine.

Slopes near the unit: Also the slopes of regressions in which log items are explained by s, an effect common to them all, exhibit a regular behaviour. They remain close to 1.

	D	WC	CL	NW	CA	S	C	W	EB	FA	FL	I	DB	mean
CLOTH	0.91	0.83	1.04	0.96	0.96	0.99	1.02	1.07	0.85	1.07	0.99	1.13	0.97	0.98
ELECT	0.98	0.93	1.06	0.87	1.01	1.03	1.12	1.05	0.97	0.93	0.95	1.01	0.97	0.99
PLANT	0.97	0.92	1.02	0.87	0.94	1.10	1.04	1.04	1.02	0.90	0.97	1.12	1.07	1.00
ELTN	0.96	1.08	0.95	1.04	1.02	0.99	1.00	1.00	0.93	1.04	0.96	1.07	0.95	1.00
TOOLS	0.98	1.14	0.91	0.93	1.01	1.11	1.02	1.02	1.06	0.92	1.04	1.02	0.93	1.01
LEATH	0.95	0.86	1.01	1.06	1.00	1.03	0.99	0.91	1.12	1.18	1.06	1.07	0.87	1.01
WOOL	0.96	1.02	1.02	1.03	1.08	0.93	0.99	0.98	1.16	0.97	1.11	0.95	0.95	1.01
PAPER	1.05	0.96	1.00	0.90	0.97	1.06	1.03	1.04	1.03	0.95	1.05	1.03	1.08	1.01
METAL	0.94	0.93	0.98	0.99	0.93	0.95	1.01	1.09	0.97	1.08	1.05	1.12	1.11	1.01
MOTOR	1.00	1.02	0.95	1.00	0.98	0.98	0.97	1.04	0.98	1.08	1.01	1.02	1.24	1.02
FOOD	0.94	0.95	0.97	1.04	0.97	0.94	0.98	1.06	1.07	1.12	1.08	1.07	1.12	1.02
MISC	0.97	0.94	0.98	0.98	1.00	0.98	0.99	1.03	1.15	1.04	1.12	1.09	1.15	1.03
BUILD	0.98	1.04	0.98	0.99	0.98	1.00	1.01	1.03	1.04	1.05	1.06	1.07	1.28	1.04
CHEM	0.90	1.05	0.94	1.13	0.99	0.91	0.97	1.06	1.11	1.25	1.11	0.99	1.20	1.05
mean	0.96	0.98	0.99	0.99	0.99	1.00	1.01	1.03	1.03	1.04	1.04	1.05	1.06	

Table 12: The slopes (mean of the five years observed) of regressions in which s, a proxy for the size effect, explains 13 items. Rows are industries and columns are items.

Let us suppose that the log variability present in items is the result of two or three independent effects. Such effects would be differently mixed up to form each item. In that case the above regressions would yield very different slopes for different items. Each slope would reflect the particular co-variance between the predictor, s, and the mixed variability present in each item.

This is not the case. All the observed items yield the same slope when explained by the same variate. Hence the variability of our log items comes mainly from a unique source.

When considering all items during the same period of five years and the 14 industries one by one, 95% of the obtained slopes are larger than 0.9 and smaller than 1.2 (see table 12).

Sales, Inventory, Debtors, Creditors, Current Assets and Liabilities and Wages are the nearest to 1. Next come Net Worth, Gross Funds From Operations, Fixed Assets and Earnings, with slopes ranging from 0.85 to 1.3. Finally, Working Capital, with 0.7 to 1.3 and Debt with slopes ranging from 0.6 to 1.4.

Notice that the slopes being near this particular value — the unit — stems directly from applying log transformations. In log space any scaling becomes a translation. If we have several variables which are proportional to each other, after applying logs they become translations of each other. But translations can be accounted for by using a single point in the space of the observed variables — an intercept or the constant term of regressions. In other words, the main source of variability of log items can be modelled by a simple mean adjustment.

In log space any regression between closely proportional variates yields slopes which are near the unit and intercepts which are the logarithm of the expected value of such a proportion. Unit slopes are not real slopes, as a unit scaling is not really a scaling.

In order for items to be proportions of each other — in log space translations of each other — there must be a multiplicative common source of variability commanding them all.

S	W	I	D	C	CL	FA
0.63					13 cases	
0.41	0.33					
0.54	0.39	0.61				
0.53	0.38	0.51	0.49			
0.47	0.33	0.45	0.45	0.48		
0.48	0.35	0.48	0.46	0.45	0.47	
0.60	0.41	0.53	0.54	0.49	0.48	0.67

S	W	I	D	C	CL	FA
0.32					23 cases	
0.24	0.28					
0.32	0.28	0.47				
0.28	0.26	0.32	0.30			
0.23	0.22	0.28	0.24	0.23		
0.25	0.23	0.28	0.26	0.23	0.26	
0.24	0.26	0.25	0.26	0.23	0.24	0.33

Figure 16: Two Σ matrices obtained from the same sample. On the left, cases with negative Working Capital. On the right, cases exhibiting negative EBIT.

Negative cases: The displayed Σ matrices and the experiment explained above only consider positive cases. The few negative cases present in Δ items were excluded from the sample on a case-wise basis. In other words, all the firms used for building the observed matrices belong to the same group — They are healthy.

The observation of Σ matrices formed with logs of the absolute values of negative cases is, of course, not feasible. The sample would be too small. Also, these matrices would mix up situations which are very different from one another.

Instead, we isolated groups of firms exhibiting the same particular illness: Liquidity problems, poor profitability. Then we formed Σ matrices with each one of such groups. For the few significant samples we could find, the behaviour of all the items is clearly distinct from the same matrices for healthy cases. It seems dependent on the class of financial problem the negative items reflect.

For example, for the same group and year, firms with liquidity problems (Negative Working Capital) have a pattern of co-variance clearly different from the corresponding pattern of those having profitability problems. And the negative-EBIT matrix is different from the healthy one.

The matrices displayed in figure 16 belong to the Electronics industry in 1986. On the left all cases had negative Working Capital. On the right all cases had negative EBIT. These patterns are homogeneous but different. Samples with negative Working Capital generate more spread than those with negative EBIT. And both differ from the pattern usual in positive cases.

Notice that the number of cases in the samples used for calculating the above matrices is smaller than the desirable. We obtained a statistic, Σ, which actually has got more parameters on it than the number of cases in the sample. Such an analysis is case-dependent or even misleading.

Comparing simulated and real cases: Simulation can be carried out in order to establish, avoiding the burden of analytical developments, which is the ideal multi-variate pattern of such matrices when they are calculated from the logs of absolute values of negative cases. This allows us to understand which features of the above matrices are due to a particular behaviour of firms — liquidity or profitability problems for example — and which are due to the mechanism of subtracting two log variates.

We used the same group, Electronics 1986, and starting conditions similar to those found in

S	CA	FA	CL	
0.510				S
0.478	0.504			CA
0.490	0.480	0.599		FA
0.462	0.470	0.467	0.460	CL

Figure 17: Σ matrix used for simulating Working Capital, 1986, Electronics.

	S	FA	CL	CA	WC		S	FA	CL	CA	WC
S	0.516			W. Cap. > 0			0.462			W. Cap. < 0	
FA	0.505	0.627					0.416	0.475			
CL	0.464	0.477	0.459				0.419	0.401	0.423		
CA	0.475	0.484	0.465	0.489			0.422	0.400	0.424	0.429	
WC	0.494	0.492	0.474	0.532	0.672		0.399	0.407	0.416	0.392	0.680

Figure 18: Σ matrix from the simulated Working Capital. Electronics, 1986.

positive deviates. Such conditions were:

Item	Mean	Standard Deviation
Current Assets	4.213	0.723
Fixed Assets	3.757	0.779
Sales	4.577	0.720
Current Liabilities	4.005	0.680

In figure 17 we display the Σ matrix used for introducing in the simulation the co-variance of these items. After generating 2,000 cases with multivariate lognormality we obtained 185 with negative Working Capital.

Both positive and negative Working Capital were lognormal. When we slice such a sample along a principal axis, in general the positive cases exhibit positive, though small, skewness. The negative ones have negative, very small skewness and larger kurtosis. The resulting variance and co-variance matrices for both groups are displayed in figure 18.

Such matrices are not especially different from the expected. It seems as if no real differentiation between positive and negative cases exists caused just by the subtraction of two lognormal distributions. The strong differences observed in matrices obtained from real cases are most probably due to mechanisms internal to the firm — the lack of liquidity or a poor profitability.

We carried out simulations of other groups. One more example is in appendix A. In general, the only feature we could recognize as particular to simulated negative cases was its larger spread and the described skewness and kurtosis.

Σ matrices of negative cases are not stable: A feature of data obtained from samples of negative cases is that variance and co-variance matrices seem not to exhibit the slightest traces of stability. In most of the industries matrices change to different ground values from one year to the other. Even this lack of stability is not a stable feature since it is absent in one or two industrial groups. All this could just result from the small number of cases in the observed samples.

The only constant feature is the overall similarity of cells. Matrices vary but all their cells remain fairly similar. This means a certain amount of dependence on size.

2.3 The Subtraction of Two Accounting Items

Many important items from accounting reports, namely those representing flows, are obtained by subtracting two other items. In this section we try to answer two questions related to the statistical modelling of flows. These questions are:

- Why are flows lognormal? In general, there is no reason why the subtraction of two lognormal variates should remain lognormal.
- If an accounting item has negative cases the log transformation cannot be applied. This is the case for McLeay's Δ variables like Earnings, Working Capital and Funds Flow. How can they be used in statistical models?

After explaining the lognormality of flows we show that the problem of negative cases is not specific to the log transformation. We suggest alternative solutions for statistical modelling.

2.3.1 The Statistical Distribution of Flows

Samples containing positive cases only: The apparent lognormality of positive differences between two lognormal variables is a consequence of a strong correlation between them. In a subtraction, $z = y - x$, of two correlated items, cases in which y is large also have proportionally large x. And cases in which y is small are expected to have proportionally small x.

The extent to which z follows x and y is dependent on the correlation between x and y. For a pair, $\{y, x\}$, of exactly synonymous items, z would exhibit exact proportionality with them. If s is the effect common to both x and y we can write in such an extreme case

$$\frac{y_j}{s_j} = R_y \quad \text{and} \quad \frac{x_j}{s_j} = R_x \quad \text{for any firm } j.$$

R_y and R_x are constants. Therefore $z_j/s_j = R_y - R_x$ for any j.

The more general case of correlated items can be written in a similar way. In fact, if we introduce in the above expression small f_j^y and f_j^x so as to reflect the variability particular to x and y we obtain

$$\frac{y_j}{s_j} = R_y \times f_j^y \quad \text{and} \quad \frac{x_j}{s_j} = R_x \times f_j^x \quad \text{for any } j.$$

These f_j account for departures from a strict correlation with s in case j. If such departures are small, they should be near the unit. Considering $f^y = 1 - \eta^y$ and $f^x = 1 - \eta^x$ we can write

$$\frac{z_j}{s_j} \quad = \quad R_y \times f_j^y - R_x \times f_j^x$$

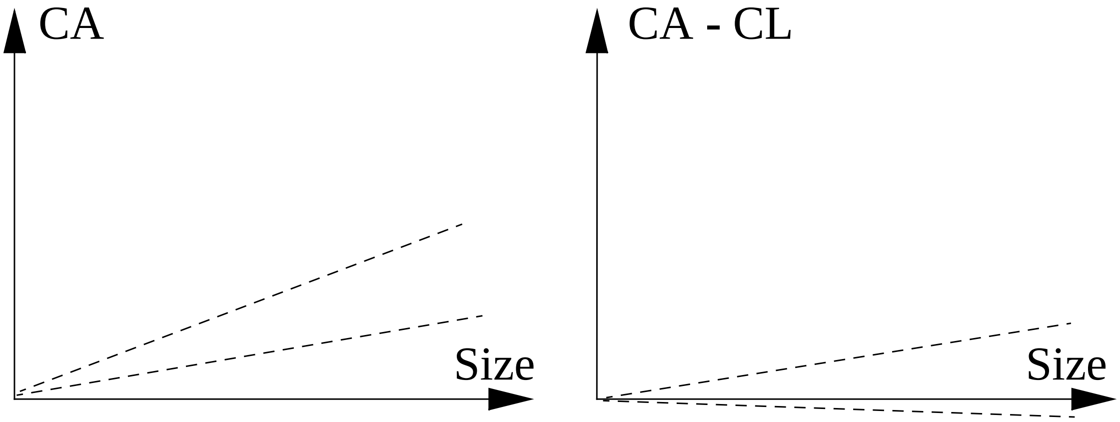

Figure 19: Schematic representation of a subtraction of two accounting items yielding a new one with some negative cases.

$$\begin{aligned}\frac{z_j}{s_j} &= R_y \times (1-\eta_j^y) - R_x \times (1-\eta_j^x) \\ &= R_y - R_x - (R_y \times \eta_j^y - R_x \times \eta_j^x) \\ &= R_y - R_x - d_j.\end{aligned}$$

The departure from proportionality is now represented by $d_j = R_y \times \eta_j^y - R_x \times \eta_j^x$. Since the f_j are near the unit the η_j will be small. And for values of R usual in ratios d_j will be a difference between two small values as well.

z will be near lognormality for the same reasons x and y are lognormal, whenever d_j will not modify significantly the shape of s. This could occur because the variability introduced by the f_j is negligible when compared with the one of s, or because their distribution, when combined, will not distort the shape of s. The first case requires s to be strong as we know it is (section 2.2.2).

The same reasoning could be expressed in many ways, all based on the existence of a strong effect common to the items being subtracted. The lognormality observed in flows is a clear clue for the existence of a common effect. Lognormality only propagates across sums or differences of variates which are strongly correlated. The difference of two independent or weakly correlated lognormal deviates yields a distorted shape.

Graphical subtraction of items: Graphically, any subtraction of two lognormal variables $y - x$ strongly correlated with s is approximately a clockwise rotation of y in a two-variate distribution of y with size. If we build a plot in which y, a lognormal variable, is represented against s and then we subtract to each y the corresponding x, the result will be a rotation of the original shape downwards.

In fact, to large y it will correspond large x and to small y, proportionally small x. When subtracting them, the sliding down of large y will be large and the sliding down of the small y will be proportionally small. The resulting movement is approximated by a rotation rather than by a

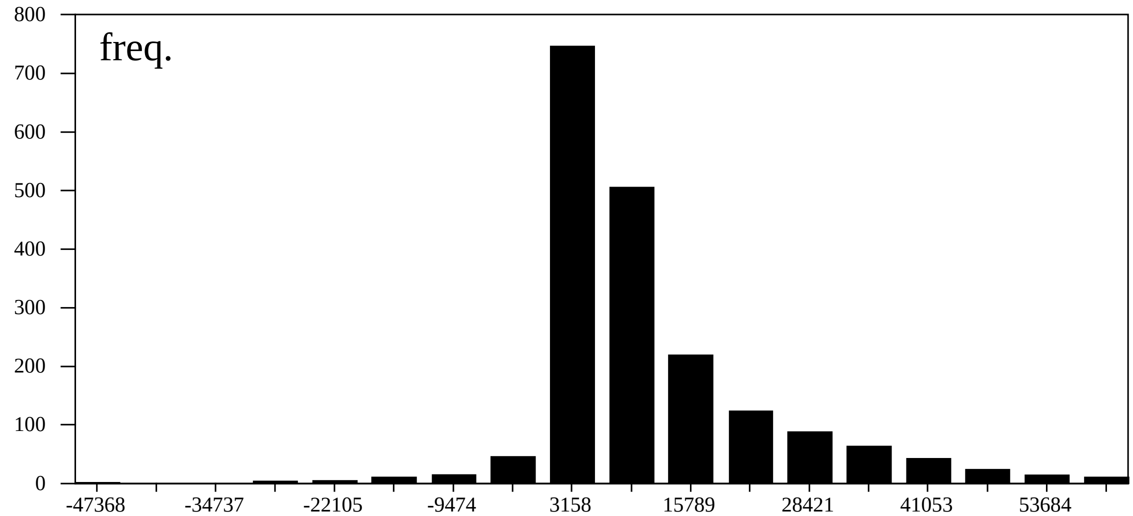

Figure 20: Detail (near zero) of a frequency distribution obtained by simulating the difference between two items. This distribution is a juxtaposition of two lognormal ones.

translation. Figure 19 on page 49 illustrates this.

Rotations are important because they will not modify the distributional characteristics of the variables involved. If the overall sliding of cases is a rotation then the resulting distribution will remain close to lognormality. Empirically we know that this is so.

Samples containing both positive and negative cases: In chapter 1 we showed that, when taken separately, both the positive and the negative cases of the observed Δ items are as lognormal as any other items. We can see why by subtracting graphically two lognormal shapes, both correlated with size, as in figure 19.

Given $WC = CA - CL$ we start with a scatter-plot of CA on any measure of size and then we subtract CL from CA. The values of WC were obtained from CA by sliding them down a value which is CL. But the CL are proportional to size too. Hence, a large firm's CA is expected to be largely modified and a small firm's CA is expected to change proportionally less.

Notice that the X-axis now slices the two-variate distribution. There are now two regions separated by the X-axis. But such two regions preserve the proportional nature of the data.

If the slicing were made along a line not passing through zero, or in less correlated items, the result would be a truncation. But whenever the slicing of a two-variate distribution of very correlated lognormal items is made along an axis of the distribution itself, the resulting two scatters will project their values in the Y-axis in a way that preserves lognormality.

In fact, when we go along growing size we find at our right hand an increasing spread in the positive WC direction. And the corresponding for the region below $WC = 0$. The projection of such scatters in the Y-axis yields approximately lognormal distributions by the same reasons non-sliced scatters do. The result of a subtraction of two accounting items so that a few negative cases emerge generates a juxtaposition of two approximately log-normal distributions. One contains the positive cases. The other one, the mirror-image of the absolute value of the negative ones.

The simulations we carried out corroborate this fact. Figure 20 on page 50 displays the central part of the distribution of a simulated difference between two accounting items. It is a juxtaposition of two lognormal distributions, one with much more cases than the other. As referred, empirical tests show that both such distributions are lognormal.

2.3.2 Modelling Samples Having Negative Cases: Discussion

What is the interest of considering — in a cross-sectional context — the whole sample of positive and negative cases? What do we loose by taking separately one sample with positive cases and another one with the negative ones as we have done so far?

At first, the existence of negative cases in a sample seems unsatisfactory for modelling purposes. By taking both sets separately we break the continuity of the sample and loose the information describing the passing through the zero value. It would seem desirable to bridge this lack of continuity and be able to work with the whole set of cases as a unique variable in the sample.

It is easy to recall important pieces of research in which the used ratios had values passing through zero. Beaver's classic study on the importance of ratios for tracing firm's failure [7] shows how revealing a ratio of Cash Flow to Total Debt can be when sliding down from positive to negative values during an observed period.

The consequence seems to be that the ratios like the one above should be considered as a unity and taken as a whole. Breaking them into two samples, one for positive Cash flows and another for negative ones, would apparently damage the most important part of its information content. Accounting research has devoted some effort to the assessment of the distribution of such ratios (see, for example [86] and [87]).

The cross-section context: In fact, we loose nothing by using split samples in cross-sections. Beaver's ratios draw a trend during several time periods. One unique object — the average ratio for a group of firms — is observed for consecutive intervals. Cross-sections are not about one unique object. They capture the behaviour of many objects ideally at the same instant in time. The concern referred to above stems from picturing time-series and transposing it to cross-sections.

Ratios and the log transformation: The lack of continuity between positive and negative cases is not a problem specific to the log transformation. And it is not a problem either. McLeay's Δ variables also break the continuity of ratios or regressions. They break it in a so large extent as the log transformation does. Ratios — or regressions — will fail to model correctly exactly the same samples the log transformation is not apt to model.

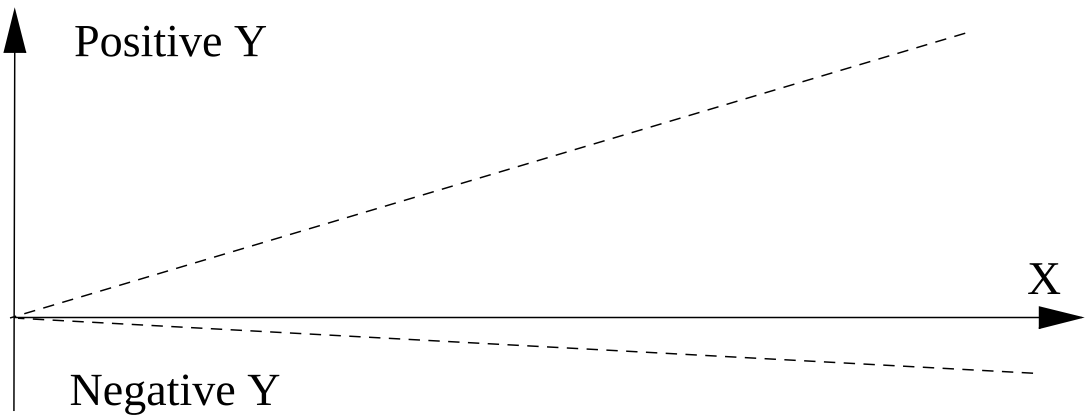

Figure 21: A schematic representation of a scatter-plot for the ratio y/x when y can have negative values.

For the ratio y/x in which y is a variable having both positive and negative cases, when we go along decreasing values of y and pass through zero, the corresponding evolution of x change in direction. It ceases decreasing and begins to increase. It is impossible to model such a sample using one ratio. For each x there are two possible y. That's why practitioners calculate standards by considering only positive values. They can't find a consistent standard for samples with both signs.

Figure 21 illustrates this fact. It represents a cross-sectional sample. Many firms and two observations: For example, the Y-axis could measure Cash Flow and the X-axis Total Debt. Each firm would be represented by a point in this scatter. In our example there are firms with large positive Cash Flows and large Total Debt. But there are also firms having negative Cash Flows and large positive Total Debt. Now let's produce a ratio for explaining the joint behaviour of these two items. The firms with negative Cash-Flow push the expected value of such ratio towards values more near zero than otherwise. Hence, we obtain an estimation for the Total Debt any particular firm should have — given its Cash-Flow — which is larger than it should be.

The expected value for the ratio could even approach zero or become negative. When approaching zero, the amount of Total Debt predicted by such ratio would rise to infinity. After passing through zero the ratio would predict infinitely large negative Debt.

Positive and negative cases form two groups: There is a breakage of continuity when passing through zero. Each sample — positive Cash Flow and the negative one — really represent different groups and should not be mixed up.

Clearly, the problem is not in the use of logs or ratios. The problem is simply that two different groups cannot be modelled by the same parameter. We couldn't use one unique regression to account for the data in figure 21. We could but we shouldn't. We would need two regressions. And this is what it is all about.

Of course, in the case of ratios or logs, the algebra itself precludes one single model. Proportional co-movements cannot pass from one quadrant to the other one except by going both together through the origin. This stems from being proportional or taken as such. Ratios, the same as logs, entail an assumption of proportionality. Ratios, a finite difference one. Logs, a differential one.

Lev and Sunder [79] devote to this breakage of continuity a large comment. After presenting a reasoning similar to ours, they conclude that "A change in Earnings which has a favourable effect on the ratio before the change of sign, will have an unfavourable effect after it. This loss of continuity is a frequent cause of problems in interpreting ratios computed from negative numbers. (...) This problem renders ratios a hazardous instrument of controlling for size in the presence of negative numbers and the researcher would be well advised to seek alternative means of exercising such control whenever feasible." Remarks similar to this one are usual in text books on ratio analysis.

Ratio analysts and practitioners, instead of considering the positive and negative cases together, recommend avoid doing so. We see no reason for considering such variables as a continuum when building statistical models.

Does it make sense to consider jointly the positive and the negative cases? In a strict cross-sectional context the referred lack of continuity seems to make sense also on grounds of financial analysis. Positive and negative cases represent different groups of firms. They are to be accounted for by grouping variables, not by continuous-valued ones.

We recall the patterns of joint variability displayed in section 2.2.2 for samples reflecting liquidity or profitability problems. They are very particular. This, contrasting with the general rules governing the positive ones.

Negative cases reflect firm illness. On the contrary, the positive ones reflect a healthy state. In statistics, situations as those require a grouping variable. The groups formed in cross-sections by firms having positive and negative cases can be modelled and compared provide they are considered as two groups.

2.3.3 Alternative Variables

Despite the above remarks, there are cases in which it would be interesting to use continuous-valued information as the one conveyed by Δ items. It is the case of the building of trajectories of a firm's position for several time intervals. Here we suggest some solutions.

Substitution of items: The problem of statistical models requiring Δ items can be solved in most of the cases by using their components instead. The information conveyed by negative-valued items can be introduced in a relation by other items appropriate for the log transformation. The variability of Earnings can be brought to the model by Sales and Expenses. Working Capital can be substituted by Current Assets and Liabilities.

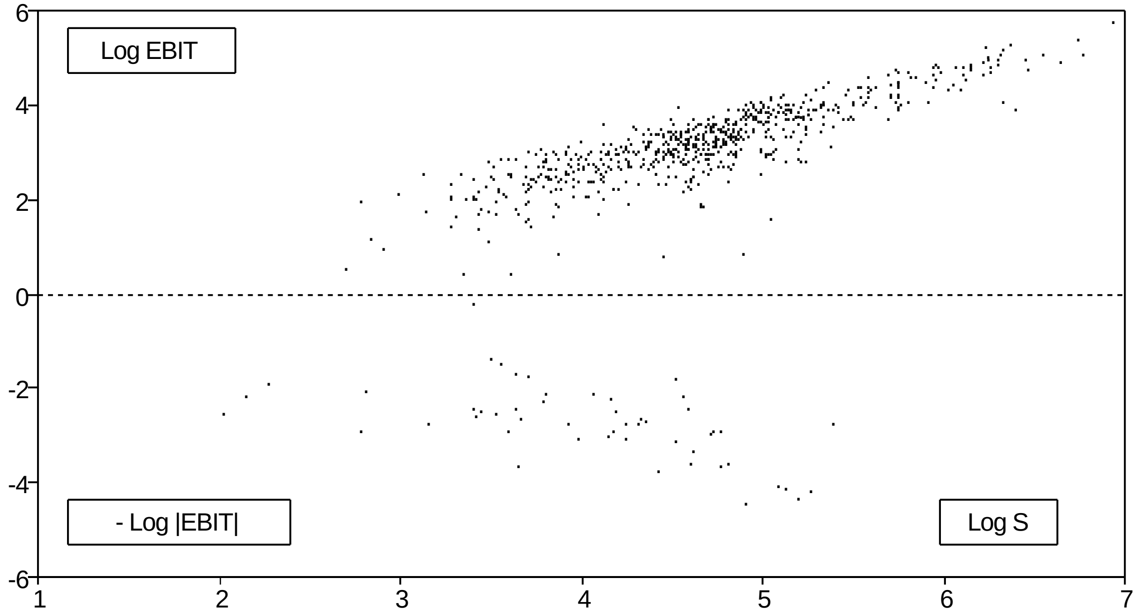

Figure 22: The effect of a symmetrical transformation on the EBIT to Sales relation. All groups, 1984.

For any accounting item z, resulting from a subtraction of two positive items y and x, the pair $\{y, x\}$ will obviously contain all the information z would contain, and a bit more. In fact, there is a unique z for any given pair $\{y, x\}$. But for a given z, there exist an entire line of pairs $\{y, x\}$ able to yield such z. Therefore z could never bear information not contained in $\{y, x\}$.

Scaled differences: Another alternative solution for the modelling of Δ variables is the use of a scaled difference. If we scale one of the components of the difference so that negative cases cannot occur in the sample we obtain a new item which is also a difference but has only positive cases.

Scaling is equivalent to a translation in log space. Samples will not change their shape by introducing a scaling. Therefore, if we use $z' = y - x \times S$ instead of $z = y - x$ (S being a constant) we can work with this new variable knowing that the shape of z' is the same as the one of z, its standard deviation in log space didn't change and even the outliers are all there. Only the mean value has emerged a bit.

Symmetrical transformations: Another possible way of dealing with both signs is the use of a symmetrical transformation:

$$x \mapsto \log(x), \quad \text{for } x > 0 \tag{2}$$

$$x \mapsto -\log(x), \quad \text{for } x < 0 \tag{3}$$

or the equivalent creation of a dummy variable, d, retaining the information regarding the sign of x:

$$\left.\begin{array}{rl} x \mapsto & \log(x) \\ d = & 1 \end{array}\right\} \text{for } x > 0 \qquad\qquad \left.\begin{array}{rl} x \mapsto & \log(x) \\ d = & -1 \end{array}\right\} \text{for } x < 0$$

Such transformations correspond to the fact that negative cases are also lognormal and correlated with size. They will be useful and valid, provide no attempt is made to fit a unique model to the transformed data.

Symmetrical transformations preserve the relative size of Δ items. They allow the building of maps for drawing trajectories. And such trajectories will be appropriate for financial diagnostics except in the neighbourhood of zero.

The addition of constants: Indeed, there is one transformation which should not be applied for it severely distorts distributions. It consists of adding a large positive constant to all cases in the sample so that the negative ones become positive. This practice has been reported in a few studies.

Modelling Debt: The log transformation cannot be applied to items having cases with values of zero. We could avoid the problem of variables having zero values, like Debt, simply by using, instead of $\log 0$, a very small number: $\log 1 = 0$. The use of $\log(x + 1)$ instead of $\log x$ is recommended by Snedecor and Cochran [119] for proportional samples with zero cases. This is acceptable if in the sample there are no cases with values near zero. For instance, if a scaling of one million pounds is used instead of the usual scaling of one thousand, it is very likely to have in the sample cases with values near zero, both positive and negative.

Even so, we should avoid — whenever possible — to use Debt directly as an input variable for statistical modelling in samples having both leveraged and non-leveraged firms. The leveraged firms are lognormal but the group of non-leveraged firms forms a cluster of cases all with the same value. This set of cases would be influential.

When Debt is required, it is supposed to bear important continuous-valued information only in the case of leveraged firms. Therefore, we divide the sample in two. One contains leveraged firms. The other one the non-leveraged ones. Inside each sample, estimation and inferences based on continuous-valued models can be carried out. Inferences about differences between these two sets are also possible.

2.4 Summary

In this chapter we extracted the most obvious consequences of the lognormal nature of the observed items. A few drawbacks widely discussed in the literature seem to have been accounted for. We also described the multi-variate behaviour of log items and the consequences of lognormality in the case of differences between items.

Outliers and Heteroscedasticity: The lognormal distribution is likely to generate what seems to be an outlier. By the same reason, it will also generate non-homogeneous variance in regressions.

Regressions should not be used to model relations between lognormal variables. Lognormal distributions generate large residuals which monopolize the minimization of square errors. The results are then dependent on one or two influential cases.

Weighted Least-Squares is not an appropriate recipe for dealing with the above problem. It simply transfers the influence from the largest to the smallest cases in the sample. The heteroscedasticity would not vanish in any of these cases. The log transformation is adequate, in a first approximation, for rendering additive residuals. But appropriate models ought to be developed that fully explore the existence of non-proportional and non-linear relations in accounting items.

Also the trimming of outliers becomes a useless exercise for two-variate lognormal data. The shape like a "<", typical of two-variate proportional relations, will not change across scales. It will generate more and more outliers if successive trimming was to be attempted.

The common component of the variability of items: An important point this chapter outlines is the existence of a common source of variability in the observed log items. In log space these variables can be viewed as the addition of two processes. The first one is common to all items and seems to reflect the relative size of firms. The second one, particular to each item, reflects its uniqueness.

A consequence is that accounting items should be explained in terms of size and deviations from size. Instead of viewing each item individually as an independent source of variability — eventually correlated with other few items — we should first account for an effect common to them all and then take the residual variability as the particular contribution of each item.

Our results seem to erode or smooth the distinction between Σ and Δ items regarding the influence of size. Earnings or Gross Funds From Operations are strongly correlated with size. As correlated as Fixed Assets and Net Worth.

The item showing a distinct behaviour is Debt. But even in this case size is present. The negative cases of Δ items seem also correlated with size, as far as it was possible to observe them.

Cross-sectional samples with negative cases: We also studied the problems posed to statistical models by items having both positive and negative cases. We pointed out that there is no continuity between the positive cases and the negative ones. We further suggested that negative cases should be viewed as a different group since they represent situations of the firm which are specific and not related to those observed for positive cases. Nevertheless, we suggested alternative transformations to make statistical models, when necessary, able to deal with log-transformed data in the case of samples of positive and negative values.

Chapter 3

An Extension of the Ratio Model

So far ratios have been used as input variables for statistical modelling in Accountancy. In this chapter we question their use. Ratios cannot account for non-proportional and non-linear features. On the other hand, the lognormality observed in items suggests the use of multiplicative or proportional models of which ratios are the simplest example.

In what extent is the lognormal nature of the observed items compatible with non-proportional and non-linear relations between them? The development of new models rely on the ability to answer this question. Therefore, the first task we undertake in this chapter is the answering of the above question as a necessary step towards the building of appropriate tools.

We first recall a known mechanism for generating the probability distribution observed in our data. Then we study the extent in which financial ratios and multi-variate relations are consistent with such a mechanism. Finally we introduce on it conditions leading to non-proportional and non-linear relations.

We show that there is no contradiction between proportional mechanisms and a class of non-proportional relations. Financial ratios emerge as a particular case of more general descriptors. They can be extended so as to include non-proportionality. The chapter finishes with examples of use of the new ratios.

3.1 Introduction

Empirical observation suggests that cross-sectional samples of many accounting items are approximately lognormal. McLeay [86] observed lognormality in large samples of accounting items which are sums of similar transactions with the same sign like Sales, Stocks, Creditors or Current Assets.

Along with the items already studied by McLeay, we found that lognormality cannot be rejected also for stock variables like Fixed and Total Assets or Net Worth and non-accounting items related to size like the number of employees.

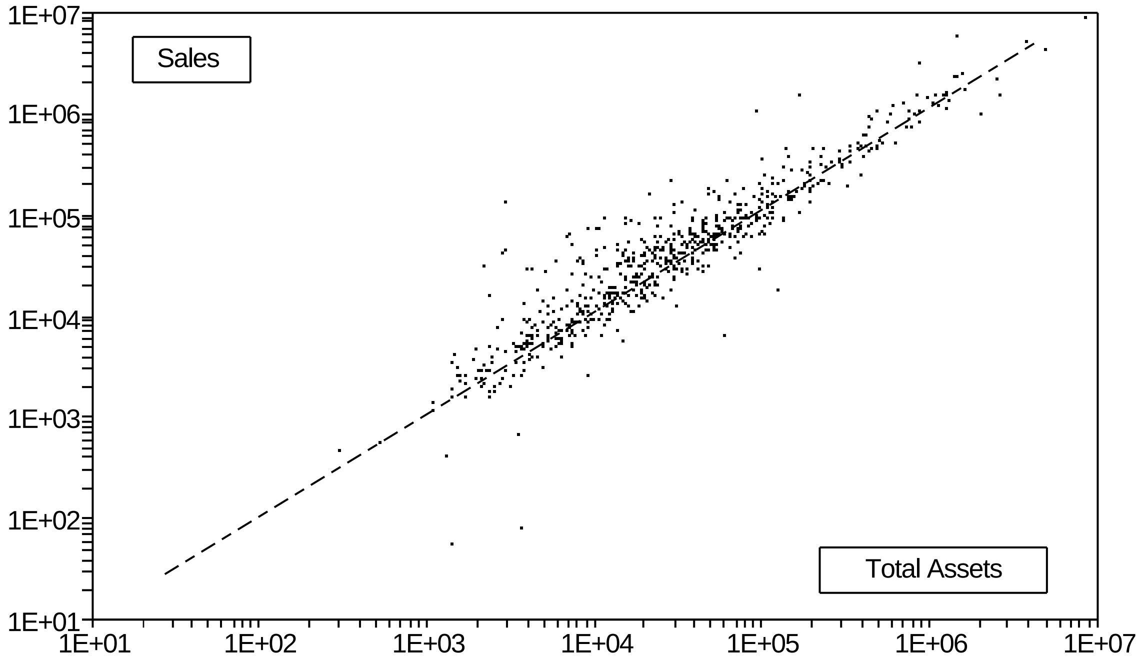

Figure 23: The relation between Sales and Total Assets in log space. All groups together 1984. The dashed line is the axis $y = x$.

Positive values of accounting items having both positive and negative cases, like Working Capital, Earnings, Gross Funds from Operations and the absolute value of the negative cases of these items, are also lognormal.

We also gathered evidence on the lognormality of small, homogeneous, samples. We examined 18 accounting items for a period of five years (1983-87) belonging to 14 industry groups in the U.K. The results are displayed and discussed in chapter 1. We concluded that lognormality seems to be a general quality associated with the statistical behaviour of the observed items.

Empirical models: The observed items are lognormal. How far can we go in the building of appropriate models for accounting relations just by considering this empirical finding?

The first consequence of our study is that, instead of an entire sample, we need only to consider the estimated central trend and scatter. In ordinary space, financial variables are lognormally distributed. A particular observation cannot be simply described by the mean of the distribution plus a deviation from that mean. But once we move into logarithmic space the resulting variable is normally distributed. Any observation can now be explained as the mean of the transformed variable plus a deviation from that mean. That is, any lognormal item, x, is described in log space as an expected value, μ_x, plus a residual, e^x:

$$\text{For the } j^{th} \text{ firm in a sample,} \qquad \log x_j = \mu_x + e_j^x$$

When we deflate one item with the median of the same item for that industry the residuals are Gaussian in log space. Lev and Sunder [79] discuss appropriate estimators for the central trend of several possible distributions. Amongst others, the median is also analysed.

Since $\overline{\log x}$, the mean value of $\log x$, is a good estimator of μ_x, $\exp(\overline{\log x})$ will be a good estimator of the median of x. Then,

$$\text{for the } j^{th} \text{ firm in a sample, the quotient} \qquad \frac{x_j}{\exp(\overline{\log x})} = \exp(e_j^x) \tag{4}$$

will reflect the number of times the case x_j is larger or smaller than the standard for its industry. If x_j is Sales of firm j, such quotients would reflect $j^{'}$s relative position and relative progress. In general these position quotients seem not to be especially useful in accountancy. They measure size instead of controlling for it.

Financial ratios: We could also say that, since our items are lognormal, financial ratios y/x can be written in log space as a difference of two position quotients defined in 4:

$$\text{For the } j^{th} \text{ firm in a sample,} \qquad \log(y_j) - \log(x_j) = (\mu_y - \mu_x) + (e^y - e^x)_j \tag{5}$$

This expression is obtained just by subtracting two log items. It is similar to

$$\frac{y_j}{x_j} = R \times f_j$$

with $R = \exp(\mu_y - \mu_x)$ and $f_j = \exp(e^y - e^x)_j$. Here, we arbitrarily used natural logarithms.

R will be the expected proportion in which y differs from x. Proportional effects lead to variables related by percentages rather than by additive displacements. We say that FA is expected to be, say, 25% smaller than TA if $R_{\frac{FA}{TA}} = 0.75$.

A good estimator for R is $\exp(\overline{\log y} - \overline{\log x})$, the median of the ratio — in log space, the difference between two mean values —.

The multiplicative nature of ratio residuals: f_j is a multiplicative residual deviation and accounts for the particular case of firm j. It shows in what proportion the relative magnitude of y, in the j^{th} firm, diverges from its expected relative magnitude as predicted by observing x.

The use of multiplicative residuals contrast with the practice but it seems reasonable. Ratios are multiplicative models. The residuals should be taken as multiplicative too. In other words, if the expected value is a proportion, deviations from it are likely to be proportions as well. As an example, it would be possible to say: "For values of B ranging from half an inch to ten miles, A is expected to be twice the length of B with an error of plus or minus twenty inches". But it would be difficult to imagine an error mechanism yielding such deviations. The usual would be to consider errors of, say, 3%.

The ratio model: Discussion Ratios are simple proportions. Lognormal items become homogeneous in a proportional space and their difference is a proportion too. These facts seem to match. But, is the ratio model adequate beyond this apparent matching?

The sole consideration of the lognormal nature of individual items seems to be enough to conclude about one appropriate estimator for ratio standards and also about the multiplicative character of deviations from such standards. These are interesting points in their own right. But the sole consideration of lognormality on accounting data is not enough to validate the financial ratios themselves as appropriate models. Such an empirical basis cannot prevent the ratio model from being questionable. Ratios are just the simplest relations allowed by the lognormal nature of items. Are ratios able to model all the relations important for financial analysis and knowledge acquisition?

Accounting research seems to give a negative answer to the above question. It is usual to find in the literature a tone of pessimism about the usefulness of ratios. The existence of non-proportional and non-linear relations between items are the main causes of concern. Whittington [134] explains that

> ... in an empirical relationship between a pair of accounting variables, two of the conditions necessary for proportionality are quite likely to be violated. Firstly, there may be a constant term in a relationship (...). Secondly, the functional form of the relationship may be non-linear.

The potential convenience of more elaborated models like regressions has also been stressed by Barnes [5]. He showed that in any regression $Y = A + BX$ the distribution of Y/X will be skewed whenever $A \neq 0$. Ratio standards would be likely to misinform since no central trend would exist.

The role of a generative mechanism in this study: Most of the concern about the validity of ratios is based on the possibility of non-proportional relations between items. In order to study causes for intercept terms it is usual to describe plausible mechanisms able to generate intercepts. For example, in the accounting literature it is frequent the use of arguments based on the existence of fixed costs.

In this study we also use a generative mechanism. Ours is not an accounting mechanism. It is not intended to describe the internal features of the firm like liquidity or financial structure. It describes items in terms of size and deviations from size. Items, according to our mechanism, are endowed with two qualities: They are lognormal and they have in common a strong component of their variability.

The reason for using a generative mechanism is not the claim that such a mechanism is actually the real cause and explanation of the behaviour of accounting items. Our claim is that it is possible to interpret the evidence referred to in a consistent way and develop new models bearing the same consistency.

3.2 Ratios and Lognormality

In this section we use the well known proportional effect as a basis for explaining ratios. The usual financial ratio emerges as a simple consequence of a strong, common, effect.

The proportional effect has been quoted by McLeay [86] as a mechanism able to explain the existence of lognormality in a few items. Also a recent study [128] uses it. Both studies seem to suggest a basic or qualitative distinction between two kinds of items. The first kind would include items reflecting size. The second one, items which cannot "be treated as size measures [86]". In this study and in another one we published [129] the proportional effect explains size and deviations from size. No attempt is made to specify the particular behaviour of any item. Items having negative cases are considered as a subtraction of two positive ones and explained as such.

The Constant Effect: The Gaussian distribution is often interpreted as the result of many independent elementary perturbations. This approximation entails the strong assumption of a constant effect. For example, the probability of getting odds, when tossing a fair coin, is a constant value of 1/2 no matter the number of games or the size of the coin. And the probability of getting particular proportions of odds when tossing a coin in several sequences of games draws a Gaussian distribution. This constant probability of 1/2 governing the game referred to is what we call a constant effect. It leads to normality.

The Proportional Effect: If, however, any random change dx suffered by a variable x is proportional to the value of x itself, the effect is no longer constant. It is a proportional effect and a Gaussian generative process will not be able to explain it.

Gaussian variables spread their final realizations around an expected value. They are bounded: It is most unlikely to find cases many orders of magnitude larger or smaller than the expected. This is because the random changes leading to such realizations are expected to be similar — a constant effect. Contrasting with such a mechanism, when the random changes leading to any final realization are expected to be similar only when taken as proportions of the momentary value of the variable, the effect is proportional. The probability distribution of such variables is unbounded. It exhibits strong positive skewness. The observed samples contain cases which are many orders of magnitude larger than the expected ones.

An example of a proportional effect would be the size of organs or animals when growing to maturity. A whale grows $dx = x_2 - x_1$ while a mouse grows $dy = y_2 - y_1$. The random changes suffered by x and y are dx and dy. Such changes are not expected to be similar. Their expected values will be proportional to the size of the animal or organ they affect.

The Gibrat Law: The lognormal probability distribution can be viewed as a result of a generative proportional mechanism. This fact is known as the Gibrat law [48]. Let x be the position of an accounting item. If dx, the random transactions affecting x, are expected to be proportional to x itself,

$$\text{the quotient} \quad \frac{dx}{x} \quad \text{will be expected to be independent of } x.$$

So, if we can find a function

$$z = f(x) \quad \text{such that} \quad dz = \frac{dx}{x} \tag{6}$$

then the new variable z will obey the assumption of a constant effect. In the case of dz being many, independent, perturbations $f(x)$ is the logarithmic function. Aitchison and Brown [1] contain a detailed explanation of this reasoning. Singh and Whittington [118] explore the growth of firms as governed by the Gibrat law.

Notice that the logarithmic function emerges as a result of the Central Limit theorem. The normality of the process governing dz is not required as an assumption, whenever the dz are many, independent, changes.

The relative growth: Any elementary perturbation dz will produce a small change dx which is expected to be a proportion of x. Therefore dz can be seen as an elementary relative growth and z as an expected relative growth. For example, when x increments in average from 5 to 6 thousand pounds, dx is a growth of 1 thousand and z is 20%. The same relative growth of 20% impinging upon different firms would make x increase in average from 10 to 12 millions or from 15 to 18 pounds.

Gaussian final realizations $z_j = \log x_j$ are explained in the same way. Firstly, by a central trend μ_x which is the expected one for the average relative growth affecting all cases in the sample. And secondly, by each particular departure from μ_x, the e_j^x, affecting only firm j. These e_j^x are residual average relative growth: When back in multiplicative space, the e_j^x are the proportion in which the average relative growth of firm j is above or below the expected.

For example, the expected value for Sales in the Food industry was estimated as $\overline{\log x} = 4.9218$, (83,521 thousand pounds) in 1987. We say that G. F. LOVELL PLC and UNITED BISCUITS are positioned at similar distances from the standard. UNITED BISCUITS sold 1,832,400 — about 22 times more than the standard — and G. F. LOVELL sold 3,722 — about 22 times less. For both, the relative departure from the standard is $e = 1.35 = \log 22$. Only the signs are different.

The generative mechanism responsible for the cross-sectional distribution of sales in the Food industry describes these cases as a sum of two components. One which is common to all the sample — an expected relative growth of 4.9218. And another which is a residual relative growth particular to each company: +1.35 for UNITED BISCUITS and -1.35 for G. F. LOVELL PLC.

3.2.1 Financial Ratios

Now we study the joint variation of more than one item. The notion of financial ratio as a descriptor stems from the assumption of an effect common to all items for the same firm.

As we saw, $dz = dx/x$, the elementary relative changes of x, have the structure of a relative growth. z is Gaussian as dz is commanded by a constant effect.

Modelling a common effect: We now assume that in the case of accounting data this Gaussian relative growth is the sum of two components. A common and strong component, σ_j, which accounts for random changes acting over the firm j as a whole and therefore is the same for all the $1, \cdots, i, \cdots, M$ items belonging to the same report. And a weak residual, ε_j^i, particular to item i.

Let x and y be the position of two accounting items for firm j. dx_σ and dy_σ are random changes in x and y caused by σ, a disturbance influencing both. Considering the way such common source of variability affects the relative growth which is about to generate x and y we can say that

$$\frac{dy_\sigma}{y_\sigma} = \frac{dx_\sigma}{x_\sigma}$$

This basic mechanism yields final realizations of x and y obeying general expressions of this kind:

$$\log y_\sigma - C^y = \log x_\sigma - C^x$$

C are constants depending on the initial values of x and y. Here, the superscripts are used for identifying corresponding items, not as exponents.

Since we defined $\varepsilon^x = \log x - \log x_\sigma$ and $\varepsilon^y = \log y - \log y_\sigma$ as the variability unexplained by σ,

$$\text{we have:} \quad \log(y) - \varepsilon^y - C^y = \log(x) - \varepsilon^x - C^x$$

If the cross-sectional distribution of the common effect is dictated by the Gibrat law it will be lognormal. In this case, when we consider the whole sample of $1, \cdots, j, \cdots, N$ firms, it is easy to see that the statistical model describing the relation between y and x for firm j is

$$\log(y_j) - \log(x_j) = (\mu_y - \mu_x) + (\varepsilon^y - \varepsilon^x)_j \tag{7}$$

a form similar to equation 5, the one based on empirical manipulation. μ_y and μ_x are the expected values of $\log y$ and $\log x$. Therefore ratios can be viewed as specific models describing the common component of the variability of y and x when both x and y are supposed to be final realizations of a unique proportional mechanism. The residuals are independent of the common effect.

Notation: Equation 7 can be written in the form of a ratio:

$$\frac{y_j}{x_j} = R \times f_j$$

with $R = \exp(\mu_y - \mu_x)$ and $f_j = \exp(\varepsilon^y - \varepsilon^x)_j$.

For expressing the differences between expected values we use the notation $\mu_{y/x} = \mu_y - \mu_x$ or, for the ratio standards, $R_{y/x}$ and so on. We write the differences between residuals as $\varepsilon^{y/x} = (\varepsilon^y - \varepsilon^x)$.

Superscripts are intended to avoid too many subscripts and should not be taken as exponents. They are used only in the C_j, ε_j, e_j and f_j.

A good estimator for $\mu_{y/x}$ is $\overline{\log y} - \overline{\log x}$, the difference between the mean values of y and x in log space. It is, of course, coincident with the median of the ratio expressed in logs. If an homogeneous sample of accounting reports is to be taken as a reference for the building of standards, the value of $R_{y/x}$, the ratio standard, should be calculated as

$$R_{y/x} = \exp\ (\ \overline{\log y}\ -\ \overline{\log x}\ )$$

or directly as a median. And if we want to build a new variable from the residuals of the fitted model we can calculate each $\varepsilon^{y/x}$ as

$$\varepsilon_j^{y/x} = (\log y_j - \overline{\log y})\ -\ (\log x_j - \overline{\log x})$$

or each equivalent proportion, $f^{y/x}$, as

$$f_j^{y/x} = \frac{\dfrac{y_j}{\exp(\overline{\log y})}}{\dfrac{x_j}{\exp(\overline{\log x})}}$$

Both $R_{y/x}$ and $f^{y/x}$ are ratios. $f_j^{y/x}$ is a ratio of two position quotients (defined in 5 on page 59). Notice that the ε^y or the ε^x are different from the e^y or the e^x in 5, the empirical formulation. However, $(\varepsilon^y - \varepsilon^x)_j = (e^y - e^x)_j$ for any j.

The weak, particular, effect: $\varepsilon^{y/x}$, the difference between residuals, can be interpreted as the weak effect particular to y when x is taken as a proxy for the common effect. Unless we know σ, we cannot determine exactly the real weak effects associated with individual items. We know $\varepsilon^{y/x}$ but we don't know each ε^x and ε^y, the components of such a residual difference.

Conversely, it is impossible to determine the value that σ, the common effect, assumes for firm j unless we know the components of the residual difference. Therefore, both the common effect and the particular one are not directly measurable. Ratios reveal what is different in their components, by concealing what is common in them. In chapter 5 we describe a model able to reveal what is common and conceal what is different in its components.

The lognormality of the residual ratio, $f^{y/x}$, is not required for the validity of the model itself. Because x and y are lognormal, it seems reasonable to expect that $f^{y/x}$ would bear the same multiplicative nature. But the validity of the ratio model is not dependent on any assumption about the nature of this log difference. Such distinction is important for understanding the distribution of ratios. We study this point in section 4.2.

Ratios as functional relations: As described here, ratios are functional relations. That is, they are not intended to explain one item in terms of the other one. Ratios yield a contrast between two items both affected by errors. Such a contrast measures how big are the discrepancies between the ratio components. Therefore, the above description is intended to assess deviations from standards, not to prediction.

3.2.2 Assumptions Underlying the Ratio Model

An usual topic in accounting literature is to call the attention for the assumption of strict proportionality underlying the use of ratios. Such a statement is descriptive. We could now enumerate the assumptions attached to the ratio model in a generative, rather than in a descriptive way:

1. Accounting items are final realizations of elementary random changes. Such changes, when expressed as proportions of the item they affect, are, in average, the same. This is the Gibrat law.

2. The elementary random changes leading to final realizations of accounting items are, when expressed as proportions of the item they affect, a sum of two components: One which affects in the same way all the items in the same report and another which is particular to each item.

As we remarked before, the normality of the process governing dz is not required as an assumption.

The advantage of using a generative description is that we can now develop models other then simple ratios which are also consistent with this basis. Here, ratios emerge as models obeying to the statistical or expected proportionality of random effects. Proportionality at the end of a growth process is just a particular consequence of a given generative mechanism. Which other models are allowed by such a mechanism?

3.2.3 Ratios With More Than Two Items

In section 3.2.1 we noticed that the ratio model emerges when we consider the common variability of two relative growth processes in a generative mechanism. By considering more than two items an obvious extension of the usual ratio emerges.

Let $x_1, \cdots, x_i, \cdots, x_M$ be the position of M items for firm j. $dx_{i\sigma}$ are random changes in x_i caused by σ, a common source of variability. Considering the way such common disturbance affects the relative growth which is about to generate the set of x_i we can say that

$$\frac{dx_{1\sigma}}{x_{1\sigma}} = \frac{dx_{2\sigma}}{x_{2\sigma}} = \cdots = \frac{dx_{M\sigma}}{x_{M\sigma}} \tag{8}$$

For example, we may want to consider two groups of items instead of two simple variables. Given $y_1, \cdots, y_k, \cdots, y_K$ and $x_1, \cdots, x_l, \cdots, x_L$ and reasoning in the same way as in previous section the mechanism described by 8 leads to the relation

$$\left[\frac{1}{K}\sum_{k=1}^{K}\log(y_k) - \frac{1}{L}\sum_{l=1}^{L}\log(x_l)\right]_j = \frac{1}{K}\sum_{k=1}^{K}\mu_k - \frac{1}{L}\sum_{l=1}^{L}\mu_l + \left[\frac{1}{K}\sum_{k=1}^{K}\varepsilon^k - \frac{1}{L}\sum_{l=1}^{L}\varepsilon^l\right]_j$$

Despite its outlook, this model is very simple and can be seen just as an expansion of equation 5. Here, instead of unique variables, expected values and residuals, we have averages of items for every j.

In ratio form the model would be

$$\frac{\prod_{k=1}^{K} y_{jk}^{1/K}}{\prod_{l=1}^{L} x_{jl}^{1/L}} = R \times f_j$$

that is, a ratio of geometric means of variables describes an effect common to them all in the same way simple ratios do.

As the estimators for the μ are the mean value of the corresponding log item, the above expression is easily computed just by subtracting to each log item its mean value and then averaging groups of items inside the same firm.

Multiplicative residuals: The above model leads to ratios like these ones:

$$\frac{\sqrt{CA \times CL}}{TA} \text{ or } \frac{X \times Y}{Z \times W} \text{ or } \frac{S}{\sqrt{N \times W}} \text{ or } \frac{A \times B \times C}{D^3} \text{ or similar.}$$

Notice that such ratios would immediately result from the usual ones when considering residuals as multiplicative. In fact, any expansions

$$\frac{y}{x} = \frac{y}{z} \times \frac{z}{x}, \qquad \frac{y}{x} = \frac{y}{z} \div \frac{x}{z}$$

or similar, only require to be statistically valid the multiplicative nature of residuals. For example, ratios are interpretable as a contrast between two deviations from size.

We must point out that conversely, additive residuals as those accepted in the literature and practice are not statistically consistent with the above expansions. If residuals were additive their spread would not retain its statistical nature after expansions such those displayed above. Well known expansions of this kind like the “Pyramid of Ratios” or the DuPont profitability triangle require to be valid the assumption of residuals being multiplicative.

Degrees of freedom involved: All the above explanatory models use only one degree of freedom. They are simple translations in log space. One free parameter is enough to account for a unique optimal value. Such an optimum is an estimator of a difference between two central trends. This fact has important implications for the assessment of ratio standards and the interpretation of departures from such standards.

The inclusion of more than one variable in each group will not account for more explained variability. The number of used degrees of freedom remains equal to one. We are still modelling a single parameter. However, more variables, if conveniently selected, can enhance the accuracy of ratios by a self-smoothing process able to make particularities cancel out. We explore this possibility in section 5.1.

A geometric interpretation of the common effect: Before finishing this section we review the ratio model using a geometric, more concrete, interpretation.

Ratios emerge as a consequence of dividing the log variability of items in two components and modelling the one which is common to both. This common component can be interpreted as the effect of the size of the firm.

If only such common source of variability would exist, firms belonging to homogeneous groups would be similar except for size. The values assumed by all accounting items would be just different magnifications of such an effect. In log space, distributions of accounting items would only differ by their position. Their shapes would be exactly the same. They would exhibit similar variance and when paired, they would also exhibit co-variance similar to the variance. Any two-variate scatter-plot would simply show a 45^o slope intercepting the y axis in $\overline{\log y} - \overline{\log x}$. This is an abstraction, of course, but it shows what a common component represents. If now we add some particular variability to each item we obtain the statistical behaviour of items as described here.

According to the developed models all accounting items should be seen just as different aspects of one unique underlying variable, the size of the firm. The particular behaviour of each variable emerges as a residual or deviation from the common variability.

3.3 Extending the Notion of Financial Ratio

The ratios introduced in the last section, despite their unusual outlook, are obvious applications to more than two items of the same principle governing the usual ones. In this section we extend the notion of financial ratio in two new directions allowed by the proportional mechanism. First we introduce non-linear proportions consisting of applying the linear model to the log space. Second, we model non-proportionality as part of the Gibrat law.

3.3.1 Non-Linear Proportionality

If we wish to model the joint behaviour of $1, \cdots, i, \cdots, M$ items after controlling for the common effect we must be able to account for differences amongst them other than the simple position or mean differences the usual ratios account for.

In order to do this we notice that the proportional mechanism is able to yield more complex relations than those developed above. Expression 8 is just the simplest case. Accordingly, we now develop similar models, but able, to some extent, to cope with the variability of individual items.

The introduction of multi-variance in the generating mechanism can be done with different degrees of complexity. The simplest approach consists of using just one parameter, b_i, individualizing each proportion. This new parameter allows the description, using the same formulation and without loss of generality, of the two components of the variability of each accounting item. A common effect would have $b_i = 1$ for all variables.

Similarly to 8, there is a relative growth, ρ, for which

$$b_1 \times \frac{dx_{1\rho}}{x_{1\rho}} = \quad b_2 \times \frac{dx_{2\rho}}{x_{2\rho}} = \quad \cdots \quad b_M \times \frac{dx_{M\rho}}{x_{M\rho}} = d\rho \tag{9}$$

In this mechanism, the b_i are gain or attenuation factors expressing different degrees of linear correlation between the generative growth rates leading to these variables. Notice that only $M-1$ of these b_i are independent.

In the simple case of b being similar across firms the consideration of two items, y and x, would yield general expressions like

$$\log y_j - b \times \log x_j = (C^y - C^x)_j + (\varepsilon^y - \varepsilon^x)_j - b \times \log b$$

or similar. Such free-slope ratios would look like this:

$$\frac{y_j}{x_j^b} = \exp(w_0) \times f_j$$

b and w_0 are now the two parameters of the model. w_0 is expressible in terms of b and the initial values C^y and C^x.

And when considering two groups of items instead of two simple items we would have

$$\frac{1}{K}\sum_{k=1}^{K} b_k \times \log(y_{jk}) - \frac{1}{L}\sum_{l=1}^{L} b_l \times \log(x_{jl}) = w_0 + \left(\frac{1}{K}\sum_{k=1}^{K}\varepsilon^k - \frac{1}{L}\sum_{l=1}^{L}\varepsilon^l\right)_j$$

or similar. w_0 is a parameter expressible in terms of the b_i and the initial values. In the form of ratio,

$$\frac{\prod_{k=1}^{K} y_{jk}^{b_k/K}}{\prod_{l=1}^{L} x_{jl}^{b_l/L}} = \exp(w_0) \times f_j$$

We can simply say that any multi-variate descriptor of this kind has, for $1, \cdots, i, \cdots, M$ items, a general form

$$\sum_{i=1}^{M} w_i \times \log(x_i) = w_0 \tag{10}$$

in which the residual is omitted. w_i are parameters expressible in terms of the b_i, M, and the initial values. In ratio form,

$$\prod_{i=1}^{M} x_i^{w_i} = \exp(w_0).$$

We can write 10 as a linear relation in log space

$$\sum_{i=1}^{M} w_i \times u_i = w_0 \qquad \text{where} \qquad u_i = \log x_i.$$

that is, in log space a simple inner product can account for linear correlations in the residual behaviour of accounting variables.

A class of non-linear ratios: At the beginning of this section we remarked that the free-slope ratios about to be introduced were intended to model relations between ratio residuals. We now discuss their direct use in the modelling of relations between items. What is the consequence of using free-slope ratios instead of the usual ones?

Firstly, let us notice that free-slope ratios contain the traditional ones. We can view financial ratios as a special case of the equation 10 in which $M = 2$ and $\{w_1 = 1,\ w_2 = -1\}$.

Free-slope ratios have some attractive features when directly modelling items: They preserve proportionality. They are also able to model a class of non-linear relations. Simple ratios only consider a linear proportionality in which the exponents affecting the items are the unit or a fractional — the inverse of the number of items gathered in the numerator or in the denominator. Here, we envisage a more general proportionality, not necessarily linear, in which items have real exponents. As long as we accept linearity in log space, proportionality will be preserved and inner products will model accounting relations.

However, free-slope ratios are no longer interpretable in terms of relative size and deviations from it. In log space, any slopes different from 1 — when relating to items, not to residuals — rely on simplified mechanisms difficult to accept and lead to functional relations with no intuitive meaning.

The evident functional relation linking all accounting items seems not to be a co-variance but a simple scaling — in log space, a displacement. Slopes diverging from 1 in log space mean a non-linear relation. As a rule, we shouldn't accept non-linearity before being compelled to reject more simple mechanisms. We based the assumption of $b = 1$ on findings described in chapter 2: For the observed items the log variance and co-variance are similar for homogeneous groups.

There is another reason for avoiding the use of more than one free parameter in financial models. Accounting data is fertile in irregularities capable of distorting the meaning of a model in which slopes are allowed to vary freely.

In figure 24 (left) on page 70 the non-linearity introduced by letting the slope of the regression in log space be different from 1, is clearly being used just to approach a few influential cases. These cases are marked with a plus sign and an arrow on the right of the same figure. In other words, the resulting model is now sample-dependent.

Using non-linear ratios: There are a few cases in which the extended ratios discussed above could be useful. The first one is when we want a proportion of size to be present in a simple ratio.

For example, we could form the ratio $\dfrac{y_j}{x_j^b} = R \times f_j.$

By selecting appropriate b we obtain a residual, f_j, contaminated with as much size as desired.

Another possible application is when we wish to introduce a second free parameter in the model because our goal is the prediction of y using x as predictor, not the assessment of a contrast between them. In such circumstances we don't need to be guided by the above, cautious, reasoning.

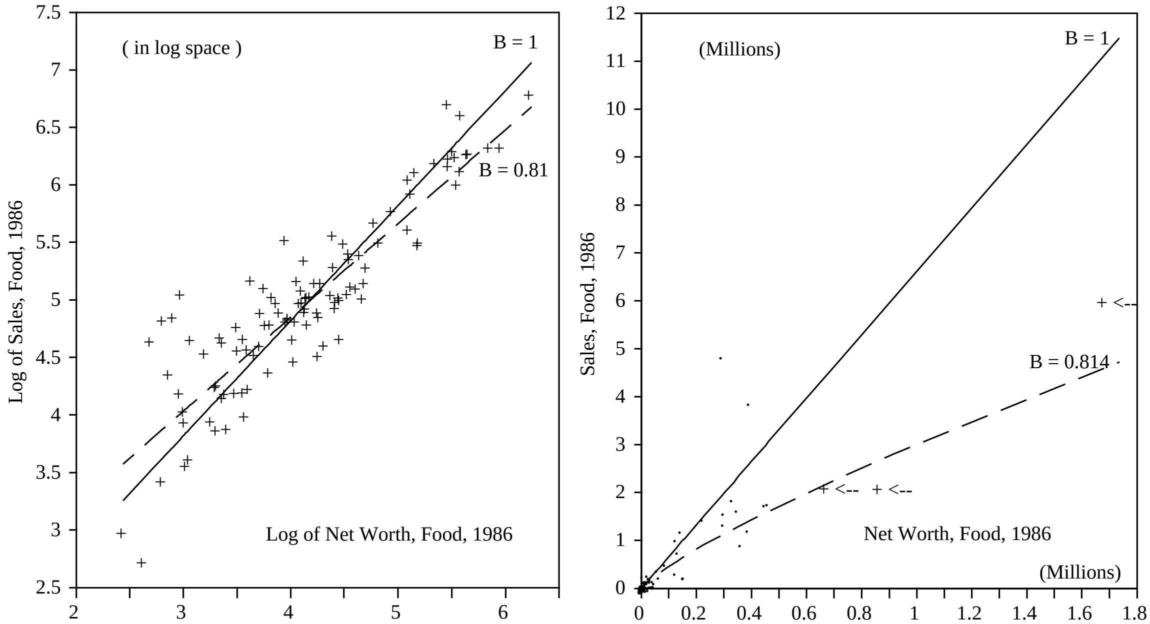

Figure 24: Comparing a regression in log space ($B = 0.81$) with the ratio model ($B = 1$).

Functional relations describe mechanisms. Mechanisms should be plausible. Free slopes in log space are not plausible for describing items since they imply the existence of a unique relative growth for the same item across many firms. Moreover, it would be inadequate to consider non-linear mechanisms as a rule. But when the goal is to predict one item from another, there are no known objections to the use of simple regressions in log space.

The class of non-linearity introduced by free-slope ratios: Figure 24 or any similar one will show that the distortion introduced in a two-variate model by allowing small departures from $B = 1$, affects mainly large firms. It will be a concavity towards smaller y, in the case of $B < 1$. And it will be a convexity towards larger values of y for $B > 1$.

The practical assessment of significant departures seems difficult since cases which could be interpreted as drawing these non-linear features also could be considered as influential.

3.3.2 Non-Proportional Ratios

The relation $dx/x = dz$ is a simplistic description of generative processes. The Gibrat Law allows a more realistic basis by admitting that the random changes dx affecting x are proportional, not to x itself, but to $x + x_0$.

We call this x_0 a base-line. Since the generative process leading to a particular realization of x starts with a non-zero value for $x = 0$ the increments x receives at this point are in average proportional to such base-line. Therefore,

$$\text{instead of} \quad dz = \frac{dx}{x} \quad \text{we should write} \quad dz = \frac{dx}{x + x_0}$$

for describing the generation of a particular item x.

Such a process leads to a class of ratios which can have many different characteristics according to the magnitude, sign and position of their base-lines. In some cases, but not in all, these base-line ratios will draw non-proportional relations between its components.

Notice that x_0 should not be taken as the initial value of x, that is, the value of x at the beginning of the process leading to its final realization. Such initial values — which in our notation are the C^x — will not induce non-proportionality in the models describing cross-sectional samples. As long as the process is strictly proportional, the outcome is proportional as well. Non-proportionality emerges only when the random changes dx are proportional to values which are not x.

Next we briefly describe some of the possible models resulting from base-lines.

An overall base-line in the denominator: In the simplest case, x_0 would be a constant value affecting all realizations of x_j for any j. That is, for a particular item all firms in the sample were expected to be affected by the same non-zero base-line.

One possible model resulting from a two-variate relation would be

$$\log(y_j) - \log(x_j + x_0) = \mu_{y/x} + \varepsilon_j^{y/x}$$

when the base-line acts on the denominator but not in the numerator. In ratio form,

$$\frac{y_j}{x_j + x_0} = R \times f_j$$

Base-lines occur when any of the ratio components is three-parametric lognormal instead of two-parametric. In the above expression and in all subsequent ones, the item affected by the base-line — in this case it is x — receives a transformation similar to the one used in formula 1 (page 8) for achieving three-parametric lognormality.

Estimating x_0: The problem of finding good estimators for x_0 can be approached in two ways:

- Firstly, this problem could be considered as concerning each item independently, as we did in chapter 1. Hence, the estimator for $\mu_{y/x}$, supposing that we know in advance δ_x, an estimated x_0, would be simply $\overline{\log y} - \overline{\log(x + \delta_x)}$. By considering the estimation of μ as independent of the estimation of x_0 we considerably simplify the formulation of estimators for base-line ratios.
- But we could as well find plausible reasons for considering these two parameters as not independent. It would be complicated to discover the analytical expressions for estimating x_0 and μ when considering their dependence. In practice, there is no problem in finding estimators for

these models. Any iterative Least-Squares algorithm, used in log space, will generally succeed in doing so.

Both the independence and the dependence models are interesting. They can explain different kinds of base-lines. But the independence model is attractive also because of its robustness. In this model only the small cases in the sample are affected by the introduction of a second parameter. For larger firms the model behaves exactly like the corresponding simple ratio. And the estimation of expected values is less affected by influential cases or non-desirable interactions between the ratio components.

Non-independent residuals: Returning to the above model — base-line in the denominator — it is clear that ratios of this sort are a non-proportional relation:

$$y_j = x_0 \times R \times f_j + x_j \times R \times f_j$$

The above form is useful just to show that such a model is not a linear regression. The non-proportional term $x_0 \times R \times f_j$ is not independent. It will introduce displacements proportional to a residual value. Distortions will vary from case to case.

The distortions introduced by this kind of ratio will be small provide $|\delta_x|$ remains small. The non-proportional term will be significant only for values of x_j near δ_x, that is, whenever the generative process leads to final realizations of items which are near their base-line. Cases far away from their base-lines exhibit proportionality since $x_j \gg x_0 \times R \times f_j$.

An overall base-line in the numerator: By considering a base-line, y_0 affecting y, the numerator of the ratio, instead of x, we get non-proportional terms which can more easily be significant. The expression

$$\log(y_j + y_0) - \log(x_j) = \mu_{y/x} + \varepsilon_j^{y/x}$$

means a ratio

$$\frac{y_j + y_0}{x_j} = R \times f_j$$

which can be written as

$$y_j = x_j \times R \times f_j - y_0$$

In this case the base-line acts as an intercept in a regression. It introduces a displacement affecting all cases in the sample. Notice that this model is still not a regression. The difference, however, is not functional. It stems from the multiplicative nature of the residuals.

Notice also that, except when assuming independence between the estimation of y_0 and μ, there are practical problems in modelling this relation directly with the usual algorithms.

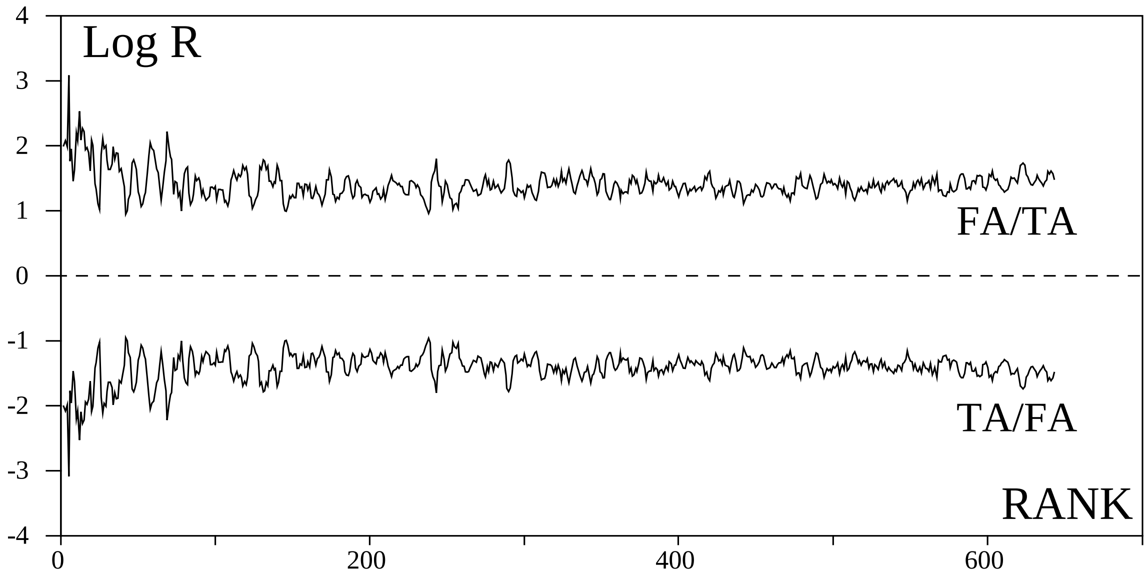

Figure 25: An example of the effect of a base-line in the denominator (TA/FA) and in the numerator of a ratio (FA/TA). The X-axis measures the rank of firms according to size. The Y-axis measures the ratio in logs. All groups, 1987.

Differences between numerator and denominator base-lines: When the base-line acts on the denominator its effect is not independent of the value of the residuals and other parameters. For instance, the overall distortion introduced depends on the value of $R_{y/x}$, the expected proportion between the numerator and the denominator. On the contrary, even a small y_0, the base-line in the numerator, seems able to induce significant displacements. Therefore, it is possible to have a ratio, y/x, exhibiting only residual non-proportionality and just by inverting it to get another ratio, x/y, now exhibiting a strong departure from proportionality.

This behaviour is expected. In log space both the distortions introduced by y/x or x/y are symmetrical about 0 and have exactly the same magnitude for equal base-lines. But since the one originated in the denominator is a distortion acting towards the range of negative log values, it will induce, when in ratio form, departures from proportionality limited by $\{0, 1\}$. Differently, the departures observed in the numerator will be, when in ratio form, directed to span the interval $\{1, \infty\}$ since they come from the positive region of the log space. There is the difference.

Figure 25 shows the distortions observed in two symmetrical ratios (FA/TA and TA/FA) by significant base-lines. The X-axis is the rank of the cases in the sample when sorted by TA in ascending order. The Y-axis is the value of the ratio in log space.

A ratio for which proportionality holds will induce an horizontal trend. The expected value for the ratio is the same no matter what value is expected for TA. On the contrary, in this case the smallest firms in the sample clearly break the assumption of proportionality. This is because of a significant base-line affecting one of the ratio components.

The scatter-plot described here is valuable in detecting base-lines. It is just a practical application of the usual plots for tracing possible trends in residuals. It is convenient to apply a small amount of smoothing for improving its clarity. In the above example we used moving averages.

Base-lines both in the numerator and in the denominator: When considering y_0 and x_0 as both significant, the amount of non-proportionality introduced results from their interaction. A reinforcement of non-proportionality will occur when y_0 and x_0 have different signs. Apart from this, the overall effect will depend on R, the expected proportion.

In a very particular case, $y_0 = x_0 \times R$, both base-lines cancel out. The remaining non-proportionality is residual.

Multi-variate base-line ratios: The general multi-variate descriptor, involving free-slopes and base-lines affecting all cases and present in several items would be written as

$$\sum_{i=1}^{M} w_i \times \log(x_i + x_{0i}) = w_0$$

or, in ratio form,

$$\prod_{i=1}^{M} (x_i + x_{0i})^{w_i} = \exp(w_0)$$

It is expected that multi-variate models of this sort will eventually generate strong departures from proportionality. The x_{0i} can easily reinforce their effects creating important joint displacements.

Proportional base-lines: The mechanism leading to the above descriptors requires an overall displacement — a base-line acting upon the whole of the sample in the same way —. Overall base-lines suppose the existence of overall costs or income.

We now consider the case of base-lines which are dependent of the size of the firm. Mechanisms internal to the firm are likely to generate base-lines proportional to size. The assumption of such internally generated base-lines being similar for the whole of the sample would be difficult to accept.

For $1, \cdots, j, \cdots, M$ firms, x_{0j} is now a particular base-line concerning the generative process of each x_j. This base-line will act as a new variable, not as a parameter of the model.

The model collapses into the no-base-line ones. In fact, if x_{0j} is proportional to the size of the firm it is similar to any other accounting item. For instance we could write $x_{0j} = x_j \times R_{0j} \times f_{0j}$ and we would have a relative growth

$$\frac{dx}{x \times (R_{0j} \times f_{0j} + 1)} = dz$$

for the generating process of a particular realization of x.

And since R_{0j} and f_{0j} are not involved in the subsequent growth of x the resulting model would be a version of the free-slope ratio we explored in section 3.3.1.

Base-lines proportional to the size of the firm will not break proportionality. However, they will induce differences in the way each item is affected by the common variability. In order to account for such differences, mechanisms similar to free slopes are required.

The described model is interesting because it has been often used in the accounting literature as an example of the plausibility of intercept terms in two-variate relations. It was an awkward choice since, as we see, base-lines acting just as another item are not likely to induce overall translations. We now analyze this subject in more detail.

3.4 The Basis for the Existence of Non-Proportionality

Base-lines can occur due to internal or external causes. The most general one is internal: when non-existent variables ought to exist. All growth processes starting with $x = 0$ must have at its origin a base-line acting like a seed since zero-valued items can't grow. This is the basic reason for the existence of base-lines. It applies to all growth processes of the kind we consider here. And it certainly applies to accounting items as well.

Due to its exponential nature, the final realizations of proportional mechanisms are likely to attain values which can be many orders of magnitude larger than this seed. In such cases, $x - x_0 \approx x$ and the non-proportional term vanishes. Growth processes in which the final realization is not far from the base-line would generate non-proportional terms which would not vanish.

3.4.1 Internal Base-Lines

The foundation invoked in some literature for the existence of significant departures from proportionality is coincident with the model we call the proportional base-line. Foster [44], for example, referring to the Earnings-to-Sales ratio, explains that

> One rationale for a negative constant term is the existence of fixed costs, which implies a loss at zero sales level. One rationale for a positive constant term in the earnings-sales relation is an income source (for example, interest income on cash investments) not related to sales.

Lev and Sunder — apart from suggesting a different sign for the base-line — argue in the same way:

> The relationship between gross profit (y) and sales (x) probably contains a positive constant term given the frequent existence of a significant fixed costs component. Accordingly, observed differences in gross margin ratios (over time or across firms) will reflect the confounding effects of differences in efficiency, reflected by β, differences in the level of fixed costs, α, and differences in sales volume, x.

This literature says that, because in each individual firm some internal mechanisms exhibit constant terms, the corresponding statistical variables, obtained when gathering many firms in a sample, would exhibit also a constant term.

It is worth emphasizing that, as we saw above, this is not so. The base-lines described by the above mechanisms could induce particular correlations between items instead of overall displacements.

This seems to be another case of picturing time-series while working with cross-sections. Returning to the example of Fixed Costs, the meaning of a cost being fixed is that it is fixed inside a firm. But a cost can be fixed inside a firm and variable across firms. As a first approximation, we can say that large firms exhibit large fixed costs, small firms exhibit small fixed costs and infinitesimal firms would exhibit infinitesimal fixed costs. On the limit, the zero-sized firm would have zero fixed costs thus yielding strict proportionality. Whittington [134] clearly distinguishes between time-series and cross-sections when addressing this problem:

> In cross-section, such an interpretation (sales-unrelated income) could not be placed on the constant term: It would now represent an estimate of the average amount of sales-unrelated income for the average firm, provided the further assumption is made that "sales-unrelated income" is strictly independent of size.

This statement is equivalent to ours. In cross-section Fixed Costs should be regarded as another item with nothing very special about it.

3.4.2 Overall Base-Lines

We noticed that mechanisms which are internal to the firm would not yield intercept terms. Only overall base-lines can do it. The question now is: Can such overall base-lines plausibly induce strong intercepts? It seems as if there is a limit for the plausibility of overall translations affecting entire samples. In fact, such translations must be small because they have to impinge upon all firms, small or large. And being small, they will be entirely un-noticed by large ones.

Clearly, if an overall cost were big enough to promote a significant displacement it would be far greater than the earnings of many firms — leading them to immediate insolvency. And if it were small enough to allow any firm to survive then it would not be noticed by most of the firms and indeed its effect would be negligible.

For example, a fixed cost of 3,722 thousand pounds over the whole of the Food Manufacturers in the U.K. would represent to UNITED BISCUITS just 0.2% less earnings in 1987. This firm would have to be content with 99.8%. But such a cost would eat up the whole of sales in G. F. LOVELL PLC. All the firms similar or smaller in size would perish (about 5% of the industry). This is not conceivable. The acceptance of overall displacements able to influence large firms would imply the existence of unreasonable or impossible mechanisms.

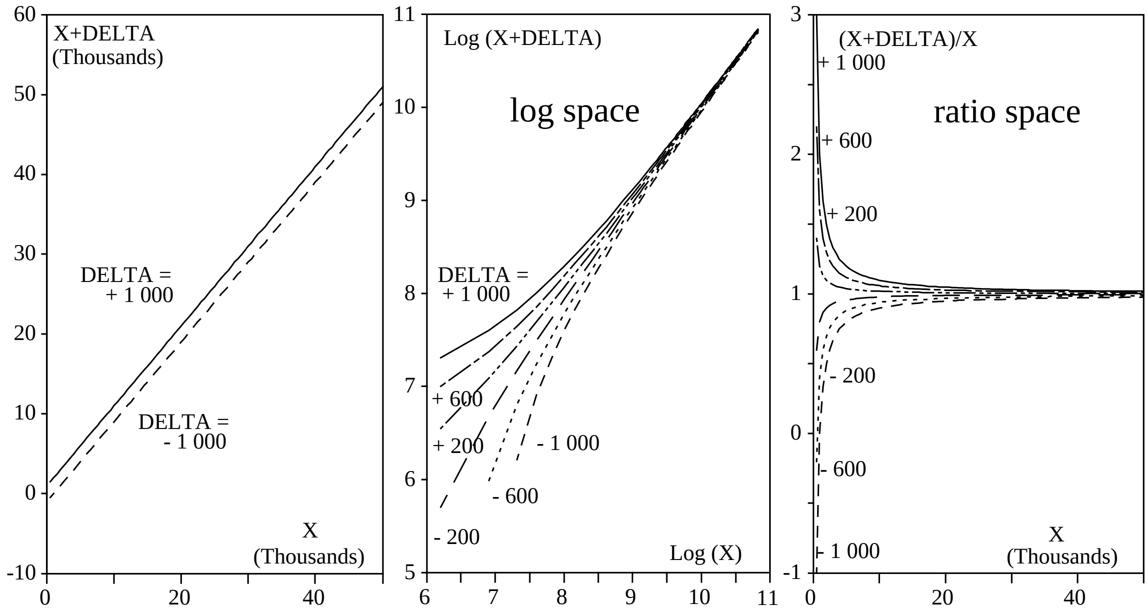

Figure 26: When $Y = A + X$ is transformed, the fact that $A \neq 0$ introduces non-linearity in the resulting relation. Such non-linearity affects only values of Y near A. On the left several $Y = A + X$ slopes with very small A. In the centre the same slopes in log space. On the right, in ratio space.

3.4.3 Assessment of Overall Departures From Proportionality

In order to assess the significance of overall departures from proportionality it is important to gain insight into the way the introduction of a constant term affects the linearity, in log or ratio space, of an otherwise proportional relation. By applying log or ratio transformations to both sides of $Y = A + X$ and observing the distortions resulting from increasing the value of A, we can acquire a precise idea of the impact of deviations from strict proportionality.

Figure 26 on page 77 shows the results of applying logs (centre) or dividing by X (right) in both sides of $Y = A + X$ (left) for small values of A — thus obtaining relations which are formally similar to the above non-proportional models.

Those considered A are $\pm 1,000$, ± 600 and ± 200. For large X, the effect of introducing such intercept terms is negligible. The effect of A becomes significant and visible whenever the order of magnitude of the X is similar to the order of magnitude of A.

Accordingly, base-lines must be taken into account only when the final realization of a growth process, x, is not far away from x_0. This could happen when the growth is weak (very small relative growth and very few random changes). Please compare figure 26 (right) with figure 25 on page 73. The last one is a practical realization of the former.

Traces of non-proportional relations in log space: The examination of two-variate scatter-plots of accounting items in log space can thus detect departures from strict proportionality when

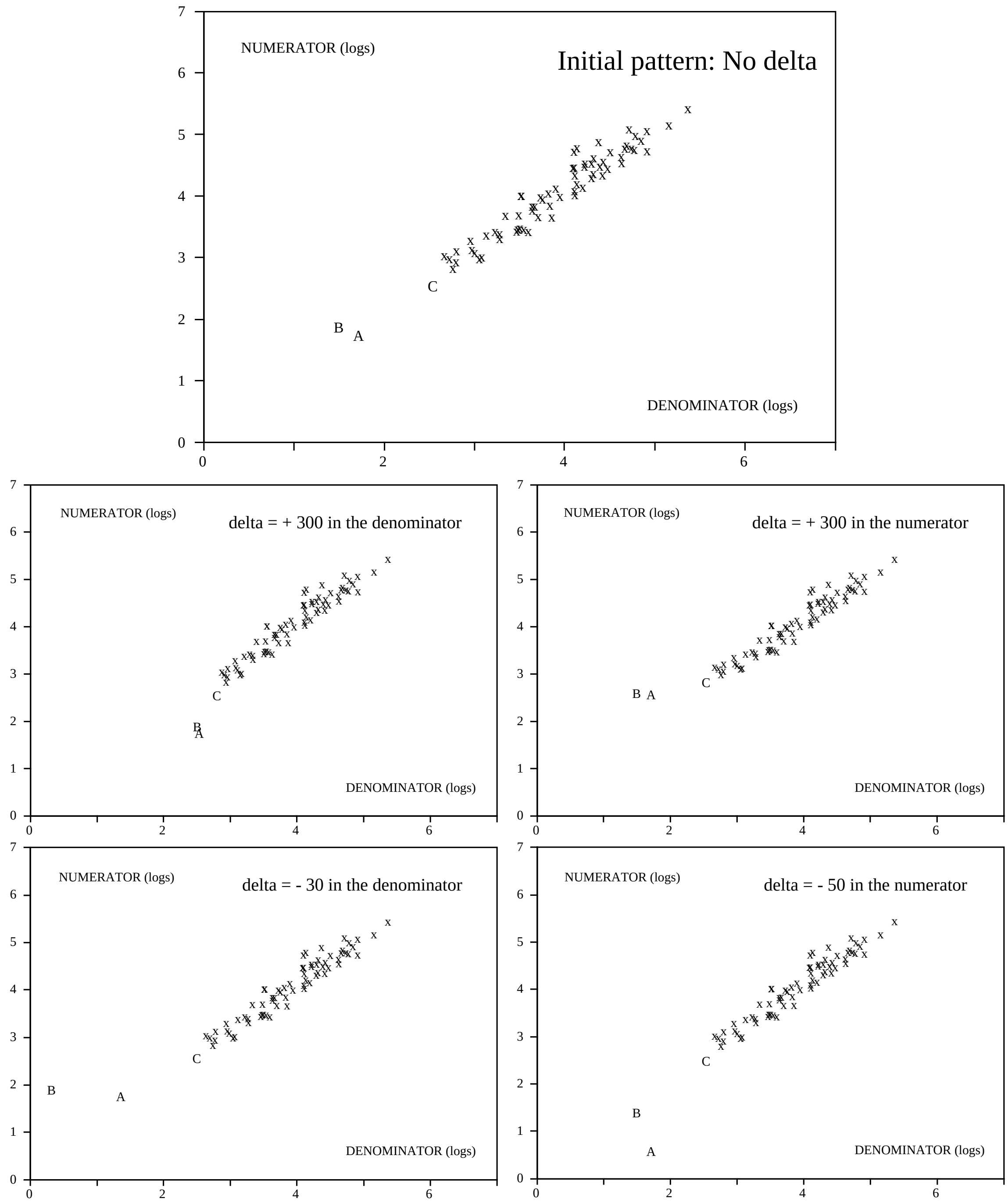

Figure 27: Overall base-lines introduce non-linearity in log space. Four possible cases. Simulated data. Above, the original scatter with no base-line effects introduced.

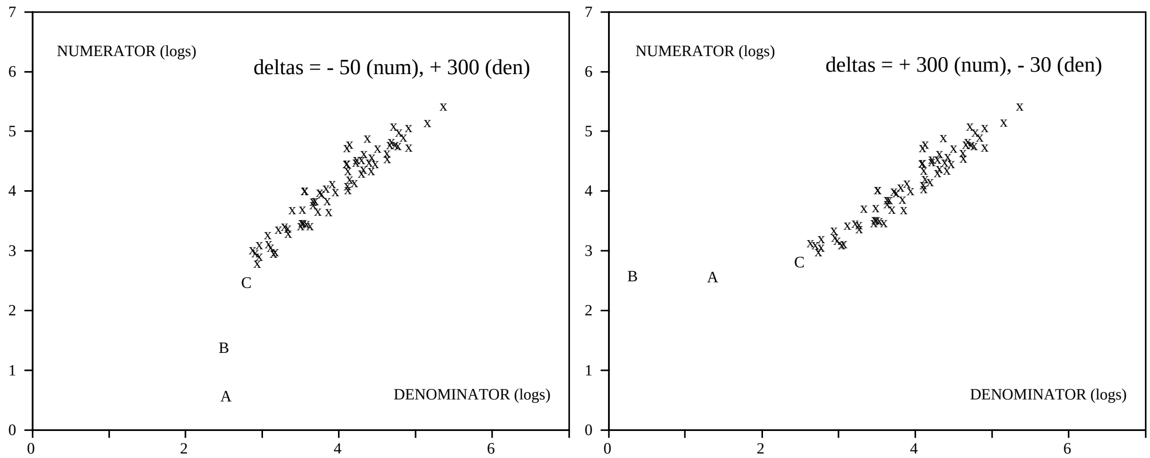

Figure 28: Overall base-lines introduce non-linearity in log space (Cont.). Two more cases. Simulated data (continuation).

they turn out to be significant. As seen above, the log transformation — and also the ratio one — produces a trade-off between non-proportionality and non-linearity so that even small departures from proportionality result in clear departures from linearity. Figure 8 on page 31 is an example.

Figures 27 and 28 on pages 78 and 79 systematically explore the different possible patterns observable in scatter-plots in log space when overall base-lines are in the numerator or in the denominator of ratios. The Y-axis assess the numerator and the X-axis the denominator of the ratio. The data is a set of 71 cases simulated so as to replicate the ratio CA/CL using log mean values and correlation similar to those found in the Electronics industry (1987):

ρ	$=$	0.980			
$\overline{\log CA}$	$=$	4.343	$\min CA$	$=$	53
$\overline{\log CL}$	$=$	4.192	$\min CL$	$=$	32

The base-lines (δ in the figures) were introduced in the simulated data after the generation of correlated variates. In appendix A we use this sample to display some methods available for the estimation of base-lines.

The observed shapes agree with those of figure 26. They can be used for gaining insight into the effect of overall displacements in the small firms's deviations from standards.

3.5 Using Extended Ratios

In this section we show examples of the applicability of the devised extended ratios. Two items are used that often exhibit significant overall base-lines. They are Current Assets and Earnings. Both samples belong to the Electronics industry. We formed the ratios FA/CA and $S/EBIT$ and explored their behaviour.

As usual, x represents the denominator of the ratio and y the numerator. For each example, five different models are compared: Model 1 to 5. The numbering refers to figure 29 on page 81 and to tables displayed later on. Next we describe each model.

Model 1: The usual ratio. It engages one degree of freedom. Its unique parameter, the median of the ratio in log space, is estimated as a in the equation

$$\log y_j = a + \log x_j + \varepsilon_j^{y/x} \qquad \text{yielding the ratio} \qquad \frac{y_j}{x_j} = 10^a \times f_j$$

The graphical representation of this ratio has the label 1 (solid line) in figure 29. It is a 45^o straight line in log space. In ordinary space it is also a straight line passing through the origin.

Model 2: The free-slope ratio. It engages two degrees of freedom. Its two parameters are a and b, the slope. They are estimated by the equation

$$\log y_j = a + b \times \log x_j + \varepsilon_j^{y/x} \qquad \text{yielding the ratio} \qquad \frac{y_j}{x_j^b} = 10^a \times f_j$$

The graphical representation of this ratio has the label 2 (dotted line) in figure 29. It is a straight line in log space. It is not linear in ordinary space. It goes through the origin.

Models 3 and 4: The base-line ratios. They engage two degrees of freedom. But in the former both parameters (a, and δ) are estimated jointly. In the last, a is taken as known. Then δ is estimated based on this assumption. The slope is not allowed to vary freely. The parameters are estimated in the equation

$$\log y_j = a + \log(x_j + \delta) + \varepsilon_j^{y/x} \qquad \text{yielding the ratio} \qquad \frac{y_j}{x_j + \delta} = 10^a \times f_j$$

The graphical representations of these ratios have the labels 3 and 4 (dashed lines) in figure 29. They are non-linear in log space and straight lines in ordinary space. They don't go through the origin. In log space both models yield curves which are parallel to each other. Model 4 converges to model 1 for medium-sized and large firms.

Model 5: The base-line plus free-slope ratio. It engages three degrees of freedom. The parameters are a, b and δ. It is the result of considering free slopes and δ together. The parameters are estimated in the equation

$$\log y_j = a + b \times \log(x_j + \delta) + \varepsilon_j^{y/x} \qquad \text{yielding the ratio} \qquad \frac{y_j}{(x_j + \delta)^b} = 10^a \times f_j$$

This ratio is not displayed in figure 29.

The above description is complemented with figure 29. It represents graphically the typical shape of each model. In this representation there are three plots. Above, the models in log space. The

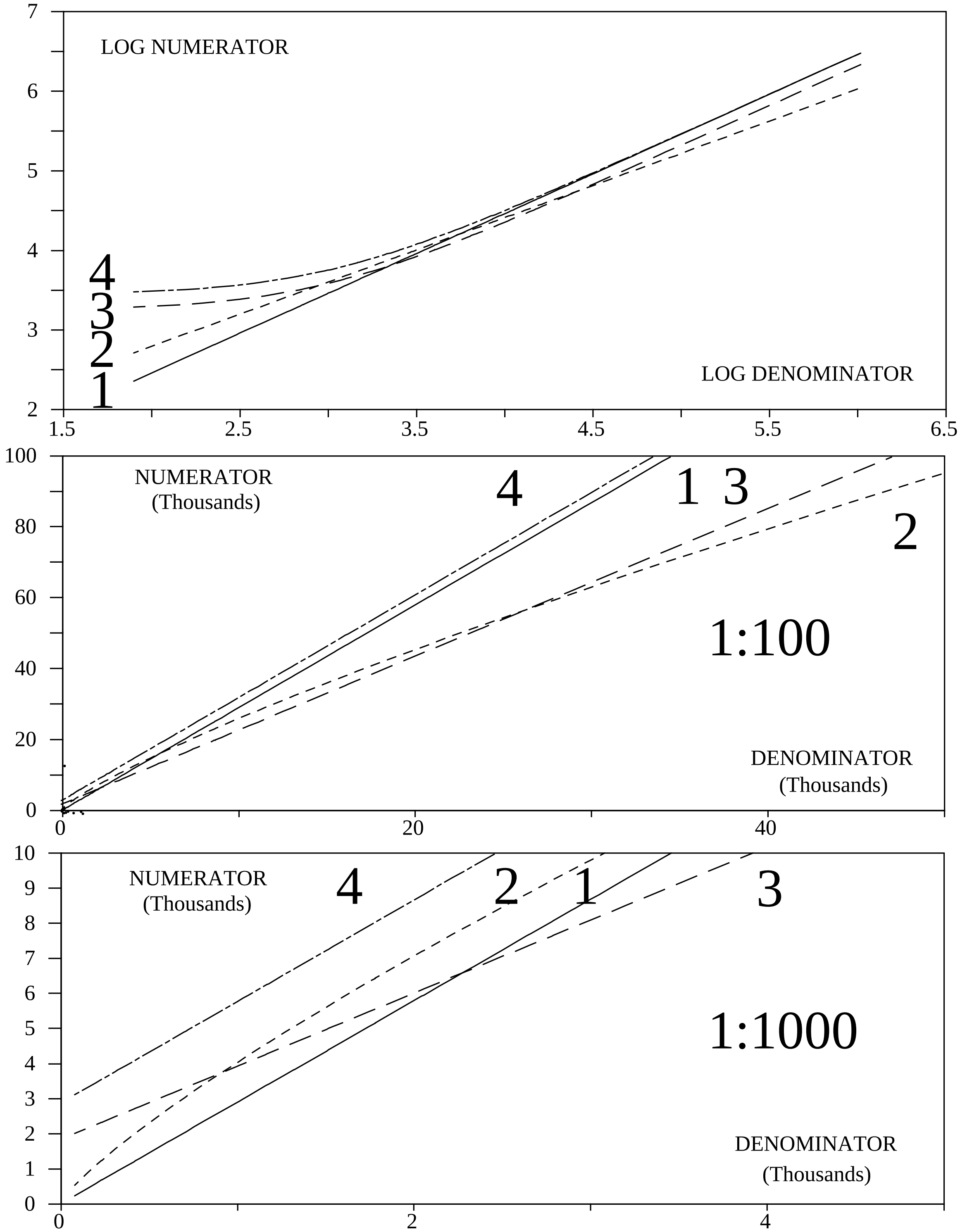

Figure 29: The typical shape of extended and base-line ratios. Above, log space. Below, two magnifications of the region near the origin in ordinary space. (1) is the financial ratio. (2) is a free-slope ratio. (3) and (4) are base-line ratios.

two plots below this one represent each model in ordinary space. They explore the region near the origin using two magnifications.

Methods similar to those discussed here but using simulated base-lines can be found in appendix A. Some important practical details to bear in mind when attempting to estimate base-lines are only referred to there.

3.5.1 Fixed Versus Current Assets

For some industries, CA generates clear base-line effects when combined with other variables. We selected Electronics, 1987, and tried several combinations of items. Finally we decided to use FA as the numerator. Fixed Assets is a complementary item to Current Assets. No external constraints resulting from accounting identities will distort this relation.

Next table relates the tested models to the variability explained, the skewness and the kurtosis of the ratio residuals in log space (the $\varepsilon_j^{y/x}$).

Model	a	b	δ	R^2	Skewness	Kurtosis
1	-0.46			72%	0.27	2.18
2	-0.21	0.94		76%	0.07	1.30
3	-0.49		+528	78%	-0.12	1.19
4	-0.46		+327	76%	-0.12	1.78
5	-0.40	0.99	+403	78%	-0.12	1.19

We now observe the features of each one of the obtained models. Model 1, the simple financial ratio, accounts for 72% of the variability. By letting the slope b vary freely in log space, further 5% of the variability are explained. Both the skewness and the kurtosis of the residuals are improved. In the 3^{rd} and the 4^{th} models the gain in explained variability is similar to the one obtained with free slopes. The explained variability didn't improve in the 5^{th} model.

Discussion: In this example, the variability explained is smaller than the usual in ratios. It is common to find ratios able to explain between 80% and 95% of the variability on its components. This is because, apart from Debt, FA is the item with the largest amount of unique variability.

The base-line ratio with two free parameters (the 3^{rd}) explains as much variability as the three-free-parameters model (the 5^{th}). It is interesting to notice that the free-slope model (the 2^{nd}) is using it to approach the non-linear effect of the base-line. Once such base-line is accounted for, the slope returns to the value of 1 (the 5^{th} model).

The value of the base-line is itself very small and its effect will vanish except for the smallest firms in the sample. When using the method described in chapter 1 to estimate the base-lines in CA and FA we obtained values which agree with those displayed in the above table. CA is three-parametric lognormal: Significant departures from normality vanish for $-580 < \delta < -300$. The maximum W is obtained with $\delta = -570$. FA has no significant departures from the two-parametric hypothesis.

The skewness of residuals is not very strong in this example. And it is clearly controlled by the introduction of a base-line. The kurtosis is strong and it will not vanish with base-lines or free slopes, as expected. In chapter 4 we discuss the distribution of ratio residuals.

From the four extensions of simple ratios we present here, the most attractive one seems to be the 4^{th}. It accounts for the base-line but approaches the usual ratios for larger values of its components. Such a independently-estimated base-line ratio is also simple to implement. It will exhibit the same kind of robustness regarding the influence of particular cases or external constraints simple ratios have. And it seems able to account for a significant increase in explained variability. The estimator for δ produced by this model is smaller than the predicted by the other ones.

3.5.2 Earnings Versus Sales

The relation between Earnings and Sales has been frequently quoted as an example of the plausibility of an intercept term. We examined such a relation for all industrial groups introduced on page 5 (chapter 1), for a five-years period (1983-1987). In most of the industries no traces of non-proportionality were found. In others, the apparent effect of an overall base-line turned out to have a very different explanation.

We first selected the Food Manufacturers group in 1987. Figure 30 shows the log scatter-plot of Earnings versus Sales for this group. It seems a case of non-linearity consistent with an overall base-line. However, four of the firms, the smallest when measured by Sales, have Earnings larger than Sales and shouldn't be considered as typical industries. These cases are marked with a "o" instead of a "x" in figure 30. Without such cluster the non-linearity seems much less impressive.

This example shows how vulnerable and potentially misleading the financial relations become if more than one free parameter is used.

Figure 31 on page 85 shows some other industries. The visual examination of samples in log space and the modelling of base-lines were particularly difficult in this case because one of the components of the ratio is bounded by the other.

In the Electronics industry, a small departure from proportionality could be observed in 1986 and in 1987. The results of applying extended and base-line ratios to the 1986 sample are displayed in next table. We modelled the ratio $S/EBIT$, not the $EBIT/S$ one.

Model	a	b	δ	R^2	Skewness	Kurtosis
1	1.03			77%	0.73	0.99
2	1.39	0.90		78%	0.49	0.66
3	0.97		+273	79%	0.54	0.70
4	1.03		+148	78%	0.62	0.84
5	1.10	0.97	+207	79%	0.50	0.68

The same features observed for CA are replicated here but not so strongly. The improvement in explained variability over the usual ratio is not significant. When the base-line is estimated independently of the median, its value is smaller than the one for joint estimations.

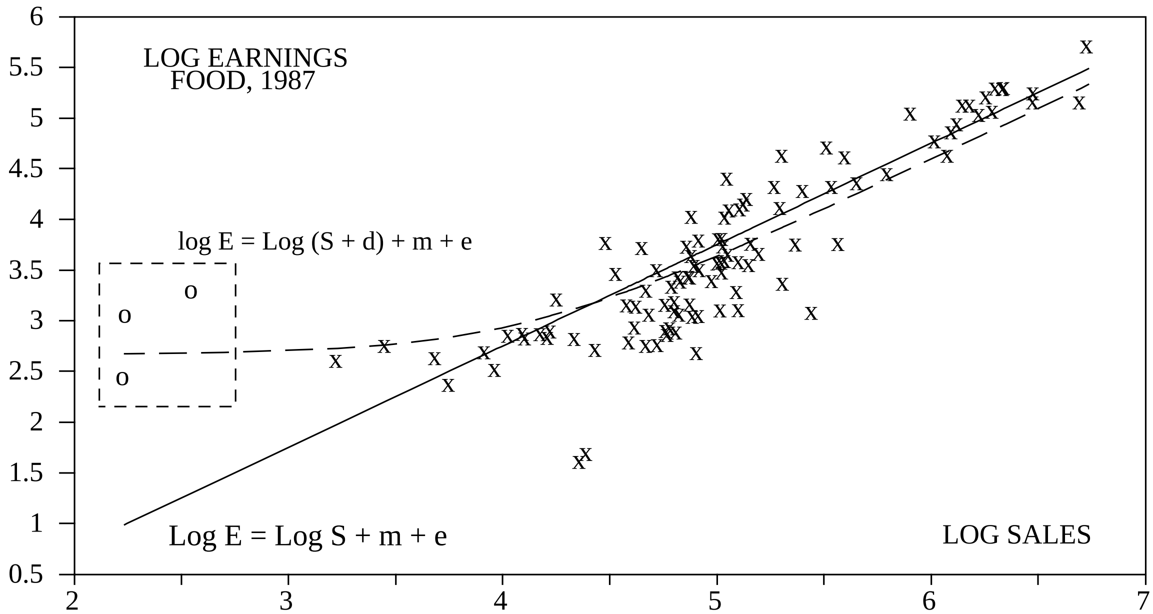

Figure 30: Log Earnings versus Log Sales for the Food Manufacturers, 1987. The cluster inside the dashed square has $EBIT > S$.

3.5.3 Other Non-Linear Relations Between Items

Apart from the mechanisms described above, non-linearity could emerge in accounting models due to other causes. Two of them seem plausible:

Higher order effects: It could occur, for example, when modelling financial risk. Leveraged and non-leveraged firms can behave in opposite directions if they belong to some specific industries. In this case, the significant interaction would emerge owing to the presence of two groupings: Financial structure and industry. A statistical version of the "exclusive-OR" problem, that is, a second-order effect, can arise when modelling an outcome for more than one grouping. Such possibility is important when using linear techniques. Linear tools wouldn't be able to separate effects other than first-order ones. In chapters 5 and 9 we explore this possibility.

Non-proportional non-linear relations between variables: Proportional non-linearity is just a particular class of non-linear relations between items. It would require the use of free-slope ratios instead of the usual ones. However, many other kinds of departures from linearity are possible.

Whittington [134] reports quadratic relations in profitability ratios. He suggests that this could be explained by saturation effects. Saturation is the kind of distortion free-slope ratios could broadly model since it affects mainly the largest firms in the sample in a way similar to free-slope ratios do — with $b < 1$, as in figure 24, page 70 —.

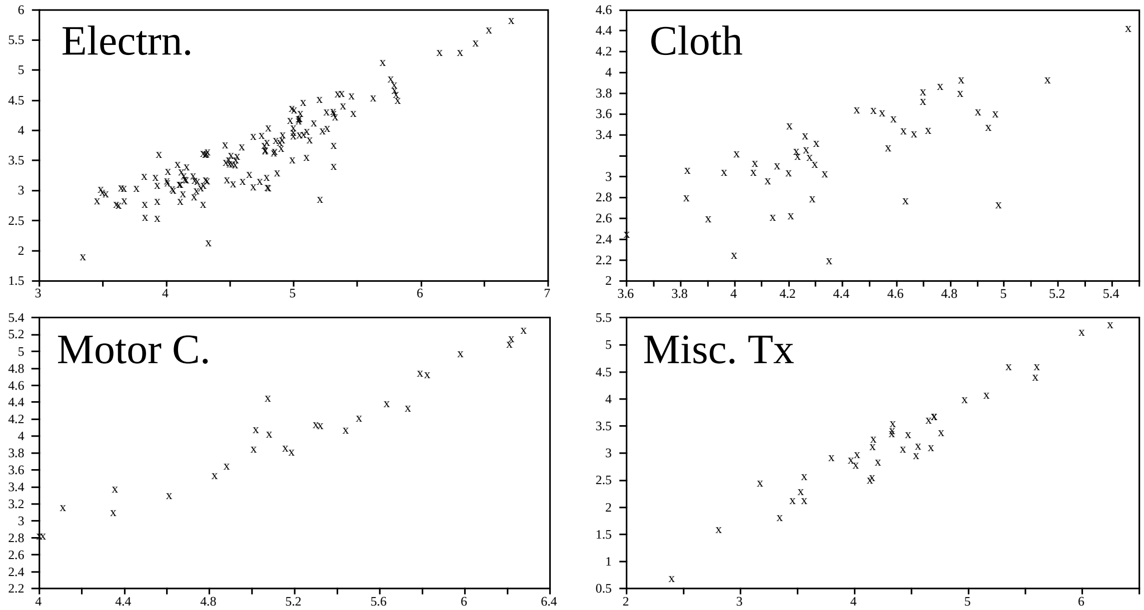

Figure 31: Earnings versus Sales for four industries in 1987. The X-axis is $\log S$ and the Y-axis is $\log EBIT$.

However, non-linearity affecting, for example, the smallest cases in a sample, wouldn't be modelled by free-slopes. Notice that base-lines represent linear displacements. They only yield non-linearity in log or ratio space. And it is possible that, along with genuine base-lines, other influences affect small firms.

Existing accounting statistical models seem not to be aware of potential sources of non-linearity. Base-line and saturation effects are not very imposing and the second order effect can be avoided by increasing the dimension of the input space, which accounting models implicitly do.

3.6 Discussion and Conclusions

We described a generative mechanism for the probability distribution observed in our data. The traditional notion of financial ratio stems from considering a common relative growth impinging over the two items forming it.

Ratios can be extended in several ways consistent with such a mechanism. Firstly, they can have more than two components. The sole requirement for the statistical validity of such ratios is the use of multiplicative residuals. Ratios can also be viewed in log space as a regression. Such free-slope ratios preserve proportionality. They introduce non-linearity in the large firms in the sample.

Finally, the existence of base-lines in the generation of items will eventually introduce non-proportional relations between the components of a ratio.

The reasons generally invoked for expecting significant intercepts don't lead necessarily to the emergence of non-proportional relations. Only an overall cost or income impinging upon the whole of the sample seems able to yield non-proportional relations in cross-sectional samples. This overall base-line couldn't be far away from the smallest case in the sample though. And, in that case, the effect of such a translation wouldn't be noticeable except in its neighbourhood.

Distortions in proportionality resulting from overall base-lines depend on several factors. They are maximal for base-lines in the numerator of the ratio or when the signs of the base-lines of the numerator and the denominator are different.

The validity of ratios: We conclude that it is not necessary to abandon the basic notion of ratio to model non-proportional and a class of non-linear relations. Ratios can be extended so as to include these features. This isn't the same as saying that the simple ratios are always acceptable.

Base-line ratios seem promising for ratio analysis and statistical manipulation of accounting data. They are robust, easy to estimate and it is likely that they will be able to gather in one simple model the correct relation between two items for firms of very different sizes.

This study also supports the idea that the use of ratios was not just an arbitrary choice amongst many possible ones. Ratios are consistent with a trend towards lognormality. They are, however, too simple relations having this quality. Since the non-linearity modelled by free-slope ratios mainly concerns the largest firms and the non-proportionality generated by overall base-lines affects only the smallest ones, it seems as if there is room left amongst the medium-sized firms for the traditional, non-extended, ratio to be used more or less accurately.

It is worth emphasizing that all the developed models require the acceptance of the multiplicative nature of deviations from standards.

The proportional mechanism: Discussion. As referred in chapter 1 lognormal distributions are not a mandatory outcome of the proportional mechanism. The Gibrat law is at the origin of a whole class of positively skewed distributions of which the lognormal is just one member. Also, the proportional mechanism is not the only one capable of producing lognormal variables. However, the other known mechanisms are not plausible for reflecting growth processes. Aitchison and Brown [1] offer a more detailed development of these topics.

As mentioned, we are not using the Gibrat law for explaining the distribution of particular items. We don't think Fixed Assets is more lognormal than Earnings owing to the proximity of the former process to the assumptions of the Gibrat Law. We picture proportionality as a stochastic effect present in all items: Not as a mechanism internal to the firm but as a mechanism explaining differences in size observable in cross-sections containing many firms.

The proportional mechanism is clearly not intended to describe the history of particular items. It strictly applies only to positive stocks in periods of exponential growth. Our choice of the Gibrat law for explaining and exploring the empirical findings of previous chapters could be summarized by saying that it is consistent with lognormality, the common effect and the existence of base-lines. It is also plausible as a mechanism underlying any growth process. Therefore it fits in the role assigned to it, the one of ensuring the consistent development of models based on previous findings.

The common effect: Discussion. The developments presented in this chapter are consequences of empirical observations. They are in line with previous research carried out mainly by Whittington [134] and Lev and Sunder [79] who pointed out the limitations of ratios. By Barnes [5] who stressed the potential applicability of models in which items could be used directly. And by McLeay [86] who described the lognormality of some items, amongst others. These authors open up a path for viewing more general models behind ratios.

However, the central assumption of this study is likely to eventually cause surprise. We based our approach on the existence of a common effect. According to it, all the items belonging to an accounting report should be expressed in terms of the same effect, size, and deviations from it. We have shown that non-proportionality is compatible with this assumption. The common effect greatly simplifies the formal treatment of the general problem of modelling with lognormal data.

The extension of such an effect to items like Working Capital, Earnings and Funds Flow is not usual — it has not been explicitly denied either —. The common effect itself is also a somehow new way of referring to size. Our empirical findings dismiss any strong statistical differentiation between items which are accumulations and those which are not. Items are lognormal because they reflect size, not because their internal generative mechanisms lead to lognormality.

The effect of size is clearly present in the observed items. Only Long Term Debt shows a less strong but not negligible correlation with size. In our opinion, the assumption of independence from size is not tenable. If Earnings were independent from size there would be no place in the economy for firms other than those having a particular dimension. We think that only firms in distress can exhibit accounting features not correlated with their sizes. Firms are so distinct in their sizes that all their accounting features have to reflect, to a smaller or larger extent, this basic characteristic.

Chapter 4

External Constraints and the Cross-Sectional Distribution of Ratios

The residuals of ratios, the $f^{y/x}$, should be positively skewed due to their multiplicative nature. The same should happen for the ratio output, $R_{y/x} \times f^{y/x}$. In practice, some ratios are Gaussian or even negatively skewed. How is this possible? And which causes influence the distribution of ratios?

This chapter applies the findings of the previous ones to the problem of the distribution of ratios. Our aim is to find the rules governing them. We base our approach on the models introduced in chapter 3 and on the effect of external forces.

Previous research and contents: We introduced the literature on this subject early in this study (see chapter 1). The problem of the distribution of ratios has suffered from the same drawback other problems involving ratios suffer. It concerns a vast amount of different situations. There are many possible ratios, many possible choices for the definition of samples, many tests and criteria for analyzing the results.

The studies on the distribution of ratios typically try to avoid dispersion by using Deakin's set of 11 ratios. Despite this effort, the results turn out to be difficult to interpret. Apart from the positive skewness — which is not a general rule — no widespread behaviour was found in the cross-sectional distribution of ratios. Ratios seem able to assume any possible distribution, from an extreme positive skewness consistent with their multiplicative nature to an almost perfect normality and even a negative skewness.

Indeed, just by observing ratios, it is very difficult to discover the rules governing their distribution. As stated at the introduction to this study, our method consists of reckoning that ratios

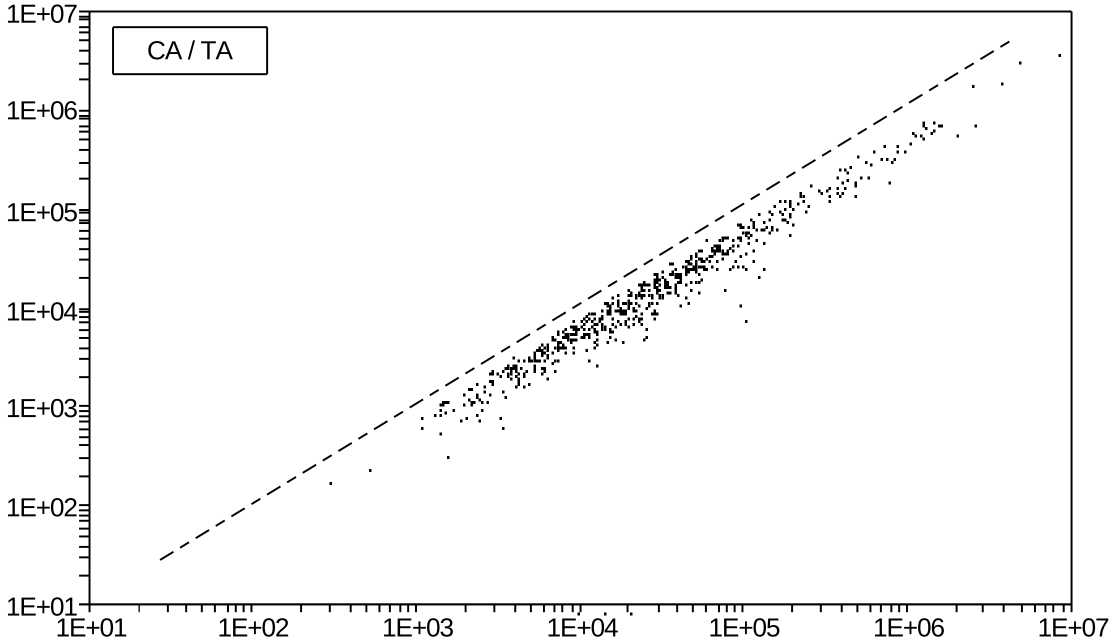

Figure 32: Scatter-plot in log space showing the effect of a strong constraint imposed on CA (Y-axis) by TA (X-axis). All groups together, 1984.

are two-variate relations. Their distribution is determined by two main effects and one interaction. Eventually, this interaction plays an important role.

In section 4.1 we examine the influence of external forces on the two-variate lognormality of items. This allows us to identify and predict which ratios will be near normality, which ones will have negative skewness and which ones will remain broadly lognormal. Our predictions are supported by the published research.

Once solved the most puzzling situations, we are in good position to find real regularities in the behaviour of residuals. We describe and discuss the ones we found in our data in section 4.2. As a result, we suggest practical procedures for the selection and pre-processing of input variables in statistical models.

The method: The goal of this study is the establishment of convenient procedures for the statistical modelling of accounting relations. We are mainly concerned with common characteristics, not with the behaviour of particular ratios. Therefore, we selected the items to be examined according to criteria which somehow differ from those adapted in the research concerned with the distribution of ratios. In general, combinations of items were selected so as to form as many relations as possible with a small number of deflators.

However, we didn't go too far in the differentiation from the published studies. It seems desirable to compare our results with other's. For example, we used the ratio output, $R_{y/x} \times f^{y/x}$, instead of

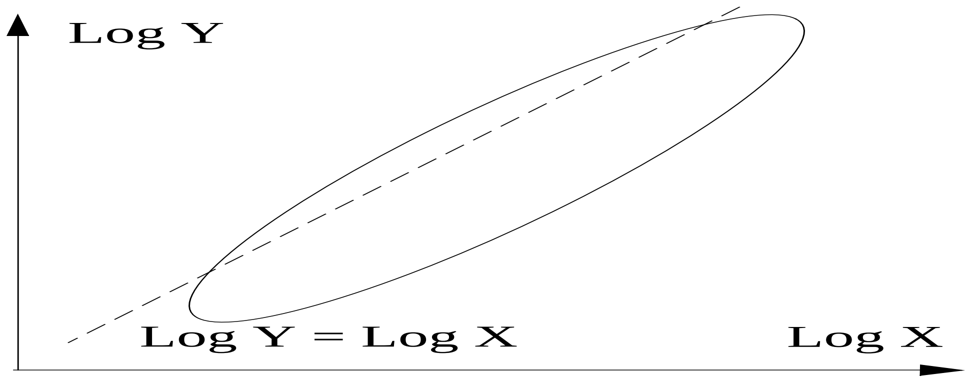

Figure 33: How log residuals of the ratio model $y/x = R$ are affected by the identity $x_j > y_j$ for all j. The two-variate distribution is confined to assume values below the line $\log y = \log x$.

ratio residuals, $f^{y/x}$. And some of the selected ratios are standard in the literature. We think that the results of this chapter are interesting also in the context of ratio analysis.

4.1 The Effect of External Constraints

Usually ratios exhibit strong positive skewness. This is consistent with their multiplicative nature. However, the literature often mention ratios which are Gaussian or even negatively skewed. Typically, TD/TA is reported as being Gaussian [27] [38]. How is this possible?

The reason is straightforward. Accounting identities like $TA = CA + FA$, make it impossible for some two or higher variate relations to have all the values a skewed distribution would allow. Such identities act as a constraint introduced in the normal course of a multi-variate spread. This effect, accountants often mention to explain why some ratios are bounded, had never been associated with the strong departures from positive skewness observed in the distribution of these ratios. In log space it turns out that this effect is clearly observable and self-explanatory.

Gaussian ratios and accounting research: The finding of Gaussian ratios had a negative effect in the way accountants picture important problems. For example, it is possible that this observation is in the origin of the conviction about dividing the items in two categories — those which are dependent on size and those which are not. Gaussian ratios would denote the existence of additive effects in accounting data. In that case independence from size could be possible. Since Gaussian — and other not positively skewed — ratios are the result of external forces, one of the objections to the acceptance of a widespread common effect is removed.

Quantifying the constraint: The two-variate case is the most important one for it directly affects financial ratios. In order to quantify it we shall write accounting identities in a non-equality form like this

$$NW < TA, \qquad CA < TA, \qquad C < CL, \qquad \text{and so on.}$$

We say that there is a constraint if, due to any accounting identity or other external force, the two-variate relation $y_j/x_j = R \times f_j$ is bounded so that one of the next non-equalities hold.

$$\text{for any } j, \quad x_j \ > \ y_j \quad \text{or} \quad y_j \ > \ x_j$$

The non-equality on the left can be found in constrained ratios where the numerator is bounded by the denominator. An example could be Debt to Total Assets. The non-equality on the right arises in ratios in which the denominator is bounded by the numerator. This is not as frequent as the first case. But, of course, it is possible to create such a situation just by inverting one of the former ratios.

The consequences for the distribution of ratios are different in one case and in the other.

When the constraint is $x_j > y_j$: Taking first the case on the left we must have $\overline{\log y} - \overline{\log x} < 0$. This constraint affects the distribution of $\varepsilon^{y/x}$, the residuals of the ratio model in log space. Since

$$\varepsilon_j^{y/x} \ + \ [\ \overline{\log y} - \overline{\log x}\ ] \ = \log y_j - \log x_j$$

we must have, for any j,

$$\epsilon_j^{y/x} \ < \ -[\ \overline{\log y} - \overline{\log x}\ ]$$

In log space this constraint imposes a frontier on the values residuals can attain. The frontier is the line $\log x = \log y$, the anti-clockwise 45^o axis. Figure 33 is a geometrical representation of a two-variate relation in log space under this kind of constraint. The residual is no longer free in its scatter. It is now constrained to have only small positive values.

Figure 32 (page 89) shows, in log space, a real situation in which a numerator (Y-axis) bounds the denominator (X-axis). A case of no constraint is displayed in figure 23 (page 58). Figure 22, on page 54 shows another example of constraint but in smaller degree.

The effect of constraints on the distribution of $f^{y/x}$, the multiplicative deviations from the ratio standard, is that of not allowing the spread out of its otherwise skewed distribution. Instead of a clear tail towards the positive values, such ratios will exhibit a smaller or much smaller tail. This fact explains why some studies didn't find positive skewness in a few ratios. We shall see that this constraint can be very effective in creating Gaussian-like distributions.

When the constraint is $y_j > x_j$: When the numerator of a ratio is bounding the denominator we have a constraint imposed on the values residuals can attain, described by

$$\epsilon_j^{y/x} \ > \ -[\ \overline{\log y} - \overline{\log x}\ ]$$

for any j. Now the residuals are constrained to have only small negative values since large negative deviations from the expected value are not allowed. This will increase even more the skewness of $f^{y/x}$, the multiplicative residual.

Which ratios are affected: The above considerations are enough to predict the conditions under which a constraint created by accounting identities will visibly affect residuals of ratio models.

The variance of $\varepsilon^{y/x}$, the residual in log space of the ratio y/x, can be expressed in terms of the variance and co-variance of $\log x$ and $\log y$ as

$$\mathrm{VAR}(\varepsilon^{y/x}) \;=\; \mathrm{VAR}(\log x) + \mathrm{VAR}(\log y) - 2 \times \mathrm{COV}(\log x, \log y) \tag{11}$$

Since, in general, an existing constraint will not allow $\varepsilon^{y/x}$ to spread across $|\overline{\log y} - \overline{\log x}|$, this difference can be used, along with the above measures of spread, to estimate in what degree a constraint will affect the symmetry of ratio log residuals. We define

$$\zeta \;=\; \frac{\overline{\log y} - \overline{\log x}}{\sqrt{\mathrm{VAR}(\log x) + \mathrm{VAR}(\log y) - 2 \times \mathrm{COV}(\log x, \log y)}} \tag{12}$$

as the distance, expressed in standard deviations, separating the constraining frontier from the expected value of the ratio. Then, for

$$\begin{cases} \quad\quad\;\; |\zeta| > 3 & \text{the constraint will be very small.} \\ 3 > |\zeta| > 2 & \text{the constraint will be small.} \\ 2 > |\zeta| > 1 & \text{the constraint will be strong.} \\ 1 > |\zeta| > 0 & \text{the constraint will be very strong.} \end{cases}$$

That is, given an existing constraint, the severeness will increase with two factors:

- The proximity between the mean values in log space of the items under constraint. Ratios like Current Assets to Total Assets or Net Worth to Total Assets are more likely to exhibit a severe constraint than, say, Inventory to Total Assets.
- The spread of $\varepsilon^{y/x}$, the residuals in log space. Smaller spread means smaller constraint. Such a spread will depend on the spread of the items and on their correlation. Notice that some combinations of items, similar in their variability, can create residuals with a very small spread.

Both factors are expressible by measuring, in standard deviation units, the distance, ζ, between the constraining frontier and the expected value of the ratio.

An example: The mean values and standard deviations in log space of some items of the Balance Sheet (all groups together, 1986) are displayed next.

Item	Mean	Standard Deviation	Item	Mean	Standard Deviation
Current Assets	4.258	0.698	Fixed Assets	3.900	0.829
Total Assets	4.454	0.706	Net Worth	4.080	0.734

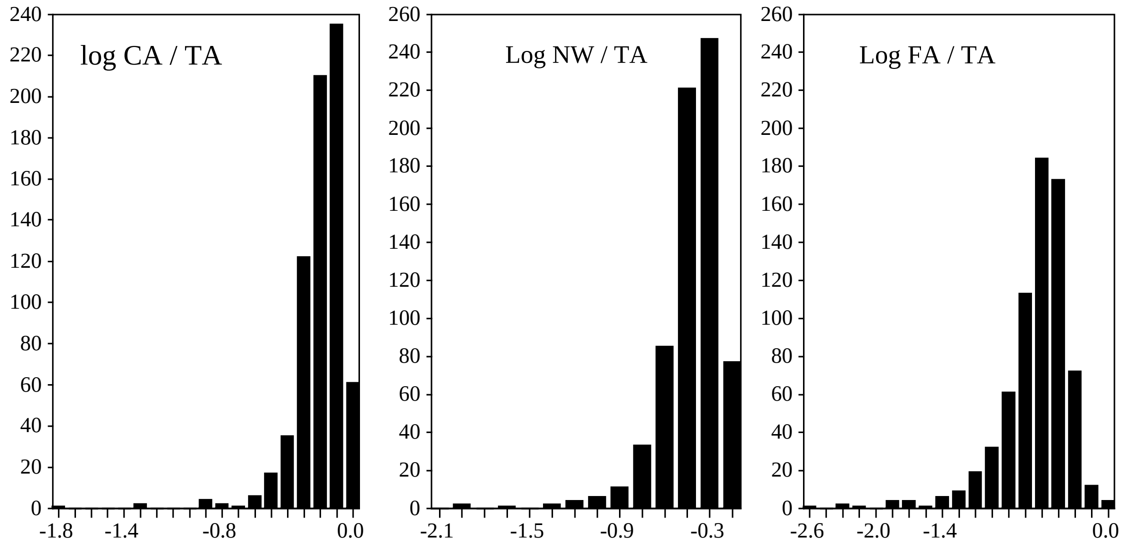

Figure 34: Frequency distributions in log space of three ratios differently affected by constraints.

From the table above, and accepting as a first approximation that the differences in variance and co-variance amongst these variables will not introduce significant new information, we can predict that, due to the existence of the identities $TA = FA + CA$ and $TA = NW + DEBT + CL$ the items most affected by such constraints will be,

- in first place Current Assets because its mean value is the nearest to TA.
- in second place Net Worth,
- in third place Fixed Assets, the item with a mean value away from the one of TA.

In fact, the ratios CA/TA, NW/TA and FA/TA have their distributions strongly affected by constraints. They form a scale illustrating different degrees of severeness.

- CA/TA is so strongly affected that its distribution, instead of being positively skewed, becomes skewed in the opposite direction. Lognormal variates are two-tailed. If the cases in the largest one are constrained so as to make it vanish, the remaining one generates negative skewness.
- NW/TA is affected in a way that makes it almost Gaussian, despite NW and TA being as lognormal as any other items. The long tail of the distribution of residuals is constrained to become much shorter so that the resulting one is almost balanced by the small-values tail.
- FA/TA remains positively skewed but less than the expected for multiplicative residuals. Its distribution is symmetrical to the one of CA/TA.

Figure 34 shows the frequency distribution of the above three ratios in log space. The bounding effect of the denominator is clearly visible.

In the next table we show the skewness and kurtosis of the logs of the items involved in this example. The skewness and kurtosis of the items in ordinary space is so high that their computation causes overflaw problems.

Item	Year	1983	1984	1985	1986	1987
Log Net Worth	SKEW	0.174	0.163	0.253	0.289	0.289
	KURT	0.468	0.402	0.255	0.263	0.353
Log Fixed Assets	SKEW	0.097	0.177	0.119	0.124	0.159
	KURT	0.421	0.110	0.114	0.113	-0.008
Log Total Assets	SKEW	0.301	0.351	0.404	0.343	0.425
	KURT	0.546	0.276	0.228	0.349	0.309
Log Current Assets	SKEW	0.237	0.349	0.056	0.295	0.345
	KURT	0.372	0.374	1.840	0.345	0.480

The conclusion is that the above variables, when considered individually, are not far away from lognormality. Appendix A contains more detailed statistics.

Next we display the values of the skewness and kurtosis obtained for the ratios used in this example. They were not transformed.

Ratio	Year	1983	1984	1985	1986	1987
CA/TA	SKEW	-0.410	-0.434	-0.352	-0.502	-0.592
	KURT	0.332	0.261	-0.098	0.332	0.528
NW/TA	SKEW	-0.146	0.012	0.009	-0.015	-0.063
	KURT	-0.233	-0.137	-0.136	-0.039	-0.036
FA/TA	SKEW	0.410	0.434	0.352	0.502	0.592
	KURT	0.332	0.261	-0.098	0.332	0.528

The values obtained for the skewness agree with the constraint mechanism: CA/TA, the most affected ratio, has a negative skewness. NW/TA comes next and finally FA/TA exhibits positive skewness. Notice how small is the skewness for the ratio NW/TA. In general, the skewness and kurtosis displayed in the above table cannot be considered as far away from normality. But a powerful test like the Shapiro-Wilk rejects normality in almost all cases displayed. The skewness is far smaller than the expected for ratios of correlated lognormal deviates having spreads similar to those of the above items. Table 14 on page 98 and 15 on page 99 show typical values for the skewness and kurtosis of ratios selected so as to avoid constraints.

Finally we display the skewness and kurtosis of ratios similar to the above ones but inverted.

Ratio	Year	1983	1984	1985	1986	1987
TA/NW	SKEW	17.712	15.462	22.287	12.891	12.453
	KURT	356.027	260.067	536.404	192.986	179.477
TA/FA	SKEW	11.211	21.927	23.788	12.704	11.214
	KURT	147.527	519.812	592.486	197.060	151.238
TA/CA	SKEW	18.714	5.044	6.286	20.685	17.955
	KURT	377.719	37.058	71.563	483.578	337.921

Ratio	1983		1984		1985		1986		1987	
	skew	zeta	skew	zeta	skew	zeta	skew	zeta	skew	zeta
CA/TA (1)	-0.41	1.29	-0.43	1.05	-0.35	1.58	-0.50	1.34	-0.59	1.14
CL/TA (x)	0.636	2.2	0.587	1.84	0.616	2.21	0.56	2.24	0.597	2.26
C/CL (2)	-0.09	1.54	-0.15	2.02	-0.21	1.56	-0.27	1.63	-0.27	1.72
C/TA (3)	1.247	2.88	1.227	2.61	1.357	2.94	1.191	2.87	1.137	2.96
DB/TA (4)	1.873	2.37	1.92	2.29	2.103	2.36	1.774	2.17	2.056	2.21
FA/TA (5)	0.41	1.69	0.434	1.96	0.352	1.76	0.502	1.76	0.592	1.81
I/CA (6)	0.264	1.53	0.102	1.74	0.218	1.57	0.433	1.44	0.294	1.33
I/TA (7)	0.569	2.08	0.473	1.9	0.867	2.16	0.913	2.04	0.812	1.89
NW/TA (8)	-0.14	1.61	0.012	1.85	0.009	1.73	-0.01	1.8	-0.06	1.82
Q/CA (9)	-0.26	1.34	-0.10	1.17	-0.21	1.81	-0.43	1.4	-0.29	1.3
Q/TA (0)	0.504	2.07	0.583	1.92	0.51	2.41	0.387	2.17	0.455	2.26
TD/TA (o)	0.349	2.31	0.173	2.19	0.222	2.23	0.198	2.17	0.177	2.21
EB/S (+)	2.03	3.06	1.63	3.16	2.06	3.12	1.824	3.18	1.480	3.25
W/S (*)	0.42	2.14	0.42	2.01	0.412	2.08	0.395	2.13	0.371	2.13

Table 13: The values of $|\zeta|$ (zeta) and skewness for 14 ratios likely to suffer a constraint in their distributions. $Q = CA - I$; $TD = DEBT + CL$.

The above ratios were not transformed. Ratios which are the inverse of the ones affected by the constraint $y_j > x_j$ for all j will obviously be affected by the constraint $x_j > y_j$ for all j. That is, ratios like TA/NW and so on, should exhibit positive skewness because their tail is now free from constraints. In fact, the constraint is working in the same direction as the lognormal skewness, not against it.

Not only the values obtained for these statistics are lognormal-like. The mirror-image effect linking the distributions of CA/TA and FA/TA also vanished. This simple example has shown how accounting identities explain the strongest deviations from a multiplicative behaviour mentioned in the literature.

Other constraints: There are other external forces likely to condition the distribution of ratios. Instead of defining frontiers which are impossible to bridge, as in the case of accounting identities, these other forces impose frontiers in which only a gradient in the density of cases is observed. For example, the non-equality $CA > CL$ defines one of such gradients because firms avoid, if they can, negative Working Capital. And, at least in industrial firms, the non-equalities $S > OPP$, $S > EBIT$, $S > W$ and so on, will be almost equivalent to real accounting identities.

One of the most interesting consequences of the lognormal nature of items is the possibility of directly observing two-variate relations by building simple scatter-plots in log space. The constraints described above and many other features become clearly visible with these tools and can be identified.

4.2 Comparing Constrained and Non-Constrained Ratios

The last section was devoted to the identification and description of the effect external constraints can have in the distribution of ratios. We discussed a limited example, showing how an accounting identity can hide the multiplicative nature of ratios. In this section, apart from providing a more systematic empirical evidence on such effect, we show that ratio outputs are broadly lognormal.

The data: We examined 14 ratios (see list in table 13) formed with items from the Balance Sheet and 2 from the Profit and Loss Account. For such ratios there is an accounting identity or at least a strong constraint which should influence their distribution in a variable extent.

We also examined 20 other ratios (and their inverse) for which there is no direct accounting identity constraining their distribution. These ratios are listed in tables 14, page 98 and 15, page 99. In both cases ratios were selected so as to share as many items as possible but representing an approximate random choice amongst all possible combinations.

We gathered in the same sample all the 14 industrial groups listed in table 1, page 5. This sample was then examined for a period of five years (1983 - 1987). Some industrial groups were also examined individually.

Constrained ratios: Table 13 lists the 14 constrained ratios. It also displays the values obtained for their skewness during a period of five years. The same table shows the value $|\zeta|$ assumes in each case. In figure 35 (page 97) we reproduce table 13 as a scatter-plot. The marks identifying each ratio can also be found in table 13, on the left. Apart from the usual abbreviations for items, Q stands for $CA - I$ and TD for $DEBT + CL$.

Figure 35 shows that the ratios I/CA, NW/TA, C/CL, Q/CA and TD/TA are not far away from the Gaussian distribution in what concerns their skewness. Others like I/TA approach a skewness of 1 and DB/TA or C/TA have their skewness well above the unit.

The results also show that ζ, as expected, predicts the skewness of the constrained ratios, though it cannot be considered as the only factor influencing it. For the displayed cases ζ accounts for 65% of the observed variability of the skewness ($r = 0.8$). The ratio DB/TA seems to depart from this rule. Its skewness is larger than predicted by $|\zeta|$.

Gaussian ratios: We identified two ratios with almost Gaussian distributions: NW/TA and TD/TA. These ratios are the ones Deakin and other studies also identify as Gaussian, thus breaking the rule of positive skewness [27] (see also [38], a recent study with U.K. data). We now understand why this happens and in which cases it is likely to occur.

However, the most important finding is that constrained ratios behave in a way which can be very different from the non-constrained ones. In order to see this we must explore the behaviour of non-constrained ratios first.

Non-constrained ratios: The 20 ratios which — as far as we know — should not be affected by constraints and therefore should be expected to be positively skewed were built with the same variables used in the 14 preceeding ones and a few more. Tables 14 (page 98) and 15 (page 99) contain

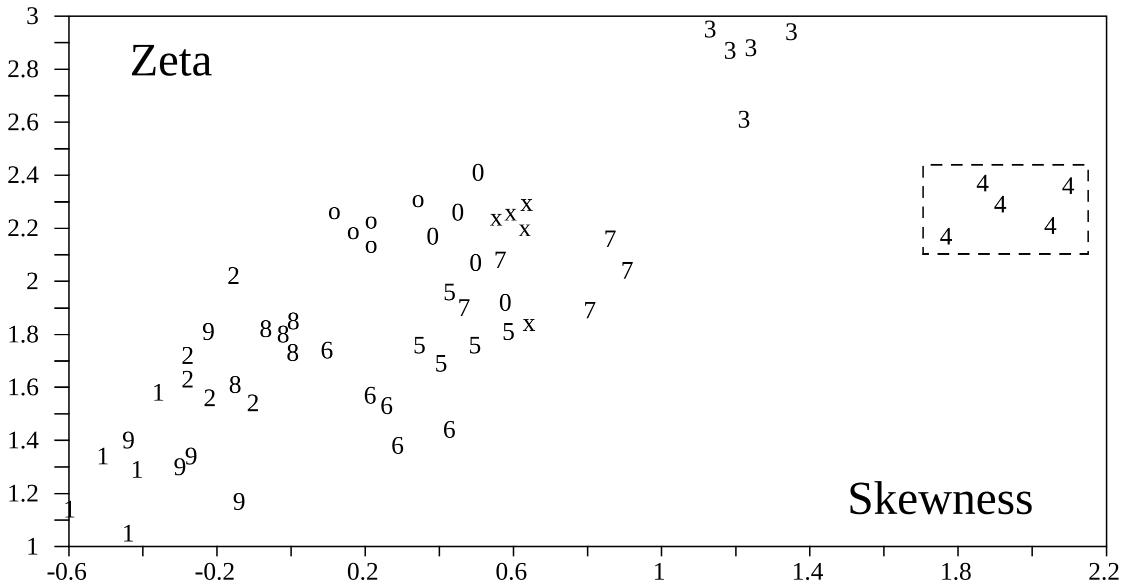

Figure 35: Correlation between the observed skewness in 12 ratios built with items from the Balance Sheet and the value of ζ, the standardized distance measuring the expected degree of constraint.

a listing. The number of employees was used because there is no obvious constraint affecting its relation with other items. In ratios built with $EBIT$ and other items having zero or negative cases only the positive ones were used.

Identifying the broad lognormality of non-constrained ratios: In order to highlight the existence of a broad lognormal trend in non-constrained ratios we plotted their skewness against their kurtosis. Lognormal variables should have their skewness and kurtosis functionally related.

Lognormal distributions have skewness and kurtosis depending solely on their variance. In lognormal variates the skewness and the kurtosis are not two independent statistics. One of them is a function of the other one. The book by Aitchison and Brown [1] contains this formulation.

Figure 36 on page 98 is a graphical reproduction of tables 13, 14 and 15. It displays the regular curve non-constrained ratios and their inverse form when their skewness is plotted against their kurtosis. Below we can see in detail the lower values of this characteristic pattern approaching the random scatter produced by constrained ratios around the origin.

The regularity observed in figure 36 follows the outlined relation which exists between the skewness and the kurtosis of lognormal variables for different spreads. No variate other than the lognormal or a distribution of the same class could yield such a regular spread of cases.

The particular values that the skewness and the kurtosis assume in each case are meaningless. The variance of a lognormal variable is not independent of the mean. A given log spread, when associated with a large mean value, will produce more spread — and the corresponding skewness and kurtosis — than when associated with a smaller mean. Therefore, the resulting skewness and

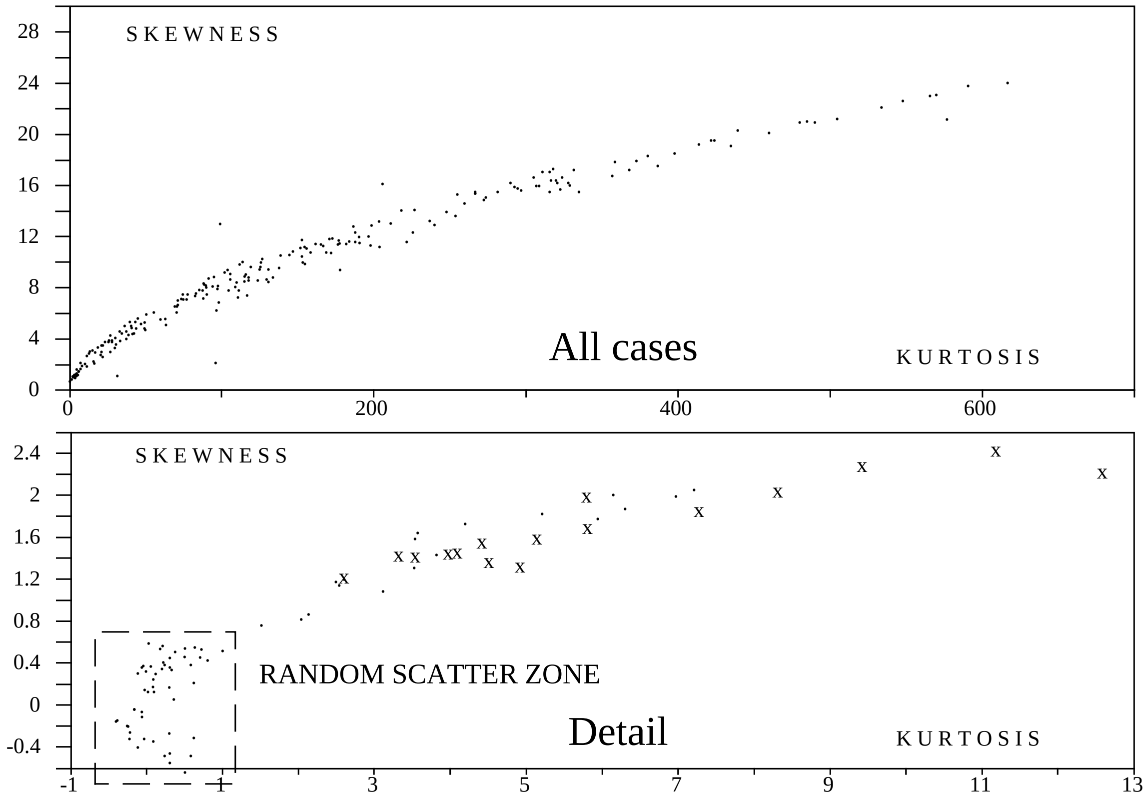

Figure 36: Two scatter-plots of the skewness with the kurtosis of all the studied ratios and their inverse. Above, the typically lognormal relation drawn by non-constrained ratios. Below, the same scatter near zero, with the constrained ratios forming a disordered scatter around the origin.

Ratio	1983		1984		1985		1986		1987	
	skew	kurt	skew	kurt	skew	kurt	skew	kurt	skew	kurt
D/C	5.7	44.5	5.6	48.3	14.0	255.0	15.0	261.0	4.8	35.5
C/D	18.7	381.1	20.5	461.0	22.5	535.0	14.3	249.0	12.4	198.0
CA/CL	2.0	5.8	4.2	29.2	5.0	38.4	11.6	205.0	4.8	43.6
CL/CA	19.6	415.0	4.2	34.5	3.0	22.9	9.2	135.0	21.6	506.0
C/I	9.9	139.0	9.8	132.0	24.2	592.0	15.8	268.0	21.3	491.0
I/C	6.5	71.5	3.3	18.0	8.9	119.0	9.2	119.0	7.9	91.3
Q/CL	0.0	14.2	3.1	21.8	4.0	31.6	7.3	99.3	5.5	64.7
CL/Q	11.9	192.0	2.5	97.2	-21.0	486.0	6.6	97.8	-18.7	418.0
W/N	18.9	399.0	1.5	4.1	1.5	4.0	1.4	3.3	1.6	4.4
N/W	11.7	199.0	2.2	12.6	7.6	112.0	7.8	118.0	8.9	116.0
S/TA	7.0	72.3	13.6	238.0	13.3	241.0	6.3	51.7	8.5	95.3
TA/S	8.8	111.0	19.5	436.0	7.6	89.0	15.9	336.0	3.4	22.3
S/FA	9.8	105.0	21.4	486.0	10.4	115.0	9.2	96.1	9.6	103.0
FA/S	2.3	9.4	15.9	317.0	8.9	125.0	5.1	50.8	5.9	60.9
S/NW	16.4	310.0	13.2	188.0	23.0	549.0	16.1	324.0	16.5	207.0
NW/S	1.6	5.2	4.4	38.6	2.5	17.6	12.0	223.0	4.7	39.8
S/I	11.7	168.0	12.0	189.0	17.0	306.0	23.5	571.0	21.3	481.0
I/S	8.9	132.0	17.1	358.0	17.9	388.0	9.8	179.0	1.4	4.5
EB/TA	2.0	8.3	2.4	11.2	1.4	3.6	1.9	7.3	1.2	2.6
TA/EB	16.6	291.0	12.2	172.0	11.1	173.0	7.5	76.0	18.3	374.0

Table 14: The skewness and kurtosis of ratios selected so as to avoid constraints. All groups. 1^{st} table. $Q = CA - I$; $TD = DEBT + CL$.

Ratio	1983		1984		1985		1986		1987	
	skew	kurt	skew	kurt	skew	kurt	skew	kurt	skew	kurt
EB/NW	19.9	425.0	17.5	317.0	16.6	322.0	8.4	110.0	16.4	330.0
NW/EB	15.7	256.0	17.6	369.0	12.7	227.0	7.0	71.9	23.4	567.0
D/I	10.4	127.0	16.6	329.0	21.6	578.0	24.4	618.0	19.9	423.0
I/D	10.7	128.0	11.8	166.3	16.0	298.0	10.0	126.8	7.5	74.6
W/I	15.8	268.0	9.3	116.2	7.9	79.0	13.6	204.5	14.4	219.3
I/W	8.6	89.7	4.6	28.1	4.5	31.2	3.1	12.7	3.3	14.0
EB/FA	9.4	117.0	11.2	169.8	10.4	154.7	9.0	107.0	8.5	98.8
FA/EB	12.1	154.0	9.8	126.0	11.6	155.7	7.7	83.8	16.8	320.8
S/N	12.7	189.0	12.3	174.0	7.8	75.6	8.7	89.3	11.1	159.8
N/S	1.7	5.8	1.1	1.3	3.7	31.0	2.6	16.9	3.3	27.9
EB/N	9.1	92.5	11.7	177.7	5.2	42.0	3.4	14.6	5.2	45.0
N/EB	15.9	268.0	9.0	130.7	6.0	46.1	8.0	84.2	13.4	212.4
NW/N	8.4	90.9	5.0	34.1	3.5	16.1	3.9	23.0	4.3	29.1
N/NW	6.4	56.4	4.8	42.7	18.2	359.8	4.2	24.5	8.2	112.6
W/TA	1.3	4.9	1.5	5.8	1.4	4.3	1.8	7.2	1.5	6.4
TA/W	9.5	107.0	11.5	152.9	5.7	50.6	4.2	27.0	3.7	19.7
DB/NW	14.5	228.0	10.9	140.0	8.6	90.7	7.5	78.0	5.4	37.4
NW/DB	13.3	199.8	17.6	332.7	12.0	184.9	11.0	145.6	16.1	295.7
DB/S	2.5	8.4	6.0	64.2	15.3	273.4	8.3	98.7	10.3	155.8
S/DB	10.0	120.4	16.3	293.6	12.1	178.2			10.2	113.2

Table 15: The skewness and kurtosis of ratios selected so as to avoid constraints. All groups. 2^{nd} table. $Q = CA - I$; $TD = DEBT + CL$.

kurtosis can be to a large extent influenced by $R_{y/x}$, the expected proportion between the numerator and the denominator of the ratio. The understanding of this mechanism accounts for a few remarks found in the literature.

In case it would be convenient to compare the spreads of several ratios, residuals, $f^{y/x}$, should be used instead of the ratio output, $R_{y/x} \times f^{y/x}$. Since the $f^{y/x}$ have a constant expected value of 1, we would make the skewness and the kurtosis relate to the variance alone.

When reproducing this experiment, notice that the SPSS-X package we use to compute the skewness and kurtosis in this study does it in a way that is not exactly the one found in text books.

Departures from this relation: In the 200 examined samples (20 ratios and their inverse during 5 years) only three yielded values of the skewness and kurtosis which wouldn't obey the above formulation. They were from the same ratio, CL/Q, or its inverse, during the years 1983, 1985 and 1987. We further formed a few more ratios with $Q = CA - I$ and we found three other cases of irregular behaviour. They were the ratio $EBIT/Q$ in 1987, and W/Q, in 1986 and 1987.

Comparing constrained and non-constrained ratios: The distinct behaviour of constrained ratios also emerges when observing figure 36 (page 98). Whilst non-constrained ratios obey the formal relation between skewness and kurtosis for lognormal variates, the constrained ones form a random scatter of cases around the small values of these statistics.

In figure 36 the cases marked with a dot in the detailed scatter (below) represent ratios which are known to be to some extent constrained by an accounting identity. The ones marked with a "x" represent the position of non-constrained ratios. A few constrained ratios are near lognormality owing to its small constraint.

The special character of constrained ratios could also be traced by inverting them. Constrained ratios change completely their characteristics when inverted and actually become broadly lognormal. The above ones remain near lognormality in both situations. They will just move along the formal line linking skewness with kurtosis. When not considering the effect of the spread of the ratio itself, this movement would be commanded by the new value of the expected proportion, $R_{y/x}$.

4.3 Persistent Departures From Lognormality

In this section we return to the log space. Despite the findings of previous section suggesting a broad lognormal behaviour for ratios, hardly any of the studied ratios is exactly lognormal. When examined closely in log space, a persistent departure from the Gaussian curve can be observed.

4.3.1 Leptokurtosis In Log Space

As a rule, log residuals exhibit positive kurtosis of varying severity.

A few ratios also show asymmetry. But this is not a widespread feature. Skewness in log ratios can be explained and accounted for. A careful examination of secondary effects of constraints, managerial practice or base-lines, leads to plausible mechanisms able to generate asymmetry. The use of extended ratios yields non-asymmetric residuals.

On the contrary, the persistent kurtosis observed in log space seems to be quite a general feature. It was observed in all but one of the studied ratios and their inverse. Ratios formed with non-accounting items related to size like the number of employees also exhibit such a feature. When sampling by industry the residual kurtosis will not vanish.

Kurtosis cannot be accounted for by base-line ratios. And, of course, the modelling of the spread of log items allowed by the free slope ones will not make it vanish. It is unlikely that external forces generate kurtosis. This feature is more likely to be related to internal mechanisms of the firm. As a result of this kurtosis, the Shapiro-Wilk test seldom finds a non-significant departure from normality in the $\varepsilon^{y/x}$ residual differences. This fact contrasts with the strong consistency of results when assessing the lognormality of items. Log items exhibit positive kurtosis as well but in a much smaller degree.

Table 16 contains the usual log statistics for the non-constrained ratios used above. In log space there is no difference in the behaviour of a ratio and its inverse. Distributions are a mirror-image of each other. Therefore, the skewness of the ratios which are the inverse of those displayed in table 16 will simply be the same value with inverted sign. The kurtosis will be the same.

Ratio	1983		1984		1985		1986		1987	
	skew	kurt	skew	kurt	skew	kurt	skew	kurt	skew	kurt
D/C	-0.7	9.2	-0.8	8.6	-0.1	10.0	0.2	9.3	-0.6	6.2
CA/CL	-1.2	11.0	0.4	2.9	0.8	3.1	0.6	5.9	-0.5	10.5
C/I	0.4	2.1	0.6	2.2	1.1	8.3	0.9	6.7	1.8	11.0
Q/CL	-1.9	12.0	-0.3	4.3	-0.3	3.2	-0.5	5.2	0.1	2.8
W/N	0.3	12.0	-0.1	0.8	-0.4	2.8	-0.3	2.3	-0.6	4.4
S/TA	0.8	4.5	0.8	7.0	1.0	5.4	1.0	5.5	1.2	4.9
S/FA	1.8	5.4	2.0	8.5	1.9	6.9	1.9	5.9	1.6	4.6
S/NW	1.5	4.5	1.6	5.6	1.8	7.8	1.3	4.2	1.6	6.4
S/I	1.3	4.7	1.2	4.6	1.7	8.8	2.3	13.9	3.4	21.0
EB/TA	-2.1	11.2	-1.3	4.9	-1.2	2.9	-1.1	3.1	-1.7	7.9
EB/NW	-1.6	11.4	-0.1	6.1	-0.3	3.6	-0.5	2.7	-0.8	7.3
I/D	-0.6	6.5	-0.2	6.7	-0.3	4.8	-1.6	13.4	-2.0	15.0
W/I	0.1	6.0	0.2	3.2	0.6	3.1	1.2	6.4	1.4	7.8
EB/FA	-0.9	6.0	-0.3	2.8	-0.5	2.8	-0.8	2.2	0.1	3.4
S/N	1.7	5.4	1.8	5.8	1.2	3.3	1.5	4.3	1.3	3.6
EB/N	-0.3	3.9	0.0	1.6	-0.3	2.2	-0.2	1.2	-0.5	2.5
NW/N	0.3	2.2	0.5	1.1	-0.4	3.1	0.0	0.9	0.0	0.9
NW/TA	-1.6	5.6	-1.5	5.6	-0.9	2.2	-0.6	1.5	-0.7	1.5
NW/DB	0.7	1.7	0.8	2.3	0.5	0.5	0.6	0.3	0.8	0.8
DB/S	-0.9	0.9	-1.2	2.4	-0.7	0.7	0.9	1.3	-0.8	0.9

Table 16: The skewness and kurtosis of ratios in log space for a period of five years.

The findings of previous studies: One of the most puzzling findings of previous studies is that no transformation seems to improve the normality of ratios. In our opinion this is a result of using precise criteria to assess phenomena which are only broad trends. For example, if we would use accurate tests like the Shapiro-Wilk's for measuring the lognormality of ratios, we would get the general impression that ratios are far away from lognormality. Its precision conceals broad trends.

The use of all sorts of transformations to assess the distribution of ratios only complicates things. For example, if we replicate with ratios the experiment carried out in chapter 1 — which consisted of using progressively higher fractional exponents for transforming items and observing the results of applying the Shapiro-Wilk or other tests — it is clear that the results would be confusing. For non-constrained ratios, the skewness would probably diminish with increasing roots but the kurtosis would emerge after some improvements. For most of the constrained ratios the skewness would change sign, becoming negative. Ezzamel *et al.* [39] observed this.

Items are lognormal and ratios should be broadly lognormal since they are multiplicative deviations from expected proportions. Only when knowing this in advance is it possible to notice the departures affecting such trend. In the literature, the realization that transformations apparently wouldn't improve the normality of ratios [45] led to a cautious attitude towards transforming data and to a renewed interest in the trimming of outliers, which, as we saw, didn't work either.

4.3.2 Departures From Lognormality: Discussion

In this section we show that the kurtosis observed in the log residuals of ratios is not inconsistent with the lognormality of items. Then, we show that the strong common effect is the source of the

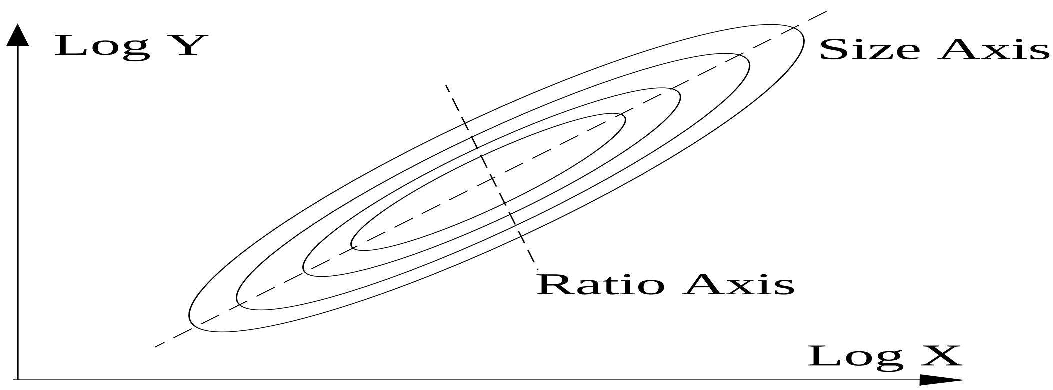

Figure 37: Two-dimensional view of a two-variate density surface in log space.

Gaussian behaviour of log items.

The difference of two similar Gaussian distributions: Let us recall the expression for estimating the residuals of the ratio model in log space:

$$\varepsilon_j^{y/x} = (\log y_j - \overline{\log y}) \ - \ (\log x_j - \overline{\log x})$$

Clearly, the $\varepsilon^{y/x}$ are a difference of two log items, both with their central trend accounted for. Their distribution is the result of the subtraction of two Gaussian distributions with the same mean.

Also, these two Gaussian distributions which, when subtracted, yield the ratio residuals, are very similar in spread. A large fraction of the variability of items comes from the strong common effect they share. Ratios account for this effect and yield — as a residual — a contrast between two weak effects. Such a residual variability is the source of the positive kurtosis in log space. But, in items, it is so small a proportion of the total one that it could have any distribution without greatly affecting their lognormality.

We can have a graphical view of this problem by considering the two-variate log distribution drawn by the components of the ratio. Such distribution is an oblong hill-shaped surface oriented in the 45^o direction and centred in $\overline{\log y} - \overline{\log x}$. The density of cases determines the height of each point in the surface (see figure 37).

Such a surface would be very thick in one of its main dimensions and very thin in the other one. The largest dimension accounts for most of the variability. In figure 37, the largest dimension is labelled the "Size Axis" and the smallest one the "Ratio Axis". The variability of ratio residuals is explained by the smallest dimension, the ratio axis. It is orthogonal to the size one, which accounts for the variability introduced by the common effect.

The variability along the smallest axis happens not to be Gaussian. When the considered surface is observed so that the largest dimension becomes parallel to the horizon, the surface shows a Gaussian aspect. On the contrary, when it is observed transversally, it yields a leptokurtic shape.

As we see, there is no contradiction in the fact that log items are Gaussian and log ratio residuals are leptokurtic.

According to the above explanation we should consider the weak effect as the source of leptokurtosis and the strong effect as the source of the Gaussian behaviour in log data. For example, the small amount of kurtosis observed in log items would denote the influence of their own variability superimposed to a much larger Gaussian spread. In chapter 5 we return to this topic.

4.4 Ratio Standards and Departures From Standards

What is the appropriate estimator for ratio standards? How should the individual deviations from standards be interpreted? In this section we suggest an answer for these two questions.

The calculation of ratio standards has been the object of widespread discussion in the literature. This is because each distribution apparently suggests its own standard. And "in non-normal distributions, location measures are far from being unanimous" [38]. Lev and Sunder [79] enumerate a few possible answers.

We are in a good position for giving an answer too. Ours is that as long as items are lognormal, the distribution of ratios doesn't matter. Neither for the estimation of standards nor in assessing deviations from standards. However, it matters — for practical reasons — for the building of statistical models.

The lognormality on items makes the choice of estimators simple. The acceptance of a plausible mechanism for the generation of the common effect makes it meaningful. But it remains a matter of opinion, of course. However, trying to assess the best distribution for each ratio individually and discussing the most appropriate estimator for each ratio individually, does not seem reasonable either. Doing so is to deny any sense in accounting data as a whole. It is equivalent to treating ratios as if they were a collection of isolated random variables with nothing in common. One ratio would be more like the rate of telephone calls, the other one more like the size of white mice's tails and so on.

Standards are not disturbed by the distribution of ratios: An intuitive way of noticing this is given by the formula for calculating the median. $\overline{\log y} - \overline{\log x}$ (or the same value obtained by finding the logs of ratios and averaging) is not disturbed by any external disturbance or interaction between y and x. When standards or central trends in ratios are estimated by the median, the actual distribution of the ratio itself is irrelevant for the estimation.

Distortions in the distribution of ratios are a result of its two-variate nature. The distribution of individual items is not affected. Only when considering two or more variate distributions will such distortions emerge.

The spread of residuals is not accounted for by the ratio model: Financial ratios are one-degree-of-freedom models: They engage only one degree of freedom from those avaliable in the sample. For example, a sample containing two observations, x and y carried out over N cases, would have $2 \times N$ degrees of freedom before any assumption were made about its behaviour. When fitting the ratio model into this sample, just one degree of freedom, the one corresponding to knowing $R_{y/x}$, would be engaged. By using only one degree of freedom it is possible to model just a unique optimal point, not a spread. The spread remains in possession of perhaps hazardous, non-modelled forces.

The one-degree-of-freedom model only requires, as an assumption for being correct, that their components must have a central trend — and hence, an expected value. Items comply with this requirement in proportional space. Accordingly, we use ratios — proportional adjustments — instead of mean adjustments.

The expected value assumption implies that the distribution of ratio components ought to be symmetrical in its proper space. But not any special symmetry is required. There is nothing either in the assumptions underlying the ratio model or in the model itself able to assess the spread of cases around the standard and as we pointed out, in practice the lognormality of the components is not enough to ensure the lognormality of the residuals.

Skewness in the log components of a ratio would affect its symmetry. The above assumption would be affected. But the existence of kurtosis in the log components wouldn't affect by itself any assumption of the ratio model. Kurtosis wouldn't break its symmetry. Hence, kurtosis is not a departure from ratio assumptions. There is nothing in ratios saying that they shouldn't be leptokurtic.

In the case of external constraints, since the components of the ratio are not affected, the model's assumptions are not violated. And a constrained ratio should be able to yield all the information any other ratio can provide. But this is only possible if the information conveyed by ratios is ordinal.

In short, the discussed facts lead us to two conclusions:

Ratios are ordinal. Strictly speaking, ratios cannot provide a scaling. Ratios only provide a measure of the deviation from standards by saying that one deviation is larger or smaller than the other.

The distribution of ratios is irrelevant. The actual shape of the residuals of the ratio model is not called upon to play any role in the ratio model itself. Constrained or distorted ratios, leptokurtic or not, yield correct estimates of standards. They also yield ranks as measures of departure from a central trend. And this is, strictly speaking, what we can expect from any ratio. The condition for the validity of ratios as models lies in the symmetry of its components taken individually, not in the ratios themselves.

Ratios and internal features of the firm: Financial ratios are about size and deviations from size. They rank contrasts between deviations from size. But they cannot provide distances between such contrasts. The distribution of ratios could be elucidated by building explanatory models for the joint behaviour of accounting items inside one firm. Only after understanding the internal mechanisms governing the features of the firm would it be possible to contrast such a theoretical basis with empirical observations. In other words, the explanation for the leptokurtosis or other regularities observed in ratios is not contained in the ratio model. It is an interesting point but ratios cannot answer it.

4.5 The Use of Ratio Residuals in Statistical Models

In the last section we concluded that the distribution of ratios wouldn't matter and the calculation of a central trend, when using the median, would also be independent of this distribution.

For financial analysis this seems interesting. However, in statistical models the use of ranks is not considered as the best solution.

Accounting statistical models should avoid the burden of rank statistics. Firstly, because statistical models are free to select the variables and can avoid, to some extent, the combinations of items which would produce constraints and other severe departures from homogeneity. Secondly, because the statistical tools based on ranks are very limited in issues regarding estimation. For example, the non-parametric equivalent of the test allowing a comparison of two means is the Mann-Whitney test. This tool will not compare two estimations. It is only intended to assess the significance of the non-similarity of two distributions. Since estimation is important in accounting practice and statistical modelling, non-parametric methods should be seen as altogether not satisfactory.

Therefore, it seems as if the problem of obtaining homogeneous residuals in ratios and their extensions still remains. Despite the remarks of the last section, the spread of the $\varepsilon^{y/x}$ can be made reasonably homogeneous and the influential points are rare.

This section suggests a few procedures to avoid the major departures from an homogeneous behaviour in ratios. Our concern is mainly the statistical modelling of accounting relations.

4.5.1 Avoiding Asymmetry

The proportionality of ratios is understood as a statistical quality related with the non-existence of significant base-lines in cross-sectional relations between the numerator and the denominator. Here we recall a different meaning, concerning the formal relation between numerator and denominator, not any statistical quality. A quotient is said to be a proportion when the numerator is a part of the denominator. Relative frequencies or probabilities are proportions.

All the ratios bounded by the denominator are proportions in this non-statistical sense. Such proportions, when taken as statistical variables, will be more or less constrained, thus yielding asymmetric distributions. Therefore, it seems wise to apply to such ratios the well-known recipes generally accepted for dealing with similar cases. The simplest of such recipes is the odds transformation.

The Odds transformation: Ratios like FA/TA have the numerator as part of the denominator. An odds transformation incorporates the underlying accounting identity thus yielding an unbounded variable.

$$\text{For any proportion} \quad p_i = \frac{x_i}{\sum_i x_i} \quad \text{there are odds defined as} \quad o_i = \frac{p_i}{1 - p_i}.$$

Hence,

$$o_i = \frac{x_i}{\sum_i x_i - x_i}$$

For example, the odds of FA/TA is the ratio FA/CA. The odds of NW/TA will be $NW/(DEBT + CL)$. Clearly, the information contained in both such ratios is exactly the same. The difference between an odds-like ratio and the corresponding probability-like one is the same existent between probabilities and odds. It is just functional. FA/TA expresses something like probabilities. FA/CA, something similar to odds. But both say the same. Therefore, it seems possible to avoid some of the ratios affected by constraints by using the corresponding odds instead.

This solution only applies strictly to constrained ratios from the Balance Sheet. In the Profit and Loss accounts, Sales is not a total. There are other possible sources of income. But it acts as if it were. At least in industrial firms the amount of Sales constrains all the other items of the P. & L.

We noticed that odds can also be used with totals which are not strictly dictated by accounting identities. Instead of OPP/S we can use $OPP/COGS$. This new odds-like ratio will not exhibit constraint effects and its information content will be similar to the former one.

4.5.2 The Selection of Input Variables For Statistical Models

As we stressed before, log items can be used instead of ratios as inputs for modelling accounting relations. Any linear relation in log space will be equivalent to a free-slope multi-variate ratio. And since statistical models are to some extent free to select input variables, they can avoid the use of items which impose constraints in other items.

Accordingly, partial or detail items are preferable to items expressing totals. Elementary pieces of information directed to a specific subject like Inventory, Creditors, Properties, Wages, should, as a general rule, be used instead of Fixed Assets, Total Assets, Current Assets. Obviously they will form ratios which are away from causing constraints.

It is especially important to avoid mixtures of both kinds of items or the inclusion in the same model of items leading, by means of some accounting identity, to the emergence of constraints.

Strong, clear, base-lines should be accounted for before the inclusion of items as input variables in multi-variate tools. This is especially important when using linear algorithms.

The problem of leptokurtosis: When using ratio residuals as input variables for the modelling of accounting relations, kurtosis is generally considered as far less damaging than skewness. Widespread financial models like the CAPM deal with variables which are leptokurtic too. So far, we didn't find any special difficulty in dealing with limited amounts of kurtosis. This is because it doesn't generate influential cases.

The neutralization of the extra kurtosis by means of transformations would be difficult and damaging for the interpretability of models. We think that the correct procedure, in face of this phenomenon, is to investigate its causes and build models able to account for them.

4.6 Summary

In this chapter we studied the distribution of ratios. We found a clear trend towards lognormality, as expected. However, a few factors affect the final distribution that particular ratios assume. Firstly, accounting identities and other external forces can act as constraints, hiding their multiplicative nature. This factor induces the severe deviations from lognormality reported in the literature for ratios like NW/TA and TD/TA. Apart from accounting identities, ratios are also affected by managerial practice and by other external forces.

Secondly, when observing in log space residuals which are broadly lognormal, a persistent leptokurtosis emerges. The weak, particular, effect is the source of this log positive kurtosis. The strong, common, one can be identified as the source of the Gaussian behaviour of accounting data.

The selection of input variables for statistical models: Since statistical models are to some extent free to select input variables, they can avoid the use of items which impose constraints in other items. Partial or detail items are preferable to items expressing totals. It is especially important to avoid mixtures of both kinds of items.

Ratios and robustness: Since the correctly estimated expected values are not disturbed by constraints, ratio standards can be estimated just by finding their mean values in log space. More sophisticated models — like the free-slope and, to a smaller extent, the base-line ratios — would suffer misleading influences when in the presence of constraints and other forces. They would also become very dependent on the sample used for building the model.

This fact is another example of the relation between robustness and simplicity. Ratio standards are not affected by any anomalies in the distribution of ratios because they only use one degree of freedom. In other words, no consideration of the spread of items is required to model with ratios. Conversely, no disturbances in their spread can affect ratio standards.

We should be alive to the fact that by using free slopes and, to a smaller extent, base-line ratios we loose one of the most attractive features of financial ratios, their robustness.

Chapter 5

The Modelling of Size And Grouping

Size and grouping seem to be the main sources of variability in our data. In this chapter we study both. The correct solving of estimation problems require, prior to any other task, the acceptance of assumptions regarding the characteristics of the input and output spaces. The two main goals of this chapter are therefore the development of appropriate variables for defining an input space and the assessment of the complexity introduced into it by random effects groups often carry with them.

First we discuss, based on empirical evidence, the most appropriate proxy for the common effect, the one which reflects the size of the firm. We conclude that it is possible to build such a variable and that the residuals obtained from it have attractive features for defining input spaces. Next we study the problem of reducing the dimension of such spaces in accounting models. Appropriate methods are developed that incorporate the corollary devised in first place.

Finally we suggest a few concepts and techniques for the assessment of the homogeneity and complexity of effects present in the input space because of existing groupings. We first recommend a standardized measure of the relative degree of homogeneity of each group one by one. We show that in our data industrial grouping cannot be ignored. Next we show that this grouping effect is complex. We conclude that linear tools should be used with suspicion and eventually discarded.

5.1 Selecting an Appropriate Proxy for Size

This section is concerned with the finding of a general deflator reflecting size. Multi-variate accounting models often require size as an input variable. Also ratios intended to reflect departures from expected size could become comparable if their deflator was the same. Such a general deflator would produce easily interpretable residuals.

As we saw in chapter 2, log items can be viewed as a unique common process with some particular variability superimposed. This is true for all the observed items. However, on practical grounds, not all of them are equally adequate for extracting the common effect.

- Items like Sales or Current Assets are almost synonymous in log space. Their particular variability is small when compared with the total variability other similar items exhibit.
- Inventory, EBIT or Funds Flow have more variability of their own. And items having both positive and negative cases exhibit a different behaviour in each of such situations. Positive cases are identical to other variates. The negative ones have a very particular behaviour, as far as we could see. As a few negative cases are always present in samples, the homogeneity of such items, when considered across the whole sample, is bad.
- Finally, Fixed Assets, Working Capital and especially Long Term Debt have large variability of their own. And the non-leveraged firms form a cluster of identical cases. They would severely damage the homogeneity of residuals when considering such items as deflators.

A proxy for the common effect should therefore be selected from the items mentioned in the first place. However, such a proxy would always have, along with the common variability we are interested in, a particular scatter superimposed — the weak effect corresponding to the selected item.

How to isolate the common effect? Is it possible to build a variate reflecting only size and having no particular variability of its own? As we saw in section 3.2.1 the common effect is not directly accessible. However, there is a way of isolating it by building a model which performs the function inverse of ratios. Ratios conceal the common variability and reveal the particular one. This model would conceal the particular variability, thus revealing the common one.

5.1.1 The Case-Average Model

Items like Current Assets, Net Worth, Wages and other expenses, and Sales, can be pulled together to form one unique variate. If we build, for each case in a sample, geometric means (in log space, averages) of these items we can ideally self-smooth their particular components so that the common effect emerges. This is the basis of our method.

Considering a group of items $x_1, \cdots, x_i, \cdots, x_M$ selected as appropriate, and a common effect, s, we explain their variability in log space as the result of an effect, $\sigma = \log s$, common to them all, plus a residual, ε^i particular to each item. In the case of firm j,

$$\begin{array}{rcccc} \log(x_{1j} + \delta_1) & = & \sigma_j & + & \varepsilon_j^1 \\ \log(x_{2j} + \delta_2) & = & \sigma_j & + & \varepsilon_j^2 \\ \vdots & & \vdots & & \\ \log(x_{Mj} + \delta_M) & = & \sigma_j & + & \varepsilon_j^M \end{array}$$

The δ_i are the base-lines eventually present in x_i. We now average the $1, \cdots, i, \cdots, M$ items case by case. For firm j,

$$\sigma_j = \frac{1}{M}\sum_{i=1}^{M}\log(x_{ij}+\delta_i) - \frac{1}{M}\left(\varepsilon_j^1 + \varepsilon_j^2 + \cdots + \varepsilon_j^M\right)$$

Since an average of independent random deviates tends to zero with $1/M$, the number of components, we would have for a large M

$$\sigma_j \approx \frac{1}{M}\sum_{i=1}^{M}\log(x_{ij}+\delta_i)$$

or the equivalent, in ratio form,

$$s_j \approx \prod_{i=1}^{M}(x_{ij}+\delta_i)^{\frac{1}{M}}$$

Once obtained, s could be used as a proxy for size as the denominator of ratios. $\sigma = \log s$ would also be welcome as an input variable for statistical modelling or in the building of tools for visual inspection of residuals.

The difficult point here is the fact that the collection of averaged ε^i are not necessarily independent. Therefore some precautions are required before attempting to build this model, especially when the number and variety of available items are limited.

- The items should not be correlated. A few accounting items are correlated beyond the common effect. Their residuals, after being deflated by a proxy for size, exhibit significant levels of correlation. Sales and Operating Expenses are strongly correlated — and in some industries it is a non-linear correlation —. Wages are correlated with the number of employees. Creditors is correlated with Debtors. The introduction of correlated pairs would reinforce the residual variability common to both instead of smoothing it out.

- The final s should not generate constraints in other items. This is the most difficult condition to achieve. For one reason or another accounting identities seem to propagate across other relations and make themselves present im some unexpected situations.

Practical criteria for building s: We used two criteria for finding the set of items appropriate for building s. The first one is intended to the selection of items. The second one is an overall test of the applicability of the resulting s.

- After the introduction in the case-average leading to $\log s$ of each new candidate, we compute the resulting variance of the average. If it suffers a decrease, the new item is accepted. If it increases, we remove one by one the items already included. For each removed item, if the resulting variance decreases beyond the original value, we replace it by the new one. If the variance never decreases we reject the new item.

	S	NW	W	D	CA	CL	N
1983 VAR:	0.5290	0.5475	0.5473	0.5161	0.5009	0.4999	0.5515
SM:	0.5290	0.5013	0.4971	0.4877	0.4843	0.4800	0.4762
1984 VAR:	0.5807	0.5963	0.6023	0.5349	0.5229	0.5390	0.5972
SM:	0.5807	0.5429	0.5412	0.5251	0.5195	0.5171	0.5138
1985 VAR:	0.5263	0.5591	0.5541	0.4977	0.4829	0.4999	0.5626
SM:	0.5263	0.5032	0.4993	0.4868	0.4812	0.4796	0.4779
1986 VAR:	0.5211	0.5356	0.5419	0.4943	0.5030	0.5009	0.5582
SM:	0.5211	0.4934	0.4920	0.4806	0.4782	0.4771	0.4772
1987 VAR:	0.5318	0.5113	0.5401	0.5003	0.4734	0.4889	0.5646
SM:	0.5318	0.4873	0.4869	0.4795	0.4733	0.4700	0.4716

Table 17: Building s: The evolution of the variance of $\log s$ for an increasing number of incoming items. VAR shows the variance of each log item. SM shows the variance of $\log s$ after introducing each item, when all the items to the left of it are already in.

	C	I	EX	TA	TC	FA
1983 VAR:	0.5350	0.5856	0.5478	0.5051	0.5670	0.6766
SM:	0.4780	0.4835	0.4842	0.4843	0.4870	0.4924
1984 VAR:	0.5776	0.6325	0.6174	0.5317	0.6289	0.7019
SM:	0.5167	0.5232	0.5250	0.5231	0.5258	0.5303
1985 VAR:	0.5405	0.5689	0.5742	0.5049	0.5828	0.6996
SM:	0.4811	0.4843	0.4858	0.4849	0.4882	0.4957
1986 VAR:	0.5384	0.6075	0.5575	0.5119	0.5606	0.6951
SM:	0.4798	0.4862	0.4855	0.4879	0.4918	0.4978
1987 VAR:	0.5314	0.6100	0.5627	0.4835	0.5408	0.6871
SM:	0.4741	0.4790	0.4798	0.4808	0.4855	0.4922

Table 18: The items in this table weren't selected for building $\log s$. SM shows what would have happened to the overall variance of $\log s$ if they were allowed in.

- After finding a model for s with minimal spread we build two-variate scatter-plots in which $\log s$ is compared with each one of all the remaining log items in order to find out if constraints or other asymmetry emerge.

The first criterion would induce misleading models for s if residuals were to be expected to exhibit negative correlations along with the positive ones. In fact, the variance of the case-average could decrease because of existing negative correlations. However, we didn't find so far any traces of negative correlations amongst the residuals of the used items.

An example: In the case of all groups together, the variance of $\log s$ decreased whenever the items S, NW, W, D, CA, CL were introduced in the model, for all the five samples examined — corresponding to reports from 1983 to 1987.

Other items, (C, I, TC, FA) had the opposite effect for all the years. They made the variance of $\log s$ increase. And a few, (EX, N) either made it increase or decrease, depending on the years: N was generally associated with a decrease whilst EX would make it increase except in one year.

Table 17 gathers these results in detail. When reading any row labeled SM from left to right we will get a description of the evolution of the variance of $\log s$ for an increasing number of items allowed in the case-average. For example, by accepting CA, the variance of this average decreased

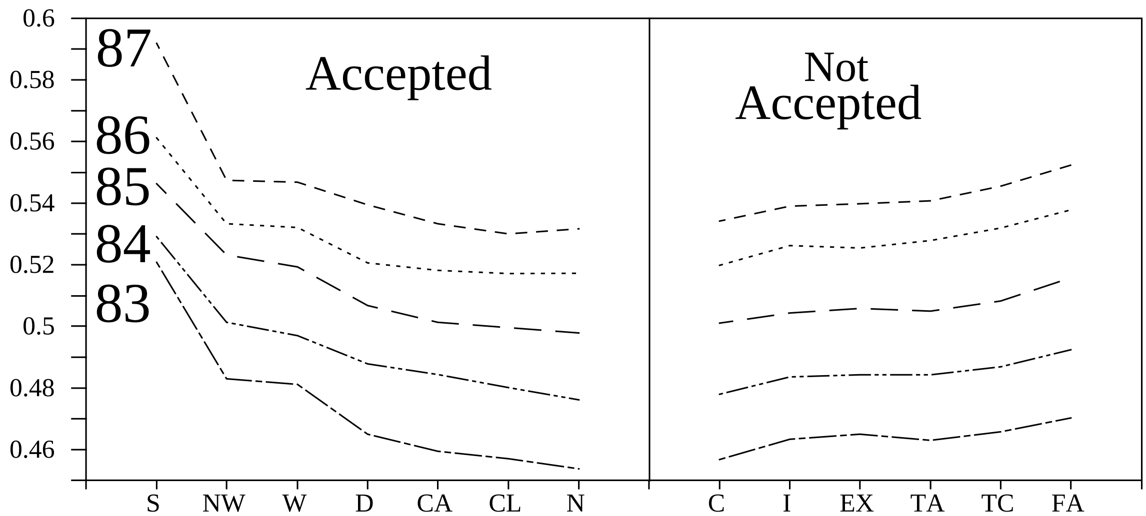

Figure 38: The decrease in variance of s, a proxy for size, for several incoming items.

from 0.4877 to 0.4843 in the 1983 sample. Notice that the acceptance of CA (and the variance of $\log s$ achieved with it) supposes the previous acceptance of S, NW, W, D, that is, of those items to the left of CA.

The individual variance of each item is also displayed (rows VAR). Figure 38 on page 112 is a graphical representation of table 17. It clearly depicts the effect of averaging together more and more items. Notice that, for avoiding the overlapping of the curves, the variance displayed in figure 38 suffered a different translation for each year.

Two items emerged as non-adequate for building s despite not being correlated with others. They were Inventory and Fixed Assets. The variability of s increases when we introduce any of them in the case-average. Fixed Assets is the item with the largest variance amongst the eligible. Its non-adequacy stems from apportioning more variability than the smooth it produces. Inventory has also a large variance but it is not as distinct as Fixed Assets.

Creditors was expected to be non-adequate since it is correlated with Debtors. The same for EX, which is strongly correlated with Sales, and Wages, which is correlated with the number of employees. It is indifferent to select one or the other from these pairs, provide both are not present in the final model.

5.1.2 Results and Discussion

Using the outlined procedure we tried several combinations of items selected from the limited set we displayed in table 2 on page 6. For each combination we observed the behaviour of the resulting s when deflating all the other items in our set. Despite the significant decay in variability obtained, about 10%, none of these combinations turned out to be completely satisfactory since it produced non-exactly symmetrical residuals when deflating items from the Profit and Loss Account. We noticed that when Total Assets deflates such items the asymmetry seems to be smaller than when using s. But, of course, TA performs badly with all the items from the Balance Sheet whilst s will not introduce any asymmetry.

For the set of items we could use, the best $\sigma = \log s$ seems to be

$$\sigma = \frac{1}{7}\left[\log S + \log NW + \log W + \log N + \log D + \log CA + \log CL\right] \tag{13}$$

In the following, any use of s or σ in this study refers to this particular case-average.

When building models like this one, care must be taken to avoid the accumulation of base-lines. Each item must be checked for really significant base-lines (see page 8 in chapter 1) and the corresponding δ accounted for before applying logs. It is worth mentioning that models like the one above involving many variables which are allowed to accumulate their individual effects can display a magnitude of problems unknown in simpler cases.

Why not Principal Components? For building a proxy for size we suggested an average of several items case by case. Such a procedure contrasts with the usual one in accounting research where tasks like this one would be carried out by a Principal Components (P.C.) rotation. We shall comment on the use of P.C. later on. At the moment, we report the following result:

- When comparing the measure of size obtained by averaging with the measure of size the first P.C. yields for the same sample and set of items we very often obtain two variates which are similar except for scale.
- For small, non-homogeneous samples, the results of a P.C. rotation and those of averaging can be very different. P.C. scatters the size effect along the largest two or three axis.

Notice that averaging, summing or any linear combination of items in which the multipliers have equal values is equivalent to a 45^o multi-dimensional rotation. When the P.C. algorithm explores homogeneous samples it finds in first place this 45^o axis in log space, the one corresponding to the proxy for size obtained above. On the contrary, when samples are not homogeneous, the existence of clusters can distort the meaning of the resulting main axis.

Averages are robust regarding the problem of non-homogeneous samples and are easier to compute as well. But they should be used instead of the P.C. rotation mainly because they are functionally correct. Group averages model the functional relation linking individual items with the main source of variability in multi-variate distributions generated by a proportional mechanism.

The need for a large number of components when building s: Another problem with this general deflator is that, if a variable is present both in the numerator and in the denominator, that is, if we deflate with s any item already used for building s, the result is the same as if we were using, instead of the entire numerator, a fractional exponent of it. For example, when deflating item x_k with s we would have

$$\frac{x_k}{\prod_{i=1}^{M} x_i^{\frac{1}{M}}} = \frac{[x_k]^{\frac{M-1}{M}}}{\prod_{i=1}^{k-1} x_i^{\frac{1}{k-1}} \times \prod_{i=k+1}^{M} x_i^{\frac{1}{M-k+1}}} \tag{14}$$

When M is large, $(M-1)/M \approx 1$. But if the number of components of s is small the exponent affecting the numerator will model a non-linear relation. Therefore it would be interesting to find a large number of items for building s. Also because the self-smoothing would improve.

We succeed in finding seven items gathering all the desired requirements. This is all our set can yield. It would be better if we were to have a few more. For a set with more items and especially with items reflecting details, not totals, it would be possible to increase the number of the components of s. Anyway, the set displayed in equation 13 performed remarkably well in the many applications it was called upon during our research. In the second part of this study we often use it.

The distribution of s: The resulting s is two-parametric lognormal for all groups and years. This is not surprising since the individual δ of the candidates are calculated on the basis of maximizing the Shapiro-Wilk's W. In general, $\log s$ has also smaller kurtosis than that observed for individual items. In a few industries negative kurtosis can be observed.

The reason for this small kurtosis is straightforward. The variability of $\log s$ is the one of the main axis of a multi-variate distribution. Since this main axis is supposed to be the source of the Gaussian behaviour of log items, $\log s$ should exhibit a kurtosis even smaller than items.

A two-variate version of this reasoning was introduced in chapter 4. We recall figure 37 (page 102). In this graphical representation, the variability of $\log s$ would be the one along the "Size Axis". The source of positive kurtosis is the "Ratio Axis".

Items are 45^o projections of this multi-variate distribution. They contain some kurtosis. Ideally, $\log s$ shouldn't be correlated with the source of kurtosis.

When used as a deflator in ratios, s yields the same kind of lognormal residuals other deflators produce. The observed leptokurtosis is neither larger nor smaller than the usual.

s and the common effect: A remarkable feature of s is the small departures from the unit in the slopes of regressions in which $\log s$ explains individual log items. For models like

$$\log x_j = a + b \times \log s_j + \varepsilon_j$$

in which x_j is an accounting item, the estimated values of b are, as a rule, very near 1. In table 19 we display the slopes and the proportions of explained variability in our sample for all groups together. Figure 39 on page 116 is a graphical representation of this table intended to facilitate its reading.

Long Term Debt emerges as the item with less explained variability and also the largest departure from the simple ratio model ($b = 1$). However, even in this case, such departure is very small. And

Item	1983		1984		1985		1986		1987	
	slope	R^2	slope	R^2	slope	R^2	slope	R^2	slope	R^2
S	1.006	0.91	1.006	0.89	0.994	0.90	0.991	0.90	1.004	0.89
W	1.027	0.91	1.046	0.91	1.040	0.92	1.033	0.93	1.036	0.93
NW	1.009	0.89	1.010	0.88	1.008	0.88	1.008	0.90	0.991	0.89
I	1.047	0.88	1.054	0.88	1.029	0.86	1.055	0.86	1.043	0.82
D	0.991	0.91	0.972	0.91	0.977	0.91	0.974	0.91	0.985	0.91
C	1.008	0.90	1.015	0.92	1.022	0.92	1.016	0.91	1.017	0.91
CA	0.996	0.94	0.985	0.95	0.995	0.94	0.992	0.94	0.976	0.94
FA	1.075	0.81	1.061	0.82	1.099	0.82	1.103	0.83	1.100	0.82
CL	0.981	0.92	0.986	0.93	0.984	0.93	0.992	0.94	0.984	0.93
FL	1.047	0.84	1.052	0.85	1.033	0.87	1.046	0.85	1.035	0.86
EBIT	1.036	0.78	1.032	0.81	1.024	0.82	1.022	0.83	1.008	0.83
N	0.987	0.84	1.004	0.85	1.010	0.86	1.009	0.87	1.020	0.86
DEBT	1.157	0.62	1.096	0.61	1.100	0.60	1.107	0.56	1.142	0.56
EX	1.000	0.86	1.006	0.84	1.001	0.84	0.990	0.84	0.997	0.83
WC	0.996	0.74	0.976	0.80	0.967	0.75	0.978	0.74	0.934	0.76

Table 19: The slopes and explained variability (R^2) obtained when $\log s$, a proxy for size, is used to explain several log items. All groups together.

the common effect still explains 55% to 65% of Debt in relative space. These values mean a strong correlation and cannot be ignored.

Working Capital comes next, with an explained variability of 75% to 80%. All the other items can be explained by the common effect in a 80% to 94%. And their slope will not be significantly different from 1.

As seen in chapter 2 the displayed results are an argument in favour of the overall proportionality of accounting items in cross-section. This proportionality leads, in log space, to a unique, strong effect. The slope emerges as a non-important parameter. Its value is predictable and departures from such a prediction are very small. They can be explained by the bias introduced in the estimation of b when using regressions instead of functional relations.

A consequence of using s: Ratios with s in the denominator no longer yield contrasts between two departures from size. Ideally, they reflect the real departure from size of the item in the numerator. Using our notation, we could ideally access each ε^x or f^x instead of the $\varepsilon^{y/x}$ or $f^{y/x}$.

As a consequence, we could also use size-adjusted Sales, Working Capital or Debt, along with s as input variables for statistical models. Such models would be self-explanatory to an extent so far not attained in accounting research. Their interpretation would be immediate. In the second part of this study we show examples of this use of s.

5.2 Dimension Reduction for Statistical Modelling

The excessive number of input variables in accounting statistical models and the consequent need for a reduction on the dimension of the input space stems in a large degree from the use of ratios as

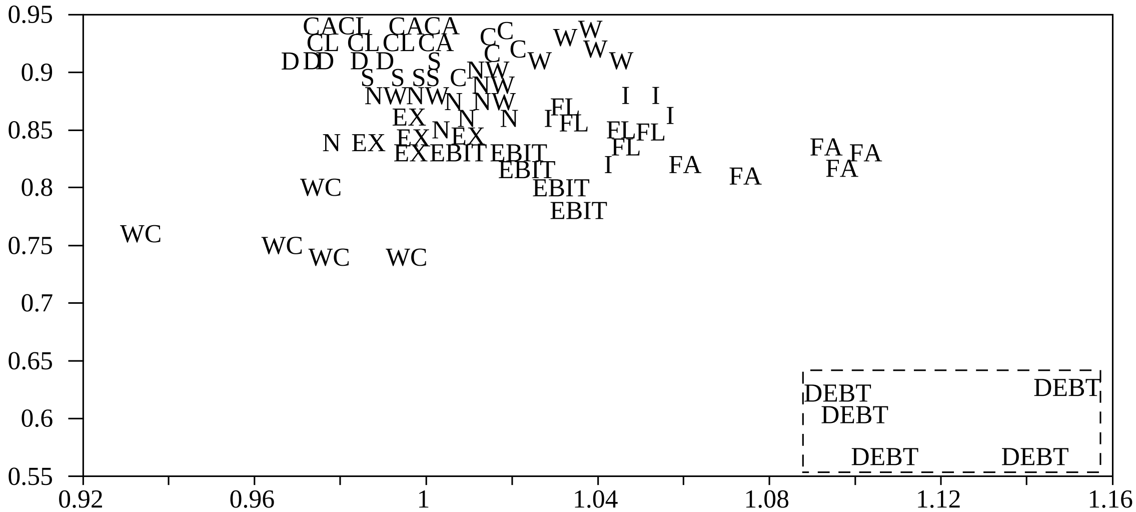

Figure 39: The slopes (X-axis) vs. the R^2 (Y-axis) when $\log s$, a proxy for size, is used to explain log items. Debt emerges as having a particular behaviour. All groups, five years.

inputs. It is difficult to know to what extent a feature of the firm is being conveniently modelled by a given combination of ratios. Some research solves the problem by using all possible combinations.

Since a few residuals (size-adjusted log items) can be used instead of ratios the problem of an excessive number of inputs should now be seen in a different light.

Nevertheless, it is possible that in some cases even the dimension of the input space achieved by using log items instead of ratios will be excessive. This section is devoted to the description of a tool, the Hadamard rotation, intended to dimension reduction and especially well suited for accounting items in log space.

In order to understand why the Hadamard rotation is interesting we must explain first why the usual one, the Principal Components rotation, can eventually become inadequate for dimension reduction with log items.

5.2.1 The Use of the Principal Components Rotation: Discussion

Factor Analysis has been widely used in accounting research. The log space could be understood as an appropriate field for achieving dimension reduction with it. However, when doing so there are some dangers we should be aware of. Here we highlight two of them.

The first danger is specific to accounting items in log space. It consists of the possible finding by the Principal Components algorithm of a main axis which is not a 45^o multi-variate slope. The second one is common to all algorithms based on optimization principles.

P.C. and the modelling of the common effect: P.C. algorithms standardize each variable individually before the rotation takes place. In other words, whenever we use such programs we are

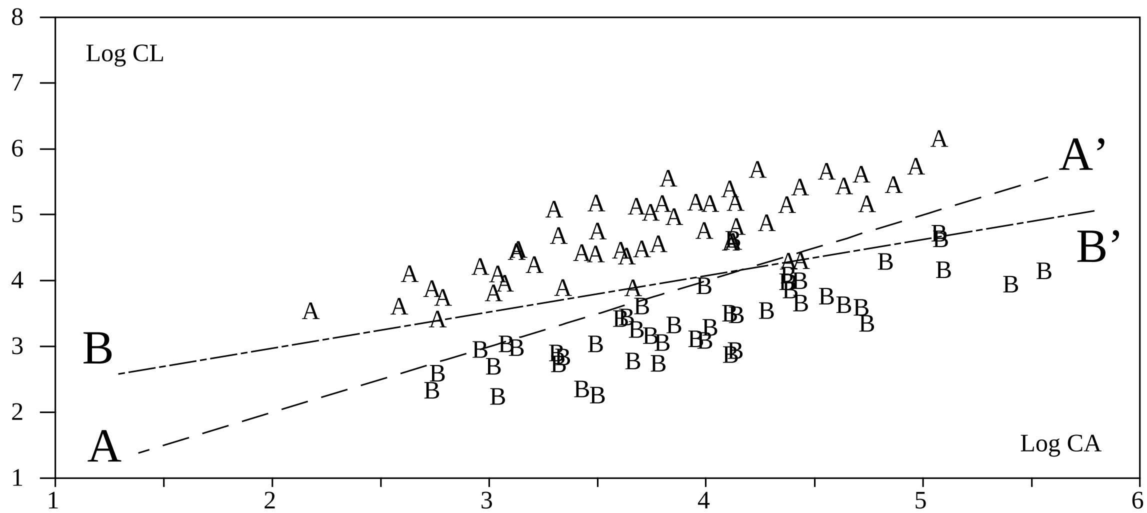

Figure 40: Two clusters of data having parallel principal axis can induce two significant dimensions where only one should be considered.

implicitly making a strong assumption — that for each variate there are two free parameters which ought to be taken into account: The estimated expected value and the variance.

What is the effect of assuming, in log space, that each accounting variable has an independent variance? It is the same as a multi-variate free-slope ratio. We would be using free-slope ratios before controlling for size, thus assuming non-linear proportions between items as a rule, not as an exception.

P.C. is also very sensitive to non-homogeneous samples. The real dimension of the data, that is, the number of independent sources of variability, can be magnified by the existence of clusters inside the sample. Clusters will give the wrong impression of different sources of variability. In figure 40 we represent two clusters of cases having only one significant dimension. P.C. would most probably induce two significant dimensions instead of one. Either lines like AA' or BB' would be candidates for appearing as the first axis. Then, a perpendicular to it would emerge as a second one. None of these axis maximizes its explained variability. None of them is adequate for capturing the effect of size in log space.

In short, the common effect requires a 45^o rotation to be accounted for. Not an optimal rotation.

Notice that the referred problem can be avoided by using the P.C. rotation after manually controlling for the common effect.

The problem of the negligible variability: Other potential dangers of using P.C. are not specific to accounting models. A clear one stems from these tools being intended to the extraction of features. Dimension reduction becomes a by-product of the extraction of features, not the main goal.

Features extraction makes sense in an identity context. The negligible variability is the one unexplained by the discovered features of the data. Dimension reduction makes sense also when modelling relations between the data and observed outcomes. It is a constraint imposed from outside in the flow of information allowed to explain a relation.

These two tasks are different. In an identity map the input variables try to explain themselves. When modelling a relation they try to explain the relation between them and the outcomes.

When a set of features are trying to explain themselves in the most economic way, the common variability will emerge. The particular one can eventually be considered as negligible. Clearly, such a particular variability — negligible in an identity map context — could be important in other contexts. In fact, the task required when performing discriminant analysis or regression is not for the input variables to explain themselves but the outcomes.

Figure 40 also depicts this. Supposing that the maximum-variance axis was correctly found, it would lie in between $AA^{'}$ and $BB^{'}$, in a way that would separate both groups. If we select such axis — and therefore we consider as negligible the second axis in order to reduce the dimension of the input space — and then we apply some discriminant technique to try to distinguish between group A and B, we would have thrown away the axis containing precisely the useful information for such a task — the second one.

5.2.2 The Hadamard Rotation

For achieving dimension reduction in log space we suggest the use of the Hadamard rotation [107]. Its remarkable quality is that it is not optimal in any sense. The Hadamard rotation could be briefly described as a multi-dimensional 45^o anti-clockwise rotation. The results of applying it will not depend on the particular statistic of the data being rotated. In other words, the Hadamard rotation exhibits the same kind of robustness ratios have.

For M items to be reduced, the effect of applying the Hadamard rotation is twofold.

- First, all the variability along the multi-variate 45^o axis — defined in the M-dimensional space of the items to be reduced — is accounted for and placed in the first extracted factor. Thus, the first factor will contain, up to a constant value of $1/M$, the strong, common effect, s, as modelled by the M items to be reduced.

- Second, the remaining variability, the weak effect particular to each item, is re-distributed by all the other $M-1$ factors according to simple combinatorial laws. Each item will be present in these new variables either summing or subtracting to the total.

We consider the Hadamard rotation as adequate for achieving dimension reduction in the input space of the log accounting items because it is able to isolate the only clear feature of the data — the common effect — and then it re-distributes the remaining variability by the dimensions we want to use in a way that will not privilege any particular piece of information.

It is up to the next step — the modelling of the relation — to determine how many variables are to be used. This will determine the information flow.

The Hadamard rotation is also easy to implement. It will not require any special algorithm.

How to build Hadamard matrices: Hadamard matrices are square arrays of only plus and minus ones disposed in such a way that the rows or the columns are orthogonal to one another. The simplest case is the matrix of order two:

$$H = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

If we are examining just two accounting items in log space, $\{\log y, \log x\}$, we would obtain, by applying H, two new variables, $h_1 = \log y + \log x$ and $h_2 = \log y - \log x$. Therefore, h_2 is the ratio y/x in log space and h_1 is the product $y \times x$ in log space. h_1 can be identified — up to a constant value of $1/2$ —as the effect common to both items.

The Hadamard rotation decomposed the total variability so that the first new variable contains the variability common to the original items and the second new one the variability unexplained by the first new one — which is the ratio of the original items. This was achieved by rotating the original axis 45^o anti-clockwise. The new X-axis is now the main dimension of the distribution. The new Y-axis is orthogonal to it. Again, we recall figure 37 (page 102). In this graphical representation, the variability of h_1 is the one along the "size axis". The one of h_2 is the one along the "ratio axis".

For more than two items, the i^{th} row and j^{th} column component of one possible M dimensional H matrix can be found by applying the formulas

$$a_{ij} = (-1)^{b_{(i,j)}} \quad \text{where} \quad b_{(i,j)} = \sum_{l=1}^{M-1} i_l \times j_l$$

The terms i_l and j_l are the bit states (+ or -) of the binary representation of i and j respectively. For example, in the case of $M = 8$ one possible H matrix would look like this:

$$H = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 \end{bmatrix}$$

Notice that, when converting the new factors obtained in log space back to the anti-logarithmic space summation become multiplications and divisions are now subtraction. Hence, the Hadamard rotation will produce a set of $M - 1$ new variables which are ratios. These ratios contain all the original items appearing either in the numerator or in the denominator.

For example, in the case of $M = 4$, the three ratios formed with items A, B, C, D would be:

$$\frac{A \times C}{B \times D}\,; \quad \frac{A \times B}{C \times D}\,; \quad \frac{A \times D}{B \times C}, \quad \text{along with the first factor} \quad A \times B \times C \times D$$

Supposing we were to decide to consider as negligible the information conveyed by the third factor — which is the ratio AB/CD — then we would get a new information content equivalent to considering

$$\begin{cases} (A^3 \times \frac{C \times D}{B})^{\frac{1}{4}} & \text{instead of} \quad A \\ (B^3 \times \frac{C \times D}{A})^{\frac{1}{4}} & \text{instead of} \quad B \\ (C^3 \times \frac{A \times B}{D})^{\frac{1}{4}} & \text{instead of} \quad C \\ (D^3 \times \frac{A \times B}{C})^{\frac{1}{4}} & \text{instead of} \quad D. \end{cases}$$

This is what is meant by saying that the H rotation re-distributes the variability of items according to simple combinatorial laws. Whenever the desired goal is just dimension reduction, not the isolation of particular features of the data, this simple re-distribution is enough.

Re-distribution avoids the emergence of factors containing very particular pieces of information. Algorithms based on optimization will easily find, should they exist, preeminent features in the data. Therefore, the dimension reduction they achieve can have the undesired characteristic of being features-oriented. As commented above we would be probably throwing away features, instead of constraining the information flow.

How to use the Hadamard rotation: In order to achieve a reduction in the dimension of the input space, the H rotation is applied to the set of mean-adjusted log-items. Notice that only summation and subtraction are required to rotate the input vectors. Then, the variance of the obtained factors is measured. Those exhibiting the smallest spread are, in general, the ones to put aside.

Since the variability particular to each item is re-distributed by all the $M - 1$ factors after controlling for the common effect, it is clear that differences observed in the spread of such factors must be the result of correlations between the weak components differently combined. As any correlation means redundancy in the information content of variables, the factors to be thrown out are those having smaller spread.

An example: Using a sample with accounts of 169 firms belonging to four industrial groups (1984) we first selected eight items — NW, W, D, C, CA, CL, S, N — and applied the H rotation to their mean-adjusted logs. We obtained eight new variables. The variance observed in each item was

NW	W	D	C	CA	CL	S	N	sum
2.72	2.76	2.66	3.04	2.61	2.81	2.77	2.53	21.91

The resulting factors had their variability distributed in this way:

Fac 1	Fac 2	Fac 3	Fac 4	Fac 5	Fac 6	Fac 7	Fac 8	sum
20.63	0.26	0.12	0.14	0.08	0.06	0.43	0.18	21.91
94.13%	1.19%	0.55%	0.65%	0.35%	0.29%	1.98%	0.84%	

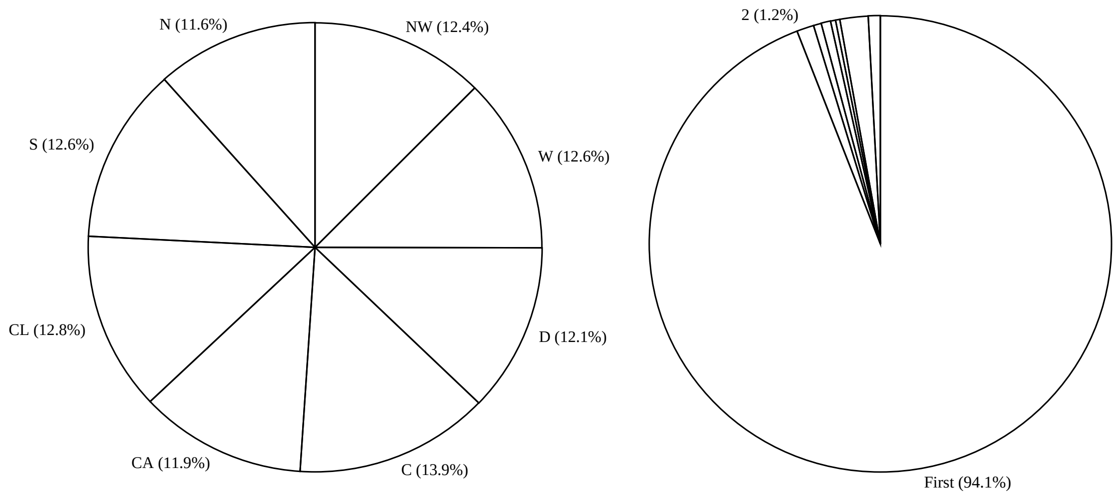

Figure 41: On the left, the distribution of the variability of eight items. On the right, the variability of factors after the Hadamard rotation.

Figure 41 is a graphical representation of the above data. The displayed proportions are, of course, only approximations. The factors resulting from the H rotation are not independent and their variance will not sum up to a total.

As we see, the first factor accounts for approximately 94% of the variability of the sample. This is a typical value attained by the common effect. Given that the smallest variability was that of factors 5 and 6, we considered such factors as negligible. If we were modelling a particular relation we would now use the six remaining factors as input variables.

In this case we were interested in reconstructing the original items from just the six accepted factors. Hence, we filled with zeros or constant values the two negligible factors for all the cases in the sample and applied the rotation inverse of H — which, in this particular case ($M = 8$) is the same as H itself, up to a constant value —. In doing so we obtained the original items reproduced after a reduction of two in the dimension of the input space.

The differences between the correlation matrices of the original and reproduced items are shown in table 20 (page 122). Since they are small, we conclude that the Hadamard rotation succeeds in reducing the dimension of the original items with no significant loss in their information content.

A P.C. rotation with the same data didn't succeed in finding the 45^o axis as the main dimension. Instead, it apparently scattered the common effect over the three largest factors. This is not a particularly clumsy result since the sample was drawn so as to be quite non-homogeneous. The four selected groups — Building Materials, Industrial Plants, Clothing and Food — are not usually found in the same sample except on purpose.

We repeated the above experiment with twelve items. The resulting factors's variance were now

NW	W	D	C	CA	CL	S	N
0							
-0.0051	0						
-0.0002	0.0111	0					
-0.0206	0.0044	-0.0069	0				
-0.0147	0.0128	0.0073	0.0094	0			
-0.0178	0.013	-0.0107	-0.0067	0.0097	0		
-0.0415	-0.0005	-0.0203	-0.005	0.0012	-0.0037	0	
-0.0236	-0.0083	-0.0202	-0.0274	-0.0069	-0.0058	-0.0193	0

Table 20: The difference between the correlation matrices of eight original items and their reproduction after reducing the dimension by two.

NW	W	D	C	CA	CL	S	N	EX	TA	FA	I
0											
-0.03	0										
-0.06	-0.06	0									
-0.07	-0.05	-0.01	0								
-0.02	-0.02	0.003	0.000	0							
-0.03	-0.04	0.002	-0.00	0.038	0						
-0.02	0.003	-0.01	0.015	0.060	0.026	0					
-0.01	0.013	-0.05	-0.06	-0.01	-0.04	0.020	0				
-0.06	-0.04	-0.02	-0.00	0.036	0.012	0.013	-0.03	0			
-0.00	-0.03	-0.00	-0.00	0.008	0.032	0.052	-0.03	0.008	0		
-0.00	-0.02	-0.05	-0.03	-0.04	0.008	0.046	0.007	-0.00	-0.01	0	
-0.03	-0.00	-0.04	-0.03	0.009	-0.00	0.000	0.046	-0.03	0.020	-0.02	0

Table 21: The difference between the correlation matrices of twelve original items and their reproduction after reducing the dimension by two.

more irregular since a few of the emerging ratios are non-balanced. We call non-balanced a ratio with a different number of items in the numerator and in the denominator.

The ratios emerging after the Hadamard rotation are all guaranteed to be balanced only when the dimension of the input space is a power of two. However, for most of the remaining M it is possible to build Hadamard matrices yielding balanced ratios. M must be even, of course.

Fac 1	Fac 2	Fac 3	Fac 4	Fac 5	Fac 6	Fac 7	Fac 8	Fac 9	Fac 10	Fac 11	Fac 12
31.85	0.265	0.057	0.131	3.973	0.091	0.243	0.232	3.233	0.227	0.220	0.240
78.13%	0.65%	0.14%	0.32%	9.75%	0.22%	0.60%	0.57%	7.93%	0.56%	0.54%	0.59%

After identifying the two factors with the smallest variance we obtained reproduced items. The differences in the correlation matrices observed between the original and the reproduced items are displayed in table 21.

The Hadamard Symmetrical rotation: Apart from the outlined features, when the number of items to be reduced is a power of two, the Hadamard rotation becomes symmetrical, that is, the matrix is its own inverse up to a constant value.

If $M = 2^m$, for any input vector X with M elements,

$$Z = \frac{1}{\sqrt{M}} \times X \cdot H \quad \text{and} \quad X = \frac{1}{\sqrt{M}} \times Z \cdot H$$

Another characteristic of this particular case is that H matrices can be obtained from others of smaller dimension just by using the simple formula

$$H_{2M} = \begin{bmatrix} H_M & H_M \\ H_M & -H_M \end{bmatrix}$$

For $M > 12$ the difference between symmetrical and non-symmetrical H becomes non-important in practice. It is then possible to find a Hadamard matrix yielding balanced ratios. Anyway, the effect of having non-balanced ratios is no longer noticeable in high-dimensional input spaces.

Conclusion: The Hadamard rotation seems to be a simple and well-fitted method for achieving dimension reduction in the input space of accounting data. It yields a proxy for size and a collection of ratios corresponding to different combinations of the original log items. In chapter 8 we use the same basic principle in the building of graphical tools for financial diagnostics.

5.3 The Homogeneity of Industrial Groups

The modelling of accounting relations in the presence of groups cannot avoid two important questions concerning the input space.

- Is a particular grouping significant so that it should be taken into account? If the data is more similar inside groups than from group to group this is the case.
- Can significant groups be taken as similar in their effects upon the features of the data? For example, is liquidity affected in the same way as, say, profitability in the presence of an industrial grouping?

We answer the first question by comparing the variability inside groups with the one between them for the most important features of the data. As a result we obtain an overall measure of the importance of a grouping for each variable involved. The second question could be answered at several levels of accuracy. The simplest procedure would consist of just ranking a measure of homogeneity of each feature by group and then verify if these rankings were consistent across different features. We use a basically similar method.

In this section we develop procedures for assessing the importance of groupings and also the complexity they introduce in the input space regarding the features of the data. The tested grouping is the SEIC industrial classification. However, any other grouping can be explored by these or similar methods.

5.3.1 Introduction and Related Research

Accounting reports do not contain all the information necessary to uniquely characterize the important features of firms. The very basic problem of ratio analysis is the existence of similar accounting patterns which are not neighbours in the space of the real features of firms. In order to correctly map firm features accounting data is not enough. External information is also required.

A clear example of this is the industrial classification. The similarity of firms as perceived by the SEIC can be different from the similarity of accounting reports. A non-standard piece of information, the number of employees, turns out to be important when checking the homogeneity of industrial groups. Other non-accounting variables, eventually also important, could be the patterns of consumption of energy, area requirements for plant or stores, the age of the firm and its location.

The use of a limited amount of information like the one contained in accounting reports alone, generates extra unexplained variability in models. Here we are not concerned with the amount of variability. We are interested in its complexity. The complexity of accounting models becomes higher when the input data reflect facts we cannot account for.

For example, the Leather or the Wool industries could introduce more than one effect in the input space if the accounting numbers of each firm were strongly influenced by its location. This location would act as a hidden grouping: Say, the north and the south, each one with particular characteristics. In such case, the complexity of the SEIC grouping would be larger than expected.

Limitations when answering the first question: In chapter 7 we shall evaluate the discriminatory power of accounting reports when classifying firms according to the SEIC. Notice that such a problem is different from the one we are exploring now. Several very homogeneous groups of firms can also be very similar in their features as perceived by accounting reports alone. When groups are homogeneous and also overlap, they cannot be correctly separated by accounting information.

Similarity can only be measured in a comparative way. We can rank groups according to their homogeneity but we cannot say that an attained degree of homogeneity is acceptable while another isn't. We cannot say that the SEIC succeeded or failed in creating homogeneous groups. We can only say that one group is more or less homogeneous than another one regarding particular features.

The second question: Grouping and features. Given a grouping, some features of the firm will be sensitive to it, varying from group to group. Others will be insensitive to the grouping. The tracing of both kinds of behaviour is potentially important for ratio analysis. Ratios reflecting sensitive features are interesting because they can be used to discriminate between groups. And ratios reflecting insensitive features are also interesting. They yield robust standards or benchmarks.

Ratio Analysis is concerned with mean values. Statistical modelling is mainly concerned with the sources of spread. Are the different sources of spread equally sensitive to a grouping?

Fixed and random effects: Some groupings are defined a-priori by an accepted institution like the SEIC in the U.K. Others are the result of objective causes. The grouping of firms into leveraged and non-leveraged or into failed and non-failed has a statistical nature which is different from the

SEIC grouping. The former introduces in the population a simple partition. The later introduces real variability.

Whenever we consider grouping variables that can make themselves present — in one way or the other — in the input space, the first point to clarify is its statistical nature. Simple partitions are known as fixed effects. Groupings which introduce random effects are known as such.

Groupings that introduce random effects in a population can indeed introduce more than one. The assessment of the number of independent sources of spread a grouping carries with it is eventually important. For example, if a particular grouping contains two random effects it is likely to induce higher order relations between input variables.

Related research: Firm grouping is itself not a very homogeneous body of research. It includes simple industry comparisons of ratios, tests on widely accepted groupings of firms and the search for clusters of firms according to similarities of ratios and other data. The former topic has been explored from very early in finance literature. Foster [44] offers an overview. There is an established evidence on differences between some ratios for well known industry groups.

The search for clusters of firms has been carried out by Elton and Gruber [35] [36], Jensen [67], and Gupta [54] amongst others. The problem with hierarchical cluster analysis is that the used algorithms seem able to always find clusters. The interpretability of the results suffered with this. It is difficult and not particularly revealing.

Another aspect of this research is the test of separability of groups using accounting data. This has been tried for the SEIC by Sudarsanam and Taffler [124].

A not well known body of research tested the homogeneity of existing groupings according to accounting measures of financial risk. Equivalent risk class hypothesis tests began with Wippern in 1966 [139] and vanished after Martin et al. in 1979 [83] (see also [49] and [99]). Foster doesn't even mention them. This research is an attempt to test a basic assumption in Finance, the one that groups are similar before risk.

5.3.2 Measuring the Significance of a Grouping

The techniques designed to divide the variability of cases in two components, inter-groups and intra-groups, have in common the basic Analysis of Variance model but differ in the assumptions. Given that the 14 groups selected represent a sampling between a much larger amount of possible choices it would seem inappropriate to use fixed-effects models. Hence, we explore a random-effects one.

Our answer to the first question is given by a statistic, the intra-class correlation, able to yield measures of similarity comparable across different samples. Notice that this kind of tools are sensitive to deviations from the Gaussian assumption. The fact that they can now be used is a consequence of the broad lognormality of firm features.

The problem: We are interested in assessing the extent to which the used grouping of industries is effective in creating more similar subsets of firms.

The intra-class correlation coefficient, ρ, measures the proportion of the total variability that is associated with a grouping. It is a standardized way of comparing the spread within groups with the one between groups when the effects are random. If s_b^2 is the expected value of the mean-squares between groups and s^2 is the corresponding mean-squares within groups, then an estimator for ρ could be

$$r = \frac{s_b^2 - s^2}{s_b^2 + (k-1) \times s^2}$$

k is the number of cases in each group. For M groups of unequal size $n_i, i = 1, M$ and $N = \sum n_i$, k should be approximated as

$$k = \frac{1}{M-1} \times \left(N - \frac{\sum n_i^2}{N} \right)$$

It is possible to estimate confidence intervals for r. A detailed discussion of this statistic and the way it is derived can to be found in Snedecor and Cochran [119].

Real groups are expected to be more similar than the whole. The more similar groups are — regarding the sample — the more the correlation intra-classes approaches 1.

When the variability inside groups is smaller than the one in the whole sample, this measure yields a positive value. For a variability inside groups similar to the one between them the intra-class correlation yields zero. In the case of groups containing more spread than the whole, a negative correlation would emerge. When the effects governing the spread within groups and between them are independent, negative ρ cannot occur. Negative ρ emerge only in cases where the effects interact.

The data: During the usual period of five years we examined three kinds of accounting information. First, several log items and also s, our proxy for size. Second, the log residuals of the same items when deflated by s. Finally, the log residuals of a few more ratios.

In tables 22, 23 and 24 on pages 127 and next, we display the estimated intra-class correlation along with the F statistic. The number of firms involved ranges from 555 to 702 in 14 groups.

Results: Log items. The log items show a small but significant increase in homogeneity owing to the industrial grouping. The values of ρ are stable during the considered period and no negative or zero cases were observed. They are not very different from one another as expected. In fact, since log items mainly reflect size they yield similar proportions of variability associated with grouping. We conclude that the relative size of firms, as reflected by accounting data, is slightly more homogeneous inside industries than for the whole sample.

Fixed Assets, Debtors and Sales are the most homogeneous log items inside groups (10%). Inventory is the least homogeneous (4%). Size itself is similar to many other items (5%). On the whole, the homogeneity ranges between the extreme values of 3% and 13%.

Item	1983		1984		1985		1986		1987	
	F	ρ	F	ρ	F	ρ	F	ρ	F	ρ
SIZE	5.34	9.87%	4.42	6.88%	4.48	6.70%	4.86	7.15%	4.87	7.31%
S	7.21	13.53%	6.61	10.79%	6.53	10.27%	6.48	9.85%	5.93	9.11%
NW	4.00	7.09%	3.94	6.00%	3.63	5.23%	4.20	6.10%	4.17	6.15%
W	5.01	9.26%	3.51	5.17%	4.09	6.03%	4.52	6.59%	4.66	6.96%
I	3.31	5.56%	2.85	3.87%	3.03	4.08%	3.12	4.12%	2.99	3.98%
D	7.02	13.20%	6.51	10.65%	6.31	9.91%	6.47	9.85%	6.39	9.92%
C	5.68	10.58%	5.17	8.26%	5.22	8.03%	5.48	8.22%	5.45	8.35%
FA	6.45	12.08%	6.29	10.23%	6.52	10.26%	7.18	11.00%	6.65	10.35%
CA	4.62	8.42%	4.10	6.32%	3.80	5.51%	3.86	5.44%	3.60	5.07%
CL	5.42	10.03%	5.14	8.20%	4.97	7.59%	5.67	8.52%	5.06	7.65%
N	4.41	8.00%	3.55	5.24%	3.93	5.76%	4.06	5.77%	4.08	5.94%
EBIT	5.88	11.73%	4.46	7.41%	3.49	5.26%	3.74	5.61%	5.45	8.86%
FL	5.90	11.51%	4.75	7.76%	4.35	6.76%	4.36	6.60%	4.94	7.82%
DEBT	4.44	11.86%	3.04	6.10%	3.84	7.66%	3.54	6.43%	4.08	7.81%

Table 22: The F statistic and the Intra-Class correlation, ρ, when log items were used to explain the industrial grouping.

Results: Log residuals. For size-adjusted variates the contrast between different items clearly increases. Some of these residuals show a much larger homogeneity intra-groups than others.

The consistency for the considered period of five years is not affected in most of the items but it is completely lost in a few. Gross Funds from Operations and EBIT, for example, plunge from a strong similarity inside groups to a much smaller one from 1986 on. It seems as if profitability were increasingly non-homogeneous inside industries. See table 23 on page 128.

Sales and the number of employees are the most similar features inside groups. The SEIC industrial grouping seems to rely on these variates as a criterion for determining groups. Next, Debtors and Wages. Debt and Net Worth are the less homogeneous residuals. In fact, almost no influence of grouping can be detected in their size-adjusted measures. The financial structure of firms seems not to be sensitive to industrial groups.

On the whole, the homogeneity of the residuals ranges from about zero to 25%. These values denote a more diversified influence of the industrial grouping upon residuals than upon size.

Results: A few ratios. In table 24 on page 128 we display the intra-class correlations for a few more ratios. The above size-adjusted measures are also ratios, of course. But the displayed ones capture contrasts between residuals, not the residuals themselves.

The Long Term Debt to Net Worth ratio shows no traces of recognizing the SEIC grouping as such. The liquidity ratio yields measures of similarity comparable with those of non-deflated items. Ratios incorporating Sales, Wages, Debtors or the number of employees clearly recognize the tested grouping. If our goal would be the identification of ratios appropriate for recognizing industrial groups then the W/N ratio would emerge as a good choice.

A method for selecting appropriate ratios for given tasks could consist of using ρ. First, the intra-class correlations of many size-adjusted log items would be assessed. Then, the most promising

Residual	1983		1984		1985		1986		1987	
	F	ρ	F	ρ	F	ρ	F	ρ	F	ρ
S	11.79	21.41%	15.27	23.54%	16.83	24.64%	13.94	20.53%	11.39	17.45%
NW	1.54	1.35%	1.92	1.96%	1.69	1.43%	1.71	1.42%	1.61	1.25%
W	5.28	9.84%	6.18	10.12%	6.34	10.00%	6.85	10.50%	6.97	10.87%
I	4.36	7.88%	4.87	7.77%	4.35	6.57%	5.36	8.10%	3.35	4.65%
D	8.57	16.06%	11.05	17.85%	9.40	14.82%	7.18	10.98%	6.06	9.36%
C	3.46	5.85%	3.45	5.03%	2.86	3.71%	2.77	3.42%	3.15	4.22%
FA	2.23	3.00%	2.87	3.89%	3.35	4.65%	4.00	5.67%	3.69	5.20%
CA	2.79	4.36%	3.87	5.85%	4.11	6.08%	4.02	5.72%	2.04	2.09%
CL	2.93	4.65%	3.94	5.96%	2.67	3.34%	2.32	2.57%	1.69	1.40%
N	9.75	18.24%	13.42	21.23%	12.57	19.42%	12.10	18.17%	9.46	14.76%
EBIT	6.73	13.49%	5.16	8.77%	3.30	4.87%	2.98	4.11%	3.35	4.89%
FL	8.13	15.91%	6.13	10.29%	4.55	7.13%	2.79	3.62%	3.65	5.40%
DEBT	1.52	1.99%	0.97	-0.08%	1.85	2.41%	1.59	1.56%	1.91	2.44%

Table 23: The F statistic and the Intra-Class correlation, ρ, when several different log residuals were used to explain the industrial grouping.

Residual	1983		1984		1985		1986		1987	
	F	ρ	F	ρ	F	ρ	F	ρ	F	ρ
S/N	10.08	18.80%	14.22	22.30%	13.44	20.57%	12.84	19.14%	11.08	17.11%
$(S \times N)/s^2$	11.50	21.14%	15.75	24.24%	16.47	24.39%	13.70	20.25%	9.98	15.52%
W/N	11.14	20.62%	16.67	25.42%	18.43	26.62%	20.25	27.86%	16.09	23.61%
CA/CL	3.71	6.44%	4.76	7.56%	3.86	5.63%	5.04	7.51%	2.56	3.11%
$DEBT/NW$	1.65	2.51%	0.97	-0.10%	2.07	3.08%	1.73	1.96%	1.90	2.45%
$S/EBIT$	8.00	16.02%	6.98	12.14%	8.69	14.62%	6.67	10.97%	5.70	9.32%

Table 24: The F statistic and the Intra-Class correlation, ρ, when a few log ratios were used to explain the industrial grouping.

combinations of items would be selected amongst the residuals with highest ρ and tested.

Conclusions: The industrial grouping clearly gathers firms which are to a small extent more similar regarding size.

Also, a few features of the firm are more homogeneous inside industries. It is the case for Sales, Wages, the number of employees or Debtors. The financial structure of firms is not especially more similar inside groups. And the measures of profitability seem to yield very different results from year to year. In the early years of our observations the profitability of firms is remarkably similar inside the same industry. In the later ones (1986, 1987) it becomes irregular.

There is nothing in the obtained results able to defy the common-sense of accounting knowledge. The results are expected. A very simple technique yielded consistent and interpretable results.

5.3.3 Assessing the Complexity of Groupings

The methods and results of this section are not particularly interesting for fields other than the multi-variate modelling of relations. We are interested in broadly knowing if it is acceptable to consider one unique random effect in the SEIC grouping regarding accounting data. The results of this experiment are important later on, when relating sensitivity of assets to market returns.

The method: Building maps from distances. As remarked before, this problem can be treated with different levels of accuracy. Here we selected a very simple and intuitive level. It could be much improved so that we would end up with a real complex instrument for measuring complexity. Since our goal is not only the assessment of complexity itself but the support of further research, we used a reasonably simple instrument.

Our method is based on the well known possibility of re-constructing maps from distances. For example, it is possible to re-construct a map showing the relative positions of the main cities in Britain just by knowing the distances between them.

Cities would require two dimensions to be mapped. When the objects to be mapped lie in a straight line the result of this re-construction can be expressed, if desired, as a simple ranking. Objects positioned so as to form a two-dimensional map cannot be ranked.

We are interested in discover if it is acceptable to rank the industrial groups according to the spread of accounting features. If it turns out that the different groups can be ranked according to their spread, then the SEIC grouping is likely to introduce just one random effect. On the contrary, if the spread of groups resist a simple ranking — thus requiring a two-dimensional map like in the case of cities — it means that the spread present in accounting features because of the grouping is complex or higher dimensional.

In fact, if the grouping is a unique effect it will impinge upon the features of the firm in different degrees but not in different directions — thus yielding a consistent spread for several features. For example, if Chemicals have a smaller spread in liquidity when compared with the one of Food, then a unique random effect would mean that Chemicals would also exhibit a smaller spread in profitability or any other feature. But if in the former group there is a smaller spread in liquidity when compared with the Food industry and a larger spread of profitability in the same circumstance, then the randomness present in the grouping is not one-dimensional. Higher order effects are expected.

The method we developed for testing the complexity of grouping consists of:

- First, the spread of several features of the firm is measured for a sampling of groups. We used the standard deviation of log data as a measure of spread. The standard deviation is a one-dimensional measure of spread. Other statistics, like the variance and co-variance matrix, are multi-variate measures of joint spread.

- Then, joint distances between industrial groups are computed from the above measures. One typical such distance could be the Euclidean distance. Notice that the use of joint distances — one distance is measured in the space of several variables — doesn't change the one-variate character of our method.

- Finally ordinal scores are discovered that position each industrial group according to the above distances.

The final result is a map. Each industry is a position in that map. The coordinates of each industry are the obtained scores. For example, the position of Leather would be determined by a vector of scores. The first score positions Leather according to the first dimension, the second positions it according to the second dimension, and so on.

If industries lie in a straight line dimensions other than the first one are negligible. This means that groups can be ranked according to a unique measure of spread. They are affected by the grouping in different degrees but not in different ways. On the contrary, if a vector of two scores is required for conveniently positioning groups, they lie in a plane, not in a straight line. In this case groups cannot be correctly ranked using a joint measure of spread. We conclude that the variability introduced by grouping cannot be considered as one effect.

Our method has the important quality that it clearly points out which industries are likely to be contributing to an increase in complexity. Such groups should be kept out of the sample whenever we want to study reasonably simple cases.

The data: We used two sets of data. First, the standard deviations of all the log items used in previous section. Second, the same for standard deviations of log residuals.

The first set is a comparison term. Clearly, as log items reflect mainly relative size there is no room for cross-effects. The results obtained for the second set can be compared with these. The experiment was carried out for the usual period of five years.

As a sampling of groups we used the 14 industries referred to in table 1 (page 5). Indeed, it is not a random sampling. But the number of groups is large when compared with the total of industries. And the results obtained with this particular sample will condition the selection of groups in experiments carried out in the second part of this study.

Results: Log items. It is possible to compute a statistic, the usual R^2, showing in what proportion each dimension accounts for the goodness of the fitted map. And, of course, it is also possible to use the first two dimensions for building visual representations of the obtained positions like in figure 42 on page 131.

R^2 measures the proportion of variance in the ranked or mapped data which is accounted for by such ranking or mapping. The obtained proportions are very high, denoting an essentially one-dimensional map. However, even in this case, the second dimension cannot be ignored. For the five periods the R^2 were:

Year	One dimension	Two dimensions
1983	90%	98%
1984	90%	99%
1985	97%	99%
1986	92%	98%
1987	95%	99%

Figure 42 (left) shows the two-dimensional maps of industries for the five years. The X-axis is

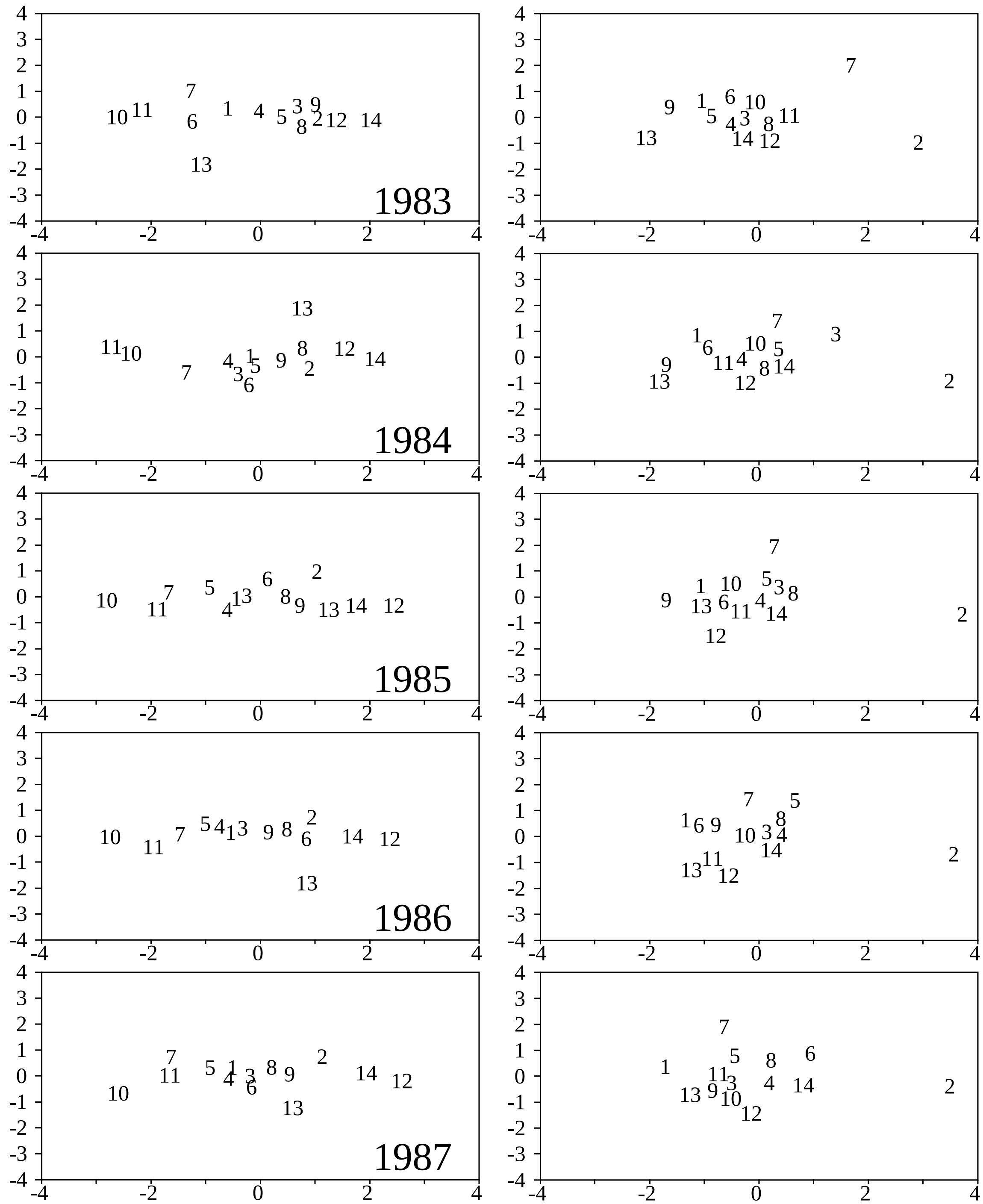

Figure 42: Each one of these scatter-plots is a two-dimensional map showing the position of industries. The X-axis is the first dimension. The Y-axis is the second one. On the left, log items for five years. On the right, log residuals for the same period.

Number	Industrial group	1983	1984	1985	1986	1987
1	BUILDING MATERIALS	-0.31	-0.16	-0.29	-0.41	-0.34
2	METALLURGY	0.85	0.80	0.88	0.80	0.88
3	PAPER AND PACKING	0.65	-0.29	-0.40	-0.20	-0.12
4	CHEMICALS	-0.04	-0.36	-0.41	-0.51	-0.41
5	ELECTRICITY	0.22	-0.16	-0.45	-0.63	-0.65
6	INDUSTRIAL PLANTS	-0.76	-0.34	0.07	0.54	-0.12
7	MACHINE TOOLS	-0.88	-0.98	-1.21	-1.02	-1.20
8	ELECTRONICS	0.72	0.54	0.33	0.31	0.10
9	MOTOR COMPONENTS	0.84	0.67	0.48	0.15	0.44
10	CLOTHING	-1.81	-1.78	-2.03	-2.02	-1.97
11	WOOL	-1.63	-2.09	-1.36	-1.47	-1.16
12	MISC. TEXTILES	1.05	1.19	1.77	1.56	1.96
13	LEATHER	-1.12	1.25	1.05	1.18	0.67
14	FOOD MANUFACTURERS	1.53	1.25	1.18	1.36	1.44

Table 25: Scores ranking the spread of log items for industrial groups.

1983	1984	1985	1986	1987
CLOTHING	WOOL	CLOTHING	CLOTHING	CLOTHING
WOOL	CLOTHING	WOOL	WOOL	MACHINE TOOLS
LEATHER	MACHINE TOOLS	MACHINE TOOLS	MACHINE TOOLS	WOOL
MACHINE TOOLS	CHEMICALS	ELECTRICITY	ELECTRICITY	ELECTRICITY
INDUSTRIAL PL.	INDUSTRIAL PL.	CHEMICALS	CHEMICALS	CHEMICALS
BUILDING MT.	PAPER & PACK	PAPER & PACK	BUILDING MT.	BUILDING MT.
CHEMICALS	ELECTRICITY	BUILDING MT.	PAPER & PACK	INDUSTRIAL PL.
ELECTRICITY	BUILDING MT.	INDUSTRIAL PL.	MOTOR COMPON.	PAPER & PACK
PAPER & PACK	ELECTRONICS	ELECTRONICS	ELECTRONICS	ELECTRONICS
ELECTRONICS	MOTOR COMPON.	MOTOR COMPON.	INDUSTRIAL PL.	MOTOR COMPON.
MOTOR COMPON.	METALLURGY	METALLURGY	METALLURGY	LEATHER
METALLURGY	TEXTILES M.	LEATHER	LEATHER	METALLURGY
TEXTILES M.	FOOD MANUF.	FOOD MANUF.	FOOD MANUF.	FOOD MANUF.
FOOD MANUF.	LEATHER	TEXTILES M.	TEXTILES M.	TEXTILES M.

Table 26: Industries ranked by spread of log items. Below, the largest spread.

the first dimension and the Y-axis is the second. Each number stands for one industry. The meaning of these numbers can be found in tables 25 or 27 on pages 132 or 135.

There is a visible trend towards a straight line. The effect of size is preeminent. Industries have different spreads but they are under the same effect. Leather is the exception. Its log items show signs of influences other than size for three of the observed years. And such influences are not stable during the period. The second dimension of Leather changes sign twice.

When the first dimension was used for ranking industries according to spread the resulting rank was stable during the period of five years. The most homogeneous industries concerning size are Clothing, Wool, Machine Tools. The least homogeneous are Miscellaneous Textiles, Metallurgy and Food Manufacturers. We show these ranks in table 26 (page 132). The scores obtained as the first dimension of the map are displayed in table 25 (page 132).

Industries like Building Materials, Metallurgy, Machine Tools, Clothing and Food show a consistent spread for the whole period. Chemicals, Electricity, Electronics, Motor Components and Wool are also regular. Paper and Packing, Industrial Plants and especially Leather are irregular. Their ranking is not consistent for the whole period. Figure 43 (above) (page 134) shows the score representing the first dimension of each industry for five years. Each year is a mark. When marks gather very close to one another the spread is consistent for the whole period.

Results: Size-adjusted items. When building a map of industries using joint measures of spread for log residuals, the obtained R^2 are:

Years	One dimension	Two dimensions
1983	91%	98%
1984	93%	99%
1985	92%	100%
1986	84%	98%
1987	82%	99%

For one dimension these R^2 are less stable than the R^2 obtained for size but the numbers are not very different. It suffers a break in 1986 and 1987, as if the spread of firm's internal features were becoming increasingly complex.

Two dimensions seem enough to account for the randomness introduced by the industrial grouping. But notice that our method is not intended to count the number of dimensions present in the data. It is intended to trace the presence of more than one.

The two-dimensional maps are displayed in figure 42 (page 131 on the right) so that they can be compared with the corresponding spread of size. They show clear differences from the spread of size. The trend towards one unique dimension is no longer visible. One industry, Metallurgy, emerges as very particular, with a much larger spread than the others. We recall from section 2.2.2 that Metallurgy was also unique in the proportion of variability size would explain in each item.

In face of these results we must conclude that the R^2 statistic is not very reliable for deciding whether the second dimension is significant. Its value is similar to the one obtained for the spread of items but we know by visual inspection that the aspects of the two maps are clearly distinct. The position of Metallurgy in the two-dimensional maps — a strong outlier — is partially responsible for this anomaly.

The scores obtained are consistent for the five years. Industries occupy positions which don't suffer clear changes. But it is clear that the effects present are not linear. They affect different features in different directions making a second dimension emerge.

Ranking the spread of industries: The internal features of firms, as perceived by size-adjusted items, yield rankings which are not in the least similar to those obtained when assessing the spread of size. Leather, Motor Components and Building Materials are now the most homogeneous groups. Metallurgy is the least. The scores seem to vary with the year. But table 28 is not a good guide to assess the consistency of results. It shows a picture which is worse than the reality. In order to have a fair idea of the evolution of the scores during the considered time period, table 27 (page 135) is more appropriate.

As in the preceding case, Figure 43 (below) (page 134) compares the scores of each industry for five years. Building Materials, Chemicals, Electronics, Clothing and Food exhibit the same score during the whole period. This means a persistent amount of spread associated with internal features.

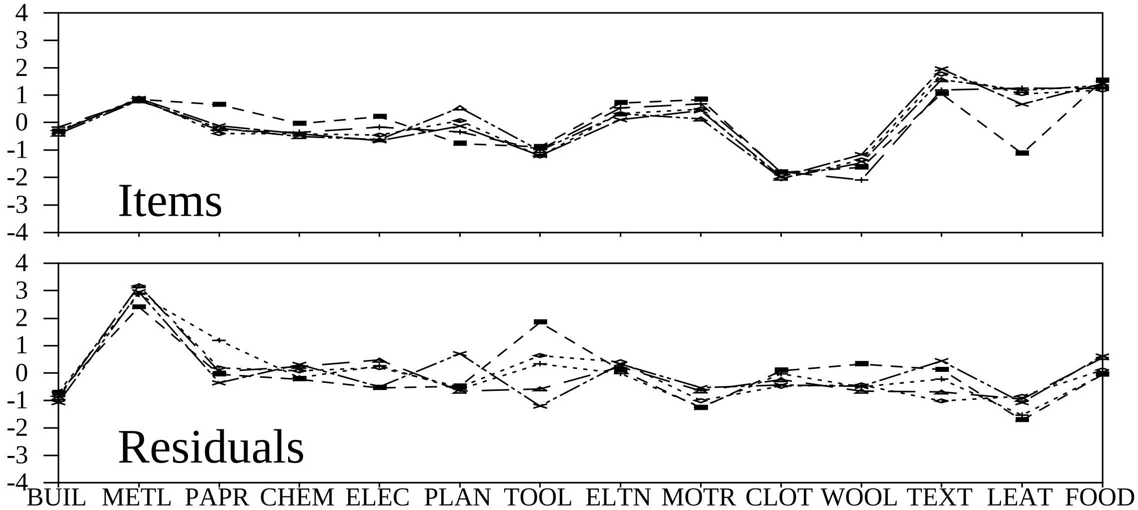

Figure 43: Industries ranked by spread. Comparing five years. The Y-axis shows the scores. A large score means a large spread. Each year is represented by a different mark.

Industries like Electricity, Metallurgy, Motor Components, Wool and Leather also show a reasonable stability. Paper and Pack, Industrial Plants and especially Machine Tools are not stable in their homogeneity. The number of non-stable industries is not larger than in the case of size.

In short: Three industries emerge as showing a consistent behaviour. Both the spread of the size measures and features determine a clear position in the case of Food, Clothing and Building Materials. Industrial Plants, Paper and Packing and Leather are examples of the opposite behaviour. And the remaining industries lie in between. An interesting finding is the time-dependence of the spread of features for some industries and not for others.

The existence of higher order effects questions the use of linear tools for modelling accounting relations. If the input space is capable of apportioning such effects is because they eventually carry information needed for explaining relations.

5.4 Summary

In this chapter we explored the main sources of variability of accounting data. First we produced a set of statements for guidance in the search for a general deflator. The use of a proxy for the common effect can enhance the interpretability of results in statistical models. We have shown that simple case-averages of selected items produce a significant reduction in the spread of the resulting variate.

Next we used a generalisation to more than two dimensions of principles discussed in chapters 3 and 4 for achieving dimension reduction in the input space. The outlined procedure, known as the Hadamard rotation, isolates the common effect and re-distributes the remaining variability by a number of factors according to simple combinatorial laws.

Number	Industrial group	1983	1984	1985	1986	1987
1	BUILDING MATERIALS	-0.72	-0.69	-0.85	-0.91	-1.08
2	METALLURGY	2.40	2.87	3.17	3.19	2.96
3	PAPER AND PACKING	-0.04	1.19	0.18	0.05	-0.36
4	CHEMICALS	-0.22	-0.17	0.08	0.24	0.31
5	ELECTRICITY	-0.54	0.27	0.19	0.48	-0.51
6	INDUSTRIAL PLANTS	-0.49	-0.59	-0.51	-0.65	0.70
7	MACHINE TOOLS	1.85	0.34	0.64	-0.58	-1.21
8	ELECTRONICS	0.10	-0.01	0.40	0.23	0.33
9	MOTOR COMPONENTS	-1.27	-1.24	-1.00	-0.65	-0.53
10	CLOTHING	0.09	-0.02	-0.48	-0.23	-0.44
11	WOOL	0.33	-0.50	-0.43	-0.66	-0.48
12	MISC. TEXTILES	0.12	-0.22	-1.02	-0.69	0.44
13	LEATHER	-1.71	-1.53	-0.84	-0.95	-1.08
14	FOOD MANUFACTURERS	-0.05	-0.06	0.11	0.57	0.62

Table 27: Scores ranking the spread of log residuals for industrial groups.

1983	1984	1985	1986	1987
LEATHER	LEATHER	TEXTILES M.	LEATHER	MACHINE TOOLS
MOTOR COMPON.	MOTOR COMPON.	MOTOR COMPON.	BUILDING MT.	LEATHER
BUILDING MT.	BUILDING MT.	BUILDING MT.	TEXTILES M.	BUILDING MT.
ELECTRICITY	INDUSTRIAL PL.	LEATHER	WOOL	MOTOR COMPON.
INDUSTRIAL PL.	WOOL	INDUSTRIAL PL.	INDUSTRIAL PL.	ELECTRICITY
CHEMICALS	TEXTILES M.	CLOTHING	MOTOR COMPON.	WOOL
FOOD MANUF.	CHEMICALS	WOOL	MACHINE TOOLS	CLOTHING
PAPER & PACK	FOOD MANUF.	CHEMICALS	CLOTHING	PAPER & PACK
CLOTHING	CLOTHING	FOOD MANUF.	PAPER & PACK	CHEMICALS
ELECTRONICS	ELECTRONICS	PAPER & PACK	ELECTRONICS	ELECTRONICS
TEXTILES M.	ELECTRICITY	ELECTRICITY	CHEMICALS	TEXTILES M.
WOOL	MACHINE TOOLS	ELECTRONICS	ELECTRICITY	FOOD MANUF.
MACHINE TOOLS	PAPER & PACK	MACHINE TOOLS	FOOD MANUF.	INDUSTRIAL PL.
METALLURGY	METALLURGY	METALLURGY	METALLURGY	METALLURGY

Table 28: Industries ranked by spread of log residuals. Below, the largest spread.

Finally we studied the importance and effect of the SEIC industrial grouping when used as an input variable for modelling accounting relations. Our results show that industrial groups cannot be ignored. Both the spread of size and the one of internal features of firms are dependent on the group to which each case belongs. But in the last case the effect of grouping is not similar across industries.

The existence of higher order effects in the space of firm's features demands the use of algorithms able to model them. Higher order effects can form "statistical exclusive-OR" structures which are impossible to model with linear tools and very difficult to account for even with the conventional non-linear ones. In the second part of this study we show how Neural Networks are able to model higher order relations ensuring good generalisation.

Part II

Neural Networks and Knowledge Acquisition in Accountancy and Finance

Introduction To This Part

The second part of our study is dedicated to show how Neural Networks implement the framework developed in previous chapters.

Despite ratios being a simple and appealing way of assessing the features of the firm, when the goal is the understanding of the statistical behaviour of accounting data it seems more reasonable to study individual items first. This was our basic programmatic statement which proved itself fruitful.

Items are much more regular and easy to model than ratios. The observed ones were two or three-parametric lognormal. Such a characteristic carries with it an explanation for the existence of outliers and heteroscedasticity so often mentioned in the literature.

Items seem to reflect mainly size and deviations from size. Regarding size there is no reason to establish a separation between the statistical behaviour of positive items and those having also negative cases. Cross-sectional samples having negative cases should be modelled as two groups.

Given the lognormality of items and their sharing of a common source of variability it is possible to extend the notion of financial ratio so as to account for non-proportionality between its components. Non-proportionality is not inconsistent with proportional or multiplicative mechanisms. And it is expected to affect mainly small firms.

The distribution of ratios becomes easy to understand by observing the way their components interact. The logarithmic space allows such a direct observation. For example, some irregular behaviour reported in the literature and so far unexplained emerges as a consequence of the numerator of the ratio being bounded by the denominator. This affects the long multiplicative tail of the ratio distribution so that it is constrained to be much shorter.

Size and grouping are the two main sources of variability in accounting data. It is possible to approach the statistical effect of size by building case-averages of several log items conveniently selected. The industrial grouping of firms introduces higher order effects in the variability of the data. Not only features like liquidity or profitability are different from group to group. They also seem able to react in different directions in the presence of the same perturbation.

As a consequence: The algorithms for modelling accounting relations will need the ability to introduce non-linearity when necessary. This stems from the existence of base-lines — which produce a concavity in the spread of cases in log space — as well as from the higher order relations in grouped data.

The introduction of a non-linear modelling capacity must be achieved without damaging the ability to generalise of the resulting model, that is, using the right assumptions and the number of free parameters required by the problem and not more than those. The second part of this study is thus devoted to explore the possibilities of the new algorithms known as Neural Networks since they seem particularly apt to respond to the requirements of a powerful non-linear yet tightly controlled modelling.

Description of this part: Chapter 6 is an introduction to Neural Networks. It contains an historical note written so as to give the necessary perspective of the development of these tools. It further explains where Neural Networks are seemingly interesting in Finance and Accountancy.

In chapter 7 Neural Networks are described as a maximum likelihood implementation of the extended ratios devised in chapter 3. We show that the problem of choosing appropriate ratios for statistical models can be avoided. Neural Networks seem able to improve our knowledge of a relation by finding the appropriate ratios to model it. When using procedures consistent with the statistical nature of accounting items the optimization process leading to the modelling of a relation can find extended ratios adequate for representing such a relation. An internal node of a Neural Network conveniently trained will build internal representations similar to ratios. Such representations are self-explanatory thus improving our knowledge of the modelled relation far beyond the information provided by the model itself.

Chapter 8 explores the possibilities of Self-Organized Neural Nets in improving ratio diagnosis power and specificity. The first part of this chapter is a discussion of graphical tools which approach ratios but are two-dimensional. In the second part, Self-Organized Maps are implemented as a means of obtaining automatic diagnostics from these two-dimensional ratios. The devised tool can be used as a pre-processor for extracting rules from databases containing accounting data. Such rules can then be used by symbol-based expert systems along with other sources of information.

Finally, chapter 9 attempts a taxonomy of risk based on the capabilities offered by Neural Networks to model complex relations. A clear improvement is obtained in the understanding of the way the market trades a particular class of assets.

The role of Neural Networks in this study: So far, expectations about Neural Networks are related to the modelling of difficult relations (pattern recognition) or the mimicking of brain functions.

A considerable effort has been devoted throughout this study to devise appropriate learning and post-processing techniques for the Multi-Layer Perceptron so that it could be used just as any other statistical modelling tool. An important result we came across is the ability these tools display, under particular circumstances, to form self-explanatory internal representations able to increase our knowledge of the modelled relation. We show that some specific statistical problems requiring self-explanatory power can greatly beneficiate from the existence of internal representations meaningful for accountants.

Therefore, the reason for using Neural Networks here is not just the need of simple and versatile yet powerful tools able to deal with the complexity of the relations. It is also the fact that internal representations turned out to be meaningful and an important way of acquiring knowledge from past experience.

This is fortunate because the traditional tools for knowledge acquisition seem not to fit well in many financial applications. The nature of accounting and financial relations, where the input variables are continuous-valued and stochastic, makes it difficult for the usual expert systems based on symbolic computation to deal with. Observations such as those found in stock returns, or data organized in accounting reports, cannot be efficiently used by actual expert systems as a source of knowledge. We expect to prove that Neural Networks can provide self-explanatory results, along with improvements in performance.

Chapter 6

An Introduction to Neural Networks

This chapter intends to be an introduction to the modern statistical modelling tools known as Neural Networks but only to the extent of their usefulness in accounting and finance research.

There are already many articles and books available on the subject. We limit this review to the kind of Neural Networks interesting for our study. We mainly focus on tasks which cannot be easily performed by tools based on symbolic computation. For example, the statistical modelling of relation having continuous-valued inputs. Therefore, we omit any reference to binary-threshold nets and those intended to mimicking the brain.

This presentation emphasizes the description of two kinds of nets we use later on in this study: The Multi-Layer Perceptron and the Kohonen's Map of Patterns. Incidentally, these are also the most used Neural Networks nowadays. And they represent the two main views or branches of Connectionism.

The contents of this chapter are neither original nor presented in an original way. We recommend, as a complement to it, the excellent book by Pao [92] on adaptive techniques. Some formalism we display when describing Back-Propagation owes its clarity to being inspired by this author. Finally, since it is usual to quote the Lippmann article [80] as an introductory piece of information on Neural Networks we obligingly do so.

6.1 Historical Notes

This section is dedicated to explaining the genesis of Neural Networks and the role they play in knowledge acquisition. Only an understanding of their origins and prospects can lead to the formation of an opinion about the interest of these tools in Accountancy and Finance. A perspective of its future becomes easier when knowing its past.

The history of Neural Networks is not a common one. Known as such from the early forties, they trod an adventurous path with periods of intense enthusiasm and almost total eclipse. The actual development, which began in 1985, is partly the outcome of a much expected discovery.

Learning theory and the learning tools known as Neural Networks are the result of two lines of research which began their paths very early — about the second world war — and remained closely associated until the decade of the sixties.

The first of such lines was technical. It included mathematicians and engineers trying to build what is known as the Optimum Filter. They used concepts extracted from Tele-communications: Linear Systems, Stochastic Optimization and Information Theory. The second line was speculative and its goal was the building of artificial machines similar to the human brain. This science was known, and still is, as Connectionism. In the next paragraphs we follow the development of both.

The Optimum Filter and other automatic learning devices: A filter is a tool able to separate, in a continuous flow of information, the real signal from the randomness attached to it.

The study of filters is typical of the Information and Communication sciences. However, the problem of filtering is very general. Optimum filters are similar to automatic controllers or predictors. And the problem of building them is also similar to the problem of building a general statistical modelling tool, able to separate pure randomness from real features of the data without the intervention of experts. For example, what engineers call a linear filter could be described as a linear regression in a time-series context. Past observations are used to predict future events.

Filters are optimal according to a criterion, in the same sense a regression minimizes the squared error. But they often include other criteria, like stability, as well.

A learning or adaptive filter is the one which adapts its behaviour (its free parameters) to changing inputs, according to one or more criteria. In the linear case it would be a sort of adaptive regression able to change the slope and the intercept when the data changes.

Such a learning system will accomplish behaviour modifications without external intervention in its operation. The number of parameters engaged in the modelling and even the amount of non-linearity introduced, are selected in such systems just by the influence of the input and responding to its requirements. An automatic learning system acts like a dedicated controller whose experience of the underlying structure of the process improves as the process unfolds. Additional information concerning its structure and features causes the controller to adapt himself to the process's behaviour.

Such tools are valuable in many areas but especially in communications and in control. They had their origin as solutions to problems posed by military automata. Nowadays they are also considered as potentially interesting for knowledge acquisition and machine learning. In fact, it is the very nature of the process which eventually emerges and becomes transparent when described by the set of parameters used for modelling it, along with the a-priori assumptions used. The amount of knowledge provided in this way often is more interesting than the model itself.

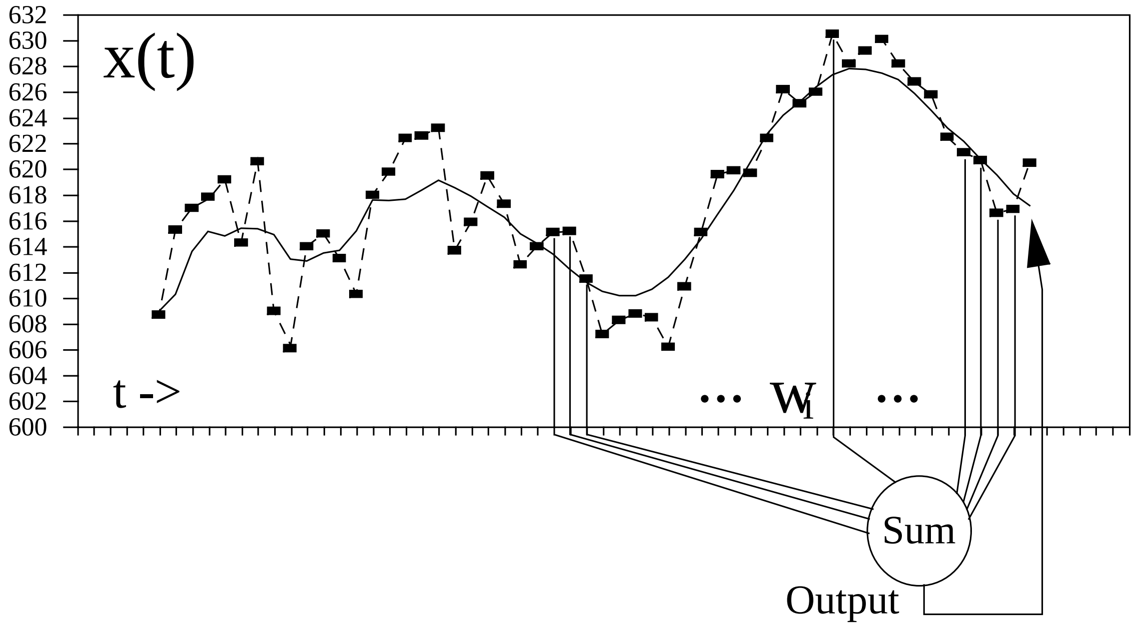

Figure 44: Schematic representation of the simplest Wiener filter. An inner product of adjustable parameters W with the past history of a time process X is used here to approach the actual event x_t. The solid line is the output of the filter.

Early studies on Learning: In October 1941, Bell Laboratories and the Massachusets Institute of Technology (U.S.) engaged Norbert Wiener and other researchers in an intense effort to design automatic devices which could track a plane or a ship, compute the main features of its trajectory and predict where it would be by the time the shell or bomb had travelled to the target area. The conceptual basis of this research became the origin of the early automatic learning systems.

Norbert Wiener had at that date a large experience on building filters. During the 30's he idealized, along with Y. W. Lee, a network of circuits able to perform convolutions of incoming signals. For a signal $x(t)$, t being a discrete time counter, such networks would produce an output $O(t)$

$$O(t) = \sum_{i=1}^{M} x_{t-i} \times w_i \quad , w_i \text{ being adjustable parameters.}$$

M is known as the delay-line size. It controls the amount of memory about the past history of the time process incorporated into the filtering (see figure 44).

With such devices Wiener and Lee were able to perform many interesting tasks like the solving of partial difference equations. They were also able to design a linear filter of any shape (any frequency response) just by modifying the parameters w_i.

For the prediction of plane trajectories Wiener used an improved version of the same basic networks attached to a mechanism of feedback. The position of a target was computed by the net and compared with its real position. From here a measure of error was obtained. Then the parameters

w_i were updated so as to minimize such error. This procedure was carried out interactively. The final result was that the trajectory of the target would be learned by the set of w_i.

Such a simple mechanism, which in practice didn't succeed, nevertheless became the basis of modern filters and the paradigm of automatic learning for more than twenty years. Some of the modern Neural Network learning rules are also based on it.

Further developments: In order to identify or to recognize the pattern for automatic learning, it is necessary to build a mathematical model of the process to be learned. Kolmogoroff [74] and Wiener [136], assumed first that the process was linear. Then, they demonstrated that the filtering and the prediction of stochastic series were special cases of the same learning problem. Their work could be described as the finding of a general recipe for the building of learning systems. The significance of such work is stressed by Y. W. Lee:

> Wiener's theory of optimum linear systems is a milestone in the development of communication theory. The problems of filtering, prediction and other similar operations were given a unity in formulation by the introduction of the idea that they all have in common an input and a desired output. Then, the minimization of a measure of error, which is absent in classical theory, was carried out. The entire theory, from its inception to the final expressions for the system function and the minimum mean-square error is invaluable in the understanding of many problems in a new light [77].

As originally formulated, Wiener's early methods are applicable only to linear time-invariant dynamic processes which are to be optimized by a Least-Squares criterion. Booton [16] extended in 1952 Wiener's work to the optimization of linear time-varying dynamic processes possessing either time-invariant or time-varying statistics. Kasakov [68], Shen [115], and others, treated nonlinear feedback control systems with random inputs using stochastic learning techniques.

Wiener's latest work on this subject, *Non-Linear Problems in Random Theory* (1958) [137] opened up the path for a theoretical approach to self-organizing and adaptive systems. In his framework the complexity of the system's repertoire of available non-linearity increases as the learning process develops so as to maximize the flow of new information about the structure of the process thus creating an internal model of it. In a restricted sense, if some input and output functions represent the behaviour of an unknown process, the Wiener system will organize itself into a model of this unknown mechanism, provided statistical regularities exist in the process. The basic tool for such organization is, of course, the ability to abstract those regularities from the stochastic series on which it is to operate.

Limitations of analytical learning systems: Being analytical, the Wiener solution cannot avoid some lack of generality. Assumptions must be made about the statistical nature of the input. If not, it would be impossible to apportion analytically the parameters between input and output.

Only Gaussian processes and a few more classes of random processes are correctly modelled by this method. And the modelling of the signals is made — in the later versions — using Volterra functionals, that is, Taylor series with some limited amount of memory of past events [131].

When a system becomes optimal only given the restriction that it must belong to a specified class, the kind of information that such a system can identify and use is also restricted. A linear system, for instance, can produce a significant improvement in mean-square error reduction, only if the spectral densities of the signal and the randomness attached to it are different, since it cannot use any other information. Therefore, an optimum linear system, in the Wiener's sense, is no better than an optimum attenuator. Higher order modelling requires non-linear systems because dissimilarities in the characteristics of stochastic series, which the linear system would ignore, can now be used to reduce mean-square errors.

Although no analytic general solution exists for the general learning model some broad cases have been explored. In the U.K. Denis Gabor built in 1960 his *Universal Non-Linear Filter, Predictor and Simulator which Optimizes Itself by a Learning Process* [46]. It was an application of Wiener's later work. In the U.S., Shen and Rosenberg [114] used the same principles.

Many military and tele-communication applications of these early attempts followed. They were analytical-based dedicated automata to be used whenever computational speed was required. These "Wiener-Volterra Systems" are not very flexible nor very good in generalisation for, after all, they use polynomials to approach the data. But being analytical they avoid problems of convergence and are easily built into very fast hardware. A good review of such systems can be found in Schetzen's book [108].

Neural Networks seek the same goal. But they are not based on analytical optimization. Stochastic search techniques are used instead to discover in the parametric space a point obeying a desired condition. They generalise better but their learning process is much slower.

Brain Mimics: It is bizarre that a practical application of Wiener's aspirations was at length provided, not by any mathematical analysis, but by a few neurophysiologists trying to build models of some brain functions like reasoning and recognition.

Artificial neural models emerged forty years ago as a broad mimic of the real neural structure of the brain. The paradigm of Connectionism's early work is the Hebb's Rule. Strongly influenced by Behaviourism and other theories accepted at that time, Donald Hebb wrote in 1948 a book, *The Organization of Behaviour* [56], proposing a plausible mechanism by which learning could take place in the brain. The Hebb's Rule simply states that whenever two neurons are excited at the same time the connection between them strengthens. Many Neural Network learning rules had their origin in this mechanism or in variations of it. However, the strong theoretical basis of connectionism has been provided by mathematicians rather than by Psychologists.

Like Wiener, Dr. Warren McCulloch was a mathematician interested in practical problems. He first met Wiener in 1942 and their collaboration lasted for a few years. McCulloch was mainly interested in the organization of the cortex of the brain. Working with him was Walter Pitts, a student of logic and biophysics. In 1943 Pitts moved to the M.I.T. for reading with Wiener. They worked together on pattern recognition until 1948.

Wiener's ideas on filtering and automatic learning must have inspired the first paper McCulloch and Pitts published. It was *A logical Calculus of Ideas Immanent in Nervous Activity* (1943) [84], formalizing the intrinsic structure of the neural process leaving aside its biochemistry. The calculus was very revealing. Using only concepts like the firing of neurons, excitatory and inhibitory connections, synaptic delays, all-or-none processes, it was possible to show that the specifically biologic aspects of the nervous system are irrelevant to the understanding of perception. But the most important aspect of this work is the parallelism it establishes between the Turing Machine and the brain.

> The concepts of Turing's Machine is significant, not only from the purely analytical standpoint of mathematical logic, but even from the standpoint of the neurophysiological understanding of the human mind ([85] page 35).

A consequence is that the mind can compute all and only those numbers a Turing Machine does.

After this, McCulloch and Pitts turned their attention to the problem of the recognition of patterns. Wiener describes it as an interrogation:

> What is the mechanism by which we recognize a square as a square, irrespective of its position, its size, its orientation? ([138] page 18).

In 1947 they offered a theoretical description of the neurophysiological mechanisms for pattern recognition in the brain, in their paper *How We Know Universals* [94]. They suggested mechanisms similar to those of modern neural nets for explaining the ability of the brain to recognize. This paper strongly influenced Wiener's thoughts on Learning and Control, being an important point of view for the work he was about to undertake.

The second generation of connectionists: In the 50's the dominant research in Neural Networks was led by Frank Rosenblatt at Cornell, U.S. Based on the theoretical developments of McCulloch, he built a class of networks called Perceptrons supposed to be able to learn to recognize physical objects by looking at them.

In 1959 Bernard Widrow [135] developed at Stanford, U.S., an adaptive linear filter called Adeline based on neuron-like elements. The Adeline and its more sophisticated successors were used for a variety of applications including the recognition of speech and characters, weather forecasting and adaptive filtering. The Adeline was also the first Neural Network to be used in a practical real-world application, the automatic elimination of echoes in phone lines. With Widrow the two branches described above made a mutual recognition: Engineers became interested in Neural Networks.

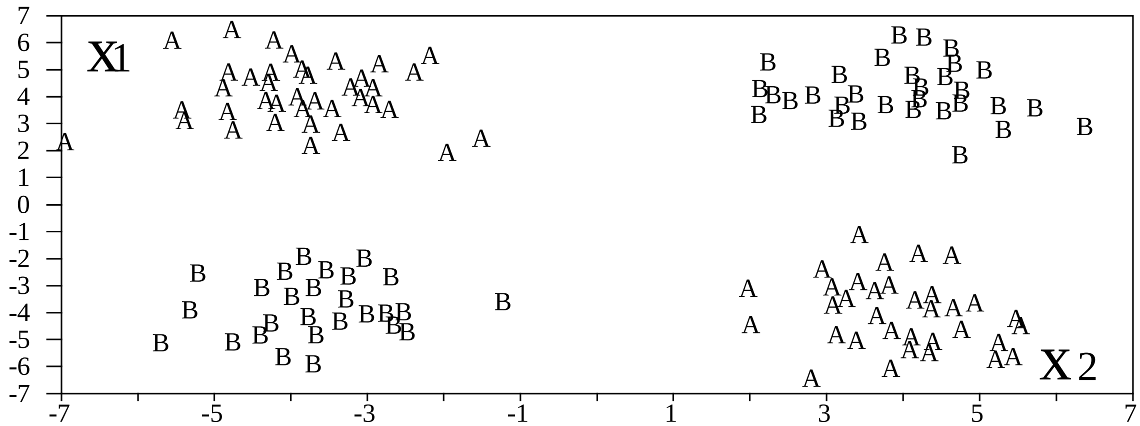

Figure 45: The statistical version of the logical exclusive–OR problem used by Minsky and Papert to discard the early Perceptron. No linear frontier in the space of x_1 and x_2 can separate the two groups A and B in spite of their clear separability.

In the sixties the results of using Neural Networks were promising enough to attract general attention. Particularly, Rosenblatt's "Perceptron" (1961, see [103]), generated real enthusiasm in the scientific community for a few years.

The emergence of Symbolic Computation: Such an enthusiasm lasted for a short period. In 1969 Minsky and Papert published *Perceptrons* [89] showing that Rosenblatt's two-layer Perceptron, being linear, would not recognize patterns involving interactive or higher order effects like those which occur in parity-detection (the logical "exclusive-OR" see figure 45). In order to correctly classify patterns using non-linear boundaries, Minsky and Papert showed that Perceptrons would need more than two layers of neurons and the introduction of non-linear transfer functions. At that time no one knew how to achieve a general learning rule able to adapt the connections between internal neurons.

After that, interest in neuro-models languished. The attention of the research concerned with Learning was directed towards the emerging tools provided by Artificial Intelligence. For twenty years Connectionism was confined to a few laboratories, mainly concerned with the brain itself: James Anderson, at Brown University, U.S., revived the Hebbian principle in his *Linear Associator*. Teivo Kohonen, in Helsinky, also envisaged a modified Hebbian principle known as *Competitive Learning*, for creating self-organizing maps of patterns.

During the seventies and early eighties the dominance of symbolic methods was overwhelming. Machine Learning and Knowledge Acquisition became synonymous with Symbolic Computation. The emergence of the Computer Sciences as an independent branch, the fast progress in the speed of conventional machines, the existence of generous funding, all this turned the attention of the scientific community concerned with these subjects towards tools based on Discrete Mathematics. Research programs based on analogue tools were discontinued and most of the earlier contributions for effective machine learning were forgotten. Machine Learning went exclusively symbolic.

The result was a delay in the course of the development of these subjects. And at length, a muddling of concepts and techniques. For example, some researchers tried seriously to use discrete, hierarchical, learning methods like Quinlan's ID3 as an alternative to simple linear regressions in straightforward problems involving the prediction of continuous-valued variables [88], [98]. These authors clearly put the finding of hierarchical rules ahead of any other considerations.

Today, the research tends to consider rule trees as a very attractive way of expressing acquired knowledge. But only when the model they express is correct. In [12] we further explore this subject.

Another result of these years was the narrowing of views and goals inside the small connectionist community. Connectionists became strictly concerned with the mimic of brain functions as a goal, frequently denying the idea that such mimic could be used also as a source of inspiration for the building of useful learning algorithms. A typical example is Stephen Grossberg. He devised an *Adaptive Resonance Theory* leading to self-organized memories with local characteristics. It is a plausible mechanism for the brain with small practical applicability.

Even nowadays, the followers of Connectionism will discard any model for learning if it is not plausible enough as a replica of the brain. One of the most demolishing things anyone can say about a new Neural Network is that it is not enough brain-like. As a consequence, many recent algorithms are local, self-organized and altogether with little interest for this study.

The return of Neural Networks: The M.I.T. was one of the few places were the interest in analogue learning never vanished. In 1978 John Hopfield initiated a collaboration with its Centre for Biologic Information Processing. As a result, he presented in 1982 a paper to the Academy of Sciences of the U.S. about a new neural model. It was the first paper on connectionism accepted in this body since the 60's.

Hopfield's model [60] introduced a conceptual basis for neural learning in terms of energy. It also established a parallelism with Ising models (Spin Glass Physics). Hopfield's net uses fully interconnected neurons that seek a minimum of energy. A few years later, Geoffrey Hinton in Toronto and Terrence Sejnowski in the John Hopkins University, U.S., developed a modified version of the Hopfield net they named *The Boltzmann Machine* [59], able to escape from local minima during learning and with the remarkable quality of being trainable even when having hidden layers. Hopfield's net and the Boltzmann Machine considerably revived the interest of the scientific community on Neural models.

The breakthrough came in 1985 when David Rumelhart, a professor of psychology at Stanford, U.S., and James McClelland, psychologist at Carnegie-Mellon, along with other members of *The Parallel Distributed Processing Group* (known as PDP) devised a learning scheme that would allow multi-layered Perceptrons to feed back deviations from correct response to more than one layer of neurons. Their scheme became known as *The Back-Propagation Algorithm* [106]. It allows the training of all nodes inside a Multi-Layer Perceptron (MLP), even the internal ones.

Much interesting research followed. For example, the link between Back-Propagation and the continuous-valued version of the Hopfield paradigm was established shortly afterwards by Luis B. Almeida at INESC, Lisbon [2]. Almeida generalised Back-Propagation so as to make possible the learning in nets with any topology. The original MLP were limited by a feed-forward topology in which the information flows only in one direction — from the input to the output layer.

This *Recurrent* back-Propagation is especially adequate for tasks requiring some amount of memory of past events like systems identification, the reconstruction of missing cases and the simulation of dynamic systems.

As predicted by the early research, the MLP turned out to be a very powerful and versatile modelling tool, able to solve Minsky and Papert's exclusive-OR problem and many other complicated ones. The MLP is, in practice, a general learning tool as Wiener foresaw it. This fact, interesting as it is, induced a sudden and somehow non-proportional enthusiasm and renewed the interest in models based on Connectionism.

Indeed, one of the reasons for such a renewal of interest in Neural Networks was the realization that Symbolic Computation, when tackling even the simplest problems involving pattern recognition, was severely limited. The task of recognizing "a square as a square irrespective of its position" turned out to be very hard for tools based on logic, rules, hierarchical structures or other concepts typical of Artificial Intelligence. Typical problems of this kind, like the recognition of voice and handwritten information, received a second chance of improving by using Neural Network techniques instead.

Neural Networks today: Nowadays, Neural Networks are being used for finding solutions for difficult problems of pattern recognition and in military applications. Sejnowski's *NETtalk System* [110] is often mentioned as a reference point for assessing what Neural Networks can do in these fields. This program converts text to speech and, connected to a speech synthesiser, it pronounces typewritten words. It learns from examples of text together with its spoken form. Over ten hours of such training it progresses from a formless babble to intelligible English.

It is clear that, again, there is no correspondence between the real possibilities offered by Neural Networks and some extravagant expectations about them. The same strong motivations which led to non-realistic views about Artificial Intelligence are now working in the direction of Neural Network research:

> The reason why the US AI community (academic as well as commercial) has taken up the neural-net model so enthusiastically is quite straightforward. It is primarily because the Department of Defense has decided that neural-net computing is a high-priority strategic technology. As an example, the UCLA (University of California, Los Angeles) AI lab has recently started ten new projects concerned with neural networks while seven symbolic AI projects are due to be terminated shortly. This switch did not come from inside the university. It happened as a result of strong prompting from DARPA and

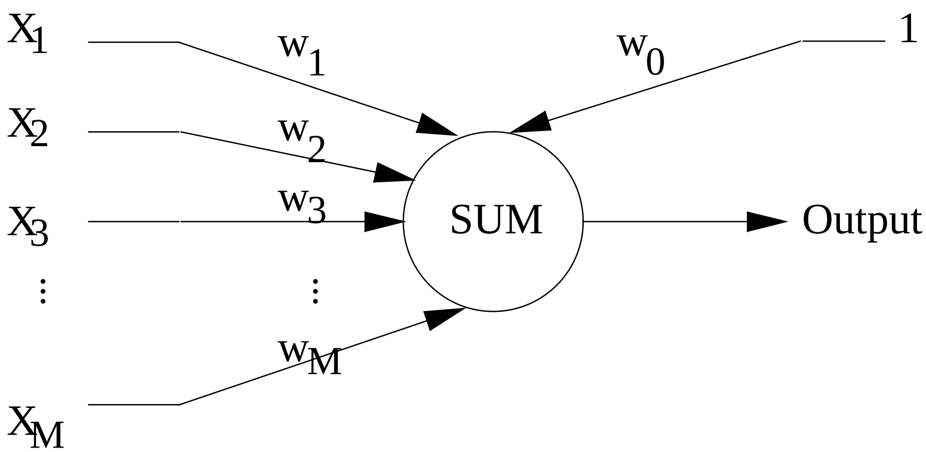

Figure 46: A neuron or node, the basic element of Neural Networks. It implements an inner product,$X \cdot W$, of an input vector, X, with another one of adjustable parameters W.

other funding bodies ([43], page 12).

Of course, the state of things in the U.S. doesn't have to extrapolate to other countries. Anyway, when assessing the possibilities offered by Neural Networks it seems important to have a clear idea about their applicability, strong and weak points and clear shortcomings.

The most quoted of these shortcomings is the non-ability of Neural Nets to produce interpretable and exportable models like those based on rules and structures. We think that knowledge doesn't have to be interpretable in all cases. The most complicated pieces of knowledge couldn't possibly be translated into simple structures.

In this study we show two examples of the use of Neural Networks in knowledge acquisition. Both are related to accounting and financial research. In the first one, the resulting model is simple and its structure is interpretable. Such an interpretability makes it attractive. But it is interpretable because it is simple. In the second one, the obtained model is complex and it cannot be decoded. But this is not a shortcoming since the model is important by itself, not because of its interpretability.

6.2 The Structure of Neural Networks

"Neural Network" is the the name of several modelling heuristics, having in common a topology inspired by the way neurons are organized in the brain and the use of non-analytical algorithms.

If a sample containing input and related outcome variables represent an unknown relation, a Neural Network will model this relation by successive approximations using interactive algorithms. Such process is known as learning or training of the net.

Topologically, Neural Networks are lattice structures of simple computational elements called neurons or nodes connected in a specific way. The connections between neurons (known as weights) can be strengthen or weakened during the training process, by means of iterative heuristics, causing

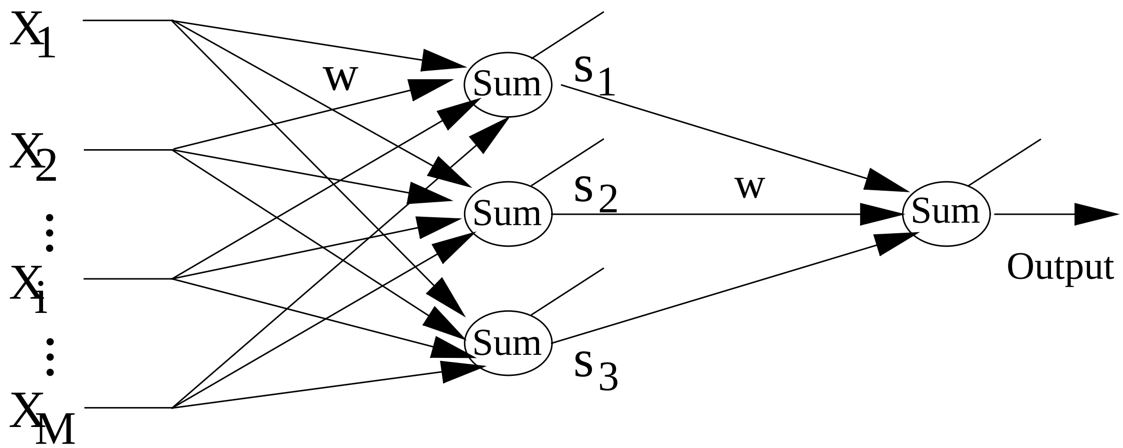

Figure 47: A simple Neural Network containing an input layer, a hidden layer of neurons and a unique output neuron.

the net, as a whole, to become a model of the data.

A typical neuron sums M weighted inputs. Then, the result will be the input for several other neurons or nodes. That is, input variables labeled $x_1, \cdots, x_M$ are applied through a set of associated weights, $w_1, \cdots, w_M$, to the neuron. The weighted sum of the inputs, $s = \sum_{i=1}^{M} w_i \times x_i$, is the output of the neuron. This output will then be used as input for several other nodes, and so on.

Figure 46 on page 149 shows the simplest artificial neuron or node.

Several nodes form a layer. There, all M inputs are fully connected to the N nodes, that is, a weight w_{ij} exists, linking each input $x_i, i = 1, M$ with each node $j, j = 1, N$. In this layer outputs of nodes are $s_j = \sum_{i=1}^{M} w_{ij} \times x_i$.

Layers can be combined into a network (figure 47). Nets formed by cascading two or more layers of nodes so that the outputs from one layer become inputs to the next one, are similar to those built by Widrow in the sixties [135]. They are limited in their applications because no non-linearity is introduced during training. It can be proved that these nets are equivalent to a simple linear Least-Squares algorithm.

Modern nets use non-linear functions associated with each node. These non-linearities are known as activation or transfer functions.

Neural Network classification: Neural Networks can be completely specified by three characteristics:

- The topology of the net: The number and position of the nodes and the way connections between nodes are allowed.
- The node's transfer function: The non-linear operation each node performs on its output before delivering it to other nodes.

- Learning rules: The particular algorithms used for training the net through a steady adaptation of its weights.

According to their topology, Neural Networks often belong to one of these families:

- Only feed-forward connections: The nodes are organized in layers and each neuron in one layer will connect only with next layer's nodes. The Perceptron belongs to this group.

- Intensive wiring: Each node's output is connected with all the other nodes's inputs. An example is the Hopfield net.

According to the transfer functions, there are two main kinds of Neural Networks: Continuous non-linearity, for analog processing; and hard-limiter threshold, for digital processing. The Multi-Layer Perceptron uses continuous non-linearity. The early Hopfield nets used hard-limiter thresholds.

Finally, according to their training rules, Neural Networks can be

- Supervised — There is a learning process in its strict sense, as in the Perceptron and Hopfield nets: Each learning case consists of a vector of inputs and a corresponding vector of desired outcomes.

- Unsupervised or self-organized, as in the Kohonen map of patterns [72] [73]: The learning process does not require explicit supervision: The net organizes itself according only to the inputs.

6.3 The Multi-Layer Perceptron

The Multi-Layer Perceptron, widely known as the MLP, is a supervised learning Neural Network. Topologically it is a layered feed-forward configuration: Nodes are arranged in layers and each node's output is connected to next layer's inputs. No intra-layer connections exist, nor any feedback paths from an output to earlier layers. Figure 48 on page 152 represents a three-layer Perceptron with only one output node.

The back-propagation of errors by means of an iterative gradient-descend algorithm allows training by minimization of the mean-squares differences encountered between the actual and desired output.

In the Multi-Layer Perceptron the output of any, j^{th}, node is not just a simple weighted sum of inputs. After the summation $s_j = \sum w_{ij} \times x_i$ the result is submitted to some continuous differentiable non-linearity $f(s_j)$ becoming $o_j = f(\sum w_{ij} \times x_i)$.

It is common to use sigmoid-like functions as $f(s)$. In this case the output of the j^{th} node would be:

$$o_j = \frac{1}{1 + \exp\left[\frac{-(s_j + \theta_j)}{\theta_0}\right]}$$

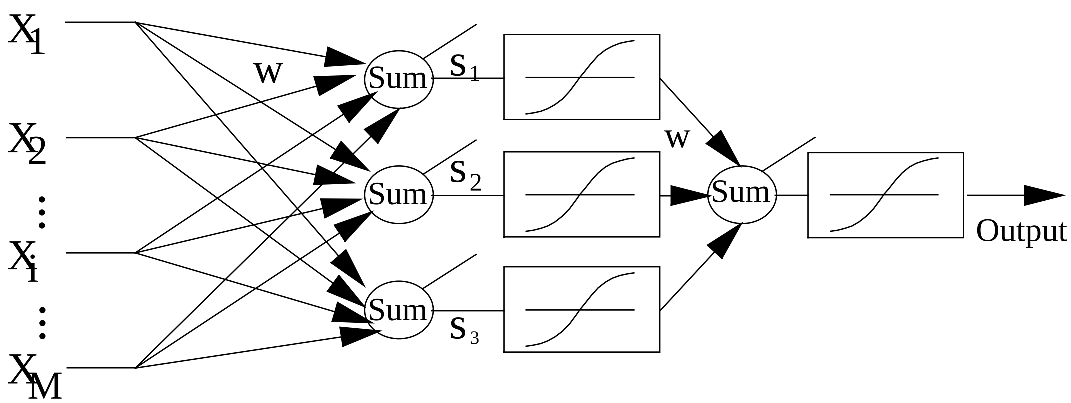

Figure 48: The Multi-Layer feed-forward Neural Network known as the MLP. In this case there is a layer of inputs, one hidden layer, and one unique output node.

Figure 49 on page 156 displays one of such functions.

The parameter θ_j acts as a threshold or bias. The effect of a positive θ_j is to shift the sigmoid to the left along the horizontal axis. θ_0 modifies the steepness of the slope in the sigmoid. Large θ_0 produce smooth thresholds. Conversely, small θ_0 produce sharp thresholds which can be almost like the hard-limiter ones. These non-linearities are known as transfer or activation functions. In the same line, $1/\theta_0$ is known as the gain.

Mean-Squares error: We now consider the last layer of nodes as the k^{th}. The layer immediately before this one is the j^{th} and the one before that is the i^{th}.

When learning, the net is shown a vector of inputs, X_p, from the p^{th} example. Then, the algorithm adjusts the set of weights and the thresholds in each node in a way that makes the effective output of the net, o_{pk}, as close to the desired outcome, t_p, as possible. Once this small adjustment has been accomplished we show another input vector and the corresponding outcome and ask that the net learns this new association as well. At length, we are making the net find a single set of weights and biases that will satisfy all the input-outcome pairs present in our learning set.

In general, the final resulting outputs o_{pk} will not succeed in approaching t_p exactly. For the p^{th} example, the squared error is

$$E_p = \frac{1}{2} \sum_k (t_p - o_{pk})^2 \tag{15}$$

and the average total error will be

$$E = \frac{1}{2 \times N} \sum_p \sum_k (t_p - o_{pk})^2 \tag{16}$$

in which the 1/2 scale is introduced for facilitating the algebra at a later stage and N is the number of examples in the learning set.

6.3.1 The Delta Rule: The Last Layer's Nodes

The Delta Rule is a stochastic version of the steepest-descent iterative optimization algorithm. It has been used in the early Perceptrons and it applies strictly to the learning process taking place in the last layer of the MLP. Since it cannot cope with a topology involving hidden nodes, the learning of these is accomplished using a generalised version of this technique. First we shall introduce the Delta rule by applying it to nodes in the k^{th} layer, the last one.

Starting with an arbitrary set of W values, every example in the learning set will be considered in a random order and its output calculated. Then, the difference between this output and the corresponding outcome will be used to correct every parameter in W by a small amount. The procedure will be repeated for all examples in the learning set again and again, until a minimum square difference exists between outputs and outcomes. In general, different results emerge depending on whether the gradient search is made on the basis of E_p or E. A true gradient search should be based on minimizing the expression 16. In practice, this is seldom the procedure adapted.

Convergence is achieved by improving the values of W. This is done by taking incremental changes Δw_{kj} proportional to $-\partial E/\partial w_{kj}$. That is, given a small η,

$$\Delta w_{kj} = -\eta \times \frac{\partial E}{\partial w_{kj}}$$

Since the error E can be expressed in terms of the outputs o_k and these outputs are a non-linear function $f(s_k)$ we can use the chain rule to evaluate the above partial derivative:

$$\frac{\partial E}{\partial w_{kj}} = \frac{\partial E}{\partial s_k} \times \frac{\partial s_k}{\partial w_{kj}}$$

Notice that

$$\frac{\partial s_k}{\partial w_{kj}} = \frac{\partial}{\partial w_{kj}} \sum_j w_{kj} \times o_j$$

We now define

$$\delta_k = -\frac{\partial E}{\partial s_k}$$

as the rate of change of the error with respect to s_k. So, we can write

$$\Delta w_{kj} = \eta \times \delta_k \times o_j$$

This expression contains the rule for updating the weights linked with the last layer of nodes. It is known as the Delta Rule. η is an arbitrary increment. It is a small value selected so as to ensure a smooth, yet fast, learning.

To compute $\delta_k = -\partial E/\partial s_k$ we use again the chain rule and obtain two terms:

$$\delta_k = -\frac{\partial E}{\partial s_k} = -\frac{\partial E}{\partial o_k} \times \frac{\partial o_k}{\partial s_k}$$

The first term expresses the rate of change of the error with respect to the output o_k and the second one expresses the rate of change of the output of any node in the last layer with respect to its input.

These two terms are easily obtained. From 15,

$$\frac{\partial E}{\partial o_k} = -(t_k - o_k)$$

The second partial derivative depends solely on the transfer function we are using:

$$\frac{\partial o_k}{\partial s_k} = f'(s_k)$$

Hence, whenever we accept a Least-Squares success criterion, the deviations $t_k - o_k$ from the desired outcome can be viewed as the rate of change of the error with respect to the node's output. Notice that when the criterion is a different one this relation has to be re-written.

For any node in the last layer we can write

$$\delta_k = (t_k - o_k) \times f'(s_k) \tag{17}$$

hence

$$\Delta w_{kj} = \eta \times (t_k - o_k) \times f'(s_k) \times o_j \tag{18}$$

The original Perceptron of Rosenblatt, having only one layer of nodes, would learn to optimize an internal model of the data with steady improvements, Δw_{kj}, of the weights w_{kj} by means of the described algorithm. Adaptive filters and some versions of the Adeline also use this learning scheme.

6.3.2 The Generalised Delta Rule

The general solution of the non-linear modelling problem came with the use of internal layers of nodes acting as intermediate maps able to apportion as much piece-wise non-linearity as needed for modelling the relation.

In the modern version of the Perceptron, the MLP, several layers of nodes are matched in cascade so that the output of a previous one is the input for the next. Theoretically, such a device is able to form any continuous smooth map if enough number of internal nodes are provided.

Of course, in this new situation the problem is to discover a suitable way for optimal parameter finding. The Delta Rule described above cannot be used in the learning of relations by more than one layer of nodes.

As mentioned, the solution for this problem has been recently provided by neuro-biologists [106]. The method is a generalisation of the Delta Rule already introduced, known as Back-Propagation.

When weights are not directly linked with output nodes we still write

$$\Delta w_{kj} = -\eta \times \frac{\partial E}{\partial w_{ji}}$$

and proceeding with the same formalism, based on the chain rule, we would have:

$$\begin{aligned}
\Delta w_{kj} &= -\eta \times \frac{\partial E}{\partial s_j} \times \frac{\partial s_j}{\partial w_{ji}} \\
&= -\eta \times \frac{\partial E}{\partial s_j} \times o_i \\
&= \eta \times \left(-\frac{\partial E}{\partial o_j} \times \frac{\partial o_j}{\partial s_j} \right) \\
&= \eta \times \left(-\frac{\partial E}{\partial o_j} \right) \times f'(s_j) \times o_j \\
&= \eta \times \delta_j \times o_i
\end{aligned}$$

an expression formally similar to equation 18. However, in the case of hidden units we cannot evaluate $\partial E / \partial o_j$ directly. But we can write it in terms of known values obtained from the last layer:

$$\begin{aligned}
-\frac{\partial E}{\partial o_j} &= -\sum_k \frac{\partial E}{\partial s_k} \times \frac{\partial s_k}{\partial o_j} \\
&= \sum_k \left(-\frac{\partial E}{\partial s_k} \right) \times \frac{\partial}{\partial o_j} \sum_m w_{km} \times o_m \\
&= \sum_k \left(-\frac{\partial E}{\partial s_k} \right) \times w_{kj} \\
&= \sum_k \delta_k \times w_{kj}
\end{aligned}$$

Therefore we can write in the case of hidden nodes:

$$\delta_j = f'(s_j) \times \sum_k \delta_k \times w_{kj}$$

That is, the rates of change of the error with each node's s can be computed from the previous node's δ. Previous in the sense that they are closer to the output. The basic mechanism of Back-Propagation consists in making it possible to evaluate all the δ throughout the net just by beginning to evaluate them for the last layer and then proceeding backwards.

Therefore, the Back-Propagation algorithm evaluates firstly the δ_k using expression 17 and "propagate" these errors backwards along the net.

Notice that, when the adapted $f(s)$ are sigmoid, hyperbolic tangents or similar logistic functions, then $f'(s)$ assumes a very simple formalism. For example, in the case of the sigmoid,

$$\text{since} \quad o_j = \frac{1}{1 + \exp\left[-\left(\sum_i w_{ji} \times o_i + \theta_j\right)\right]} \quad \text{then} \quad \frac{\partial o_j}{\partial s_j} = o_j \times (1 - o_j)$$

and the δ are given by these or similarly simple expressions:

$$\delta_k = (t_k - o_k) \times o_k \times (1 - o_k) \quad \text{and} \quad \delta_j = o_j \times (1 - o_j) \times \sum_k \delta_k \times w_{kj} \tag{19}$$

for nodes in the output layer and for nodes in the hidden layers respectively.

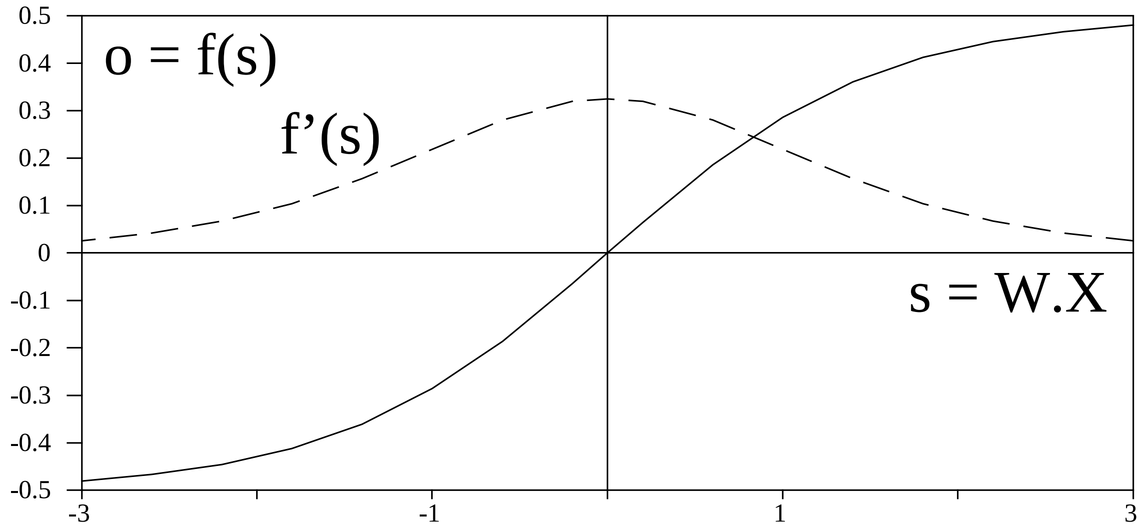

Figure 49: A sigmoid-like function can be used as activation or transfer function in the nodes of an MLP. $f(s)$ is the solid line and its derivative, $f'(s)$, is the dashed line. In this case, f is a tailored so as to span the interval from -0.5 to +0.5.

The Back-Propagation step-by-step: For implementing Back-Propagation a step-by-step procedure is required. Firstly, an example is selected at random. Using it as input, the output of the MLP is calculated. Next, the Δw_{kj} are also calculated and the weights linking with the last layer are corrected. Using these corrected weights and the δ_k it is now possible to calculate the Δw_{ji}, that is, the correction to affect the weights linking to the layer before the last one. This scheme proceeds backwards until all the weights in the net received their first correction.

The whole procedure described above is now repeated for every example in the learning set and for as many "presentations" of the whole set as necessary to bring the overall error to a minimum. In successful learning, the net's error decreases with the number of presentations and it will finally converge to a stable set of weights.

Notice that the described procedure is not a real gradient-descent algorithm since the weights are corrected by evaluating the errors E_p associated with each example (formula 15), not the overall error E of the sample (formula 16). Such a procedure could be described as a stochastic gradient descent. And, in general, it is more effective than the classic algorithm. For small values of η the difference between the stochastic and the classic versions of the gradient-descent procedure vanish.

The increments η must be selected carefully. Too small η make the learning very slow and vulnerable to local minima. Too large η produce oscillations of the overall error.

The most appropriate thresholds are learned by the algorithm just as any other parameter. The θ_j can be viewed as the weights linking a constant input of 1 with the node. The use of hyperbolic tangents instead of the logistic form has no particular relevance except in the last layer. Sigmoids in the output layer constrain outputs to span the interval $\{0, 1\}$ while hyperbolic tangents constrain

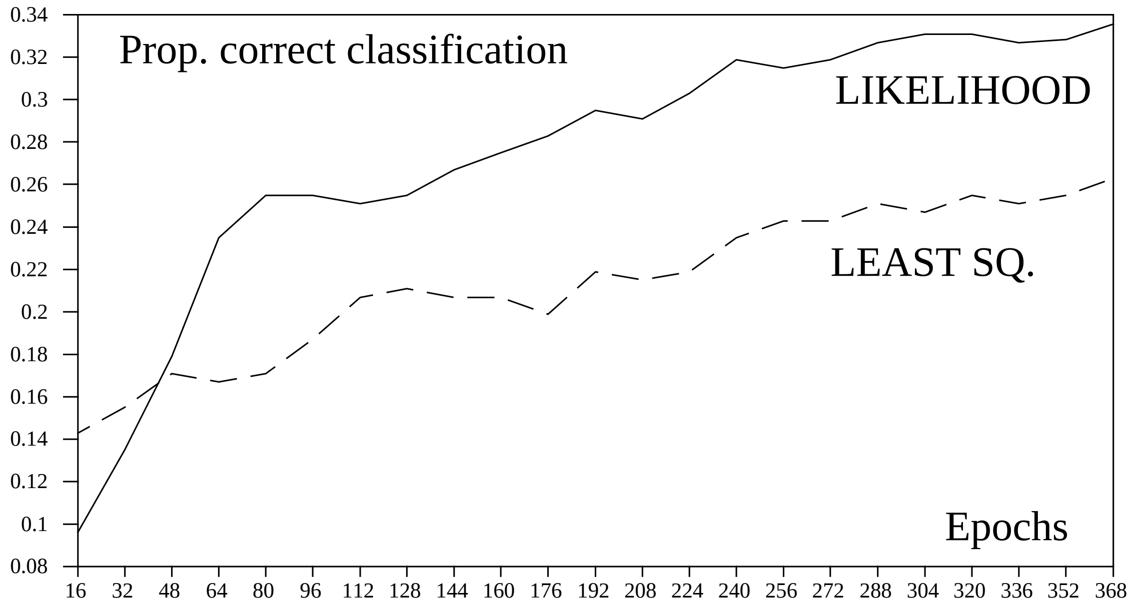

Figure 50: In the MLP the progress on learning is gradual. A given success criterion will improve steadily after each presentation of the entire training set (epoch).

outputs to the interval $\{-1, +1\}$. But, of course, a simple manipulation can produce any desired interval. Output intervals must be selected in accordance with the outcomes.

The problem of local minima: Finally, the Generalised Delta Rule requires that the initial values of the weights are set to values different from one another. It is usual to set them to small random values. If the weights were all similar the net would be in a state known as "local minimum" and the learning wouldn't take place. Several runs of the training set will generally produce a minimization of E although nothing will prevent this heuristic from reaching other local minima instead of the overall solution (see Rumelhart's original paper for a discussion on this issue). In practice, the use of advanced techniques — like having one individual η for each weight, a topic we outline in chapter 7 — can greatly lighten the problem of local minima.

6.3.3 Generalisation in the Multi-Layer Perceptron

The overall rules governing the generalisation ability in any statistical model also apply to the MLP. For instance, if the number of nodes (and therefore, connections) is large when compared with the variability to be modelled, the MLP behaves just like a checking list (a storage device) or a saturated model. No generalisation can be expected. An opposite situation, very few nodes when compared with the variability to be modelled, would make the MLP recognize only broad families of features, without detail.

Using an internal layer with a variable number of nodes it is possible to control the basic information flow used in classification: Many nodes will produce great detail or even no generalisation at all; less nodes will improve generalisation.

Nodes tend to individuality: However, the MLP often exhibits a remarkable behaviour which greatly improves the generalisation beyond that expected for a given number of free parameters.

Any change in a weight's value is proportional to $\partial o / \partial s$. Hence, the changes are maximal for values of s impinging upon the central zone of the transfer function, the one having the largest slope. Figure 49 on page 156 illustrates this fact. It shows the shape of a sigmoid-like function and its derivative. Since changes in a weight's value are proportional to the magnitude of this derivative it turns out that mid-range values introduce large changes in the weights while extreme values make the net change little.

When the s are mostly in the mid-range of their transfer functions, the node under question is not yet trained or committed. It can turn up or down. Under these conditions the weights change rapidly. On the contrary, a committed node changes their weights little since the derivative is small.

The described feature is interesting since it shows that, inside a given topology and by the influence of a learning set, nodes tend to acquire a stable state and remain there. Hence, it is expected that each node on a fortunate model will capture important features of the relation. The literature refers several of such cases. In terms of knowledge acquisition this quality is valuable.

In chapter 7 we explore the ability an MLP seems to exhibit to relate features with particular nodes. Each node seems able to learn or capture a particular characteristic of the firm in a way somehow similar to the procedures typical in ratio analysis.

Resonance: When the minimum number of nodes in any hidden layer matches the number of features important for the relation to be learned, the likelihood of each one of those nodes becoming a model of a different feature of such relation is larger. When that happens, the MLP performs an effective features extraction. As a consequence, the generalisation capacity receives a further improvement.

Some applications take advantage of the trend towards individuality the nodes in the MLP exhibit, to find the important features of a set of examples. Using a topology known as the Bottle-Neck MLP, the same set of examples are presented both as input and as outcome. This technique is similar to Factor Analysis. But when more than one hidden layer is used, also non-linear features are extracted.

6.3.4 Discrete Versus Continuous-Valued Outcomes

The MLP performs both non-linear multi-variate regression and non-linear discriminant analysis. In the first case it is called upon for approaching a continuous-valued outcome. In the second one, outcomes are discrete states. The correct classification of classes in the statistical exclusive-OR problem is an example of the second kind of task.

The MLP as a classifier: In classification, the number of nodes and layers to be used is generally critical. Following Lippmann [80] any one-hidden-layer MLP should be able to form arbitrarily complex frontiers in input space so as to obtain the best possible separation between groups.

Each first-hidden-layer node creates a hyper-plane in the input space since its input is a linear combination of input variables. In the next layer of nodes, several of these hyper-planes can be used to define a region enclosing a particular group of examples relating to one of the given outcomes. That is, in an MLP undertaking a classification, the first hidden layer needs to have as many nodes as pieces to build the frontiers for separating groups. And the last layer needs to have as many nodes as different groups (see figure 52 on page 164).

Of course, when inputs and outcomes are only statistically related, that is, when similar input vectors in the learning set relate to different states, the final error of the net, after convergence, cannot be zero. The above reasoning holds, but now we should envisage probability gradients instead of well-defined, deterministic, frontiers.

For such applications the described minimum Least-Squares criterion should be replaced by a more appropriate one. For example, likelihood maximization is often used in classification instead of minimum Least-Squares. In this case, the model selected would be the one which maximizes the probability of having obtained those samples which were actually used as the learning set. And when the number of nodes in the last layer matches the number of groups to be classified, it can be shown that an appropriate coding makes the MLP output directly the probability of obtaining such an outcome given such inputs. Solla *et al.* [120] contains the appropriate formalism.

The MLP performing non-linear regression: When outcomes are continuous-valued, no transfer function should be used in the last layer of nodes since it would limit the range outputs can attain. It is also important to bear in mind that the amount of non-linearity introduced is controlled by the total number of nodes having transfer functions regardless of its position. An exaggerated number of nodes will produce a too detailed — and hence very sample-dependent — model. Figure 51 shows the effect, in a very simple case (one unique input and output) of introducing nodes.

In the case of continuous-valued outcomes the appropriate success criterion is the minimization of the Least-Squares error. To control the learning process it is common practice to use the overall R^2, that is, the proportion of the variability of the targets explained by the model.

6.3.5 The Delay-Line MLP

An important consequence of the Wiener-Volterra analysis is that, under very general circumstances, it is possible to model the internal behaviour of any system just by using two feed-forward steps.

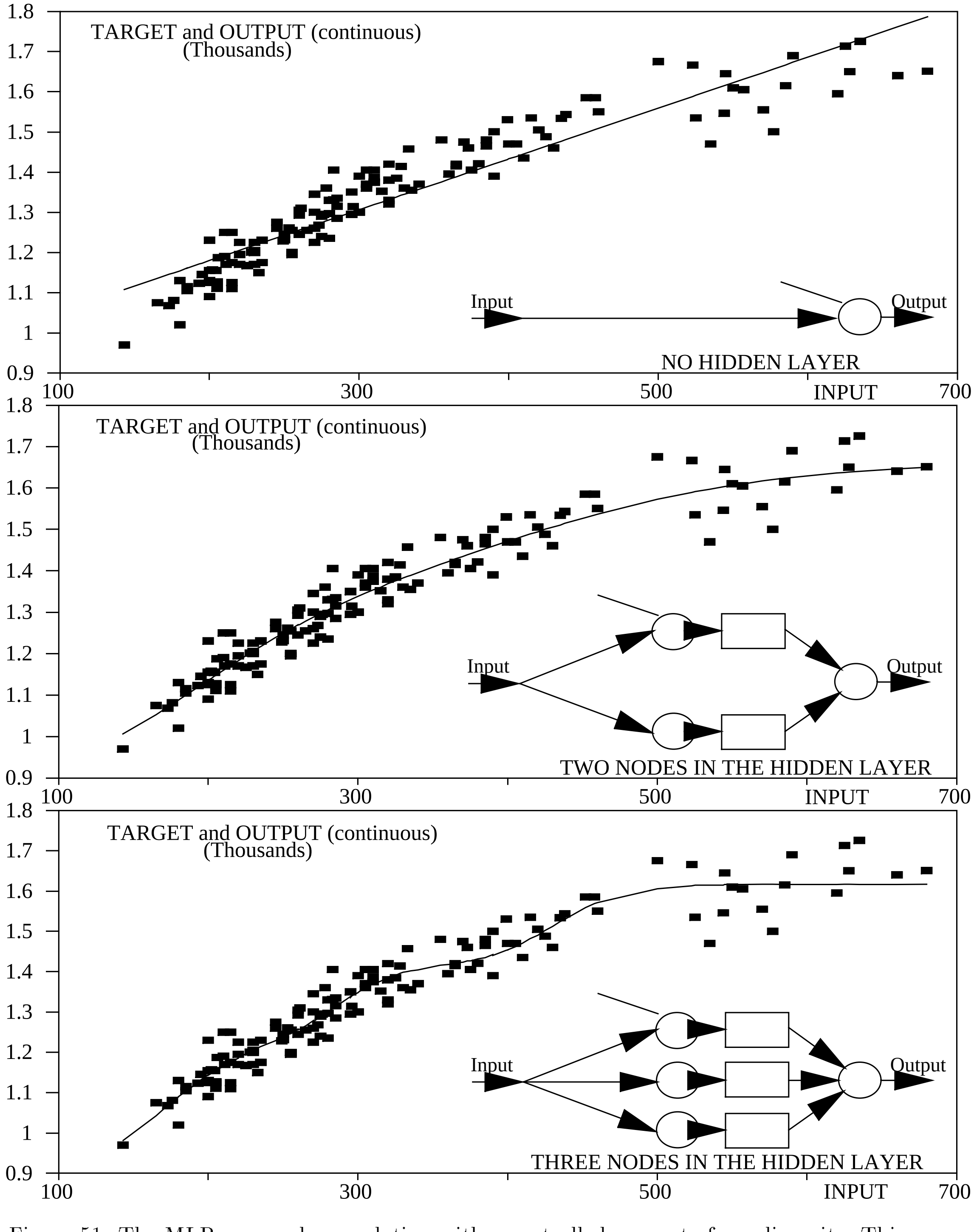

Figure 51: The MLP approaches a relation with a controlled amount of non-linearity. This amount is determined by the number of nodes.

One linear step incorporating a certain amount of memory of past events, followed by a simple non-linear map (see [108]).

This result is equivalent to say that we can mimic complex mechanisms like systems of non-linear differential equations or a chaotic attractor just by pulling together a linear filter and a non-linear function.

Lapedes and Farber [75] used this principle to show that an MLP was suitable for performing systems identification, time-series prediction and similar tasks. They simply used the input vector as a delay-line (see figure 44). Hence, the first hidden layer acts like a Wiener filter. Subsequent layers introduce the required amount of non-linearity.

When used in this fashion, an MLP becomes a very effective predicting tool. Its generalisation capacity, flexibility and ease of use makes it substantially more attractive than the equivalent analytical procedures. Especially in the prediction of time-series apparently complicated but with an underlying dynamic mechanism, the use of these delay-line MLP often mean a decisive improvement.

We used delay-line MLP to identify the underlying system governing a few random number generators. The resulting topology was very simple. We also tested their use in the modelling of first-difference chaotic series observing good predicting performances.

Notice that most of the tasks referred to here could be more elegantly performed using recurrent algorithms instead of delay-lines. Recurrent nets engage a smaller number of weights than delay-line nets. Therefore they achieve a better generalisation and require smaller learning sets. Their learning is also faster.

Recurrent nets can also cope with missing values in the learning set. In the presence of a missing value, they provide the most likely value in the context. This feature is typical of Hopfield nets and is known, in its more general form as "Associative Recall".

Recurrence, however, requires an advanced practice and is not adequate as an introduction to Neural Networks. Almeida [2] contains this formulation.

6.4 Self-Organized Neural Maps of Patterns

Contrasting with Multi-Layer Perceptrons, the self-organized maps of patterns, developed by Teivo Kohonen during the late 70's [72] [73], don't require explicit supervision for the learning to take place. Each example of the learning set contains just a vector of inputs instead of the input-outcome pair used by the MLP. Therefore, Kohonen nets don't learn any relation. They self-organize themselves according to the main features of the inputs.

These class of heuristics are a good example of strict Connectionist reasoning, inclined to non-supervised learning schemes and strongly inspired by the brain. They became popular as a simple mapping procedure in tasks related to the recognition of the human voice.

We shall use Kohonen nets in chapter 8 for mapping two-dimensional scatters suited for financial diagnosis. The final result is a set of rules able to diagnose the state of financial features of firms given its accounting numbers.

Topology: In its most basic form, the net consists of a K-dimensional lattice of nodes able to map the density distribution of the data to which it is exposed. Hence, the task undertaken by a self-Organized Map is the one of establishing a relation between a set of a large number of points, the "patterns" and another set of a much smaller number of points, the nodes.

All the nodes are supplied with the same input vector $x_1, x_2, \ldots, x_M = X$. Each element contains its own set of adjustable weights: For the j^{th} node, a weight vector $w_{j1}, w_{j2}, \ldots, w_{jM}$ will link to a corresponding input variable. Each node's output is a function of both the input vector and the weights: $o_j = f(X, W)$. For example, two frequently used outputs would be

$$o_j = \sum_{i=1}^{M} x_i \times w_{ji}, \quad \text{an inner product, or} \quad o_j = \sqrt{\sum_{i=1}^{M} (x_i - w_{ji})^2} \quad \text{an Euclidean distance.}$$

It is a requirement that these functions will yield a measure of the distance or similarity between W and X.

Non-supervised learning: The training of the self-organized map is known as non-supervised since the examples only contain input vectors, not outcomes. It takes place as follows: A learning set is supplied to all the nodes. When a new vector is shown to the net,

1. The node with the largest output is found. It will be the one with greater similarity between X and W.
2. A neighbourhood is defined around this node.
3. The weights of all nodes in such neighbourhood are updated or "rewarded" in a way that makes them more similar to the example they identified. That is, the new value of a weight, w_{ji}, linking the input i with the node j will be calculated as

   $$w_{ji}^{t+1} = w_{ji}^{t} + \eta \times (x_i - w_{ji}^{t})$$

   in which t and $t+1$ denote a sequence and η is a small increment. The nodes not in this neighbourhood receive no rewarding.

The procedure is repeated for all patterns in the learning set and then again and again. At length, specific nodes become "excited" by particular patterns. And the topological relationship between patterns is, after learning, mirrored by the spatial relationship of the nodes excited by those patterns.

Patterns are presented to a self-organized map in random order. After training, the value of each node's weight vector is usually plotted back into the pattern space with lines linking the nodes which are adjacent in the lattice as in figure 70 on page 207. The final result is thus a mapping of the many patterns onto the few nodes. Hence, each node "covers" a neighbourhood.

Specific tasks: Clearly, the learning scheme displayed above is a variation of the Hebb's Rule. Kohonen's maps owe more to Connectionism than the MLP. In the last one, the learning scheme is inspired in engineering practice rather than in the brain.

Self-organized maps are adequate to perform tasks such as:

- Dimension reduction: Find an f such $f : \mathbb{R}^N \mapsto \mathbb{R}^M$, $(N > M)$, having some optimal quality.
- Intrinsic dimension assessment: Find the smallest $M < N$ for which an f exists such $f : \mathbb{R}^N \mapsto \mathbb{R}^M$ having some optimal quality.
- Mapping as pre-processing for further MLP classification.
- Tracing dynamic features (like trajectories).

A valuable feature of Kohonen's maps is that it is unlikely to obtain an ordered map when using nodes forming a lattice of smaller dimension than the intrinsic dimensionality of the training set. For example, if a two-dimension lattice of nodes is used to map a really three-dimensional phenomena, the weight vectors will fold in waves, attempting to fully cover the 3-D space. In this case, no real relation exists between inputs and the density function. Therefore, folding can be used as a diagnostic for tracing an over-reduced dimension reduction attempt. Procedures are available to discover folding in more than 3-D maps.

Once an intrinsic dimension has been recognized and a map produced, an MLP can be used to classify each shape according to features. This procedure would be the connectionist equivalent to Factor Analysis prior to Discriminant Analysis.

Finally, if input vectors are successive events, self-organized maps will draw a trajectory, should the variables bear any kind of joint trend (cross-correlations not zero). Several different trends can be identified by its trajectories.

6.5 Neural Networks and Financial Modelling

This section compares Neural Network prospects in financial analysis with classical methods, showing how some limiting assumptions of the later can be avoided.

This section is not intended to describe what is contained in the second part of this study. Rather, it mostly describes what we couldn't do but we would like to, should enough time and appropriate data were available. On the whole, the examples explored in this study illustrate important and promising capabilities of Neural Networks in accounting and finance research.

Neural Networks are not a key to all kinds of data-analytical problems. They offer some specific advantages and they have their own drawbacks. This section mainly focus on those fields in which we think their use is advantageous.

Figure 52: Classification boundaries and number of nodes in the Multi-Layer Perceptron for two inputs. Adapted from Lippmann, R. 1987. An introduction to computing with neural nets, *IEEE ASSP Magazine*, vol. 4.

Classification: The most obvious application of the MLP is in classification. Discriminant Analysis can establish linear or quadratic boundaries between groups. This seems enough in the majority of current problems. A Multi-Layer Perceptron will draw boundaries of any shape (see figure 52 on page 164). Thus it is able to cope with complex relations involving higher order effects. But not only this. An MLP models complex relations in the most parsimonious way. The number of free parameters engaged in the modelling can be optimized as well.

Bond rating and lending decision mimics have already been attempted with the MLP [31]. A recent study on the selection of Neural Network architectures for improving generalisation uses bond rating as an example [130]. See also [57] for a review of some Neural Net applications being developed by a specialized firm. Firm distress prediction has also been modelled by an MLP. This, despite the problem being well suited for linear algorithms.

The described applications are just direct extrapolations of classic and thoroughly explored problems in accounting research. They simply substitute the linear techniques by the MLP. We think that such experiments are not the most adequate way of showing the real possibilities of Neural Networks since there is very few of specifically related to Neural Networks on them. Instead, we centre this study in what Neural Networks can do and the other tools can't.

Assessment of dimensionality and dimension reduction: Statistical tools like Multiple Discriminant Analysis or Factor Analysis are unable to clearly point out the intrinsic dimensionality of the data. N input variables or groups lead to N factors or $N-1$ scores. Despite the use of some ad-hoc tests, it is after all the intuition of the researcher who decides how many of these dimensions are to be considered as real features. Possibly, guesses of intrinsic dimensionality of processes like economic pervasive factors influencing capital markets or basic common sources of variability in ratios, based as they are on conjectures about acceptable uniqueness, are over-estimations.

One of the most promising applications of self-organizing maps of patterns lies in the fact that the real dimensionality of the data is recognized as a basic characteristic [73]. The real dimension of market expected returns or accounting statements could then be assessed.

Also the discrimination between different kinds of firm distress — should they exist — could benefit from the capacity of Kohonen maps to trace dynamic features. Since the relation between accounting data and the outcome is in this case very strong — the outcome can be predicted with small confidence limits — it seems as if a simple self-organized map of patterns would be enough to trace it. Such a map would also be able to discriminate between trajectories leading to insolvency.

Features Extraction: After the assessment of the intrinsic dimensionality, an MLP can construct new variables containing the main sources of variability present in the data. These factors can be built so as to be similar to Principal Components or, alternatively, to capture non-linear features. In the later case, the extracted factors are also representations of the data, extracted in such a way that the average missing information becomes minimal. But not subjected to the condition of being linear.

In other words, Neural Networks can, if required, extract non-linear pieces of information from the multi-variate distribution. For example, if in some two-dimensional phenomena its scatter diagram shows a clear "S" shape, the first or main feature to be extracted can be the S shape itself. An MLP will classify other S-shaped distributions as sharing that feature with the original data. This can be decisive when trying a classification of sensitivities of assets to market forces based on accounting reports and related information, or other situations where linearity doesn't apply.

Forecasting and Systems Identification: The MLP is suited for forecasting as well. In this case, the desired outcome is the same time-history as the input but placed a few periods ahead. Each input variable is related to the others so that the information fed into the MLP is a window representing a given time period. The learning takes place by showing the net many of these windows selected at random, along with the corresponding time-history a desired number of periods ahead. As a result, the MLP learns to predict the underlying phenomenon.

A description of the MLP in forecasting and Systems Identification can be found in Lapedes and Farber (1987) [75]. White [133] used an MLP to try to predict the returns of common stock.

Systems Identification is potentially interesting for assessing the extent to which some financial time-histories are dictated by a complex chaotic behaviour rather than by simple randomness. It is possible that the price of some commodities are a non-linear dynamic one as well. Benoit Mandelbrot noticed one such structure in the price of cotton [82] and other authors suggested similar behaviour in indices related to equity in the NYSE [93]. If that is so, the MLP would be a most adequate tool for capturing the underlying mechanism.

Dynamic Features, Stability and Diagnostics: Neural Networks can even cope with time-varying multi-variate data patterns as a whole. Time correlations of non-stationary data can carry important information about underlying trends.

Considering each M-dimensional cross-section input as a vector, if there is some relation between an event and its predecessors, a trajectory of such a vector will be drawn in $\mathbb{R}^M$. This trajectory can be recognized by an MLP or, in some cases, by a self-organized map of patterns after appropriate reduction. See Tattersall (1988) [126] for an explanation of this technique.

This seems a promising diagnostics tool for discussing the stability of APT factors, different kinds of firm distress or ratio information contents.

6.6 Summary

Neural Networks are versatile modelling tools likely to become useful in specific problems involving the extraction of knowledge from samples of accounting and financial data. The more promising tasks seem to be those related to:

- Complex classification and systems identification. In this case the valuable feature of Neural Nets is their power and generalisation capacity. Chapter 9 is dedicated to the exploring of this aspect.
- Features Extraction via node's specificity. In this case the feature viewed as interesting is the information apportioned by the model about the intrinsic structure of the data. Chapters 7 and 8 are dedicated to the exploring of this second aspect.

Neural Networks embody two main sources of inspiration. They are Connectionism and Tele-Communications Engineering. Examples have been given of the most representative nets in both cases: The Multi-layer Perceptron and the Self-Organized Map of Patterns.

Chapter 7

Knowledge Acquisition Using the Multi-Layer Perceptron

We introduced in chapter 6 the Multi-Layer Perceptron (MLP) as a modelling tool. Many other algorithms available are also intended to learn a relation input-outcome from a set of examples. The MLP is different in that it approaches relations by stages, not directly.

During the learning process an MLP creates new sets of variables corresponding to different stages of the modelling of the desired relation. A particular stage will use the variables from the previous one as input. Then, it will make an improvement towards the final modelling of the relation. Finally, it will output a new set of variables to be used as input for the next stage. The intermediate variables generated by an MLP are often referred to as internal representations.

In this chapter we show that the ability to create internal representations along with other characteristics of the MLP, make it able to automatically extract meaningful knowledge from raw data directly available in accounting reports and the related outcomes thus avoiding the need for searching appropriate ratios.

The next section explains which characteristics of the MLP are valuable in accounting modelling and knowledge acquisition. Section 7.2 describes how accounting items can be used as direct inputs for an MLP. Section 7.3 introduces a typical classification problem involving the prediction of the industrial group to which each firm belongs, using accounting data. The MLP was able to discover appropriate ratios for modelling such a relation. Incidentally, it also over-performed the usual statistical tools in classification power.

Using the above problem as a background example, sections 7.4 explains the departures from ordinary techniques we introduced in the training of the MLP. Two contributions are outlined: The post-processing of MLP outputs so that they can be used as scores. The random penalization of small weights for improving generalisation and obtaining meaningful internal representations.

Appendix B complements this chapter. It is a self-contained study of the performance of the MLP when compared with traditional methods.

7.1 Specific Characteristics of the MLP

In this section we summarize the characteristics of the MLP which seem valuable for knowledge acquisition and statistical modelling. This is an important issue since MLPs are very expensive in CPU time and attention from the operator. MLPs are the kind of tool which no one uses unless it is really necessary.

The characteristic specific to the MLP is the ability to model by stages, thus creating internal representations. However, the other desirable features are not easily found, all of them, in the same tool. For example, some statistical algorithms perform stochastic non-linear optimization. But they have little control over the amount of non-linearity introduced, or over the dimension allowed to model a desired relation.

We now summarize the advantages of using the MLP.

Meaningful internal representations: It is the back-propagation of deviations from the desired outcome towards more than one layer of nodes which makes the Multi-Layer Perceptron potentially attractive in knowledge acquisition tasks. The outputs of intermediate nodes, considered as new variables, can eventually bear interesting information about the process underlying the relation or its features. Such new variables, the internal representations, along with the net topology, can make the modelling self-explanatory.

The mechanism by which nodes learn has been presented in chapter 6. We saw that nodes tend to assume a stable state and behave like small learning units eager to capture a feature of the relation being modelled. This often makes them assume meaningful internal representations.

Iterative optimization: A Multi-Layer Perceptron adjusts the free parameters which ought to model a relation in small steps. Each of these steady improvements seek an advance in the minimization of the observed deviations between the produced output and the desired outcome. Therefore, the learning of a relation progresses steadily along many small steps. This allows a broad manipulation of the free parameters — known as weights — engaged in the building of the model. Such a manipulation, unavailable in analytical tools, turns out to be essential for achieving good generalisation and interesting internal representations.

Tight control over the modelling power: Another important characteristic of the MLP is that a tight control can be attained over the flow of information for modelling a given relation as well as over the amount of non-linearity introduced.

The minimum number of nodes in any layer determines the maximum dimension of the modelled relation. For example, by using a hidden layer with three nodes, we constrain the relation to be modelled to have three dimensions. On linear grounds this would mean that the matrix representing the desired relation would have one of its dimensions set to three.

This fact allows a direct control over the power of an MLP when performing classification with non-linear boundaries. Since classification with arbitrary boundaries requires the artificial enlargement of the dimension of the input space, by controlling it we define exactly the kind of boundaries we allow for classifying.

Also the amount of non-linearity an MLP apportion to the model depends on the total number of nodes in the hidden layers, no matter its topology. We can have two hidden layers, each one with two nodes, or one unique hidden layer with four nodes. The amount of non-linearity allowed would be approximately similar in both cases.

Together, the last two characteristics of the MLP make it remarkably flexible in the use of its power. Hence, the MLP is flexible in its generalisation as well.

Easy implementation of any optimization and convergence criterion: Finally, when appropriate, we can easily change the optimization criterion for it is independent of the optimization process. For example, Minimum Least-Squares deviation, as a success measure, is just one of the possible criteria. Likelihood maximization seems more appropriate for problems involving classification. In such a case, the Multi-Layer Perceptron learns to maximize the probability of having obtained the set of input-output pairs which were actually observed in the training set. This flexibility adds up to the MLP's already good one.

Also, the learning itself can be carried out using generalised hill-climbing algorithms like those explained in chapter 6, or more elaborated stochastic optimization techniques involving, for example, simulated annealing. In general, the possibility of simply acting upon the rate of convergence and the individual increments each parameter receives during the learning process can be most valuable for achieving meaningful internal representations.

7.2 Ratios as Internal Representations of a Relation

Simple ratios have been used for extracting useful experience contained in samples where reports were gathered together with known outcomes. The problem of learning from examples using ratios can be formalized in this way: Let x and y be two items forming the ratio $y_j/x_j = r_j$ in the case of firm j. For learning we have a sample containing $1, \cdots, j, \cdots, N$ examples of these two accounting observations plus t, the vector of the related outcomes. If we assume the existence of a map $\mathcal{W}$ such that $\mathcal{W} : r \mapsto t$ then we learn it by finding a $\mathcal{W}$ which is optimal in some sense.

Before doing this we would have to assume an a-priori form for $\mathcal{W}$. For example, a linear $\mathcal{W}$

would lead to N equations

$$t_j = w_0 + w_1 \times r_j$$

where the w_0 and w_1 are unknown parameters. Generally, the set of examples is much larger than the number of unknown parameters. The problem is overdetermined. However it is possible to "solve" the set of equations in a Least-Squares sense thus obtaining the values of w_0 and w_1 which minimize the squared differences between ratios and outcomes. In this case, $\mathcal{W}$ would minimize

$$\sum_{j=1}^{N}(r_j - t_j)^2$$

a measure of the error committed when predicting t from r. In this linear case $\mathcal{W} = \{w_0, w_1\}$ would be the simple regression's intercept and slope.

The above model involves only one ratio and one outcome. It is similar to the ones used by Beaver [7] for discovering ratios interesting for the prediction of firm failure. Such a formulation covers both discrete and continuous outcomes and makes no strong assumptions about the distribution of r. However, this algorithm is adequate only if such distribution is reasonably homogeneous.

Cascading two relations inside the MLP: The functional relation between accounting items — yielding ratio outputs — is different from the relation between accounting features and outcomes we now study. The last one is the goal of statistical modelling. However, these two relations are not independent. Outcomes, like distress or wealthy states, are dictated by internal features of the firm which we believe are reflected by appropriate ratios.

In the case of accounting statistical models used so far, the former relation is embedded in the choice of the input data — ratios. In the framework presented in this chapter we let the MLP form both such relations. Appropriate ratios are discovered and used for approaching the outcomes as part of a unique optimization process.

Since size is generally considered as an important piece of information for modelling some relations, we also make allowance for such an assumption to work out as an intermediate result.

In short, when modelling a relation we allow ratios to be formed as the output of nodes in the first hidden layer of an MLP, along with a proxy for size. Our approach solves the problem of finding the appropriate set of ratios for modelling a particular relation. Such problem clearly emerges when reviewing the published literature.

Forming ratios in the first hidden layer: We let the raw data be the input to an MLP. Then, we set it to model the desired relation. As a first stage in this process ratios are formed that approach the outcomes. Other stages follow. At the end, outputs are the final stage. If ratios are the appropriate way of modelling such a relation, the internal representations formed by the MLP in the first hidden layer are extended ratios.

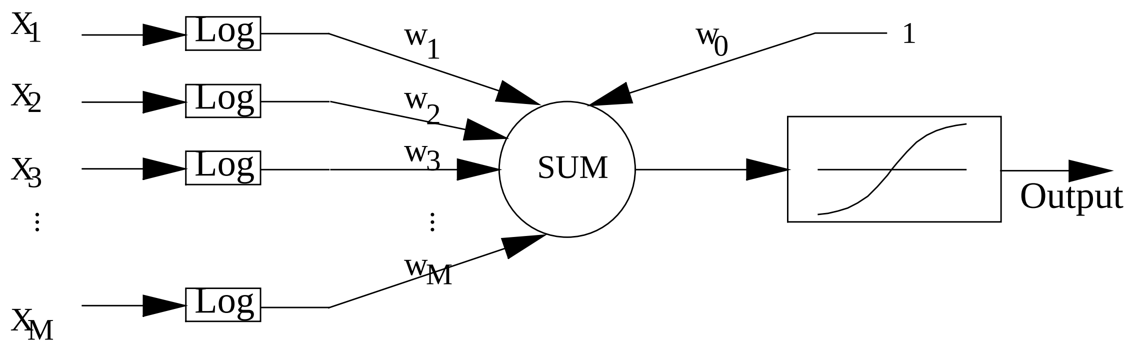

Figure 53: A node able to form a ratio in the first hidden layer of a MLP.

As seen in section 3.2.1 a multi-variate relation able to account for both common and particular components of the variability of accounting data is

$$r = \prod_{i=1}^{M} x_i^{w_i}$$

containing $1, \cdots, i, \cdots, M$ items as input. The residuals are omitted. In logarithmic space,

$$\log r = \sum_{i=1}^{M} w_i \times \log x_i . \tag{20}$$

Notice that this expression, an inner product, is the same as a Neural Network node's output.

Our approach consists of letting w_i be the adjustable connections or weights linking the inputs of an MLP with the nodes in the first hidden layer. The inputs are the logs of the accounting items, x_i, considered as interesting for modelling the desired relation. Thus we create in each node of the first hidden layer an internal representation with the form of an extended ratio. Next layers use such ratios to approach the outcomes.

We show in section 7.4.1 that by using an appropriate training scheme these extended ratios often assume a simple and interpretable form. If the overall model discovered by the MLP is optimal in some sense, it seems reasonable to expect that the discovered ratios represent an optimal choice of combinations of variables as well. Therefore, the problem of forming ratios given a set of accounting variables considered as preeminent can be avoided. The best ratios to be used are not imposed by the analyst. Instead, they are discovered by the modelling algorithm.

The transfer function: Figure 53 is a representation of a node intended to form ratios. The logistic function

$$f(x) = \frac{1}{1 + \exp(-x - \theta)} \tag{21}$$

which is standard in Multilayer Perceptrons as a transfer function, will bring back the extended ratios from logarithmic space and will also provide a controlled amount of non-linearity for the lower values of r.

$$f(r) = \frac{1}{1 + \exp(-\log r - \theta)} = \frac{r}{r + \exp(-\theta)}$$

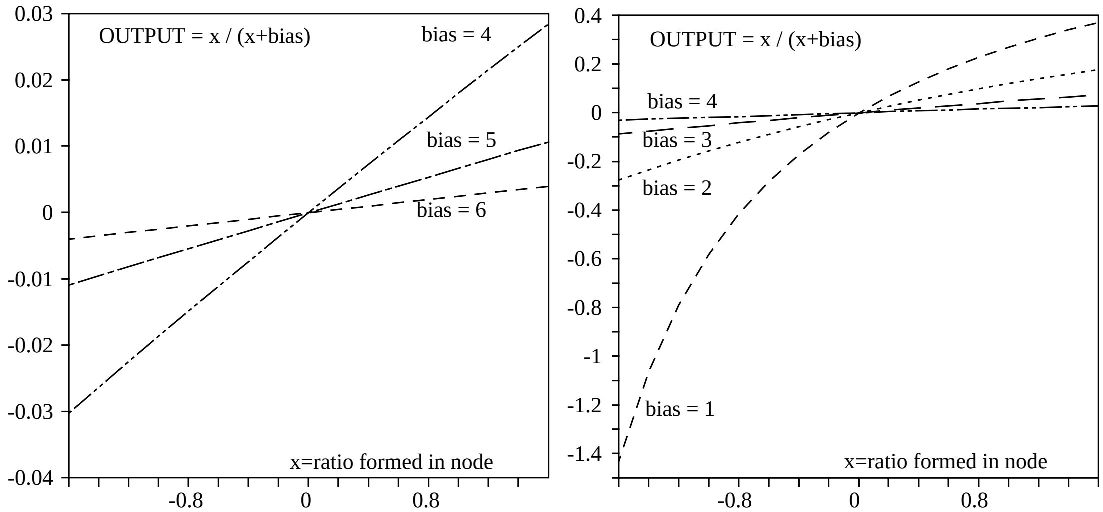

Figure 54: The output of each node in the log MLP will be a concave function approaching linearity for increasing values of the bias. On the left, a magnified view.

θ is the bias. Large negative values of θ yield a linear relation between r and the output of the node. Smaller values introduce a concavity affecting small r.

Figure 54 shows the way θ controls the output of its node. For increasing bias the node's response is linear. In general, the training of the bias is directed by the optimization algorithm so that the output is linear. Therefore, the first hidden layer is not apportioning non-linearity to the model. This can be done in next stages. However, there is a class of non-linearity which is accounted for in this layer just by allowing smaller θ (a notation also used for θ is w_0).

The modelling of base-lines and other non-linearity: The back-propagation of errors could also be used for discovering and accounting for non-proportionality in individual inputs. Appropriate base-lines could automatically be found for each input just by using the information propagated backwards, in the same way the other free parameters are estimated.

In practice, such a propagation across the log function is not stable. The MLP simply generates δ which become more and more negative during the training. Therefore, at least at the present stage of the research, we directly model the non-linearity introduced by base-lines instead of reproducing its underlying mechanism. This can be done since, as we saw in section 3.4.3 the log space introduces a trade-off between non-proportionality and non-linearity (see figure 26 on page 77) and the MLP can model such a non-linearity.

The number of hidden layers to be used and the number of nodes in each layer should be selected so as to apportion the required dimension and amount of non-linearity. For example, in order to model difficult non-linear relations with just two or three dimensions — with just two or three ratios —, one additional hidden layer of nodes is required. Depending on the number of nodes allowed, this new layer will apportion as much piece-wise non-linearity as necessary and, hopefully, not more

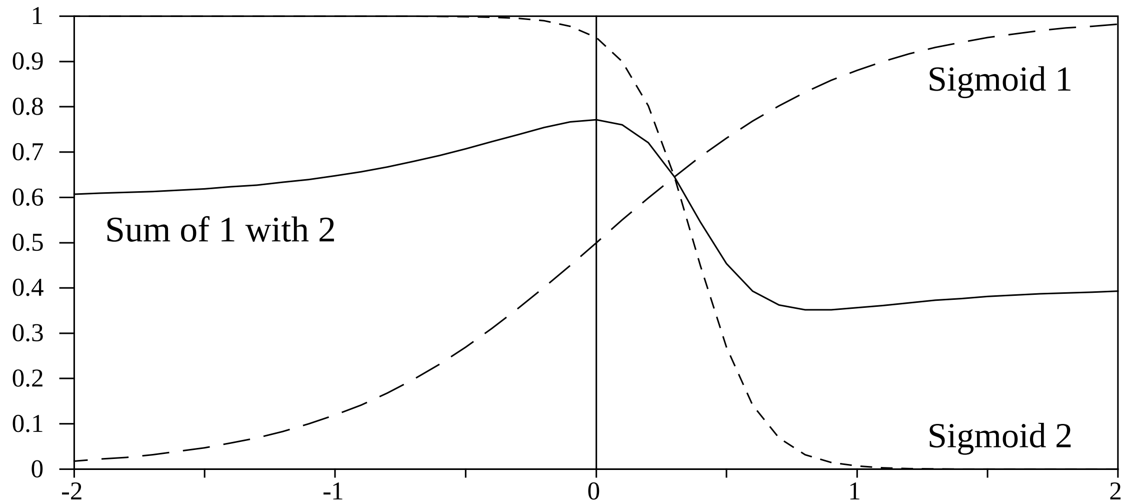

Figure 55: Two sigmoid modelling a local non-linearity.

than the necessary. Unlike polynomials, the logistic transfer function tightly controls the amount of non-linearity in use. It is a very simple, monotonic increasing curve. Its response becomes constant when the input is out of a limited range. Several of such curves can bring to the model the required pieces of non-linearity yielding good interpolation and extrapolation.

Figure 55 shows a very simple non-linear relation, the solid line, which requires two sigmoid to be accounted for (the dashed lines). This curve would never be conveniently modelled by polynomials. The non-linearity introduced by polynomials is not local. It will persist out of a limited neighbourhood.

7.3 Learning to Discriminate Industrial Groups

In this section we apply our framework to a known accounting statistical problem, the test of the separability of the components of a particular industry grouping. We compare our procedure with the traditional one and we extract some conclusions.

All companies quoted on the London Stock Exchange are classified into different industry groups according to the Stock Exchange Industrial Classification (SEIC). We selected 14 manufacturing groups according to the SEIC criteria. After discarding some firms (see below) we got accounting information on 297 firms covering a six year period (1982 - 1987) and a bigger sample (500 cases) for only one year (1984).

The data: The input variables received two different types of processing. The first, usual in finance research, consisted of "forming 18 financial ratios chosen as to reflect a broad range of important characteristics relating to the economic, financial and trade structure of industries (...) [124]" and

Feature	Ratio	Tr.	Feature	Ratio	Tr.
Operating	NW	Log	Fixed Capital	FA/TA	Sqrt
Scale	S	Log	Intensity	$S/$ Av. FA	Log
Labour-Capital	W/TA	Sqrt	Short Term	D/CA	None
Intensity	$VA/$ Av. TCE	Sqrt	Asset Intensity	D/I	Log
Profitability	OPP/S	Sqrt	Asset Turnover	DD	None
	$EBIT/S$	Log		$S/$ Av. TA	Log
	$OPP/$ Av. TCE	Sqrt		S/I	Sqrt
	$EBIT/$ Av. TCE	Sqrt			
Net Trade	D/C	Sqrt	Financial	$DEBT/NW$	Sqrt
Credit			Leverage	$DEBT/TCE$	None

Table 29: Ratios used in the original study and their transformations.

extracting from them the eight principal components. These new variables were then used as inputs for a Fisher's Multiple Discriminant Analysis (MDA). A description of these ratios and the modelling procedure can be found in [124]. Table 29 reproduces them along with the transformations applied. "DD" means the ratio Debtors Days.

Appendix B is a self-contained study of the performance of the MLP compared with traditional methods. There, a description of our reproduction of the method usual in finance research can be found along with the detailed MLP classification results.

The new approach consisted of using eight accounting variables directly, not in the form of ratios. The selected items were Fixed Assets, Inventory, Debtors, Creditors, Long Term Debt, Net Worth, Wages and Operating Expenses less Wages. All these variables were present in the original 18 ratios, along with others like Earnings, Value Added, Total Capital Employed and Total Assets which we didn't use in the new approach.

Criteria for selecting the input variables: The criteria used for selecting the new variables was threefold. Firstly, they should have been present in the original set in order to allow the comparing of results. No new information was to be introduced in the problem. Secondly, we avoided items representing totals for reasons explained in chapter 4. Finally, the input dimension should be eight or less. The number of common factors extracted from ratios in the original study was eight. Eight items or less wouldn't allow a larger flow of information.

The choice of EX and Wages instead of Sales and Operating Profit stems from the same reasoning. The discarding of Earnings stems from not being appropriate for the log transformation. The information contained in $EBIT$ could be introduced by Sales and $COGS$ but for this particular model the residual $EBIT$ didn't seem important.

The selection of cases for the samples: A major methodological difference between our approach and the usual one was the way firms were selected. In general, one-variate normality criteria is used to prune the original sample of ratios down to an acceptable number of standard deviations.

N.	Group Code	Group Name	N. Cases	Proportion
1	14	Building Materials	31	6.2%
2	32	Metallurgy	19	3.8%
3	54	Paper and Pack	46	9.2%
4	68	Chemicals	45	9.0%
5	19	Electrical	34	6.8%
6	22	Industrial Plants	17	3.4%
7	28	Machine Tools	21	4.2%
8	35	Electronics	79	15.7%
9	41	Motor Components	23	4.6%
10	59	Clothing	42	8.4%
11	61	Wool	19	3.8%
12	62	Miscellaneous Textiles	30	6.0%
13	64	Leather	16	3.2%
14	49	Food Manufacturers	80	15.9%

Table 30: Industrial groups and the proportion of each one in our sample.

We followed a case-wise method for discarding undesirable firms. It was not based on distributional considerations. Only cases known as distressed firms, non-manufacturing representatives of foreign companies, merged or highly diversified ones were excluded.

Results concerning two sets of data are reported. The first ("1984") represents a cross-sectional view. The second ("SIX YEARS") checks the regularity of firm grouping during a larger period.

Table 30 displays the proportions of cases in the 1984 set. Notice how groups are dissimilar in size, the smallest one having 16 firms and the biggest 80. These proportions entail no prior knowledge of any classification.

7.4 Improving Generalisation and Interpretability

In this section we explain the characteristics which make our MLP different from the standard algorithm. They can be summarized as:

- The use of two samples, one for learning and another one for assessing the classification performance. This is commented in 7.4.1.
- The random penalization of small weights, explained in 7.4.2.
- The post-processing of outputs, outlined in 7.4.3.
- Learning rates particular to each weight as described in Silva and Almeida [116].
- Likelihood maximization instead of squared deviations minimization, as explained in 7.4.3.

The first characteristic relates to improvements in the ability to generalise. It is a particular implementation of a known procedure, the Cross-Validation [122] [123]. The random penalization of errors and the post-processing of outputs are specific contributions of this study. They allow the use of the MLP for general-purpose statistical modelling and the interpretability of results.

7.4.1 Generalisation: Using the Test Set

In order to obtain an estimate of the generalisation capacity of a model, the original samples were divided randomly into two sub-samples of approximately equal size. All models were constructed twice, first with one half of the sample and a check carried out with the other half, and again reversing the roles of the two half data sets. Results were considered conclusive if both models, when validated with the half-sample not used to build them, produced consistent results.

All classification results reported here concern the test set, not the training set. That is, they were obtained by measuring the rate of correct classification in the half-set not used for learning. The classification performance on the set used for learning depends solely on the number of free parameters and can be increased simply by introducing more nodes on the net. Therefore such results are uninteresting and are not presented here.

The normal approach to test a model, by deleting a single observation and predicting its value with the model estimated on the rest of the data set, and repeating this procedure N times, is not feasible. This is because the training of a Neural Network is time consuming. The procedure adopted will however, with a large enough data set, produce unbiased estimates [37] [122].

The described procedure, combined with incomplete training, also allows improving the generalisation of the MLP. This is a common practice. Next we describe incomplete training.

The role of incomplete training: Since the MLP seeks an optimum iteratively, we can stop its training when an optimum is obtained in the test set rather than in the training set. In doing so we prevent this powerful algorithm from over-fitting the data.

It is generally believed that the Back-Propagation algorithm seeks the modelling of progressively smaller or less important features of the relation during the learning process. Firstly, broad features are accounted for: The mean, a linear trend. Then, more detailed ones are modelled. Hence, the effective degrees of freedom the MLP engages can be viewed as increasing during learning [132].

Assuming that the topology of the net contains plenty of free parameters, the MLP will be able to model, not only the desired features but also the undesirable random uniqueness of a particular sample. We prevent it from doing this by stopping the process before finishing. The appropriate moment for stopping is when the results, as measured by the test set, are optimal.

Figure 56 on page 177 shows the evolution of classification results in the test set during the learning process for the 1984 data. After 300 presentations the classification reaches an optimum. If the learning continues, the classification breaks down. Such a breaking down is a clear sign that, from the optimal point on, the MLP is no longer modelling any features of the population. Instead, it is modelling the variability particular to the learning set.

For a good topology, the fact that the learning stops before a minimum is reached in the learning set clearly enhances the generalisation. The difference between the generalisation performances achieved with analytic tools and the iterative ones stems from this ability to stop. In our example, if

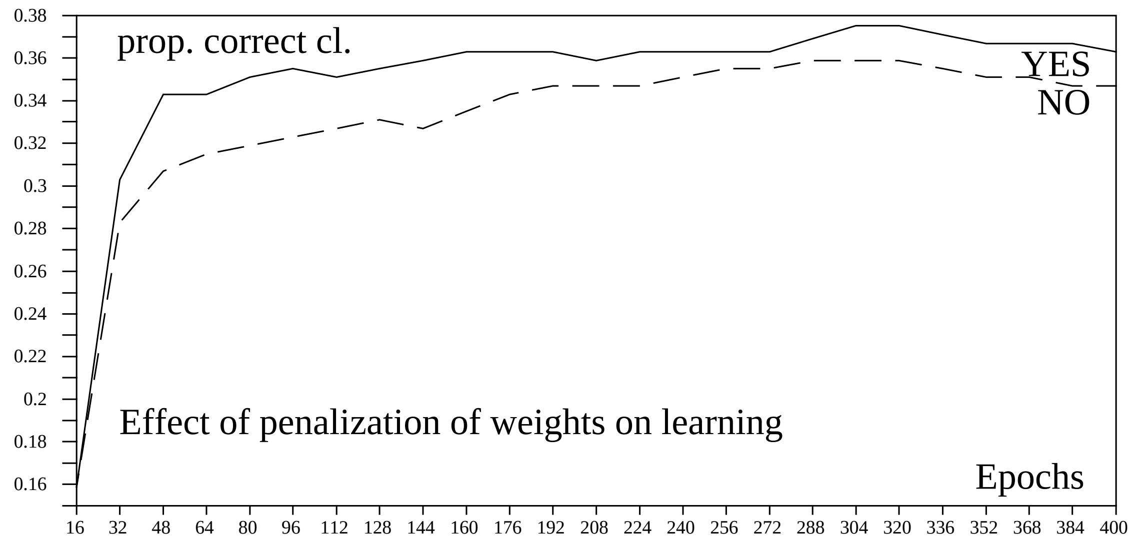

Figure 56: Typical evolution of the classification results during the learning process. 1984 (test set).

we allow the training to proceed, the generalisation obtained with the MLP is similar or even worse than the one obtained with analytic tools.

The role of an appropriate topology: We found the generalisation capacity very dependent on the topology of the net. The number of nodes in a hidden layer seems to determine, not only the dimension of the relation, but also the ability of the MLP to generalise. Persistently, we obtained good generalisations whenever a hidden layer would have six nodes. Both the 1984 and the SIX YEARS data set exhibit such a feature. Figure 57 on page 178 shows some classification results for different numbers of nodes in the first hidden layer for the SIX YEARS data. Similar patterns, though not so contrasted, were observed for the 1984 set.

7.4.2 Random Penalization of Small Weights

Another major goal of this study was to evaluate the power of Neural Networks in knowledge acquisition. Multi-Layer Perceptrons are often considered as not ideal in applications where self-explanatory power is required. However, in the case of accounting variables it seems possible to interpret the way the relation has been modelled by looking into the weights connecting input variables with the first hidden layer's nodes. These weights are the free slopes of ratios.

In order to enhance interpretability we introduced during training a random penalization of weights with small absolute values. A weight is inhibitory when its absolute value is smaller than the unit. If the input variables were very differently scaled, inhibition values in the input weights could just mean that the MLP was trying to scale down a particular variable. Since the log items used as input to the MLP are mean-adjusted and have very similar spread the only reason for any

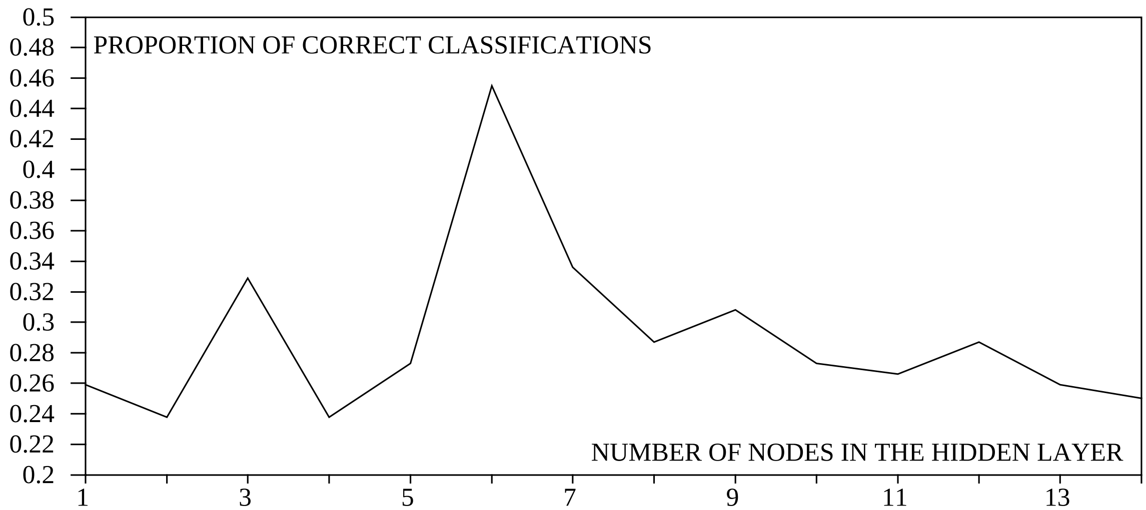

Figure 57: Classification results on the test set (Y-axis) versus number of nodes in the hidden layer for SIX YEARS data.

such weights to remain smaller than the unit throughout the learning is to try to diminish the importance of one variable in the output of the node it belongs to.

In a Neural Network each node acts as a modelling unit with a certain amount of free parameters. The same output can be obtained with very different combinations of weights. Inhibition weights connecting inputs with the first hidden layer appear when the node tries to weaken the contribution of a variable. If we randomly introduce a small penalization of such weights during the training, as the correction of weights is proportional to the input variables, the weights smaller than the unit tend to remain small. In the same way, the large weights tend to have their values strengthen.

The final result is a contrasted set of weights: The first layer now contains only very large or very small weights. The information concerning the modelled relation is concentrated in a few weights instead of distributed by all of them. If the relation to be modelled is consistent with such a contrast, then there is no reason to expect that the described manipulation will damage the performance of the model.

The algorithm: The procedure we follow to achieve interpretability involves the following steps:

- Let one node in the first hidden layer model the strong common effect and introduce it in subsequent layers. Input variables not convenient for the modelling of size (Debt is an example) have weights connecting to this node set to zero. The others have fixed and equal weights.

- During training, and whenever a new presentation of the entire training set is to begin, one of the remaining nodes of the first layer is randomly selected. Their weights are examined and those with inhibitory weights are penalized by a small factor, typically 0.98.

- Before the end of training, all the weights connecting inputs to the first layer and exhibiting very small values are set to zero and fixed.

Such procedure is applied only after the discovery of the topology yielding the best results.

Just by dedicating one node of the first hidden layer to the modelling of the strong effect we notice an improvement in speed of convergence and in the final generalisation. Adding the random penalization of inhibitory weights, both speed and generalisation receive a further, significant, improvement. When the topology is not the best, this procedure can worsen the generalisation.

Complementary remarks: The method described here was the one used for this particular experiment. In different cases we found that the performance would not suffer if all the weights below an inhibitory threshold were penalized at the beginning of each new presentation. This threshold typically would begin in 0.1 with the training and then it would be updated to larger values later on. Instead of fixing the weights just before the end of the training we also introduced their fixing during training whenever they would become small enough.

This simple procedures enhances performance and generalisation considerably. The fact that by the end of the training the real number of free parameters is much reduced is also rewarding.

Notice that we never tried this method with the usual, simple, Back-Propagation algorithm. Each one of the weights in our MLP has its own increment, adjusted as described in Silva and Almeida [116]. It may well be that our algorithm's performance is contingent on such a procedure.

Other popular methods for pruning the MLP are the "Skeletonization" [90] and "Optimal Brain Damage" [76]. The first one is intended to reduce the number of nodes, not weights. The second one is too general for this task.

Results: When the training finishes the number of variables to consider in each node is very small and characteristic. Looking at the non-zero weights it is possible to understand, in accounting terms, what the free-slope ratios formed in each node are doing.

Table 31 shows the extended ratios formed in a net with 8 inputs, 6 nodes in one hidden layer and 14 output nodes, trained with 1984 data. The emerging organization reproduces the way an expert in ratio analysis chooses variables. It is usual to build several ratios around one or two variables judged as important to capture a relation. As an example, efficiency is modelled around capital turnover, stock turnover and so on. Profitability is built around profit margin, return on equity, etc.

Experts put together several points of view around a few significant variables by using them to contrast others. Extended ratios seem to be trying the same sort of procedure. The item EX has been used in all hidden nodes to contrast others. It seems as if it were important for this problem.

The ratios the MLP discovers are not always simple. Ratios like $(C \times FA)/(W \times EX)$ are not the most familiar ones to accountants. However, in general the combinations of items which emerge as interesting are clearly visible when examining the organization of the hidden nodes.

Variable Node Number	2	3	4	5	6
Long Term Debt			-6		
Net Worth	8				
Wages	1			-6	
Inventory	8				
Debtors	2				-2
Creditors				3	
Fixed Assets	-9	-4		6	-4
Operating Expenses less Wages	-10	4	8	-2	3

Table 31: Approximate values of weights connecting input variables with nodes in the first hidden layer after training with random penalization.

Testing the Performance of the Devised Ratios: Our interpretation of the ratios formed in the hidden nodes according to table 31 is as follows:

$$\begin{cases} \text{In the } 2^{th} \text{ node: } & \dfrac{NW \times I}{FA \times EX} \\ \text{In the } 3^{th} \text{ node: } & \dfrac{EX}{FA} \\ \text{In the } 4^{th} \text{ node: } & \dfrac{EX}{DB} \\ \text{In the } 5^{th} \text{ node: } & \dfrac{FA \times C}{W \times EX} \\ \text{In the } 6^{th} \text{ node: } & \dfrac{EX}{\sqrt{FA \times D}}. \end{cases}$$

We tested the performance of such ratios when used as inputs for linear classifiers in the described problem. The five ratios plus the size effect actually classify the 14 industrial groups with the same accuracy as the original 18 variables.

The gain in performance by using the MLP is, of course, much more visible. Apart from its non-linear modelling capacity — which in this particular problem didn't seem to be very important — such a gain is due to its superior generalisation. Analytic tools cannot control the relative importance of parameters during training nor stop the optimization process before its end, to avoid over-fitting.

7.4.3 Post - Processing of Outputs

Discrimination, when overlapping distributions are present, implies a probabilistic interpretation of outputs. In accounting research, Bayesian considerations are in general independent of the proportions observed in the sample. Neural Network application to other sciences can be misleading. There, proportions observed in the sample are generally taken as acceptable prior probabilities.

Following suggestions like those of Baum and Wilczek [6] several authors advocate a direct interpretation of outputs as probabilities [61] [120] and show how the usual squared error criterion can be corrected to achieve likelihood maximization. In such case, the weights are corrected in the gradient direction of the log-likelihood rather than on the gradient of the squared error.

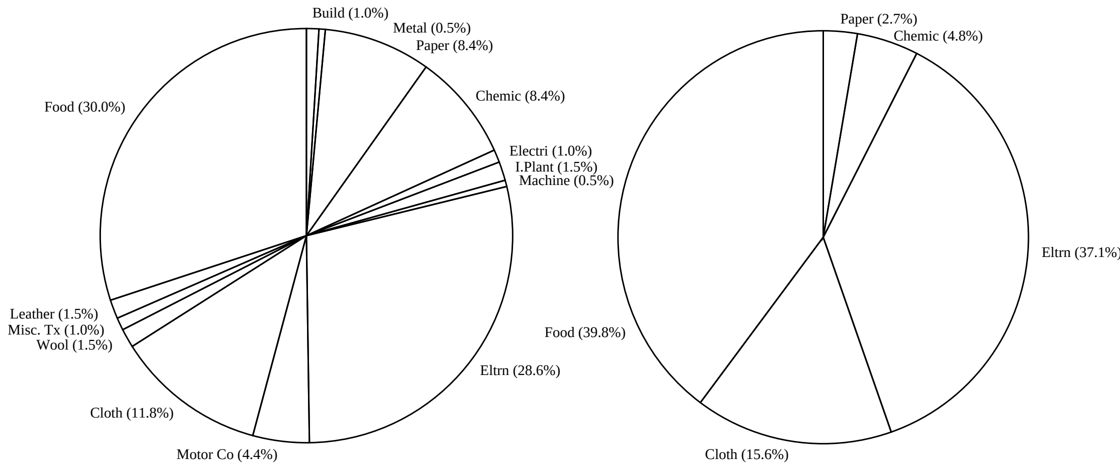

Figure 58: On the left, classification results after post-processing (1984 sample and prior probabilities proportional to the size of the group). On the right, the same with direct interpretation.

We found that node outputs — when interpreted directly as probabilities — produce a clear reduction in accuracy. The final result is a severe loss of ability to distinguish small groups.

Thus, we decided to interpret outputs of the MLP as a multi-dimensional measure of distance to outcomes. If departures from normality are not severe, this interpretation can be carried out by using conventional statistics like Chi-Square, Penrose or Mahalanobis distances. Such measures can be regarded as scores and conditional probabilities can be deduced from them, allowing further Bayesian corrections, independent of proportions observed in the sample. Of course, a Bayesian correction could be done directly over the outputs interpreted as probabilities. However, due to the observed lack of accuracy, a direct correction would lead to a very bold classification.

An experiment: Using the MLP with the 1984 data set and implementing learning schemes described by Hopfield [61] and Solla *et al.* [120] we tested the direct interpretation of node outputs as probabilities comparing it with the usual correction of node outputs based on the way linear discriminant analysis, for example, corrects scores. Results are reported in figure 58 when prior probabilities are taken as equal to the size of the group. On the left we can see the result of using post-processing. On the right, the corresponding result derived by directly interpreting node outputs as probabilities.

The post-processing gives detailed classifications. Direct interpretation ignores 9 of the 14 groups, the small ones, but finally achieves a better global performance by classifying the remaining 5 groups, which are the bigger ones, very well. Therefore, although for the sake of efficiency of convergence we adopted the likelihood cost function, node outputs were post-processed as distances. A short description of this post-processing follows.

MLP outputs as multi-variate distances: For a training set with N cases, consider o_{jm}, the output produced in node $m, m = 1, M$ by case $j, j = 1, N$. Compute K square deviations, d_{kjm}, between the m node's output and each one of the $1, \cdots, k, \cdots, K$ possible outcomes: $d_{kjm} = (t_{km} - o_{jm})^2$. The mean sum of squares in node m for the whole sample will be: $\sigma^2_{km} = \sum_{jk} d_{kjm}/(N-1)$ and the standardized distances between a node's output and all possible outcomes can now be added over all nodes:

$$D_{kj} = \sum_{m=1}^{M} \frac{d_{kjm}}{\sigma^2_{km}}$$

The minimum of these distances would identify the outcome predicted by the MLP if no Bayesian corrections were needed — that is, if the assumption of equal prior probabilities is acceptable.

This distance has been compared with a more elaborated measure, the Mahalanobis distance, and it was found that the latter would not achieve a more accurate performance. In order to introduce Bayesian considerations, D_{kj} ought to be computed as a Chi-Square distance to outcomes. The significance of this distance is the desired conditional probability.

7.5 Discussion and Conclusions

So far, expectations about Neural Networks are related to the modelling of difficult relations (pattern recognition) or the mimicking of brain functions. There has been little emphasis in their potential explanatory power. Here we argue that some statistical problems requiring self-explanatory power can take advantage from the existence of meaningful internal representations.

Numerical, continuous-valued observations such as those found in stock returns, or data organized in accounting reports, cannot be efficiently used by actual expert systems as a source of knowledge. Algorithms intended to automatic extraction of rules from examples, such as the ID3 [96] cannot perform efficiently with non-symbolic, non-hierarchical data. We explore this problem elsewhere [12].

Neural Networks can now be seen as an alternative self-explanatory tool. In our example, hidden units were able to form more appropriate ratios than those commonly used. In other cases the examination of such ratios could shed light in many important issues.

Self-explanatory models: The developments of this study are closer to Beaver's original works than its successors. Beaver tried to discover the most appropriate ratios to model a relation. The goal was not just an efficient modelling. It was mainly the discovering of simple tools for doing the job. After him, statistical modelling focuses on efficiency. The practice of using multi-variate techniques and a large amount of ratios as inputs — along with the trimming, ad-hoc transforming and rotating of inputs — made impossible any interpretation of results. Modelling became a blind automatism.

In order to return to interpretable models it is important to understand the statistical behaviour of items. And also to use tools able to implement such behaviour in a transparent way.

	1984		Six Y.	
INPUT	MDA	MLP	MDA	MLP
18 ratios	29%		30%	
8 variables		38%		45%

Table 32: The best classification results when using the MLP compared with the best ones when using MDA (Multiple Discriminant Analysis) for the two data sets.

The selected example: This specific application was chosen because it was likely to generate a relatively complex network with multiple output neurons. Complexity was seen as desirable to allow a richer investigation into the process of building and running networks. Choosing a classification problem with a better developed theoretical underpinning might have been more sensible. As it is, the also somehow complex ratios produced by the model are difficult to interpret. However, the revised training process's ability to generate simple structures promises much in other accounting applications.

Improvements in performance: The emphasis on interpretation should not hide the other findings of our study. The MLP proved able to outperform the classification performance of a traditional discriminant analysis approach. Neither method came close to adequately classifying the testing sets, but there was a substantial improvement when the MLP was used.

Table 32 shows the best generalisation results achieved with the traditional methodology (ratios) and also with Neural Networks. As can be seen Neural Networks achieved a better performance, with half the number of input variables and within a much simpler framework. Namely, the need for forming appropriate ratios was avoided as well as the blind pruning, and the extraction of a somehow arbitrary number of factors. Several accounting variables used to form the 18 original ratios were not present in our 8 variable set.

It is perhaps worth pointing out that redoing the discriminant analysis using representations of the ratios produced by the MLP captured some but not all of the MLP-based improvements. The remaining may well be related to the ability of the MLP to cope with non-linear boundaries and have a tight control over the number of free parameters.

Topology: The principle of parsimony should also be born in mind. If there are too many hidden nodes the MLP will fail to identify key features and will model the particular randomness in the data set as well. Generalisation will then be lost.

However, Back-Propagation shows a useful ability to take advantage of the topology of the net to improve generalisation. Even with a large number of free parameters, if the number of nodes in a hidden layer is in resonance with some internal feature of the data, high generalisation can arise.

Chapter 8

Improving Diagnostic Specificity With Neural Networks

In this chapter we show how Neural Networks can be used for automatic diagnosis based on accounting numbers. Our framework is intended to extract rules from large databases containing accounting data. It can be seen as a pre-processor, able to bridge the gap between continuous-valued, stochastic, data like the one found in accounting reports and symbol-based expert systems.

The rules resulting from using this technique can be fed into more general systems along with other sources of knowledge. For example, several of the developed tools could cover each one a particular feature of the firm — liquidity, capital turnover and so on — outputting rules that would be jointly processed by an expert system.

Two-dimensional ratios: Our tool allows both cross-sectional positions of individual firms and trajectories during a time period to be traced. It is close to ratios yielding similar diagnostics: Whether a particular firm is above the standards, near the expected or below the standards. But this, referred to two aspects of an accounting feature, not just one at a time. For example, whilst ratios can assess liquidity either as a contrast between the amount of current assets and the one of current liabilities or, alternatively, as a contrast between the amount of working capital and the size of the firm, our tool can show both aspects of liquidity together.

The quality of the diagnosis and its interest rely on the selected variables as happens with ratios. It is the experience of the analyst which dictates which items are to be used and how to interpret the resulting maps. Since these tools are close to ratios, all the expertize of ratio analysis can be directly implemented on them.

Given this, the expected improvements in diagnosis power and specificity are not based on any new algorithm. Rather, they rely on the common experience of practitioners and in the models devised in chapters 3 and 5. In fact, it is common practice to assess a particular feature using more than one ratio. Two or three ratios related to the same characteristic are generally called upon and compared. Our two-dimensional ratios combine two aspects of any feature as described in the mentioned chapters.

Also, it is an extended belief that the inclusion of information regarding size can be decisive for a correct diagnostic in some problems. For example, the prediction of firm failure is enhanced just by noticing that small firms are more likely to fail than the large ones. Ratios are size-adjusted variables. They cannot convey this piece of information. Graphical ratios can be set to introduce size so that the resulting tool will be sensitive to both the size-adjusted features and the position of the firm regarding size.

Contents: In the next section we develop and discuss two-dimensional representations of accounting data. Then, we implement the second step, the mapping process. Finally we show some examples. The central reasoning determining the developments presented in this chapter is explained in section 8.1.2.

8.1 Creating Two-Variate Tools

Ratios only use a small amount of the information needed for building them. For example, a collection of ratios concerning different time periods can show the existence of a trend. But the collection of items used to build them would show a trajectory in a two-dimensional space. This could be far more revealing since it would allow the distinction between several directions for the same trend.

In this section we are concerned with the manipulation of two and three-variate relations in log space. We seek the building of tools able to be a natural extension of ratios to a second dimension.

The possibility of using two-dimensional representations of accounting data is a direct consequence of the homogeneity of accounting information when examined in log space. The tools developed in this section would not be feasible in a space other than the logarithmic. As discussed in chapter 2 the result would not be usable.

In this section we briefly discuss the interest for accounting research and the adequacy for financial diagnosis of four kinds of two-dimensional representations increasingly elaborated:

- The simple scatter-plot of two log items and the mean-adjusted scatter-plot, adequate for discovering and identifying forces which are external to the firm but not for financial diagnosis.
- The residual scatter-plot, intended to jointly examine the residuals of two items when deflated by a third one. It will detect features which are internal to the firm such as correlations between residuals.

- The rotated scatter-plot, containing the same information as the simple X-Y plot but being able to display it in a way adequate for financial diagnosis. This tool is the correct choice when size is an important piece of information.

- The rotated residual scatter-plot, which stems from the previous two. It is able to show the residuals of a ratio in log space against deviations of its components from the values expected for the size of the firm. This tool is a generalisation of the ratio concept for two dimensions.

Assumptions and notation: We accept that it is possible to build a general deflator, s, reflecting the component of the variability common to all items. We further assume that this s can be used with any accounting item in a ratio model without causing asymmetric residuals. This issue has been discussed in section 5.1.

In the case of residuals from the ratio y/x we used the notation $\varepsilon^{y/x}$ (or the corresponding $f^{y/x}$ for the ordinary space) and so on. We use simply ε^x, the corresponding f^x and so on for expressing the residuals obtained from x after being explained by s, a suitable proxy for the common effect. We refer to departures from it as just residuals. Departures from a particular ratio standard are referred to explicitly as contrasts.

In using this notation we emphasize the fact that the residuals of any item, x, when deflated by s, are a proxy for the weak effect or particular contribution of x to the overall cross-sectional variability. Inside firms the statistical behaviour of one item is different from the one of others because the internal mechanisms commanding them are different. It seems potentially interesting to isolate the particular variability of items related to important features like liquidity or profitability, and then examine it. When the denominator of a ratio is any item, not a proxy for size, its residuals reflect a contrast between two particular effects, not a unique particular effect. That's why we call them "contrasts".

8.1.1 Non-Rotated Plots

In order to implement a simple visual inspection of a two-variate relation, a scatter-plot is enough. When considering the ratio y/x, $\log x$ would be the abscissa and $\log y$ the ordinate. It is practical to have the line $y = x$ drawn in the scatter-plot.

Figure 23 on page 58 contains one of these plots. Simple scatter-plots in log space reveal the existence of external forces when they affect the symmetry or the linearity of the relation. For example, in the case of constraints introduced by accounting identities, all cases gather in one side of the line $y = x$. If the constraint is strong (the axis $y = x$ is very near the main dimension of the distribution), the density of cases near the axis $y = x$ becomes really anomalous. Figure 32 on page 89 is an example.

This tool also shows that the described class of constraints is not the only external force distorting the distribution of ratios. Constraints imposed by managerial practice — for example, intended to avoid liquidity problems — are also visible. An example is the ratio CA/CL: The line $y = x$ marks the frontier between positive and negative Working Capital. Such a line also determines a gradient in the density of cases.

Non-linear relations between $\log x$ and $\log y$, revealing non-negligible base-lines, also become apparent with simple scatter-plots of this kind. Ratios like $EBIT/S$ (as those displayed on page 85) or CA/FA exhibit, for a few industrial groups, traces of non-linearity consistent with this hypothesis.

Information content: The amount of information conveyed by one of such tools contains and amplifies the information conveyed by ratios. The horizontal (or vertical) distances from any case to the line $\log y - \overline{\log y} = \log x - \overline{\log x}$, which is the axis with largest variability, measures the ratio residual. As usual, $\overline{\log x}$ stands for the log median (the mean of $\log x$). For example, the scatter-plot formed with $\log CA$ in the abscissa and $\log CL$ in the ordinate, yields, for any point $\{CA_j, CL_j\}$ representing the position of firm j, a measure of $\varepsilon_j^{CA/CL} = (\log CA_j - \overline{\log CA}) - (\log CL_j - \overline{\log CL})$ which is the ratio residual in log space. However, the way this measure would have to be carried out graphically cannot be considered as practical. Later on, we shall improve it by using rotated axis.

The mean-adjustment: A first step towards more practical tools is the mean-adjustment of the data. Financial diagnosis is based on the magnitude of deviations from standards. The value of the standard itself is important only in that it allows the calculation of such deviations. Therefore, the mean-adjustment throws away a non-important piece of information. With this, the clarity of the representations improves and the comparing of samples becomes feasible.

Controlling for joint trends: For example, mean-adjusted data is useful when it is convenient to gather in the same scatter-plot data belonging to several years. In this case we would mean-adjust separately each year. Trends like the evolution of the economy would be accounted for in this way.

In general, position measures (introduced on page 59) are often preferable to log items. Residuals or contrasts are preferable to ratio outputs. Their expected value is the same — zero in log space and one in the ordinary one — instead of varying from one sample to the other.

The residual plot: If we use, instead of mean-adjusted items, the residuals obtained after controlling for s, the common effect, we get a scatter-plot of ε^y with ε^x. The reliability of this tool depends on the quality of the proxy for size. Of course, any item could be used for deflating x and y instead of s. In that case we would have a plot relating two contrasts.

The residual plot is adequate for detecting correlations between residuals. Since the strong, common, effect has been accounted for, any residual relation becomes visible. Simple plots of log items are not accurate in detecting residual relations since the common effect, having a much larger variability, completely masks them. Figure 59 compares a mean-adjusted plot (left) with a residual one for the same data (right). Notice that in the tool displayed on the right the 135^o axis (the line

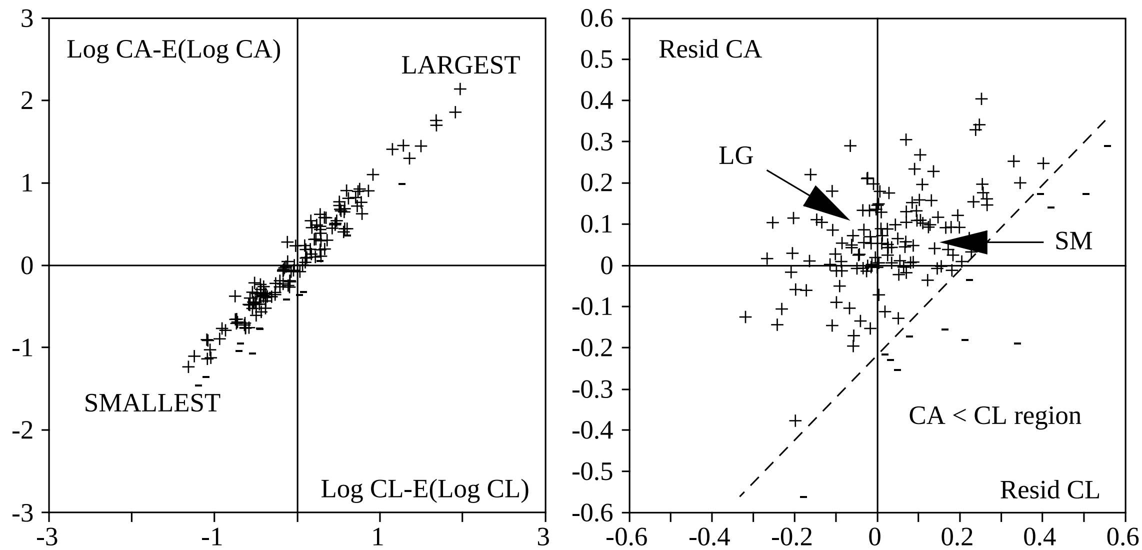

Figure 59: On the left, a mean-adjusted scatter-plot of Current Assets (Y-axis) versus Current Liabilities (X-axis). On the right, the corresponding residual plot. Electronics, 1986. The negative sign means $CA < CL$.

$y = -x$) can be used for measuring the contrast (ratio residual) $\varepsilon^{y/x}$. Notice also that residual plots and rotated residual plots are not able to display relative size. Like ratios, they control for size.

Figure 64 on page 195 displays another residual plot showing a clear case of correlation between two items after controlling for size. The items used in these plots (Wages and the number of employees) are also interesting because of their ability to separate industrial groups.

Residual plots are considered here as a step towards more elaborate tools, the rotated ones.

8.1.2 Analyzing the Information Contained in Ratio Components

One feature common to the plots presented so far is that any measure intended to mimic ratios ought to be done using the 45^o or 135^o axis instead of the natural ones. We avoid this inconvenience with a simple rotation. In doing so, we make less intuitive the detection of asymmetry and non-linearity but we facilitate the financial diagnosis.

This 45^o rotation of axis produces two orthogonal views of the main dimensions of the distribution of the ratio components in log space. Therefore, the whole of the information conveyed by the ratio components is now present in these new variables in a complementary way.

We shall see that such a pair of new variables can answer two kinds of questions that specific pairs of ratios seem meant to answer as well. From here we conclude that those pairs of financial ratios convey complementary pieces of information. They are two aspects of the same two-dimensional measurement.

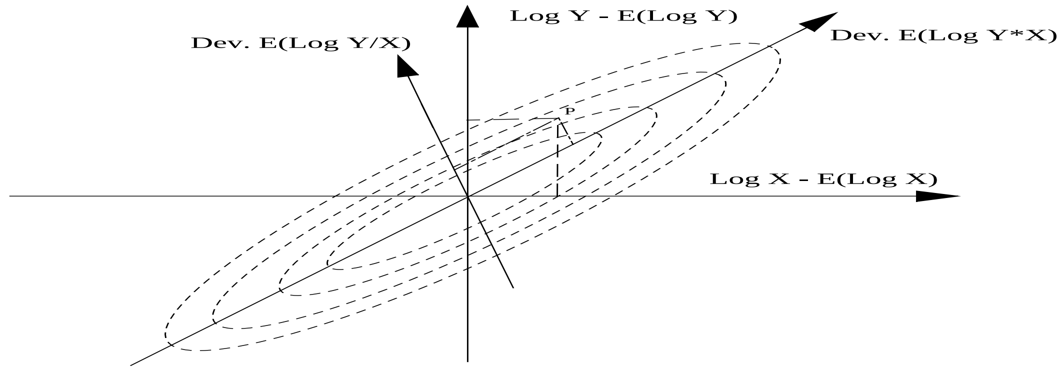

Figure 60: A graphical representation of the rotation leading to financial interpretability.

The difference between controlling for size and assessing financial features: It is often mentioned in the literature that ratios are used because of the need for controlling for size. In a cross-sectional context there are two meanings for this controlling for size, both interesting in financial analysis. Ratios seem to be called upon in order to answer two different kinds of questions, not just one.

Firstly, size can be viewed as a variable, s, reflecting the common variability as a statistical effect. And secondly, size can also mean the magnitude of one item when compared with the magnitude of another one. Therefore, two different questions can be asked when referring to size.

- What is the position of a particular item when contrasted with s, the size of the firm? This is the problem of assessing deviations from standards for size. Financial ratios are set to answer this question when the deflator is selected so as to reflect size. Sales, Net Worth or Total Assets have often been the choices for such a size proxy. In our framework the answer to the above question is given by the f^x or, in log space, the ε^x.

- To what extent a given feature of the firm, like liquidity, is far away from the expected as a feature, that is, regardless of the magnitude of its components when compared with the size of the firm? This is the problem of measuring departures from standards describing features by themselves. In such cases the deflator is selected so as to produce a contrast when compared with the deflated item. Such a contrast is in our notation the ratio residual $f^{y/x}$ or $\varepsilon^{y/x}$.

Some ratios seem more intended to answer the first question whilst others are intended to answering the second. For example, in the two liquidity measures Working Capital to Total Assets and Current Assets to Current Liabilities we notice that the first one seems to assess liquidity by referring it to the size of the firm, whilst the second one assesses liquidity by itself — the feature emerging when contrasting short term assets with liabilities — regardless of the size of the firm. Also the Debt Ratio (DEBT/TA) yields a contrast with a typical measure of size. It tries to answer the first question. But the Debt to Equity Ratio (DEBT/NW) answers the second one: It seeks a contrast of debt with equity, not with the size of the firm.

The two dimensions of the size-adjusted information: It is worth emphasizing that this list of two questions is a consequence of the assumptions and models developed in chapters 3 and 5. Given that it is possible to assess size, what kind of size-adjusted information the components of ratios can yield?

They firstly yield two residuals, ε^y and ε^x. Then, a 45^o rotation combines these residuals so as to produce two new variables which are also orthogonal — thus capturing two complementary aspects of the information contained in the original residuals.

Such two aspects are the $\varepsilon^y - \varepsilon^x$ and the $\varepsilon^y + \varepsilon^x$. The first one contains the same information accountants are used to find in ratios, but mean-adjusted. It is a contrast between two items. When the y and x are conveniently selected, such contrasts are supposed to capture desired features.

The second one contains the information originally conveyed by x and y but not captured by the ratio. Since the pair $\{x, y\}$ conveys two-dimensional information and ratios or contrasts are just one variable, when we use ratios instead of their components we put aside information. Not only the one about size. We put aside also size-adjusted information, potentially interesting on financial grounds. Such piece of information is the $\varepsilon^y + \varepsilon^x$.

It is easy to see that the two questions referred to above are answered by these two complementary variables. Therefore, such two questions should be regarded as complementary as well. And they represent the whole of the questions a pair of items can answer after being size-adjusted.

Is the second aspect important? Discussion. In his response to Barnes, Horrigan claims that the main task ratios undertake is the assessment of specified relationships. "They adjust for the data size effect only incidentally. (...) Size deflation is certainly an interesting property of financial ratios, but it is hardly their major purpose." [64]. Horrigan considers that only the $\varepsilon^y - \varepsilon^x$ are interesting. In his opinion, contrasts are the only piece of information worth considering. This is equivalent to say that the information conveyed by the $\varepsilon^y + \varepsilon^x$ is not worth taking into account.

On statistical grounds it is very difficult to sustain such a position since the two aspects carry complementary information. On financial grounds, this author would need to explain why the deviations from the expected for firms of a given size are necessarily not interesting for diagnosis. We think that this restrictive view requires some explanation since it is not evident.

In next section we discuss the specific cases in which pairs of ratio components yield eventually less interesting information.

8.1.3 Two Rotated Plots

The two rotated plots we present next are practical applications of the above considerations. The first one, the rotated plot, preserves information regarding the size of the firm. There is a growing conscience about the importance of size — not just deviations from expected size — in some specific problems. The second one controls for size. It yields deviations from expected size, not size itself.

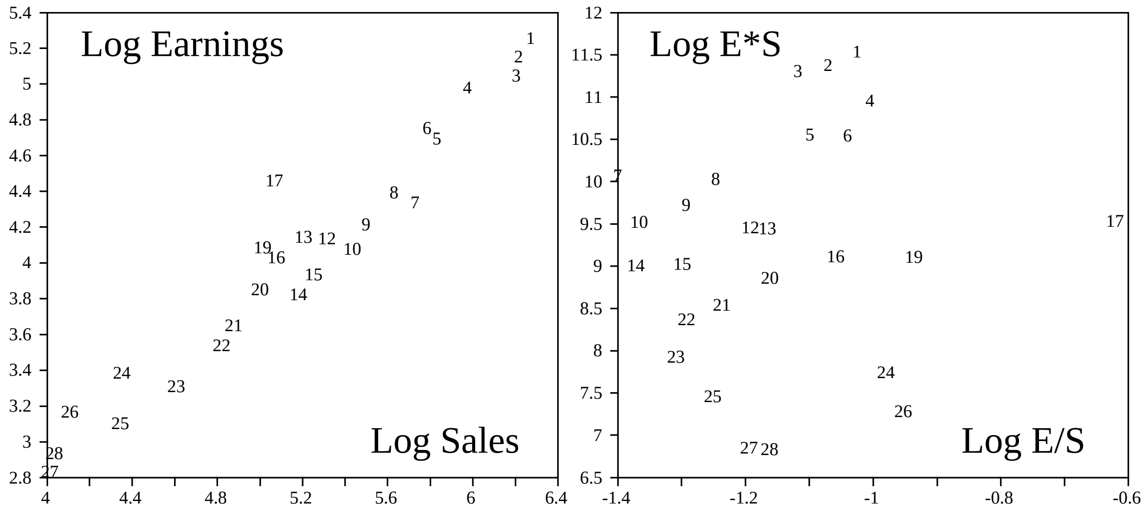

Figure 61: On the left, a scatter-plot of Earnings (Y-axis) versus Sales (X-axis). On the right, the corresponding rotated plot. Motor Components, 1984.

Taking size into account: The rotated plot. For any vector of two mean-adjusted observations, $\{\log u, \log v\}$ there is a 45^o anti-clockwise rotated vector $\{h_1, h_2\}$ obtained by applying the simple transformation

$$H = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

(a two-dimensional Hadamard rotation introduced in chapter 5) to the original log vector.

The resulting variables will be $h_1 = \log u + \log v$ and $h_2 = \log u - \log v$. Therefore, h_1 is now a proxy for the effect common to the original items and h_2 is, in log space, the residual of u/v, that is, the contrast. We can build a scatter-plot in which h_1 is the ordinate and h_2 is the abscissa. The space spanned by $\{h_1, h_2\}$ contains the same information conveyed by the usual scatter-plots but arranged in a way that makes sense in financial analysis.

This rotated plot allows the straightforward measuring of contrasts along the X-axis. It is also more accurate since the cases now span much more uniformly the whole of the neighbourhood about the expected values of their variables. However, for observing possible asymmetry or non-linearity this tool should not be used alone.

Figure 61 shows, on the left, the usual scatter-plot in log space and on the right the same data after rotation. Notice how cases now span a larger portion of space. Notice also how the small non-linearity on the left is now less evident as such.

Rotated plots like the one of figure 61 retain in the Y-axis the information regarding the size of the firm. Therefore, they are ideal for studying features influenced both by one particular contrast and by size. For example, when predicting firm distress, both the ratio Cash-Flow to Total Debt and the size of the firm seem to be revealing. We can put together both pieces of information by means of this plot.

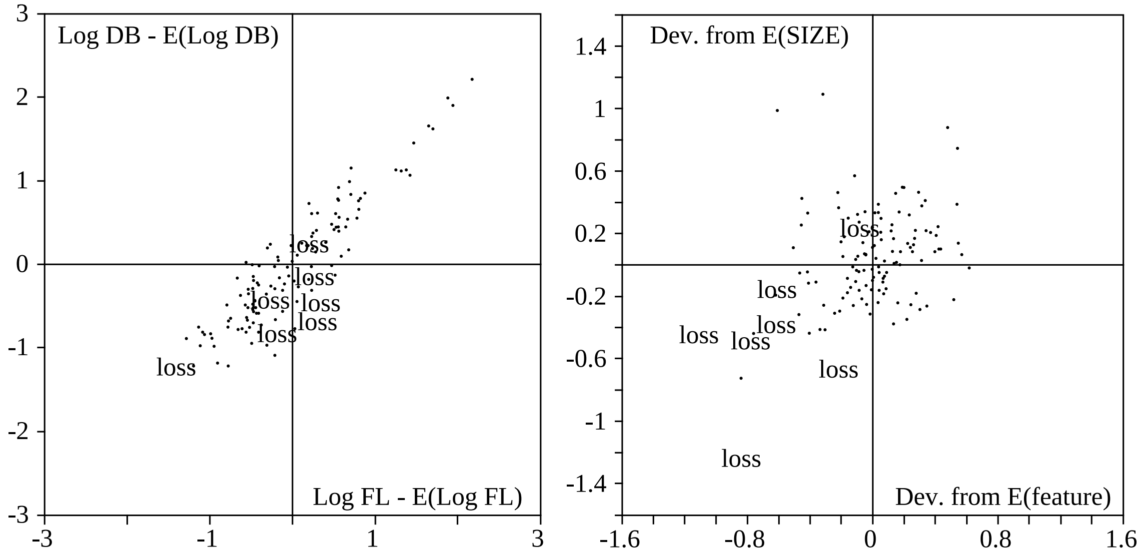

Figure 62: On the left, a mean-adjusted scatter-plot. On the right, the corresponding rotated residual plot. Funds Flow to Total Debt, Electronics, 1986. Firms having negative EBIT in one or more years in the period 1983-87 are highlighted.

The rotated residual plot: In the rotated plot the coordinates of any point, $\{h_1, h_2\}$, are ratio residuals and joint measures of position. For the tracing of features not related to size we now introduce the rotated residual plot. It uses, instead of the joint measure of position, another contrast. For example, we can introduce s, a proxy for size, as a deflator for the Y-axis of a rotated plot. In the ordinate, instead of $u \times v$ in log space, we would now have a ratio: $(u \times v)/s^2$. The abscissa would remain unchanged.

It is such a framework which allows the answering of the two complementary questions described in section 8.1.2. In fact, in this rotated residual plot,

- the X-axis measures $[\log y_j - \overline{\log y}] - [\log x_j - \overline{\log x}]$ which is $\varepsilon^y - \varepsilon^x$, the contrast or ratio residual. Therefore, the first coordinate of any point will answer the second question. For conveniently selected y and x, this axis assesses a financial feature.

- The Y-axis measures $[\log y_j - \overline{\log y}] + [\log x_j - \overline{\log x}] - 2 \times [\log s_j - \overline{\log s}]$ which is $\varepsilon^y + \varepsilon^x$. It answers the first question. It measures the joint departure from the expected for firms with a given size.

In other words, our plot will represent in log space the log residuals of any y/x ratio in the abscissa and the log residuals of a new ratio, $(x \times y)/s^2$ in the ordinate. The two complementary aspects of the size-adjusted information referred to in section 8.1.2 (the $\varepsilon^y + \varepsilon^x$ and the $\varepsilon^y - \varepsilon^x$) are thus presented as the main axis of this plot.

Notice that the rotated residual plot is just a 45^o anti-clockwise rotation (a Hadamard rotation)

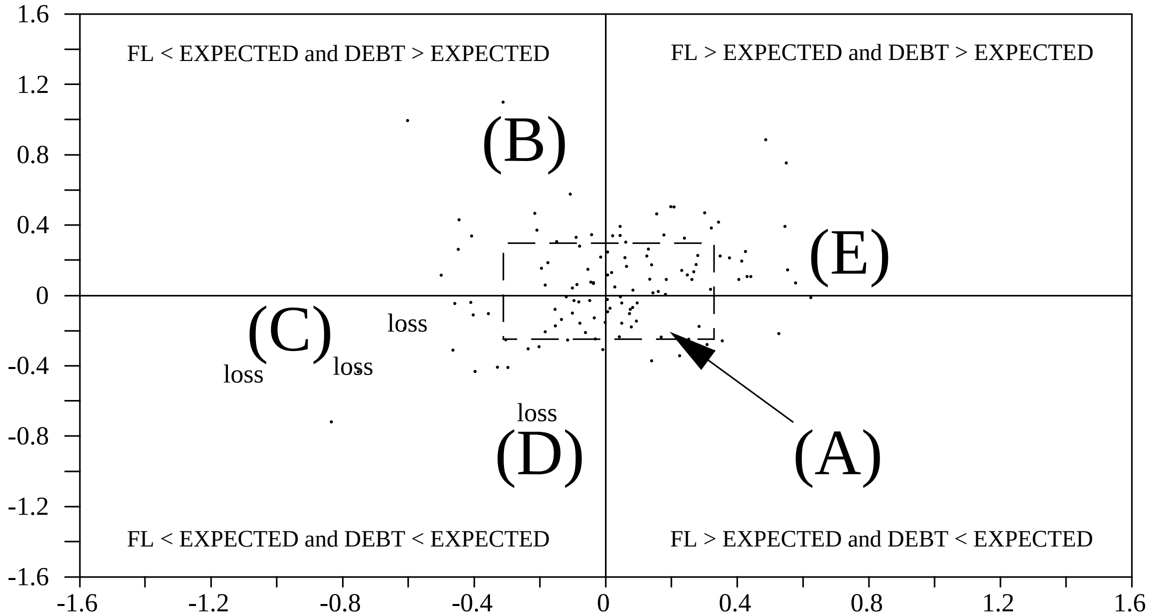

Figure 63: Diagnosis and location of firms in the rotated residual plot. Funds Flow to Total Debt.

of a residual plot in which, as we saw, ε^y is the abscissa and ε^x is the ordinate. Figure 61 on page 191 compares the usual mean-adjusted plot (left) with a rotated residual plot (right) for the same data.

Dimension reduction: The rotated residual plot can also be viewed as a way of reducing from three to two the dimension of the information we are dealing with. In fact, the whole of the potentially interesting information involving any two items y and x is a three-dimensional vector $\{y, x, s\}$. For example, when we are measuring liquidity we could hesitate between using CA/CL, thus getting an absolute (size-independent) liquidity measure, or using WC/TA in order to have insight into the position of liquidity regarding the particular size of the firm. In this case the three-dimensional nature of the desired information is depicted by the fact that only by knowing CA, CL and TA would we be able to answer the two questions.

When using this rotated residual plot we can assess the liquidity, both by itself (X-axis) and as referred to the size of the firm (Y-axis). The dimension reduction has been achieved by jointly measuring the departures from size observed on y and x instead of assessing each one. In many financial applications this reduction makes sense.

Diagnostics as a location in the rotated residual plot: Figure 63 shows the diagnostics we could infer from the location of any firm in the rotated residual plot. As usual, when we use the term "feature" we refer to any financial characteristic of firms able to be reflected in accounting information: Liquidity, profitability, financial structure and so on. Ratios are supposed to assess particular features. In the same way the diagnostics provided by the rotated residual plot for a given feature would be:

Position is A: Both the feature and its magnitude given the size of the firm are near the standards.

Position is B: The feature is near the standards but its magnitude is larger than the expected for the size of the firm.

Position is C: Although the magnitude of the feature is near the expected for the size of the firm, the feature itself is below the standards.

Position is D: The feature itself is near the standards. However, its magnitude for the size of the firm is smaller than the expected. For example, the liquidity of a firm is all right if we consider it as the contrast between short-term assets and liabilities. But both Current Assets and Liabilities are smaller than the usual for firms with the same size.

Position is in between C and D: Both the feature and its magnitude given the size of the firm is below the expected. This is a frequently observed situation with no correspondence in the other quadrant. It means an over-sized firm regarding that feature.

Position is E: The magnitude of the feature is the expected one for firms of such size. But the feature itself is above the standards.

A different way of reading positions in the rotated residual plot refers to the quadrant, not to the Cartesian axis. Since it leads to diagnostics based on log items, not on contrasts, it is not as near the accounting practice as the interpretation based on axis — the one presented above. But we think that the diagnostics based on the quadrant the firm lies in are also simple and revealing.

For the ratio y/x and s, a proxy for size, we would have:

The firm lies in the first quadrant: Both x and y, the ratio components, are above their expected values for firms with that size.

The firm lies in the second quadrant: The denominator of the ratio, x, is below the expected and the numerator, y, is above for firms of similar size.

The firm lies in the third quadrant: Both components are below the expected for a firm of that size. The firm is therefore oversized in what concerns those two items.

The firm lies in the fourth quadrant: The numerator, y, is above the expected for firms with that size. The denominator, x, is over-sized.

In the next section we give extensive examples of the use of the rotated residual plot for diagnosis.

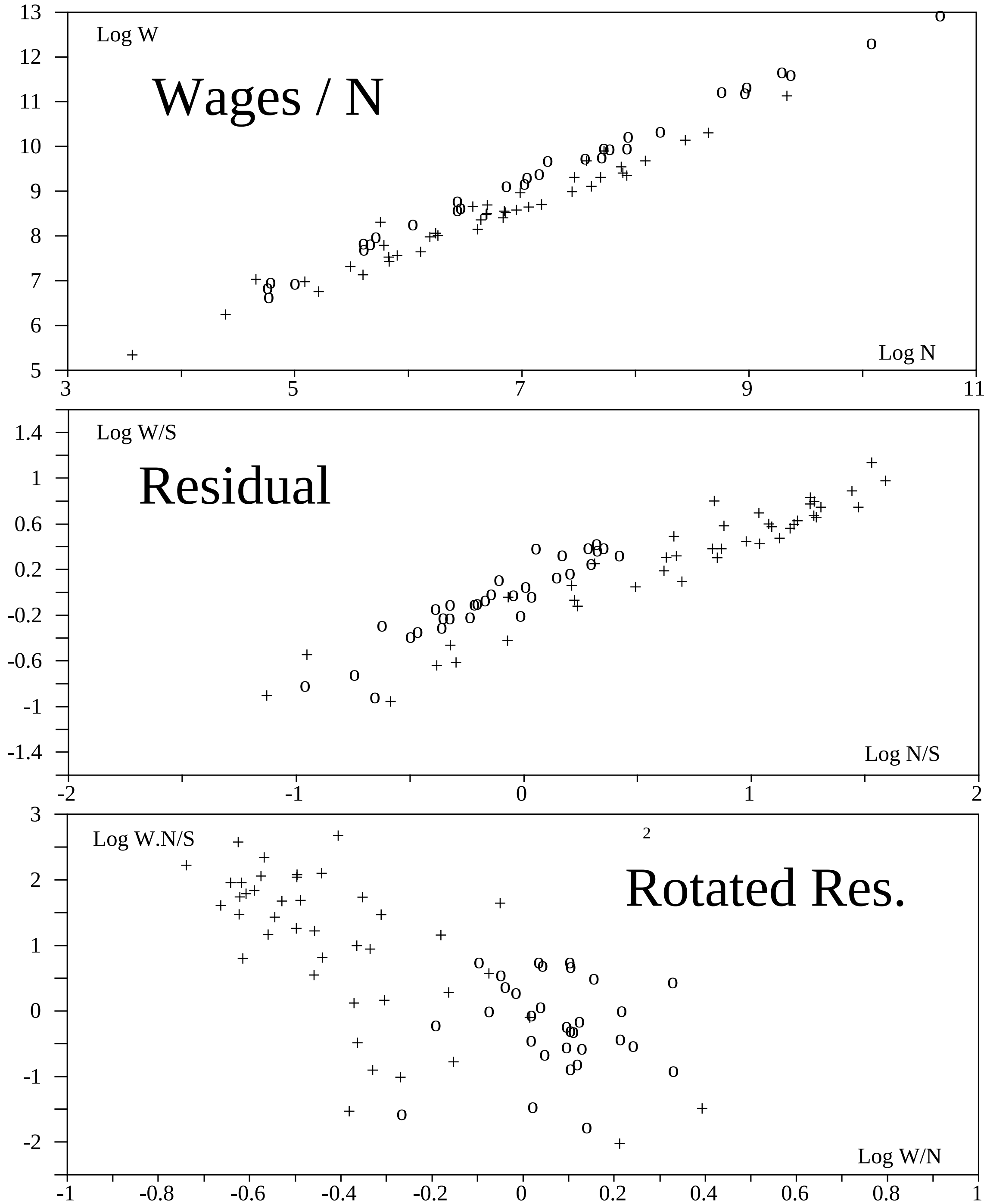

Figure 64: Comparing the simple X-Y plot in log space with a residual plot and a rotated residual plot. Building Materials (o) and Clothing (+) in the same sample (1984). The items are Wages and Number of Employees.

Applicability of the rotated plots: Discussion. Rotated residual plots will not be effective when the two main variables are related by an accounting identity. Probability-like ratios — FA/TA and similar (see chapter 4) — contain all the information the relation between its components can yield. This is because the numerator is part of the denominator. These ratios are real proportions between one part and a total. Not much improvement results from exploring the information conveyed by the two components instead of using the ratio.

Also, this plot will not show all its power when the items selected along with $\log s$ are proxies for size or taken as such.

Finally, correlated residuals make the decomposition of information described in section 8.1.2 less attractive. Correlation means redundancy. Figure 64 (below) shows the aspect of correlated residuals in the rotated residual plot. Clearly, one of the items describes the other one.

On the contrary, ratios formed with items which are not bounded by one another and contain, each one, a really original piece of information, will fully take advantage of the rotated residual plot.

In such cases, the positions and trajectories drawn in the rotated plots can be more revealing than the examination of the two ratios underlying it. Firstly, because two dimensions allow an increasing in specificity of diagnostics. Trajectories, recognized as such, are more accurate and easy to interpret than the two trends underlying them. Secondly, because in some cases the information conveyed by rotated plots is, not only more accurate and easy to interpret, but also unique. This happens whenever the scatter of cases draw, in two-dimensions, a shape impossible to reduce to a simple analytic form described by two observations.

For example, if the scatter of cases is less dense in one quadrant than in the others or if there is a comet-like shape (a two-variate tail) the two-dimensional information cannot be reduced to a pair of observations functionally linked.

8.2 The Mapping of Features Using Neural Networks

In this section we use the rotated plot or the residual one to trace trajectories of firms during a period of five years. We further explain how this plot allows a scanning procedure for the automation of financial diagnosis.

8.2.1 Modelling the Density Function

The spread of a set of stochastic variables defines a density function in the space they span. In two dimensions, the density function would be the number of cases per square unit. Two-variate Gaussian phenomena would exhibit a hill-shaped surface as its density function.

Statistical modelling techniques can be described as attempts to account for as many variability as possible using as few descriptors as possible. For the usual applications the goal is estimation.

This implies the finding of the best values for the parameters of some analytic formulation selected a-priori. Such a formulation is supposed to govern a process we want to describe.

However, in other cases we wish to model the density function itself. That is, we want to build a model — a representation as simple as possible — of the way cases spread but retaining the information regarding where they lie. These models are known as maps.

Kohonen's Self-Organized Maps are intended to undertake such a task. They output a collection of positions which reproduce the original density of cases but using a smaller number of points instead. Each one of the original cases relates to one of the new positions. The result is similar to the division of cases into classes, but for more than one dimension.

After organizing cases in classes we obtain a model: A collection of values and the corresponding frequencies. This model accounts for the density function observed in our data. The fact that it is not the usual kind of model — analytic and oriented for estimation — doesn't diminish its basic quality of model. It is a simpler representation of the data.

Kohonen's or similar maps are also models. In some cases they become the only possible model since maps are the only way of accounting for spreads which are not simple enough for being approached by analytic tools. Therefore, the reasons for using maps can be twofold:

- The information regarding the position of each case is relevant.
- The spread of the data draws a shape which is not geometrically simple.

Both reasons lead to the use of these tools as a way of modelling the rotated plots devised above. As we stressed in the last section, each zone of a rotated residual plot is assigned a financial diagnostic. It is the fact that a case lies in a particular zone that is important here. Also, rotated residual plots and other similar tools often exhibit irregular shapes. For example, in rotated residual plots reflecting profitability or flow of funds, it is frequent to observe a comet-like shape, with a high density around the origin and the surroundings almost empty except for the third quadrant. Such a shape would be difficult to account for with analytic approaches.

We shall use an abbreviation, RRP, to designate the rotated residual plot. We selected a particular industry, the Food Manufacturers and the profitability ratio to illustrate the use of maps in RRPs and the possibilities they offer in the tracing of dynamic features. On the next pages we describe, step by step, the procedure leading to automatic diagnosis.

8.2.2 Building the Rotated Residual Plot

For each year of the period 1983-1987 separately, we select two different samples. One contains the cases with positive $EBIT$. The other one, those having negative $EBIT$. Then, a symmetric log transform is applied (formula 2 on page 54). The items to be used, NW and $EBIT$, are mean-adjusted year by year. Any joint trend related to an annual effect is thus accounted for. Also the proxy for size to be used, $\log s$, is mean-adjusted separately by year.

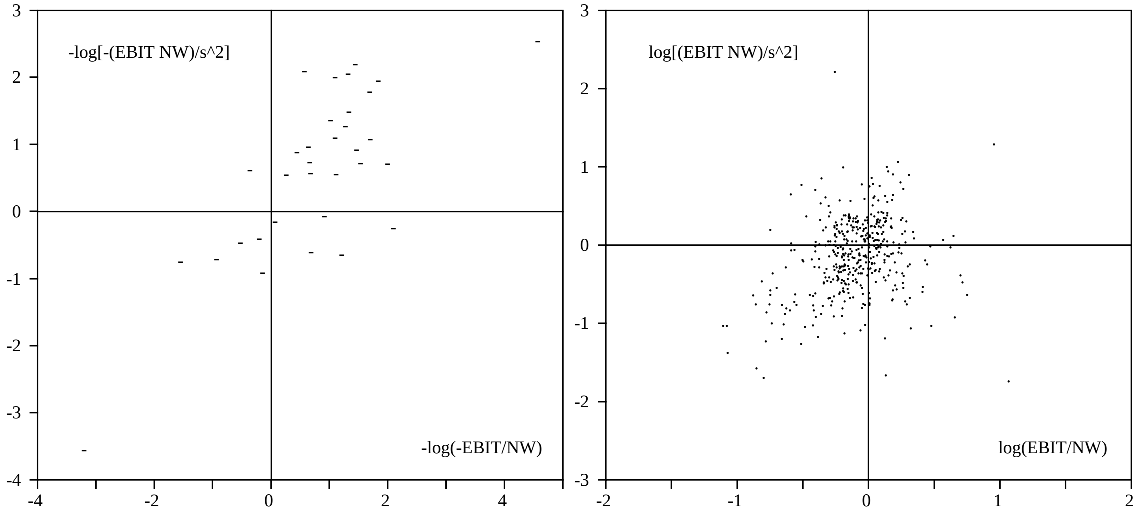

Figure 65: Two rotated residual plots for assessing the liquidity of firms in the Food Manufacturers industry, 1983 to 1987. On the left, cases with negative $EBIT$.

Next we find the Y-axis and the X-axis of the RRP, in the same way for the positive and negative $EBIT$ samples. For log residuals defined as

$$\varepsilon^{EBIT} = \log EBIT - \overline{\log EBIT} - (\log s - \overline{\log s})$$

and

$$\varepsilon^{NW} = \log NW - \overline{\log NW} - (\log s - \overline{\log s})$$

we obtain the two axis

$$y = \varepsilon^{EBIT} + \varepsilon^{NW} \qquad \text{and} \qquad x = \varepsilon^{EBIT} - \varepsilon^{NW}$$

The final aspect of the two RRP is displayed in figure 65 (page 198).

As discussed in section 2.3, these scatters represent two different things and should be modelled separately. On the other hand, when studying dynamic features, it is convenient to introduce some form of continuity between one model and the other one thus allowing the tracing of firms which $EBIT$ emerges from negative to positive values and conversely.

Drawing the continuity between profits and losses: This task is facilitated by the fact that, in the RRP of positive values, the third quadrant seems to be logically linked with the first quadrant of the RRP for the negative ones.

In figure 63 (page 193), the third quadrant is the region delimited by "C" and "D". It contains the firms below the expected both in profitability and in the magnitude of this feature regarding the size of the firm. Similarly, in the samples showing negative EBIT, the first quadrant — the region between "E" and "B" in the same figure — contains firms with less severe losses and which profitability is larger than the expected for negative-EBIT firms of that size.

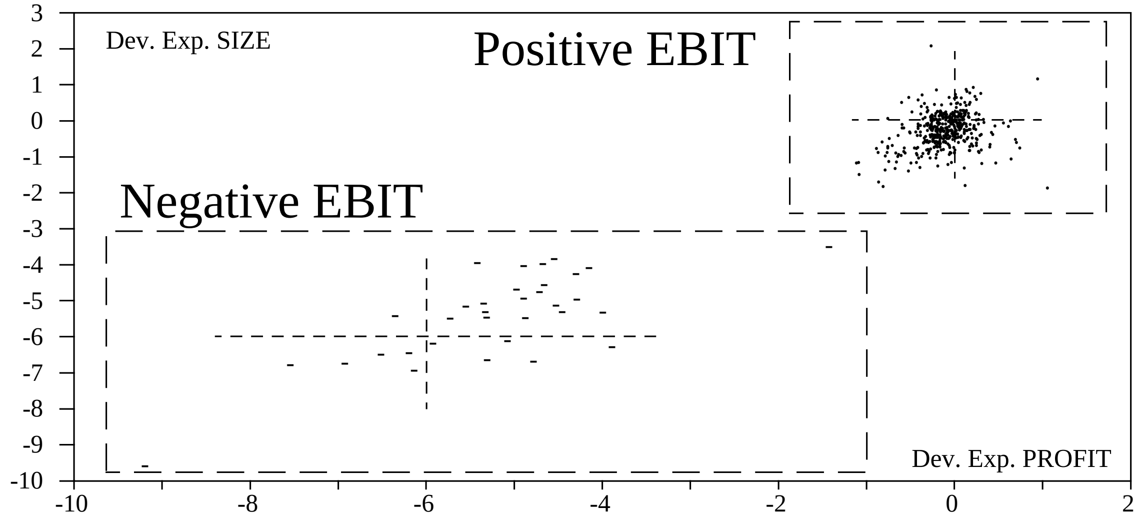

Figure 66: The joint RRP formed with the positive cases and the negative ones shifted so that they occupy a region of the third quadrant far away from the rest of the data.

Let us suppose that a firm gradually falls into negative earnings. It will draw a trajectory towards the second or the third quadrant and then into the first or fourth quadrant of the negative plot. But the path through the third quadrant seems the most logical for leading to losses since it means a firm too large for what it is worth and also for the generated profits. The path through the second quadrant would mean a firm too large for the generated profits as well, but with a balanced capital regarding its size.

Conversely, when a firm gradually emerges from a situation of poor profitability to a more healthy state its path will go along the first or the fourth quadrant of the negative plot until it reaches the third or the second one in the positive plot. But the path through the first quadrant is the logical one for getting out of a situation of loosing money since it means an improvement in both Earnings and Net Worth regarding size.

The fact that firms often fall into negative earnings from quadrant other than the third one doesn't invalidate our reasoning. The logical path linking these two situations seems to be the described one. It links the worst situation of positive earnings with the best one of negative ones.

We also observed that the position of firms having at least one year with losses during the usual period tends to be in the third quadrant. Given this, it seems as if the continuity between profits and losses should be drawn between the third quadrant of the positive sample and the first one of the negative sample.

Accordingly, we place the negative plot in the third quadrant of the positive one, far away from the rest of the data. This is done by shifting all the cases in the former by a negative amount and then mixing both plots. The result is the joint RRP displayed in figure 66. Notice how the spread of the negative cases is much larger than the spread of the positive ones.

It is now possible to set the mapping algorithm to learn the joint density of cases as it appears in figure 66. But before doing so it is convenient to gain insight in the way the RRP conveys known accounting information regarding the evolution of a firm's profitability.

8.2.3 Using the Rotated Residual Plot

In this section we compare the information conveyed by a set of ratios with the one of the RRP. The goal is to provide a means to get acquainted with this new tool by putting it side by side with the usual source of information for analysts, the ratios.

We selected nine firms from the Food Manufacturing industry. For each one of them we display the ratios Sales to Net Worth, Funds Flow to Total Debt, EBIT to Net Worth, the Current Ratio and also $\log s$, our proxy for size. The evolution of these ratios during the considered period of five years is displayed as a time-history.

For each of these nine firms we also present the trajectories drawn in the RRP described above. The whole of the information is contained in figures 67 (page 201), figures 68 (page 203) and figures 69 (page 205). This graphical description is complemented with table 33 on page 206.

How to read the displayed material: The figures mentioned above compare ordinary ratios with RRPs. For each one of the considered firms they display, on the left, a time-history of some usual ratios. On the right, the RRP as described above. Each mark on the RRP represents the position of the firm for one year. For example, "4" shows the position of the firm in 1984. In order to interpret the RRPs, firstly notice that the nearest a firm is from the centre, the less it diverges from the standards for the industry. Secondly, the scales assessing departures from standards are visible at the edges of the plots. Since we are working in log space, these scales are relative.

In this particular case, the X-axis of the RRP measures the relative deviations of liquidity from the expected. The Y-axis measures how much the magnitudes of Earnings and Net Worth taken jointly diverge from the expected for the size of the firm.

It is also possible to interpret the RRPs in terms of the quadrant a case lies. The first one means both earnings and capital larger than the expected. The second one means capital larger than the expected but earnings smaller. And so on.

We shall now comment on each one of the presented cases.

Firms occupying a steady position in the RRP: The first two firms on the top of figure 67 are an example of a steady position in the RRP.

When reading the information conveyed by ratios about the evolution of UNITED BISCUITS, a large firm, we notice that during this period their sales suffered a small decrease and their profitability was steady. The reading of the RRP says that both the position of this firm in what concerns profitability and the proportion of this feature regarding the size of the firm are the expected ones

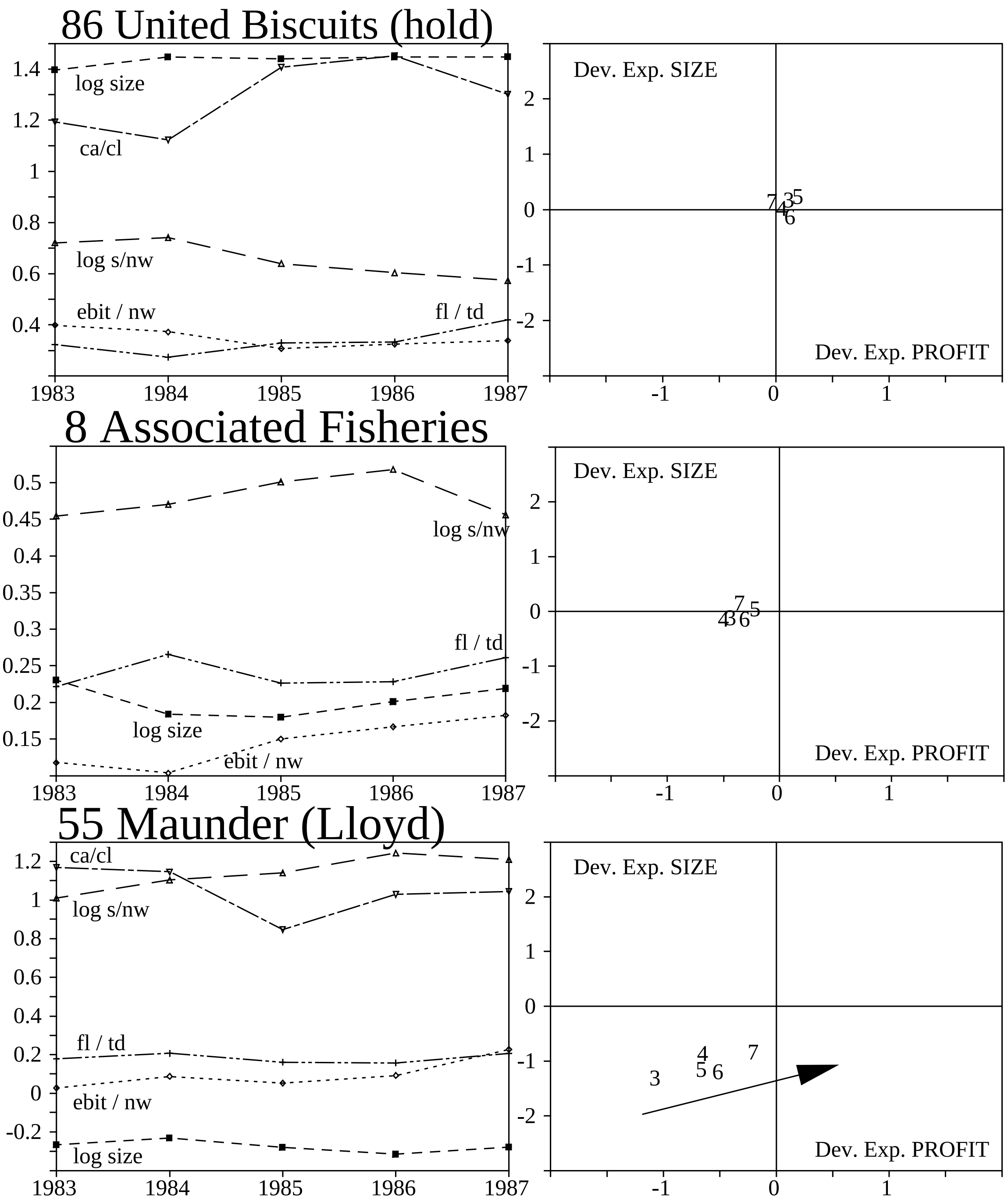

Figure 67: On the left, the evolution of a few ratios and a proxy for size during a period of five years. On the right, the corresponding rotated residual plot — EBIT to Net Worth — showing the trajectory drawn by the firm during the same period. Marks 1 to 7 indicate positions from 1983 to 1987.

for the industry. Also, they didn't change during the whole period. Both the profitability as a contrast and the magnitude of earnings or Net Worth when compared with the expected for firms with the same size stay near the standard for the industry for the whole period.

ASSOCIATED FISHERIES is another example of a steady position. Its profitability is half of a point below the expected but it agrees with the expected given the size of the firm. The information provided by ratios says that sales increased until 1986 and then broke down to the values of 1983. Profitability was steadily increasing during the whole period while the size of the firm seems to decrease — or, more likely, it is not in step with the growth of the industry.

Clear trends in the RRP: MAUNDER (LLOYD) in figure 67 and NESTLE (UK) on top of figure 68 show a clear trend towards a better performance during the five years considered. The first one is a small firm. It recovers from a dangerous position of profitability one point below the expected and over-sized regarding profits to a comfortable new one agreeing with the standard for the industry. The second firm is larger than the expected for the industry. It improved its profitability from a standard state to almost one point above such standard. This was achieved without disturbing the magnitude of this feature when compared with the size of the firm. Such proportion was kept within the standard for the industry.

More complicated trends: CAMPBELL FROZEN FOODS (figure 68, page 203) is an example of a more complicated evolution. The Sales to Net Worth ratio broke a bit after 1985 and the final picture, in 1987, is the one of a not very profitable firm. The RRP shows an increase in the importance of profitability inside the firm, followed by a severe break of one point in the profitability as a feature.

The whole of the trajectory lies in the upper two quadrant which means an excess of Net Worth when compared with the standard for the industry. The second quadrant explicitly means that such an excessive capital is not actually producing the expected profits.

OVERSEAS FARMERS is a typical case of increasingly poor profitability. Since the volume of sales didn't break down in the last two years of the period, we must conclude that other factors are affecting the performance of this firm. The RRP shows, in these two years, a sliding of almost one point in profitability and of almost two points in the importance of this feature inside the firm. This firm is too large for what it is worth and also for the generated profits.

Cases of negative EBIT: Figure 69 (page 205) focus on three firms having negative EBIT in at least one year of the considered period. Above, a medium-sized firm, BOWYERS, shows a severe break in earnings in 1985 followed by an immediate recovery. FOODANE, the next firm to be displayed, jumps between positive and negative earnings during the period. Finally, G. P. LOVELL, a small firm, shows an excursion into profitability in 1985 and 1986.

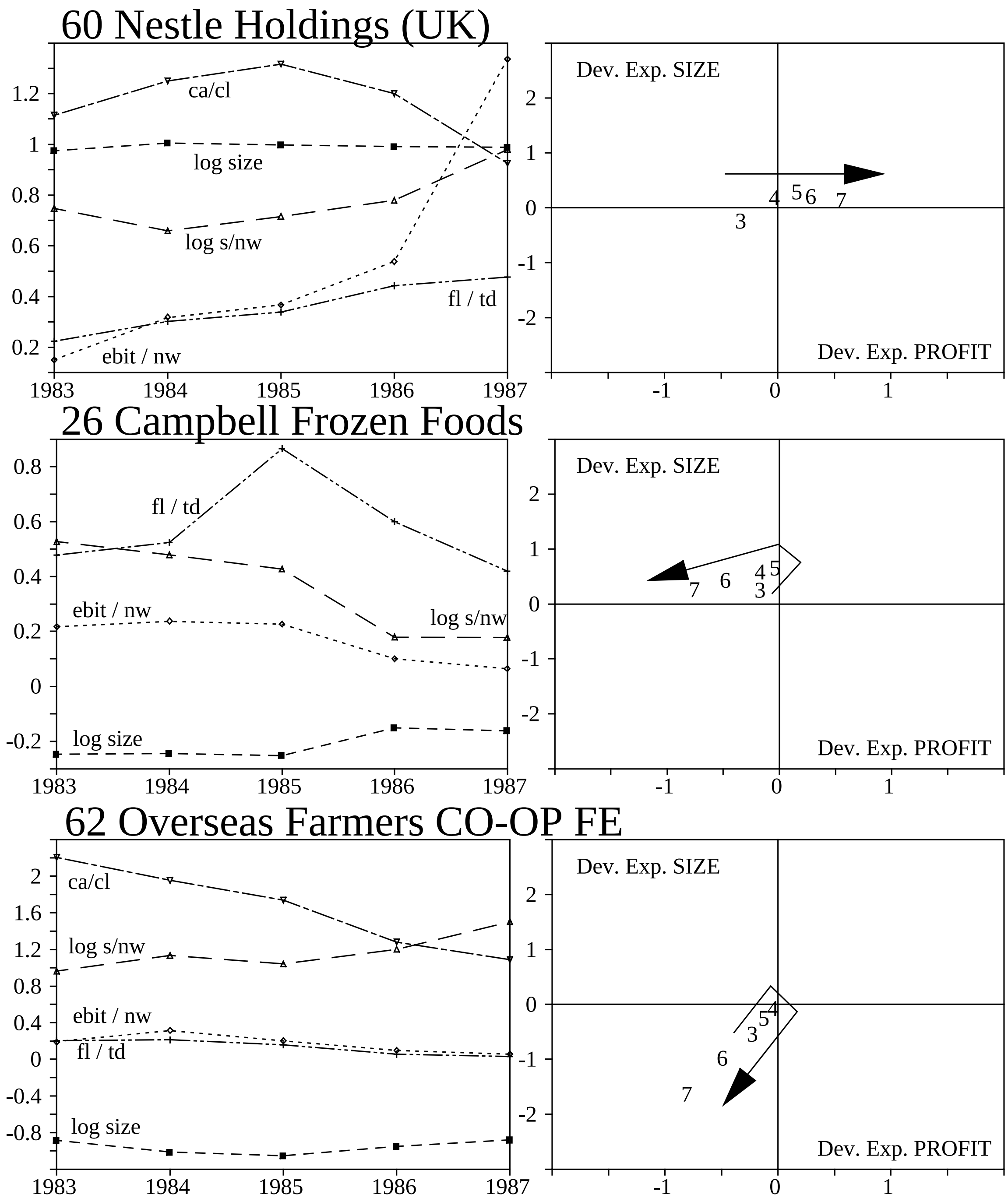

Figure 68: On the left, the evolution of a few ratios and a proxy for size during a period of five years. On the right, the corresponding rotated residual plot — EBIT to Net Worth — showing the trajectory drawn by the firm during the same period. Marks 1 to 7 indicate positions from 1983 to 1987.

When compared with the standards for the industry the above firms are all over-sized for what they are worth and for the profits they generate. As remarked, the third quadrant in the RRP seems to be a dangerous zone when both ratio components are related to positive aspects of the firm.

Discussion and conclusions: We expect to have shown that this simple tool, the RRP, is able to convey interesting information regarding the feature being analyzed. It shows itself as a different, yet familiar, way of reading accounts.

It is different from simple ratios in that it conveys more than one piece of information at a time. But it is based on the same principles: The contrasts between two magnitudes are able to capture features and the value expected for the industry sets the standard of normality.

In this particular application we focus on liquidity. The two axis were useful in characterizing the behaviour of firms and, in some cases, in drawing meaningful trajectories. Such trajectories are unique to the RRP. They reveal a certain behaviour valuable for financial analysis and less explicit when using ratios solely.

One very particular characteristic of the RRP is that it emphasizes the difference between positive and negative earnings. When we study the cases displayed in figure 69 by means of the usual ratios we notice that a jump into negative EBIT can appear just as a small or even very small break in the value of the ratio. But when using the RRP for examining the same firms this break is outlined. This is caused by the lack of continuity between the two plots forming the RRP.

Such a feature of the RRP is, in our opinion, desirable. It is good to stress the difference between two so different states as those of having profits or losses, even when the profits were small and the losses were small too.

8.2.4 Automating Financial Diagnosis

We now return to the mapping process. By exposing the spread of cases displayed in figure 66 to an appropriate mapping algorithm we obtain a smaller number of points in $\mathbb{R}^2$ reflecting its density. When the used algorithm is a self-organized mapping, each coordinate of this reduced set of points is the value a given weight has learned. And whenever a case lies in the neighbourhood of such points, the node or neuron to which these weights are linked will fire or get "excited".

Figure 70 on page 207 shows the positions, after the learning process has finished, of the reduced set of points or neurons superimposed to the scatter used to train them. These positions are only approximations of the real ones. Straight lines link nodes that are neighbours.

The number of neurons and its form was determined prior to the learning. We decided that the final map would be a rectangle with three rows of nine neurons each. Hence, it would be prone for adopting an oblong shape like the one it was to be set to learn. We refer to any position within this topography by saying that the index $i, i = 1, 9$ is a counter of the rectangle's row number and the index $j, j = 1, 3$ is a counter of its column number. Any neuron will be determined by a pair $\{i, j\}$.

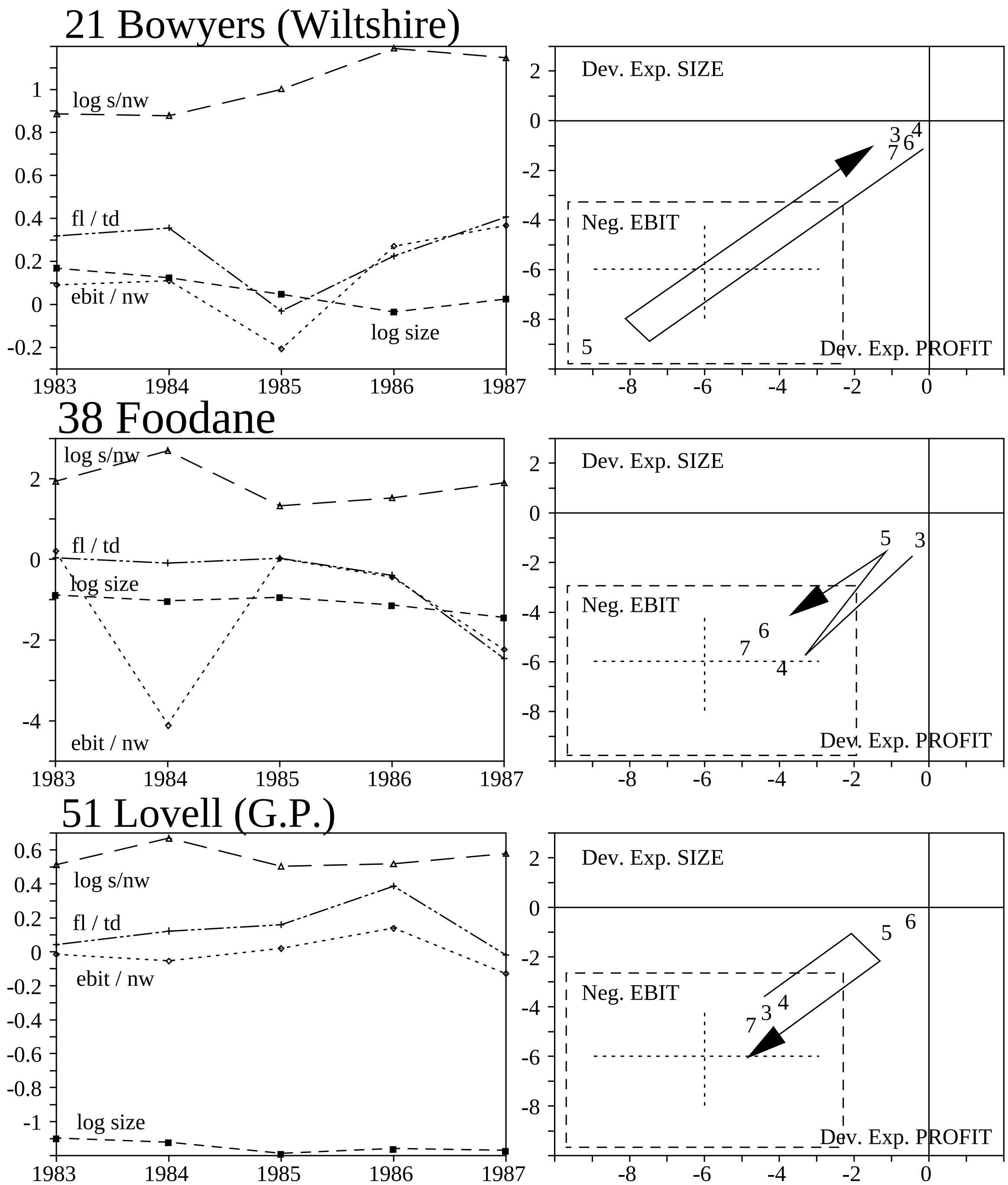

Figure 69: On the left, the evolution of a few ratios and a proxy for size during a period of five years. On the right, the corresponding rotated residual plot — EBIT to Net Worth — showing the trajectory drawn by the firm during the same period. Marks 1 to 7 indicate positions from 1983 to 1987.

Company	1983	1984	1985	1986	1987
	Mean-adjusted log size				
8 ASSOCIATED FISHERIES PLC	0.230	0.184	0.179	0.201	0.219
21 BOWYERS (WILTSHIRE) LTD	0.170	0.124	0.048	-0.03	0.025
26 CAMPBELL FROZEN FOODS LTD	-0.24	-0.24	-0.25	-0.15	-0.16
38 FOODANE LTD	-0.87	-1.02	-0.92	-1.13	-1.43
51 LOVELL (G.F.) PLC	-1.09	-1.11	-1.18	-1.15	-1.16
55 MAUNDER (LLOYD) LTD	-0.26	-0.23	-0.27	-0.31	-0.27
60 NESTLE HOLDINGS (U.K.) PLC	0.975	1.005	0.998	0.991	0.988
62 OVERSEAS FARMERS'CO-OP FE	-0.88	-1.01	-1.05	-0.94	-0.87
86 UNITED BISCUITS (HOLDINGS)	1.397	1.447	1.439	1.447	1.448
	Funds Flow to Total Debt				
8 ASSOCIATED FISHERIES PLC	0.221	0.265	0.226	0.228	0.261
21 BOWYERS (WILTSHIRE) LTD	0.318	0.357	-0.02	0.224	0.407
26 CAMPBELL FROZEN FOODS LTD	0.478	0.524	0.866	0.599	0.419
38 FOODANE LTD	0.047	-0.08	0.037	-0.38	-2.45
51 LOVELL (G.F.) PLC	0.043	0.122	0.159	0.387	-0.01
55 MAUNDER (LLOYD) LTD	0.179	0.207	0.161	0.157	0.206
60 NESTLE HOLDINGS (U.K.) PLC	0.223	0.301	0.338	0.442	0.477
62 OVERSEAS FARMERS'CO-OP FE	0.204	0.213	0.158	0.057	0.033
86 UNITED BISCUITS (HOLDINGS)	0.323	0.274	0.329	0.332	0.420
	EBIT to Net Worth				
8 ASSOCIATED FISHERIES PLC	0.117	0.104	0.150	0.166	0.182
21 BOWYERS (WILTSHIRE) LTD	0.089	0.128	-0.20	0.271	0.368
26 CAMPBELL FROZEN FOODS LTD	0.218	0.232	0.227	0.100	0.064
38 FOODANE LTD	0.208	-4.11	0.019	-0.43	-2.22
51 LOVELL (G.F.) PLC	-0.01	-0.03	0.020	0.138	-0.12
55 MAUNDER (LLOYD) LTD	0.026	0.073	0.053	0.091	0.227
60 NESTLE HOLDINGS (U.K.) PLC	0.152	0.310	0.366	0.538	1.336
62 OVERSEAS FARMERS'CO-OP FE	0.192	0.299	0.203	0.097	0.057
86 UNITED BISCUITS (HOLDINGS)	0.394	0.372	0.307	0.325	0.338
	Sales to Net Worth				
8 ASSOCIATED FISHERIES PLC	2.85	2.95	3.17	3.30	2.86
21 BOWYERS (WILTSHIRE) LTD	7.70	7.55	10.04	15.54	14.04
26 CAMPBELL FROZEN FOODS LTD	3.37	3.01	2.67	1.51	1.51
38 FOODANE LTD	87.30	502.00	21.34	34.08	79.65
51 LOVELL (G.F.) PLC	3.26	4.67	3.19	3.30	3.81
55 MAUNDER (LLOYD) LTD	10.27	12.70	13.81	17.54	16.22
60 NESTLE HOLDINGS (U.K.) PLC	5.60	4.56	5.21	6.02	9.55
62 OVERSEAS FARMERS'CO-OP FE	9.22	13.70	11.04	15.96	31.90
86 UNITED BISCUITS (HOLDINGS)	5.26	5.51	4.36	4.02	3.74
	Current Assets to Current Liabilities				
8 ASSOCIATED FISHERIES PLC	1.50	1.88	1.51	1.45	1.73
21 BOWYERS (WILTSHIRE) LTD	1.54	1.72	1.25	0.20	
26 CAMPBELL FROZEN FOODS LTD	1.75	1.98	3.08	4.60	3.05
38 FOODANE LTD	1.08	0.90			1.91
51 LOVELL (G.F.) PLC	1.81	2.14	1.96	2.04	1.79
55 MAUNDER (LLOYD) LTD	1.17	1.15	0.85	1.03	1.04
60 NESTLE HOLDINGS (U.K.) PLC	1.12	1.25	1.32	1.20	0.93
62 OVERSEAS FARMERS'CO-OP FE	2.21	1.96	1.74	1.28	1.09
86 UNITED BISCUITS (HOLDINGS)	1.19	1.12	1.41	1.45	1.30

Table 33: Some ratios and a proxy for size during the period 1983-1987 for the nine firms examined.

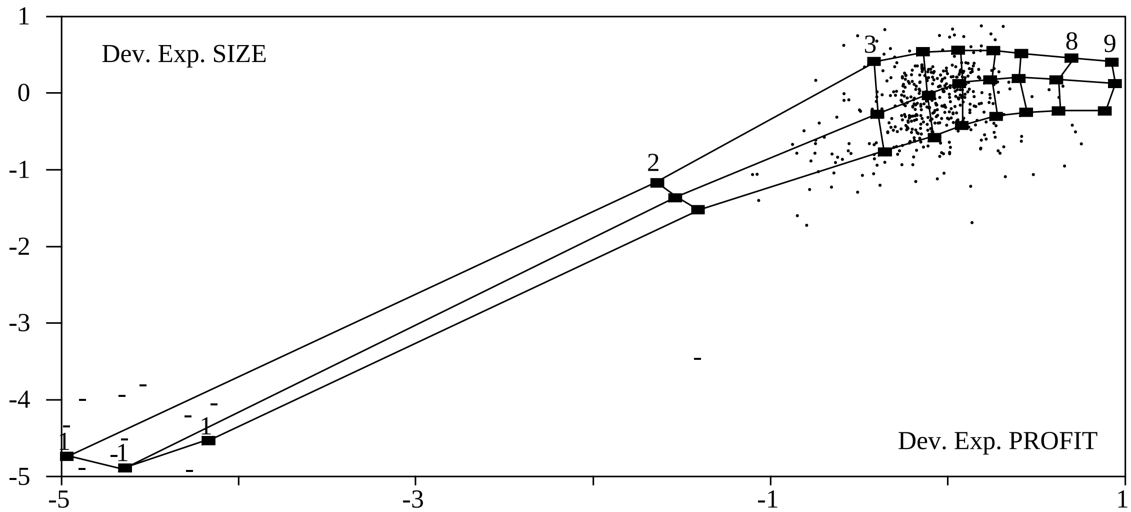

Figure 70: The position of each one of the 9×3 neurons after learning, superimposed to the shape they have learned. Neurons which are neighbours in the discrete space have been linked by solid lines.

As a result of the learning process we obtain a model. This model consists of:

- A set of 9×3 nodes, each one with two weights, $\{w_{ij}^x, w_{ij}^y\}, i = 1,9, j = 1,3$ fixed in some learned values. The weight w^x links to the X-coordinate of the input vector. The weight w^y links to the Y-coordinate.
- A neuron which measures the distance between the weight vector and the input vector and fires if this distance is the smallest amongst all the neurons. In our example the distance used was the Euclidean one:

$$d = \sqrt{(w^x - x)^2 + (w^y - y)^2}$$

Other distance frequently adapted is the inner product $X \cdot W$. It provides a measure with some interesting qualities. Tattersall [126] explores this subject.

The output of this model is the pair $\{i, j\}$ identifying the neuron which fired. For each input vector $\{x, y\}$ the corresponding output is this pair $\{i, j\}$. Hence, the described learning process can be viewed as a map of a continuous-valued space onto a discrete one. In the literature this kind of mapping is known as a quantization.

Trajectories in the reduced space: Now, if we input the model with a sequence of vectors representing the same firm during a period of five years it will output the corresponding sequence of fired neurons. This output sequence defines a trajectory in the reduced or discrete space. Figure 71 on page 208 shows several of these trajectories.

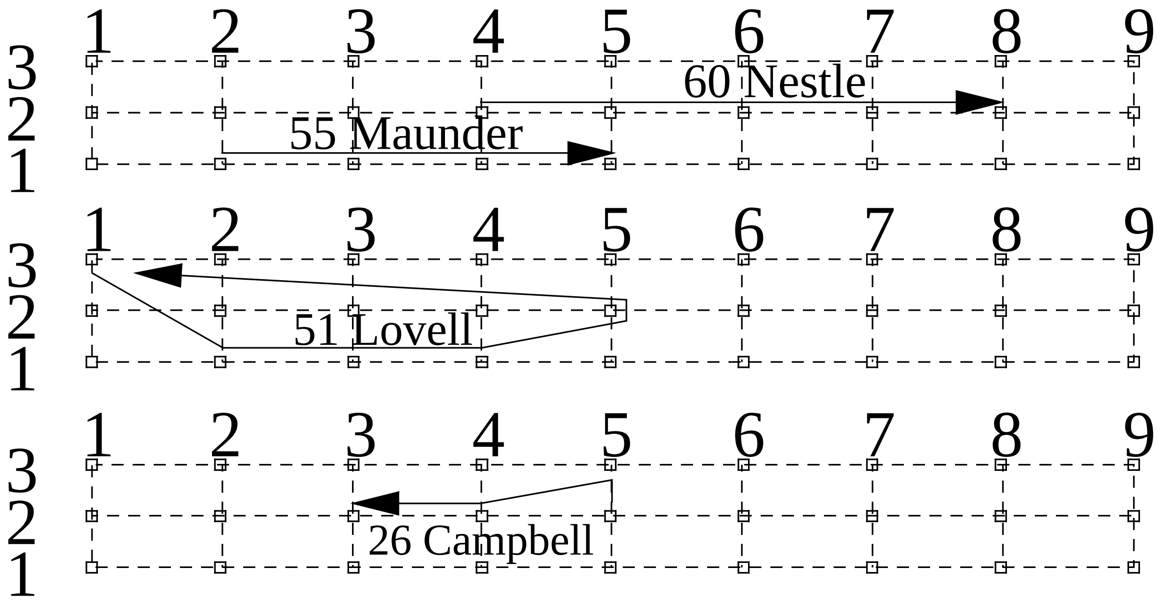

Figure 71: Three reproductions of the set of neurons supporting a few trajectories of firms in the discrete space they define.

Regions and rules: Since each neuron, after training, acquires a mapping quality, its firing has a precise meaning on financial analytic grounds. This is because each region of the mapped RRP also has a precise meaning.

Therefore, it makes sense to build a set of rules or a diagnostics table relating the i and j of a fired neuron to liquidity and the magnitude of its components inside the firm. In our example, for instance, we could set a few rules like

if $i > 2$ then

- if $j = 1$ then too much size for what the firm is worth.
- if $j = 2$ then size is correct for the worth of the firm.
- if $j = 3$ then too much capital for actual size.

and so on. A similar set of rules, or a diagnostics table, could be more or less detailed depending on the number of neurons used. If, in our example, we were to use a larger number of neurons in the j dimension (for example, 5 instead of 3) we would get more specific diagnostics. But it is not clear whether such an increase would be desirable.

Robustness of the obtained model: Since the RRP uses mean-adjusted values and the basis of each diagnostic is the extent to which each case departs from this central trend, it follows that the displayed table of correspondence is expected to be, to some extent, independent from changes introduced in the data. After building a model for a given industry it is not likely to be necessary to adjust it frequently.

It would be necessary to undergo the building of a new map only when the spread suffers itself a noticeable distortion. Strong external influences could introduce real modifications in the whole of a statistical distribution of an industry for some specific features. For example, a period of expansion or collapse would substantially modify patterns of profitability.

A drawback of this algorithm is that sometimes it is difficult to make the map span the whole of the scatter. Cases may occur that will not fire any neuron at all. But this happens mainly when, during training, the neighbourhood of neurons is defined in a simplistic way. Any attempt to reproduce Kohonen's algorithm should use more elaborated definitions of neighbourhood than those given as examples in introductory texts.

Kohonen's algorithms also seem to have difficulties when the spread to be mapped has regions with very different densities. In our case the algorithm was unable to cover the whole region of negative EBIT. Therefore, the diagnostics we obtain for this zone are very general — contrasting with the detail obtained for the positive ones.

8.2.5 Discussion and Prospects For Future Applications

Besides Kohonen's Self-Organized Maps, other algorithms exist able to perform the same task. Any of the tools known as "quantizers" — for instance, the nearest-neighbour one — could be used instead to obtain the reduced set.

We selected this particular model because it illustrates a practical use of the Hebb's Rule. Competitive Learning and in general the Connectionist approaches to known problems have some interesting features though. They are parallel in structure, which makes them ideal for implementation on future machines. And they are also robust regarding assumptions about the data.

The RRP as a pre-processor for expert systems: The rules resulting from using this technique can be fed into a more general system along with other sources of knowledge. Automatic diagnostics could be extracted from large databases containing accounting data. Several RRP with the corresponding maps could cover more than one feature of the firm outputting rules that would be jointly processed by this system along with related information.

In such an environment, the RRP plus quantizer would act as a pre-processor allowing bridging the gap between continuous-valued, stochastic data and symbol-based systems.

Other promising plots: Predicting the future failure of firms. Profitability is perhaps not the most adequate feature for testing the real possibilities of the devised tools. Unfortunately we were not able to test them in firm distress prediction. We think that simple rotated plots along with our maps are ideal for tracing strong, low-dimensional, relations such as those linking accounting information and the distress of firms.

Early attempts to diagnose firm distress using accounting information date from 1966 when Beaver [7] [8] identified the ratio that would predict failure more accurately. Altman [3] and others centred their efforts in the problem of performance. Indeed, they improved performance using multivariate techniques. Altman's last published results (the Zeta model, 1978 [4]) showed a remarkable ability to forecast failure.

From 1970 on, the number of published works on this very subject has been enormous. However, the original challenge suggested by Beaver — the identification of simple, useful, tools able to perform a task reasonably well — deserved much less attention.

A review of this research can be found in Foster [44] (chapter 15). Taffler [125] is the most representative study of this kind in the U.K.

It seems clear that little improvements in performance can now be expected when predicting firm failure. However, if the reasons leading to distress are more than one, and if these differences can be captured in a firm's accounts, it would be interesting to test the use of our rotated plots along with the dynamic tracing capabilities of Kohonen's maps to identify trajectories leading to disaster. Perhaps it would emerge that there are more than one paths leading to insolvency. In that case, a better understanding of the distress process and the mechanisms internal to the firm could emerge, balancing a common weakness of these studies.

8.3 Summary

In this chapter we used Kohonen's Maps to automate financial diagnosis. Firstly, we explored the graphical possibilities offered by the homogeneity of two-variate relations in log space. We outlined the two kinds of complementary questions ratios are called upon to answer and we described graphical tools able to give joint answers to those questions. Secondly, Kohonen's Maps were used to perform a quantization allowing the assigning of a diagnostic to each region in such graphical tools. The set of rules automatically generated in this fashion can be seen as the result of a pre-processing for symbol-based expert systems.

Rotated plots can be used as direct tools for diagnosis in the same way financial ratios are. But they yield a richer information, namely by incorporating relative size — or, alternatively, a contrast with size — and allowing the study of trajectories.

Chapter 9

A Taxonomy of Risk in Large U.K. Industrial Firms

Can Neural Networks capture the relation between the expectations of investors and the characteristics of traded assets? In this chapter we investigate this possibility using a particular kind of asset, the large and frequently traded industrial firms in the U.K. We rely on considerations similar to those of Ross's APT [104] to breakdown the market returns into four main orthogonal forces. Then we build models to explain the sensitivities of our set of industrial firms to each of these forces in terms of accounting information.

Our experiment comprises four steps. Firstly, sensitivities are extracted from the time-history of 121 returns. Secondly, the statistical behaviour of these sensitivities is described. The third step is the modelling of the relation between the sensitivities and accounting information using the MLP. Finally, we study the behaviour of the obtained models.

This research found a clear relation between some specific features of accounting reports, and a firm's appraisal by the market. A Multi-Layer Perceptron was able to approach the sensitivities of firm's returns to market forces using data reflecting stable features of a collection of firms: Size, industrial group, work force and payment patterns. It turns out that a very significant portion of the cross-sectional variability of these sensitivities can be explained by accounting numbers. It also emerged that the map relating stable features of firms to sensitivities is a complex one.

A promising characteristic of the studied relation is that, to some extent, particular forces impinge upon specific features. For example, the fourth orthogonal factor extracted from returns is the only one relating to Inventory. The third factor relates mainly to Wages and the second one to Net Worth. The first and second factors's sensitivities depend strongly on the size of the firm but in opposite directions. The third factor doesn't recognize size. All the factors's sensitivities recognize industrial grouping but the fourth one to a smaller extent. In some cases, a given force clearly reacts

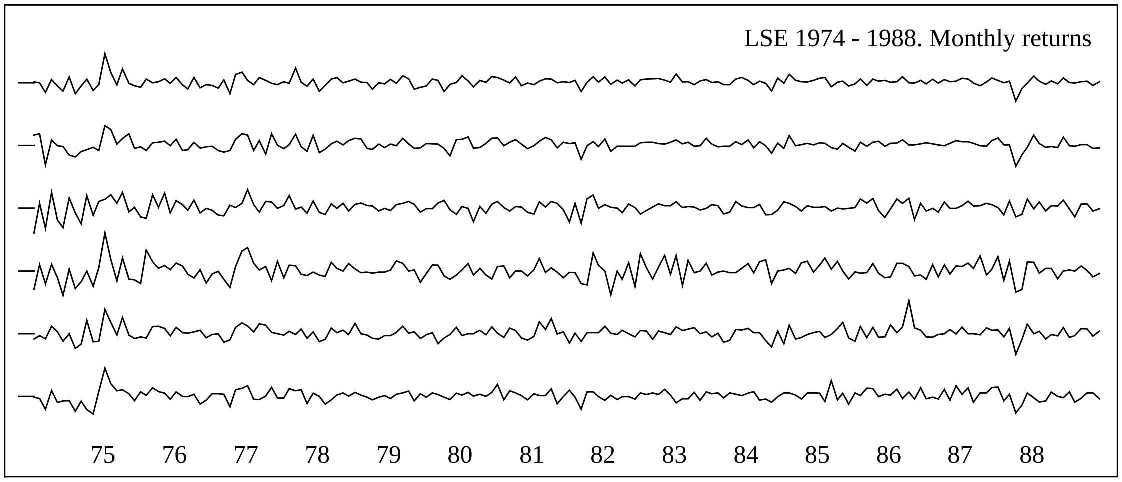

Figure 72: Time-history of six monthly returns on assets traded in the LSE. Co-variance is clearly visible.

in opposite directions for the same feature in two industries.

The method developed here opens the possibility of building a taxonomy of risk. This would be interesting for investment appraisal. Also, the knowledge about the way each market force impinges upon specific features of the firm is potentially valuable in understanding and identifying the real nature of such forces.

9.1 Introduction: Financial Risk and Asset's Features

This study relates the co-variance structure of returns on assets to the accounting features of such assets. Co-variance is a basic characteristic of returns (see figure 72 on page 212). This section explains the basis leading to expect such a relation.

9.1.1 Capital Markets and the Trading of Assets

Decisions by firms and investors are typically made under conditions of uncertainty. Unique events which frequency distribution can not be objectively specified are known as uncertain events. In the case of a known frequency distribution it is usual to talk about risk rather than uncertainty [139].

Uncertainty, as well as riskiness, can vary between a maximum and a total absence (certainty). An adequate measure for uncertainty could be the expected missing information about future mutually exclusive events. However, it doesn't take into account the values under risk.

A more useful proxy for risk is the variability future outcomes exhibit. But measures of variability with practical use are parametric-dependent. They require precise assumptions about the distribution of the data. Unfortunately, the spread observed in returns is usually leptokurtic, thus allowing the parameterization of central trends but not the one of the spread.

Finally, risk can also be seen as the sensitivity of an asset's expected return to unanticipated changes in some external factors. This proxy is free from assumptions about the spread of returns. Instead, it implies the assumption that there is some general stable model linking risk to the expected return on an asset. In a capital market this seems to be the case.

Capital markets, when available, trade-off risk and return. Actual prices of an asset are determined by the investors's expectations about its future returns.

The expected return of every asset must be proportional to the investor's perception of risk. Risk and expected return will be linearly related — otherwise arbitrage opportunities would emerge.

If the expected return on assets traded in capital markets are influenced by more than one economic force the distribution of expected returns will display more than one dimension. Absence of arbitrage opportunities will imply, in this case, that the expected return on each asset must be a linear combination of its sensitivities to such external forces.

Relating risk to return on assets: Finance theory requires that discounting of future expected cash flows ought to be made with discount rates similar to those observed in assets bearing the same risk. If there is a general and tested model relating risk, whatever it is, to expected return, and if the way assets are similar before risk is also well known, this can be done. Otherwise, discounting becomes a rather empirical exercise and probably will lead to wrong decisions.

When an asset is traded in capital markets the former requirement is fulfilled with the model outlined above, based on arbitrage, or using other models. But the relation that the market establishes between sensitivities to unanticipated changes in forces impinging upon it and the different possible characteristics of every asset is not well understood yet despite the large amount of research devoted to this issue. Even restricting ourselves to specific groups of assets like productive commitments (industries, trade or services), there is no available guidance to relate expected returns to the main features of such assets.

The information investors use in the appraisal of assets mainly includes the one contained in periodic reports and accounts, as well as in other related sources. For each traded asset there is an information content available to the market. It is natural to suppose that this content is not one-dimensional. Therefore, it could perhaps be decomposed into more simple and more general features like size and industrial group, each one of them perceived and traded by investors in a consistent way.

If that is so, the different forces impinging upon the market would produce particular contrasts when compared with each one of the features mentioned above. These contrasts would be complex, though. For example, inflation could make leveraged firms belonging to a particular industry less attractive to investors while others, belonging to a different industry, would appear as more attractive because of precisely the same reason. As a result, when modelling investor's expectations using long term debt and inflation as input variables, a second-order effect would emerge.

Diversifiable and non-diversifiable risk: Assets are traded by what they are worth in the market's eyes. But the holders of traded assets are not risking their wealth in a way that directly relates to the risk of each asset they hold. Just by holding many assets investors avoid a great deal of the variability of their wealth created by fluctuations in the price of their risky assets. The only variability they can't avoid is the background movement of the whole market, the one which is impossible to cancel out.

This background movement is a unique time-history. But it can be viewed as the result of a few orthogonal forces, each one of them capturing the largest dimensions of the joint variability of returns on all the traded assets.

Notice that we are not considering here any decomposition of the auto-correlation eventually present in the time-history of this background movement. Instead, we consider such a movement as the result of a few forces influencing all assets. Such forces are obtained from many time-histories of returns on assets by decomposition of their joint variability into its main orthogonal co-movements.

9.1.2 The APT: Diversifying the Points of View

After Ross's seminal paper in 1976 [104], arbitrage models became present in finance research but scarcely in finance practice. The first signs of practical use of the APT only nowadays seem to appear [97]. The reason for this delay may well be the intuitive simplicity and the immediate interest of the Betas, the sensitivities to non-diversifiable risk when compared with APT's sensitivities.

In our opinion the original APT, despite being presented as an alternative to the existing models for the appraisal of risk by investors, is not really an alternative. It makes interesting progress in matters investors seem not to be interested in.

The point of view of investors: For example, the APT offers four or five independent forces instead of one. But investors are only concerned with non-diversifiable risk. By definition, it is impossible to have more than one non-diversifiable force. Two non-diversifiable forces could be added and they would yield a unique one. Forces which cannot be added because they scatter in all directions are diversifiable.

The APT gives the impression of having more to say about co-variance with the market portfolio just by decomposing it into four or five components. But it hasn't. Following an example given by Copeland and Weston [26], if we are the pilots of a plane in danger, the APT would give us the latitude, the longitude and the altitude whilst the CAPM only gives the distance to the airfield. But our case is a particular one. We are running out of fuel. We can see the airfield in the distance. Only, we don't know if we should try to hold the plane or else to search for a place suitable for an emergency landing. In this case, it seems as if the distance to the airfield is the piece of information capital for getting out of trouble. And by knowing the longitude, the latitude and the altitude we are not much better off. We would have to go through an awful lot of algebra for reaching the same conclusion. In other words, it is clear that if a thoroughly calculated Beta really expresses the co-variance with the market portfolio, then investors don't need anything else. The APT cannot say more about non-diversifiable risk than the CAPM does. The co-variance with the market is the only portion of an asset's spread investors will pay to avoid. It is also the only piece of information investors are likely to pay to know.

Also, the APT can present diversifiable risk as non-diversifiable. When the market forces are extracted via a Principal Components rotation of returns, nothing prevents a small portion of diversifiable variability to creep in a factor taken as an overall force. The smallest amongst the accepted factors can be the result of the existence of large clusters of firms sharing similar sensitivities, not of any overall force.

When a factor succeeded in capturing an overall market force it's clear that such a factor represents non-diversifiable risk. But if it represents a large cluster of firms sharing similar features regarding the way they are traded, the risk they introduce in the model is diversifiable. If the method of regressions is used instead, we can't be sure to be using real overall forces. Perhaps one of the forces used has an overall effect in our sample but not to the whole market.

The CAPM is more robust regarding this problem since indices are built on purpose to avoid diversifiable risk. The idea of diversification of risk is in the core of the way CAPM betas are estimated. In the APT we can just expect it.

The point of view of individual assets: However, there are cases in which the APT could become useful for financial practice. Whenever a more precise understanding of the way assets are traded is required the APT is the choice. This is the case for the appraisal of productive commitments. It is also the case of studies concerned with the market itself.

When searching for a general asset's taxonomy of risk, arbitrage models seem very promising. They are a breakdown of an information content into features. They should allow a better discrimination, each of the orthogonal forces showing particular affinity towards some aspects of assets. The APT offers the possibility of looking onto an asset's expected returns from several points of view.

The investor's point of view is the one of an asset's holder. When the emphasis is not in the holder but in the asset itself, the APT becomes the adequate instrument for assessing the variability of returns.

9.1.3 The Trading of Stable and Fluctuating Features of Assets

A basic concept in our research is the hypothesis that the market trades differently the stable or intrinsic features of assets and the more fluctuating ones.

A rapidly growing body of research documents components which can be forecasted in asset's returns (see for example Fama [40], [41] or Keim [70]). Predictability is not necessarily inconsistent with market efficiency. Stock prices need not follow a random walk to be efficient. Given this, it is reasonable to divide investor's movements in two categories: Their expectations created by the predicted component of prices, and their reaction to unanticipated changes with regard to such prices. In other words, since the return on assets is a sum of two components — the expected or anticipated returns plus the unanticipated ones — all the perceived trends would influence the expected returns, not the unanticipated ones.

Arbitrage considerations lead to models in which sensitivities to unanticipated economic forces play a role. The APT explains the second component of returns.

It is also known that economic trends seem capable of influence accounting numbers to some extent and for particular reports. Linear models relating fluctuations in accounting figures to economic trends have been tried by Brown and Ball [20], Gonedes [50], Magee [81], Lev [78] and others. Such models seem to show that the expected component of prices — the trends — would relate to some changes in accounting figures. However, there are numbers which can't adapt themselves to economic trends because they reflect features which are stable or intrinsic. Obviously, only those numbers which can fluctuate with the economy will actually do so.

Stability and sensitivity: It is reasonable to suppose that, if fluctuations in a few features of the firm can be dictated by the expected component of economic forces, the firm's stable attributes would create in investors the sensitivity to unanticipated forces. This is because only stable features can have sensitivity. The fluctuating ones are not sensitive: They move along with the forces.

Stable features seem more promising in explanatory power than the fluctuating ones: The last ones, incorporating the co-movement with broad economic trends, would be in some extent anticipated by investors. In other words, the most stable features of firms would explain why investors see assets differently, regardless of economic trends. For example, the industrial group and size of firms would explain their sensitivity to unexpected returns for they cannot adapt their size or change sector to face inflation or other trends. Conversely, less stable — less intrinsic, more contingent — features like the financial structure or the dividend policy, would relate to trends in economic forces and to expectations observed in prices.

If that is so, this study should search for relations between the most stable features of assets and their sensitivities to the market forces. We consider industry group, location, size, operating leverage, labour and capital intensity, payment pattern (hence, the short-term debt) as candidates for explaining market sensitivity. Dividend policy, some component of capital structure, are less intrinsic.

9.2 Existing Research

Our research relates mainly to the APT. Other somehow similar topics could be the "Accounting Beta" or other econometric models relating Beta to firm features.

The APT today: After facing some queries about its testability [111] [33] [112] and partially because of it, the APT incorporated equilibrium considerations as well as market portfolio equivalences [52] [32] [24]. The generally accepted factors have being labelled in the following way: Inflation, Industrial Production, Risk Premium in the market and Interest Rate's term structure growth.

The tests of APT validity (see, for example [23], [101] and [117]) are generally considered as non-conclusive. The main issues are the intercept term of the model, the independence of residual risk, the lack of significance of the model in seasonal periods [53], and estimation problems [113].

The number of factors, their interpretation and stability have also been explored, leading to not very consistent conclusions. See [28] for a study based on the LSE and [113] for remarks on the influence of the estimation technique on the number of factors.

Three irregularities have been the subject of much interest:

- The small firm's effect. There is considerable evidence that the mean returns of small firms exceed those of large firms [30].
- The January effect. Mean returns in January exceed the mean returns in other months for small firm categories [69]. The model seems to loose significance in other months [53].
- The weekend effect [71].

An excellent yet simplified discussion of arbitrage pricing models and their evolution can be found in Jarrow [66].

Accounting measures of Beta: Much research has been published on the relation between specific characteristics of the firm like operating leverage, gearing, or size, with Beta, the covariance of an asset's expected return with the market portfolio. Gahlon [47], Hamada [55], Hill [58] are some examples.

An accounting proxy for Beta has been discussed in 1970 by Beaver, Kettler and Scholes [9], and later by Thompson [127], and Beaver and Manegold [10]. After this, Bildersee [13], Blume [15], Bowman [18], [19] and others, did the same, achieving small but significant relations between accounting and market measures of risk.

Traditional accounting measures of risk are attempts to highlight the uncertainty associated with earnings of the firm. They are surrogates for the total variability of returns. Accounting Betas reflect both the systematic component of risk and the residual one. Thus, the quality of accounting betas as measures of real market expectations depend on the existence of a strong positive correlation between this systematic risk and the residual component — the one the market doesn't contemplate. See [9].

Models of asset valuation: Another related body of research is the one concerned with the valuation of assets. Foster's chapters 9 and 12 [44] review two branches of such research. Chapter

9 is concerned with capital markets and information efficiency. For example, the "Fundamental Analysis" assumes that each asset has an intrinsic value that can be determined on the basis of earnings, dividends, capital structure and growth potential.

Chapter 12 reviews the use of financial statement information in the trading of equity. Equity valuation models focus on one or several features like predicted earnings, dividends, cash-flows or the value of the assets owned by a company. They have been used for the valuation of non-traded firms, and for supporting investment decisions.

Most of the methods referred to above are based on time-series prediction. They could be described as projections based on the past. It is expected that they will explain the market reaction to each asset's performance. Rosenberg and Guy [102] are an example. Their research relates Betas to fundamentals. For each particular firm, they attempt to increase the quality of this measure of risk by incorporating corrections based on industry group, growth, the spread of earnings over time, the financial structure, size and others.

Differences from our study: We would like to underline the fact that such models are not the kind of relation we are interested in here. We are not using individual histories of firms to extract variables and then correlate them with betas. We estimate the relation between the co-variance of many assets with market forces and cross-sections containing accounting features of the same assets. Individual histories reflect individual performance. Based on them it is possible to anticipate future performance. Hence, these models can hide sensitivity to unanticipated forces.

9.3 The Data

The firms selected for this experiment belong to the usual set of industries we refer to in many occasions during this study.

Prior to the final selection of the set of firms to be used we examined three kinds of returns. Daily returns for a period of five years (1985-1990). Weekly returns for the same period. Monthly returns for a period of 15 years from January 1974 to December 1988 and listed in the FTA All Shares Index. For these three sets, returns were checked for the frequency of non-traded cases. The daily returns yielded only 60 in a total of about 500 firms with a reasonably small number of non-traded days. The weekly set yielded 71 such firms. We considered as reasonably small a number of non-traded periods of 10% or less.

It was decided that it would be desirable to avoid the use of infrequently traded assets. The known means for circumventing this problem [29] [109] would introduce in our experiment an extra manipulation of information. Therefore, the daily returns and the weekly ones were discarded as not suited for the experiment.

In the case of monthly returns there is a natural limit for the number of assets to be used.

Fifteen years contain 180 time periods. It is impossible to extract real new information from more companies than time periods: There would be more variables than equations.

From the set of monthly returns we obtained 121 industrial firms having both the quality of belonging to our set of industrial firms and being frequently traded. A third characteristic is that all of them are listed in the FTA All Shares Index. The institution providing the monthly returns is the London Business School. Returns are calculated as

$$r_t = \log_e \left(\frac{p_t + d_t}{p_{t-1}} \right)$$

in which p stands for traded price in month t, p_{t-1} is the last traded price in month $t-1$ and d_t is the dividend declared adjusted to a month-end basis. All adjustments are based on the principle that the value of a share is unaltered by a change in capital structure.

We decided to use these 121 collections of monthly returns for the extraction of sensitivities. From such 121 firms, 77 belonging to 7 industries were used for building the model. It is possible to access the accounting reports of these 77 firms for a period of four years, from 1983 to 1986. Each example in our learning set is built with the accounts of one firm for one of these years and having as outcome one sensitivity. Therefore, there are $77 \times 4 = 308$ examples.

The 77 selected firms and their industrial group are listed at the end of this chapter, in table 36 and the one on next page. The total number of examples is 308. By industry it is:

Industry	N. Cases	Percent
Building Mats.	44	14.3%
Paper & Pack	40	13.0%
Chemicals	68	22.1%
Electrical	24	7.8%
Electronics	36	11.7%
Textiles	40	13.0%
Food	56	18.2%

The selection of the 77 final firms was made on an industry basis. Only seven industries were selected. The reason for putting aside the other industries is twofold. Some of them, like Metallurgy and Leather, were likely to introduce a very particular behaviour in the overall variability. Also, some industries were discarded due to the small number of firms represented. On the whole, since one of our goals was the assessment of the effect of industrial grouping, it didn't seem appropriate to gather in the same sample a large number of industries.

9.4 The Market Forces

Our experiment comprises four steps. First, the extraction of sensitivities from the time-history of 121 returns. Second, the study of the statistical behaviour of these sensitivities. Third, the modelling of the relation between them and accounting information, using the MLP. Finally the study of the behaviour of the obtained models. In this section we describe the two initial steps.

Notice that the used marginal returns belong to a very particular kind of asset: Large and frequently traded industrial firms in the U.K.

9.4.1 Extracting Sensitivity

Given the returns of 121 companies during the referred period, a Factor Analysis was performed on their time-history taken as stochastic variables. The aim was to rotate the variance and co-variance matrix so that a few minimum co-variance axis would emerge.

Four specific sensitivities of each asset to corresponding external forces were calculated as the loadings of the four main factors present in returns. We obtained, for the j^{th} asset's return, r_j,

$$r_j - r_{0j} = b_{1j} \times f_1 + b_{2j} \times f_2 + \cdots + b_{4j} \times f_4 + \varepsilon_j$$

in which r_{0j} is the expected return on asset j and the $b_{kj}, k = 1,4$ are estimations of the influence or sensitivity of asset's j return to the discovered forces $f_k, k = 1,4$. Large b_k mean a large influence, small b_k mean a small one. When a b_k is negative, the influence and the force go in the opposite directions. Any anticipated return will be incorporated into the r_0.

Clearly, what we obtain in this case is a breakdown of the cross-correlation between time-histories into its main components. In order to maintain a minimal level of comparability with other studies we used Maximum-Likelihood, not Least-Squares, as the criterion for the extraction of factors. This issue is not important, as far as our data is concerned.

The resulting factors are supposed to be a linear composition of a few main economic forces impinging upon the market. In our case the five larger factors explained half the total variability. Their Eigenvalues are displayed next.

factor	Eigenvalue	Percent	Acc. Percent
1	40.6	39.5%	39.5%
2	3.1	3.0%	42.5%
3	2.5	2.4%	44.9%
4	2.1	2.0%	46.9%
5	1.9	1.9%	48.8%

Apart from the first factor all the others decay smoothly towards smaller explained variability. Table 36 on page 238 and the one on the page next to that one display the commonality of each firm and the loadings corresponding to the four largest factors. The commonality ranges from 0.15 to 0.75. It is approximately Gaussian with a mean of 0.5 and a standard deviation of 0.11. Therefore, for most of the firms, half the variability of their returns can be explained by these four factors.

The factor loadings are the main object of this experiment. They are supposed to represent sensitivities of assets to unanticipated changes in market forces. In next section we examine them briefly.

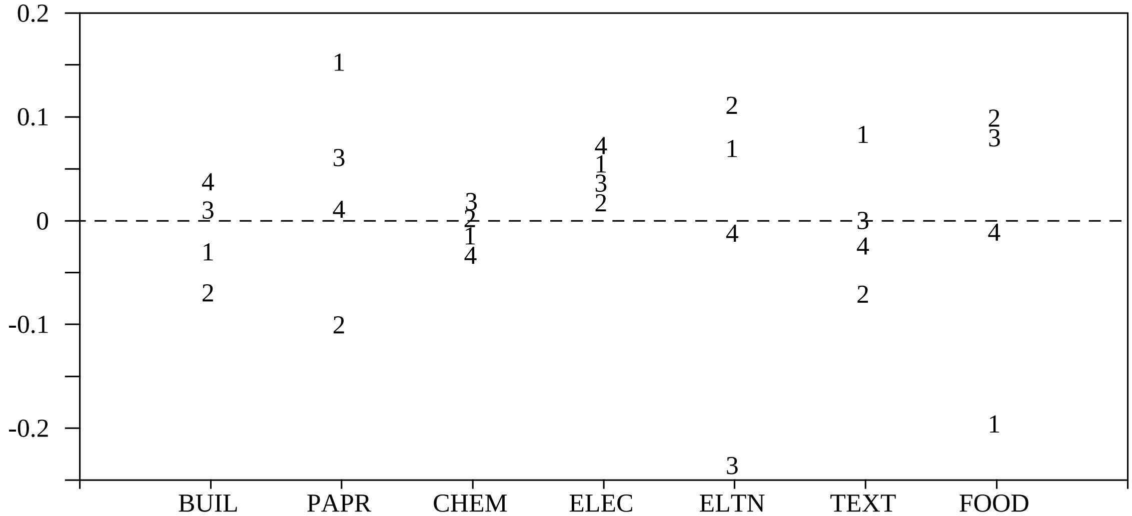

Figure 73: Mean values of the four sensitivities by industry. The displayed values show positions of industrial groups (X-axis) regarding deviations from the overall mean of each sensitivity (Y-axis).

9.4.2 Preliminary Study of Sensitivity

In order to gain insight in the broad rules governing the relation between sensitivities and features of the firm, preliminary studies were carried out. The results are important since they allow us to highlight the differences between the MLP and linear tools.

We first observed the statistical behaviour of each sensitivity individually. Then we studied their mean values by industry. Next, the co-movements of accounting items with each sensitivity regardless of the industrial group were also assessed. Finally we built a linear model explaining sensitivities in terms of both industrial groups and items.

Statistical behaviour of each sensitivity: The four variables containing the sensitivities are broadly homogeneous. No real influential points were found. In the next table we display their basic statistics.

Factor	Skewness	Kurtosis	Mean	Standard Deviation
1	-0.882	-0.122	0.41176	0.18900
2	-0.478	0.195	0.26898	0.12415
3	0.404	0.605	0.26286	0.14825
4	1.881	1.034	0.18363	0.12611

There are no strong differences between the sensitivities in what concerns these values. The fourth sensitivity is less homogeneous than the others.

These variables are not correlated. Despite the factors being orthogonal, the sensitivities they generate could be correlated. But only the fourth sensitivity seems to show traces of negative correlation with the first and second ones.

Sensitivity by industrial group: Next table summarizes the influence of industrial grouping in these four variables. It shows the mean values and standard deviations of each sensitivity by industry. The displayed values are the deviations from the overall mean for each industry. This makes the interpretation easier. Figure 73 on page 221 is the graphical representation of this table.

Industry	Factor 1		Factor 1		Factor 1		Factor 1	
	mean	st. dev.	mean	st. dev.	mean	st. dev.	mean	st. dev.
BUIL	-0.030	0.164	-0.069	0.105	0.011	0.102	0.037	0.114
PAPR	0.153	0.151	-0.100	0.117	0.061	0.125	0.011	0.143
CHEM	-0.015	0.141	0.002	0.118	0.005	0.112	-0.025	0.100
ELEC	0.055	0.184	0.017	0.113	0.042	0.164	0.063	0.101
ELTN	0.070	0.124	0.111	0.083	-0.246	0.151	-0.008	0.178
TEXT	0.084	0.160	-0.071	0.080	0.001	0.103	-0.024	0.114
FOOD	-0.196	0.171	0.095	0.066	0.081	0.071	-0.011	0.119

The effect of industrial grouping is significant for all the four variables. However, it is very strong in the first three and it is weak in the fourth. The observed intra-class correlations (introduced in chapter 5) were:

Factor 1	Factor 2	Factor 3	Factor 4
35.4%	39.9%	43.1%	3.4%

The three main factors are correlated with industrial grouping to an extent none of the accounting items observed in previous chapters seem to attain. The fourth factor is also correlated — its F is significantly different from 1 — but to a much smaller extent.

We use the convention of calling "positive" to influences which are above the expected for that sample. We call "negative" otherwise. In this fashion, the features of this effect are:

Force 1 strongly influences the prices on the Paper industry. It is also very influential for the Food industry, but in the opposite direction.

Force 2 yields a positive sensitivity in Food Manufacturers or Electronics. It has a negative or neutral effect on the other industries.

Force 3 has a strong, negative, influence on the Electronics industry. It is positive for Food Manufacturers and Paper industries.

Force 4 generates a positive sensitivity in industries like Building Materials and Electricity. It generates a negative one on the other industries.

This simple observation of means by industries is likely to help in elucidating the meaning and origin of each one of the four forces. Notice that the above description is specific to this sample since it relates to an expected value.

	1983		1984		1985		1986	
Sign of slope	+	-	+	-	+	-	+	-
Fac. 1	Size G. Fund	Invent.	Size	Invent.	Size Sales	Invent.	Size	Invent.
R^2	52%		53%		50%		49%	
Fac. 2		EBIT Sales		G. Fund Sales		G. Fund Sales		G. Fund Sales
R^2	15%		11%		19%		13%	
Fac. 3	C. A. G. Fund		C. A.		W.Cap. G. Fund	Net W.	W.Cap.	
R^2	26%		16%		29%		17%	
Fac. 4		Size Debtors		Size Debtors		Size Debtors		Size Debtors
R^2	18%		10%		13%		10%	
Comm.	Size G. Fund		Size G. Fund		Size G. Fund		Size	Credit.
R^2	46%		44%		44%		40%	

Table 34: Results of the preliminary regressions. For each period and factor, this table displays the items that turned out to be relevant and the proportion of explained variability. The sign (+) means a positive correlation with the sensitivity under question. The sign (−), a negative one.

Regressing items with sensitivities for individual periods: We selected four years of accounting reports — 1983 to 1986 — matching the firms which sensitivities we had extracted. For each one of these periods we examined the linear models obtained when explaining such sensitivities in terms of accounting information.

We first used size-adjusted residuals, $\varepsilon^x = (\log x - \overline{\log x}) - (\log s - \overline{\log s})$, along with our proxy for size, $\log s - \overline{\log s}$, as the input variables. The outcomes, the sensitivities, suffered no transformation.

Table 34 on page 223 displays the results. For each period, the items that turned out to be significant in explaining sensitivities are displayed along with the proportion of explained variability. Also possible correlations with the commonality were investigated.

The only sensitivities clearly explained by accounting numbers are the ones related to the first factor. Investors seem to reward Size and penalize Inventory. This, when no distinction between industrial groups is introduced. The remaining factor's sensitivities are very little explained by the features of the firm, at least in this linear way. The second factor must represent something negative to the economy since its sensitivities are negatively correlated with Earnings and Sales. The third factor shows some affinity with short term features of the firm, but a small one. And the fourth factor is negatively correlated with Size and Debtors.

The results are generally consistent during the observed period, the third factor being the exception: Its R^2 is dependent on the period.

The commonality is an expression of the variability explained by the four factors. It is clearly explained by Size and Earnings. This means that the largest firms and the most profitable ones are also the ones traded in a more regular, predictable, way. The market trades large and profitable firms in a way that is more similar than the way it trades other assets. An interesting study related to this issue is Roll's "R^2" (1988) [100]. It is an empirical assessment of the variability explained in

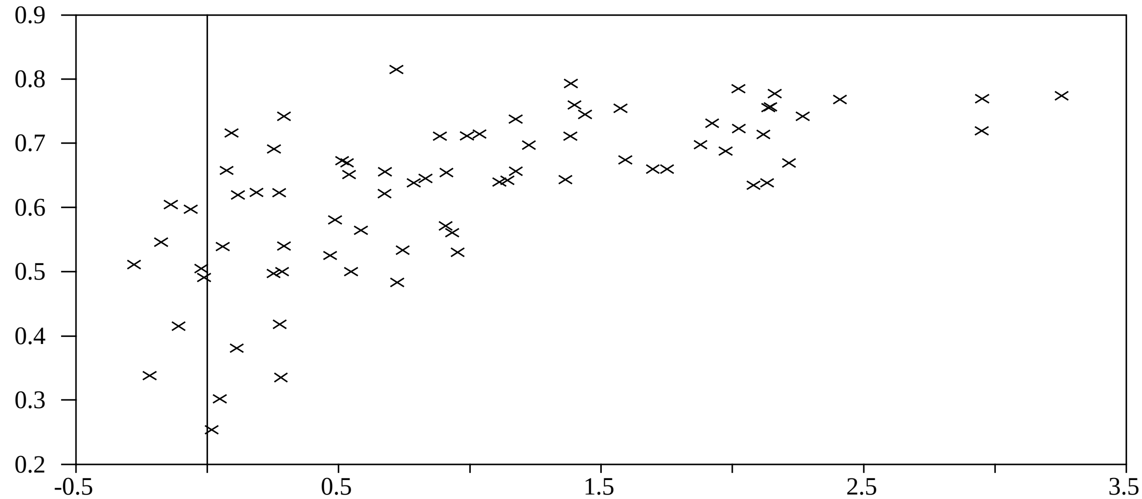

Figure 74: The first factor's sensitivities (Y-axis) against the relative size of firms (X-axis).

stock returns by the market models. Roll correlates the collection of R^2 observed in several kinds of market models with a few characteristics of the firm.

Scatter-plots of sensitivities against size-adjusted items were also observed as part of this preliminary study. Figure 74 on page 224 shows a typical shape obtained by plotting $\log s$ against the first factor's sensitivity. Clear correlations were observed between Size and the first two factor's sensitivities. The correlation between Size and the first one is non-linear. The second one exhibits a negative linear correlation with size but industrial groups form visible clusters. Also Gross Funds From Operations and Inventory show significant correlation with some sensitivities and they form clusters as well.

Regressions including industry grouping as inputs: This second group of regressions explained each one of the four sensitivities in terms of accounting items plus a set of dummy variables representing industrial grouping. It was meant to directly assess the gain in explained variability resulting from using the MLP. The input space was similar to the one we used in the MLP experiment. Therefore, we report these results later on together with the MLP ones.

9.5 Modelling Sensitivity With the MLP

Using the procedures explained in chapter 7 we modelled sensitivities in terms of accounting information using the Multi-Layer Perceptron. Since the four outcomes are almost orthogonal there is nothing to be gained from showing them jointly to the MLP. Hence, we built four independent models, one for explaining each sensitivity.

Topology: In chapter 7 the problem was one of classification. In this case it is one of regression. The only difference in what concerns the MLP is the use of linear transfer functions in the last layer of nodes and Least-Squares as the success criterion.

The used topology and input variables were similar in the four cases. Two models were built for each of them. First, one with 17 input variables corresponding to 10 residuals and 7 groups. The topology was 4 nodes in the first hidden layer, 1 node in the second hidden layer and one output node with a linear transfer function. Second, another one with 8 input variables corresponding to Size plus 7 groups. Similarly, the topology was 4 nodes in the first hidden layer, 1 node in the second hidden layer and one output node with a linear transfer function.

Description of the input variables: Table 35 shows these 17 input variables. Ten of them correspond to items and seven are dummy variables taking the value 1 when a case belongs to a particular industry and 0 otherwise.

One major difference with respect to the experiment carried out in chapter 7 was the nature of the input variables used. Instead of raw data in log space we used the residuals or the size-adjusted items, ε^x, along with Size itself. The procedure for obtaining this proxy for size is described in chapter 5. We recall that in chapter 7 the size component was not introduced explicitly as an input variable. Size was formed in one of the first hidden layer's nodes. Here we introduce size explicitly as one of the inputs and all the other items are size-adjusted.

The reason for not using log items is related to the goals of this experiment and the nature of the problem. In chapter 7 our goal was to test the ability of the MLP to form meaningful structures interpretable in terms of ratios. The relation itself was not difficult to model since it was near linearity. Therefore it made sense to use afterwards the obtained ratios as predictors instead of the MLP. In this case we face a complex model, highly non-linear. What would we do with the ratios the MLP would form in the nodes of the first hidden layer? We couldn't use them instead of the model. The weight of the model in this problem is heavy.

Therefore we decided to use the size-adjusted items. They offer a ready interpretability of results. They are also mean-adjusted so that the exploring of the model is straightforward.

Description of the training sets: Data from four time-periods (1983 to 1986) were gathered in the sets to be used in the learning and test of the generalisation performance. For each firm — and outcome — there were four input vectors. The total number of cases was enlarged in this way to a reasonable value.

The randomization process leading to the division into two samples — the test and the learning set — was carried out by blocks on an industry basis. As a result, firms were divided randomly in two groups but since the randomization was made inside the same industry the two sets didn't yield large differences in their proportion of cases for each industry.

Input variable		Number of parameters engaged				Total
		Factor 1	Factor 2	Factor 3	Factor 4	
Chemical	dummy	3	1	1	1	6
Electronics	dummy	2	3	1		6
Size		2	2		2	6
Textiles	dummy	2	2		2	6
Wages		3	1	2		6
Current Assets		3	1	1		5
Debtors		2	1		2	5
Electrical	dummy	1	1	2	1	5
Net Worth			3	1	1	5
Paper and Packing	dummy	3	1		1	5
Building Materials	dummy	3			1	4
Sales		1	2			3
Fixed Assets		1		1		2
Food Manufacturers	dummy	1	1			2
Gross Funds From Ops.		1	1			2
Debt		1				1
Inventory					1	1

Table 35: The number of free parameters (first hidden layer) engaged in the modelling, by input and by factor. The table is ranked by total degrees of freedom.

The learning itself was performed at random too. Each example, comprising the inputs for a given period and the outcome, was randomly selected before being present to the MLP. This procedure avoids the modelling of any auto-correlation present in the data due to the mixing of four year's accounts.

As usual, the checking of the performance of the model was made in the test set, not in the learning set. Since in this case the outcomes were continuous-valued the adequate measure to assess the quality of the fit was the proportion of explained variability, R^2, corrected for the degrees of freedom engaged.

9.5.1 Allowing Complexity to be Accounted For

The adapted topology is somewhat austere. Since the number of cases to be used as learning and testing sets was smaller than desirable, the number of free parameters should be kept as small as possible as well.

The random penalization of weights linking input variables with the first hidden layer of nodes further reduced them in a very significant way. During training we allowed this penalization to continue even beyond the point at which the performance of the fitted model degrades. In other words, we allowed a few weights which values were near the inhibition frontier but who were not inhibitory themselves to be eliminated. This explains why, in the final models, some industrial groups overlap so perfectly.

Table 35 on page 226 shows, for each one of the four models, the number of free parameters remaining after the training finished. Since the number of nodes in the first hidden layer was 4, the maximum number of free parameters per input variable would be four. This never happened. The training started with 68 free parameters and finished with 27 remaining ones (the first model), 20 (the second), 9 (the third), and 12 (the fourth).

The next table displays the R^2 observed in all models. In the MLP, values are similar except for the third model in which the variability each parameter explains is higher. This model is the most efficient one.

R^2	Factor 1	Factor 2	Factor 3	Factor 4
MLP, all input variables	71%	48%	69%	36%
MLP, only Size plus industry	59%	36%	60%	22%
REG, all input variables	60%	42%	52%	21%
REG, only Size plus industry	52%	31%	42%	12%

As noticed before, industrial grouping is an important source of explained variability.

The number of parameters engaged is a rough measure of the complexity of the relation. When looking into the behaviour of continuous-valued inputs we notice that Size and Wages engage six parameters. Debt and Inventory merely engage one. In general, the dummy variables engage more parameters than the continuous-valued ones. Chemicals and Electronics require six parameters whilst Food only requires one.

Discussion: The overall proportion of explained variability in the test set is the most remarkable characteristic emerging from this experiment. Especially in the case of factors other than the first. When comparing these numbers with those of linear regressions it is clear that the MLP made a significant improvement in capturing the relation. The conclusion is that the relation linking accounting information with sensitivities is non-linear. This problem contains higher order effects, created by the different behaviour of the relation to be modelled when in the presence of industrial groups. In next section we shall highlight some of these interactions.

The variability explained just by the size of the firm and the grouping puts the above results in perspective. Size and industrial group are the two factors which clearly explain the market's sensitivities. The first regression we performed didn't have the information regarding groups. This explains its poor performance. Inventory emerged, in such a situation, as the most suitable proxy for the grouping effect.

Factor three's sensitivities are not affected by size and both factor two and four ones are negatively correlated with it. But all the four sensitivities are akin on grouping information.

9.5.2 Exploring the Models

This section describes how the exploring of the resulting models was carried out. We used an "other things being equal" approach. It consists of varying one of the inputs at the time and maintaining all the others fixed in their mean values. The observed outputs can then be compared.

The chosen approach is the only one available in this case. The model the MLP yields is far too complicated for direct interpretation. The used procedure allows an effective look into what

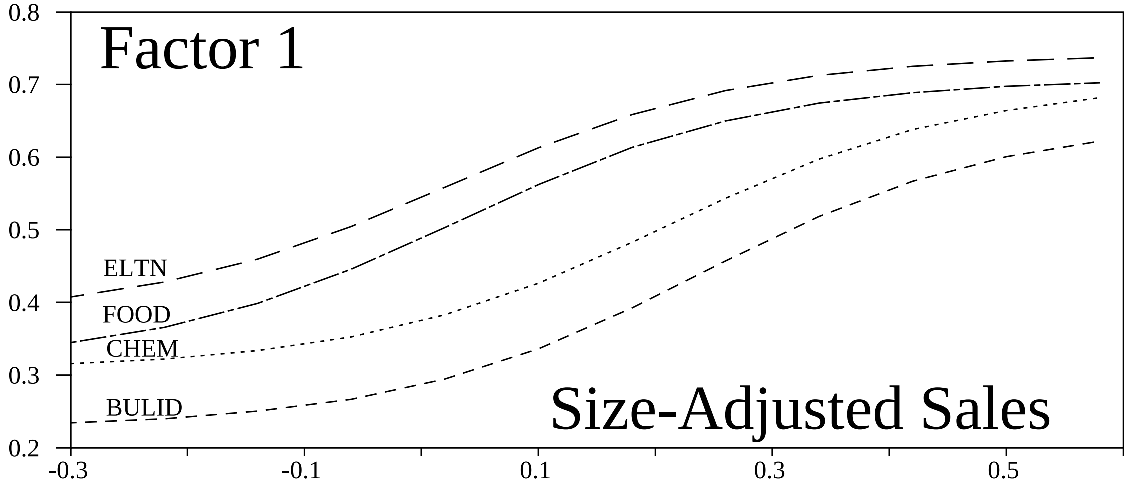

Figure 75: The first factor's sensitivities (Y-axis) against size-adjusted Sales (X-axis) as predicted by the "other things being equal" technique.

the model really does. However, the description that follows should be accepted with caution. In some cases the "other things being equal" techniques yield misleading results just because the other things cannot remain equal in the real world. No extrapolation beyond the neighbourhood of the mean values of the inputs should be taken seriously.

In all that follows it is also important to remember that our experiment was carried out with far fewer cases than the adequate. The results are interesting, not so much because of what they show but more because of what the method promises.

How to read and interpret the displayed results: The figures related to this section are at the end of the chapter, between pages 232 and 237. Figure 75 shows how these results are presented. The X-axis measures Sales and the Y-axis the first factor's sensitivity. Notice that, since we are using size-adjusted items, zero sales means the value of sales which is expected for the size of the firm. Values larger than zero mean sales above the expected for firms of that size and conversely.

This figure says that Sales and the first factor's sensitivity are positively correlated. But Electronics is the industry with the largest sensitivity while Building Materials is the one with the smallest. However, since in factor analysis the sign of the factors is arbitrary, the displayed figures yield a coherent view but not one directly comparable with studies based on correlations with Beta. For example, the trends we observe in the case of the first factor are the inverse of the ones Beaver, Ketler and Scholes [9] obtained for similar variables.

The linear behaviour of factor's sensitivities: We first comment on the simplest relations — the ones which are linear regarding industrial grouping. Next, we comment on those cases in which the model shows industries reacting in different directions to the same force.

Figure 76 on page 232 shows the relation modelled by the MLP when explaining the first factor's sensitivity in terms of Size, Gross Funds, and Long Term Debt. These are easy to explain relations. Each line is an industrial group. Sales, as seen, shows a similar pattern.

The first factor rewards positive features of the firm and penalizes Debt. None of the displayed curves deserved from the MLP the engagement of many free parameters.

Figure 77 on page 233 shows the most linear relations with the sensitivities to the second factor. They reward Wages, Debtors and penalize Earnings. But it is clear that some industries are more sensitive than others. Electronics and Paper are slightly penalized. Notice that, since we are working with size-adjusted variables, the meaning of Wages or Sales being large is that they are large for the expected given the size of the firm.

The third factor's sensitivities reward Net Worth and penalize Fixed Assets for some industries. And they reward Wages if near the expected, penalizing both higher and smaller ones (figure 78 on page 234). Both the Net Worth and the Fixed Assets of the Electronics industry seem not to be affected by this force. On the contrary, its Wages are penalized while this is not the case for the other industries.

Finally, the fourth factor's sensitivities penalize Size and Debtors and reward Inventory (figure 79 on page 235). But Inventory and Debtors of the Electricity industry seem not to be disturbed. The most affected industries are Electronics and Food.

So far, we described the linear models, that is, those in which the sensitivities of industrial groups react in the same direction for the same force.

Higher order effects: If we examine the number of parameters engaged by the MLP we notice that, except for Size, none of the relations commented above deserved the use of many degrees of freedom. Size needed more degrees of freedom because it is a very strong non-linear though monotonic relation.

We now study a few cases in which the MLP put a great deal of effort. The first one is displayed in figure 77 on page 233. It is clear that the first factor's sensitivities reward Debtors for industries like Textiles or Paper and penalize it for industries like Food Manufacturers. It also seems as if they would reward Wages in the case of industries like Textiles, Electricity and Food whilst penalizing it in the Electronics industry. Chemicals and Building Materials seem not to be affected.

The interaction of Current Assets with the first force is particularly awkward. It seems as if there is a penalization of the values expected for a given size in industries like Chemicals and Building Materials along with a rewarding of the same situation in Electronics and Food.

Factor 2 also displays this non-linear kind of behaviour. Figure 81 on page 237 shows the complicated pattern of its relation with Net Worth and Size. Again, industries seem to be affected in different directions by the same market force. It rewards Net Worth if it is reasonable for the size of the firm in industries like Textiles and Paper. But it penalizes the same situation if the industry is Food. In the Electronics industry this force simply rewards capital.

A standard size is rewarded by this factor if the industry is Textiles. It is penalized if it is Food. Small firms are, in general, rewarded by this factor. But there is one exception, Textiles. On the contrary, in the case of Electronic firms, the smaller the better — with regard to the second factor.

The remaining two factors didn't show complex relations of this sort.

9.6 Conclusions: Towards a Taxonomy of Risk

The sensitivities of assets to unanticipated market forces seem to be closely related to features identifiable and measurable in assets. Size and industrial grouping are the characteristics of assets which mostly explain sensitivities. But other stable ones explain sensitivities as well. This is the case for deviations from the expected for size in Wages, Net Worth and Sales, along with the short-term structure. However, the relation between sensitivities and features is complex. The industrial grouping interacts with features like Net Worth, Current Assets or size. The same market force reacts in opposite directions to the same feature in different industries.

The sample: The use of the MLP and its ability to model higher order relations, along with our framework, seem able to achieve a significant improvement in our knowledge of the way the market recognizes and rewards features of assets. However, the results were obtained from less cases than desirable and further experiments involving more firms should be carried out before practical uses were to be attempted.

The fact that our sample of firms is a dated one — it belongs to a period of stability and expansion — seems fortunate to us. It is clear that the relation we explored would be more difficult to model with data from difficult periods.

The firms used are also from a very particular set. They are large and frequently traded. They were drawn from well known and homogeneous industries. Therefore, the results should be always referred to such set. The contingency of the selected sample is a desired advantage. This problem couldn't be equated with all generality. It is a problem to be solved one piece at a time.

How to explain the high R^2 obtained: There are many studies of this kind in the accounting and finance literature. Though they found significant correlations between accounting and market variables, they also convey the general impression that the explained variability is not very high. Why did our results find high R^2? The main reasons seem to be:

- We defined an input space based on the framework described in the first part of this study. It may well be that such inputs are suited for statistical modelling to an extent so far unattained by other studies.
- We used homogeneous samples obtained from a stable period of growth. The homogeneity of the sample is granted by a background study (chapter 5). Too particular industries were

removed from the experiment.

- The firms allowed in the sample are large and frequently traded. It is known that the co-variance of such assets with market forces is larger than the expected.
- We modelled joint trends in cross-section, not indices extracted from individual histories. Thus we avoided using pieces of information the market regards as predictable.
- We used non-linear tools for modelling relations which are complex. The improvements over the linear ones were in some cases more than 10% of extra variability explained.
- We modelled four different points of view, not just one. This allows a better scanning of the variability to be explained.

It seems now clear that there is room in finance research for the use of MLPs. They can apportion very significant improvements in the amount of explained variability in some crucial problems. This is because they are able to model higher order relations without damaging generalisation.

Stability and unanticipated movements: We would like to underline the fact that most of the existing studies on this subject use earnings and dividends and their standard deviations to predict Betas. But these features are case-dependent and to some extent expected. They rely upon the quality of management and on his policy. They have little in common with the unanticipated components of overall forces — the co-variances with the market movement. Betas are about unanticipated overall forces, not about management policy or success. It is natural to find small correlations between these two variables.

We think that the way investors react to really unanticipated changes is very dependent on a few stable characteristics of assets. An accounting proxy for Beta should be searched mainly amongst those features which actually are sensitive to unanticipated movements, the intrinsic ones.

A taxonomy of risk: Our results show that it is possible to build, for each group of assets, a real taxonomy of risk in terms of their characteristics. Expected returns can then be calculated on the basis of such characteristics allowing a less blind discounting of future case-flows.

For example, if a productive commitment is to be undertaken, the models produced by the MLP could be fed with its budgeted numbers. As a result they would yield four sensitivities. Then the APT would predict its return based on these sensitivities.

The identification of market forces: This experiment also shows that it is possible to achieve a better understanding of the forces impinging upon the market by examining the features of firms and industries affected by each force. A further exploring of this particular subject is not in the main line of this study.

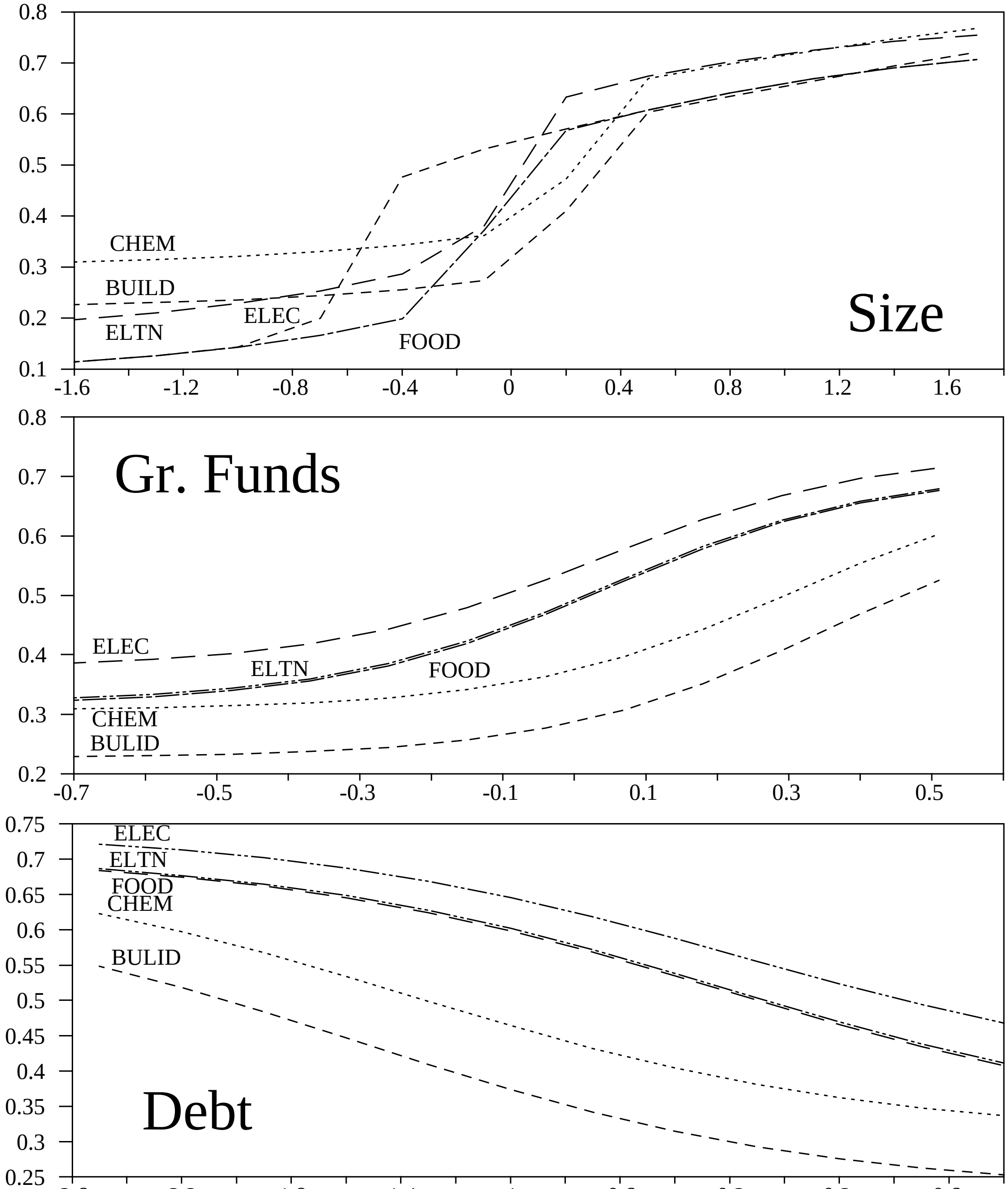

Figure 76: Other things being equal, the model obtained with the MLP predicts these relations between accounting features (X-axis) and sensitivity to the first factor (Y-axis).

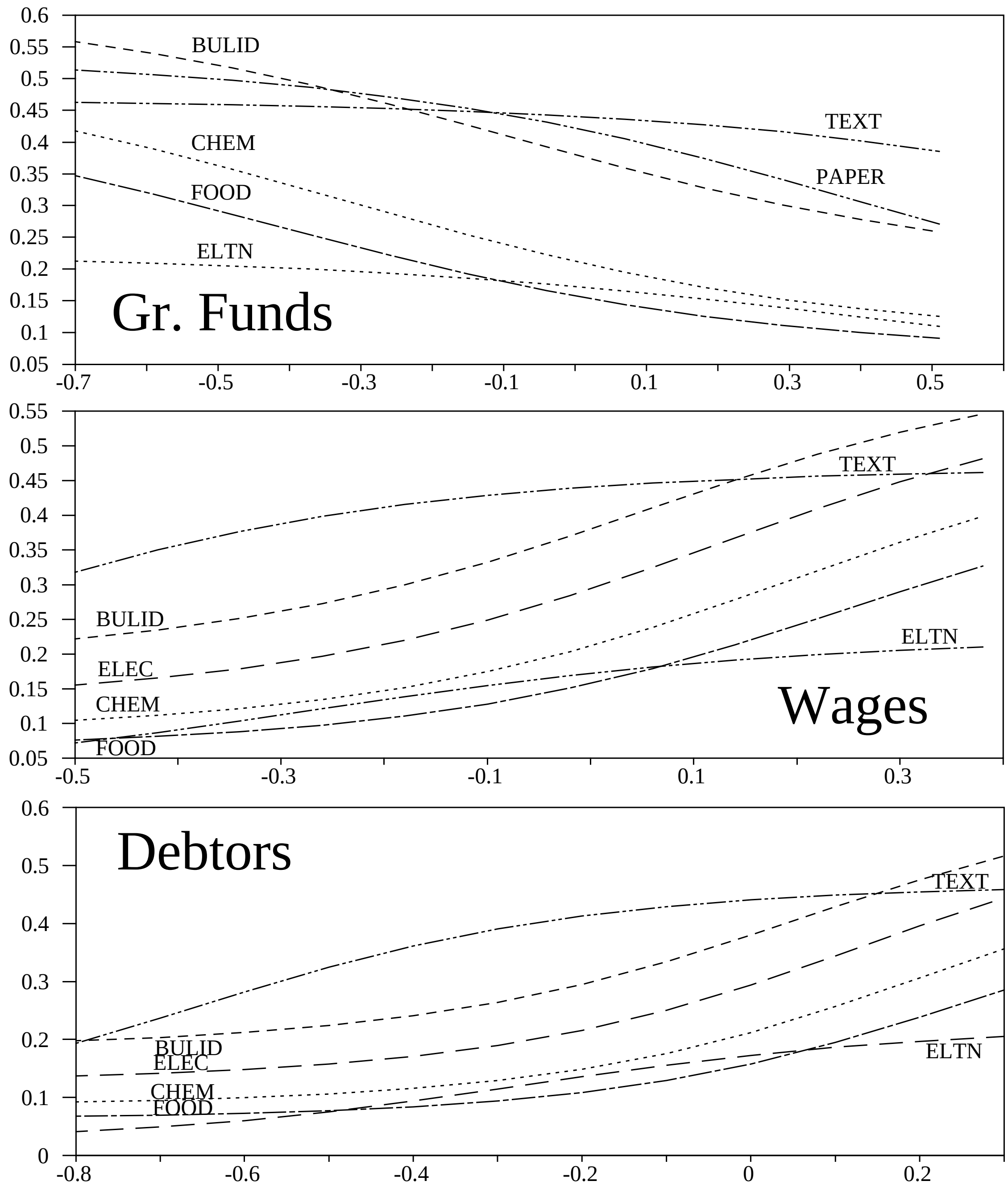

Figure 77: Other things being equal, the model obtained with the MLP predicts these relations between accounting features (X-axis) and sensitivity to the second factor (Y-axis).

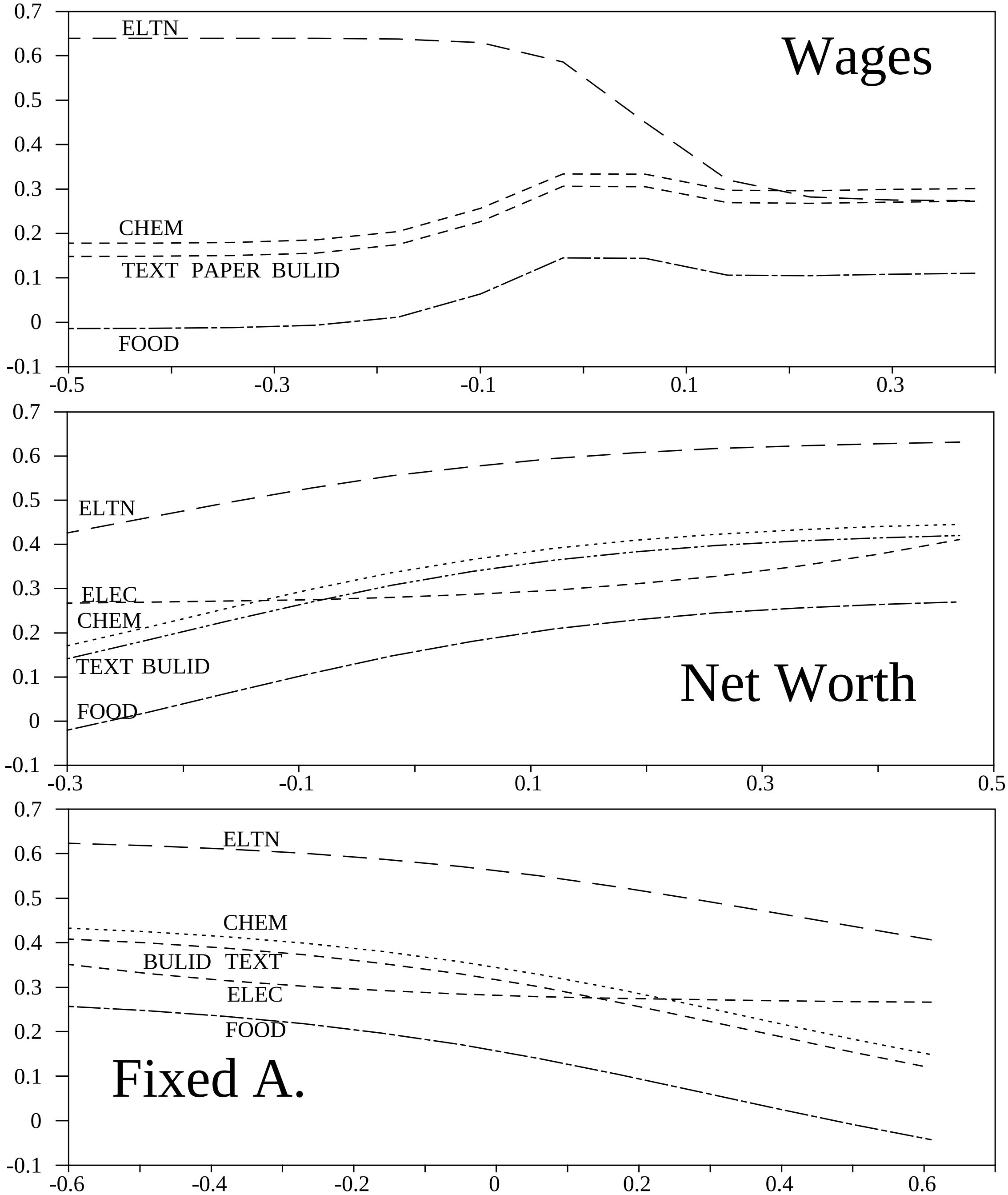

Figure 78: Other things being equal, the model obtained with the MLP predicts these relations between accounting features (X-axis) and sensitivity to the third factor (Y-axis).

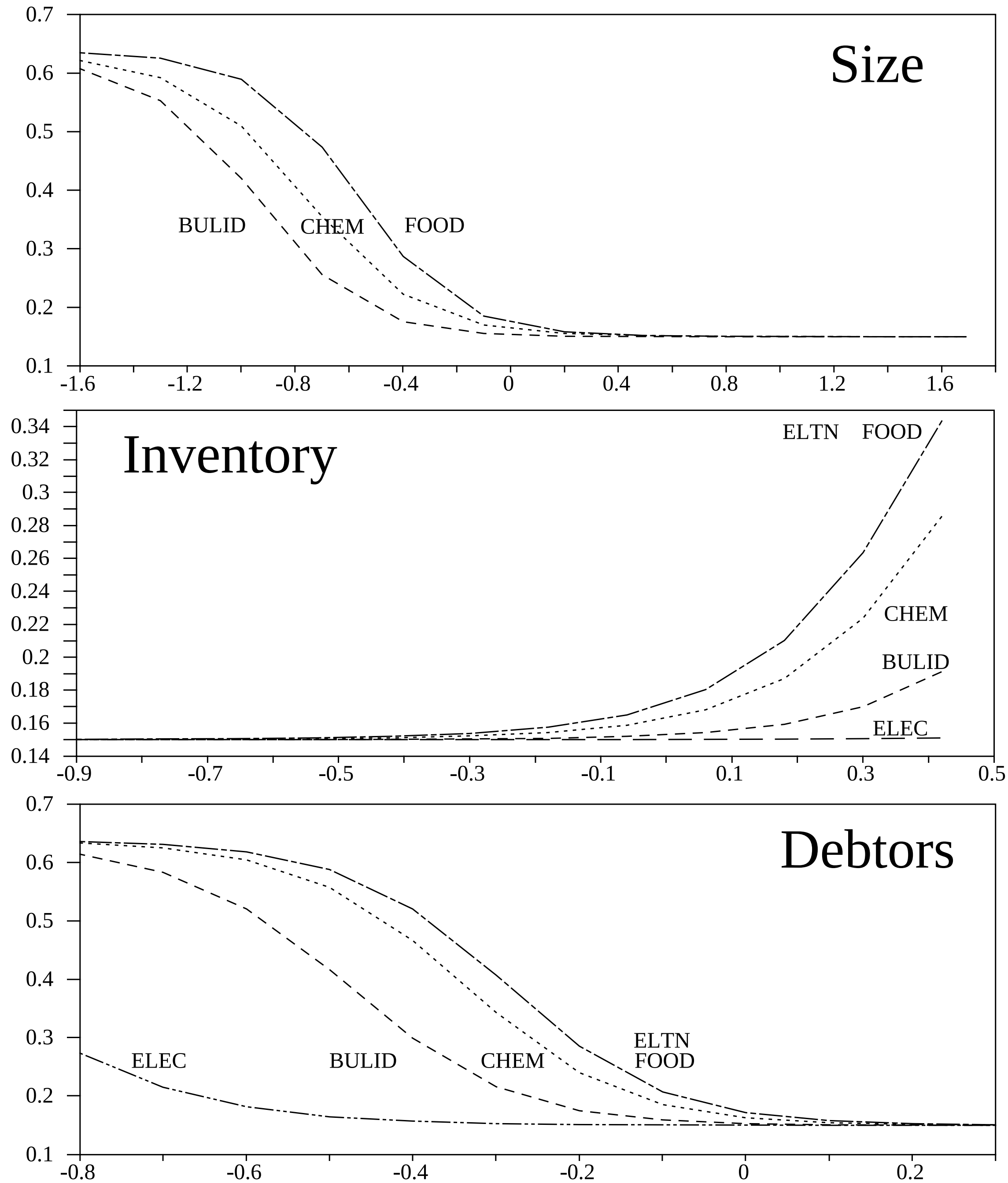

Figure 79: Other things being equal, the model obtained with the MLP predicts these relations between accounting features (X-axis) and sensitivity to the fourth factor (Y-axis).

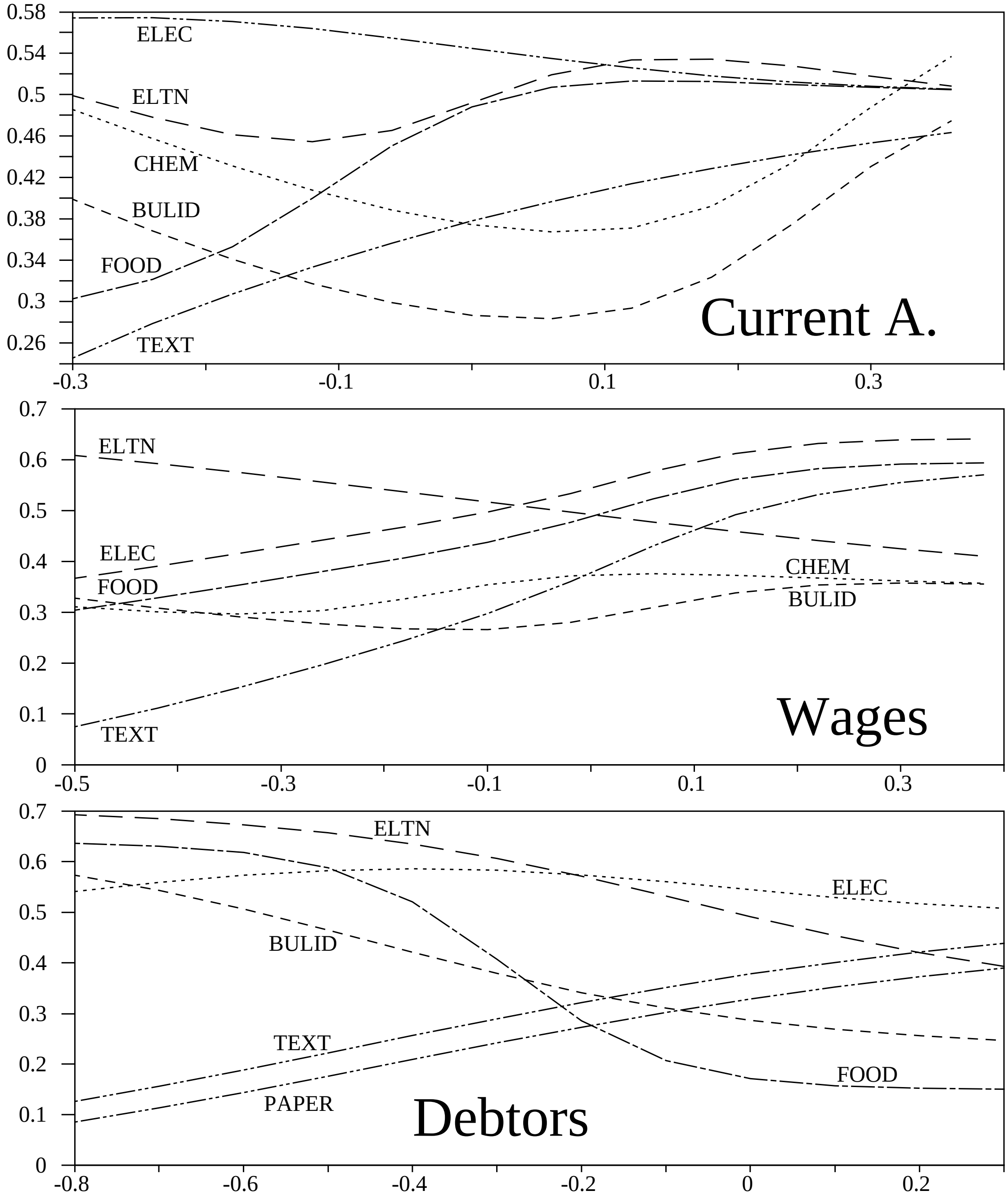

Figure 80: Other things being equal, the model obtained with the MLP predicts these relations between accounting features (X-axis) and sensitivity to the first factor (Y-axis).

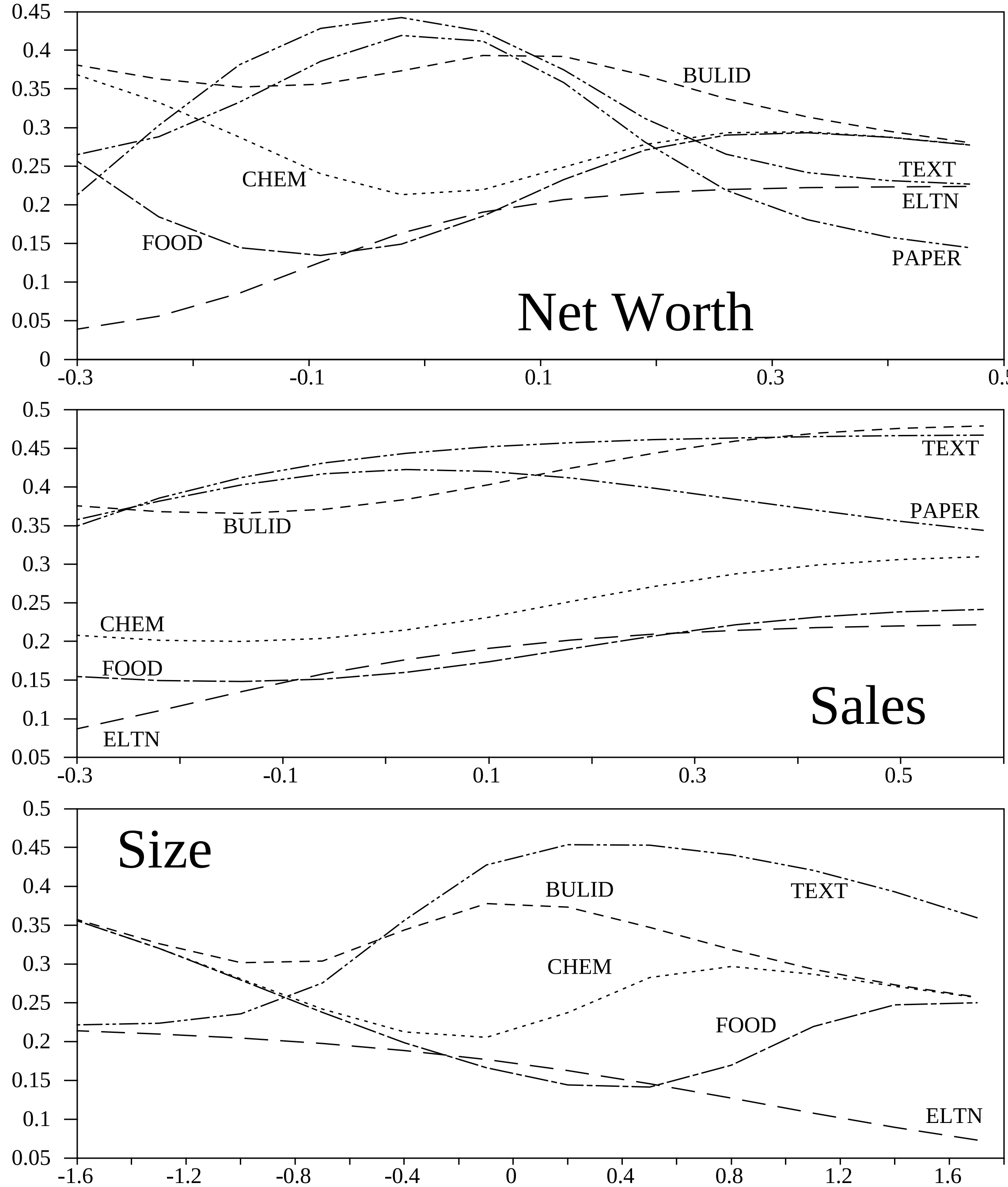

Figure 81: Other things being equal, the model obtained with the MLP predicts these relations between accounting features (X-axis) and sensitivity to the second factor (Y-axis).

ind	Name	Commonality	Size	Factor 1	Factor 2	Factor 3	Factor 4
BUILD	BPB INDUSTRIES PLC	63.6%	1.18	0.604	0.305	0.381	0.138
	EVERED PLC	24.1%	0.38	0.209	0.127	0.043	0.045
	EXPAMET INTERNATIONAL	44.7%	0.14	0.329	0.474	0.130	0.188
	HEPWORTH PLC	65.7%	1.04	0.493	0.433	0.361	0.156
	HEYWOOD WILLIAMS GROUP	40.3%	0.38	0.201	0.416	0.136	-0.030
	JOHNSTON GROUP PLC	45.4%	0.21	0.331	0.215	0.315	0.375
	MARLEY PLC	58.5%	1.15	0.599	0.295	0.283	0.138
	NEWMAN TONKS GROUP PLC	52.6%	0.40	0.321	0.358	0.264	0.309
	PILKINGTON PLC	65.3%	1.68	0.627	0.384	0.273	0.181
	STEETLEY PLC	58.8%	1.03	0.492	0.455	0.267	0.069
	TARMAC PLC	67.5%	1.51	0.656	0.259	0.322	0.041
PAPER	ASSOCIATED PAPER INDUS.	43.0%	0.12	0.269	0.209	0.256	0.062
	BLAGDEN INDUSTRIES PLC	47.9%	0.31	0.113	0.528	0.346	0.022
	BUNZL PLC	49.0%	1.09	0.463	0.310	0.281	0.204
	FERGUSON INDUSTRIAL	45.3%	0.54	0.249	0.406	0.358	0.287
	LOW & BONAR PLC	49.6%	0.77	0.279	0.578	0.015	0.234
	MACFARLANE GROUP	25.8%	0.06	0.012	0.327	0.015	0.260
	METAL CLOSURES GROUP	48.0%	0.42	0.437	0.246	0.271	0.064
	ROCKWARE GROUP PLC	43.9%	0.56	0.442	0.332	0.269	0.020
	SMITH(DAVID S.)	36.2%	-0.78	0.084	0.287	0.084	0.484
	WADDINGTON(JOHN)PLC	32.4%	0.28	0.240	0.469	0.127	0.088
CHEM	ALLIED COLLOIDS GROUP	50.5%	0.34	0.490	0.112	0.361	0.089
	BOC GROUP PLC	67.7%	1.81	0.677	0.268	0.246	0.203
	BRENT CHEMICALS INTERN.	51.8%	0.16	0.323	0.375	0.459	0.179
	BTP PLC	39.7%	-0.13	0.332	0.415	0.265	0.186
	CANNING(W.)PLC	46.6%	0.09	0.228	0.487	0.257	0.301
	COALITE GROUP PLC	55.8%	0.96	0.474	0.051	0.273	0.329
	COATES BROTHERS PLC	53.4%	0.72	0.235	0.231	0.363	-0.036
	CRODA INTERNATIONAL PLC	56.2%	0.95	0.471	0.246	0.129	0.295
	ELLIS & EVERARD PLC	52.3%	0.03	0.228	0.273	0.233	0.391
	EVODE GROUP PLC	45.3%	0.22	0.421	0.128	0.249	0.310
	FOSECO PLC	57.6%	1.14	0.600	0.370	0.135	0.150
	HICKSON INTERNATIONAL	68.6%	0.56	0.485	0.436	0.362	0.186
	IMPERIAL CHEMICAL INDUS.	64.7%	2.42	0.642	0.304	0.258	0.121
	LAPORTE PLC	62.0%	0.92	0.535	0.249	0.251	0.169
	LEIGH INTERESTS PLC	38.8%	-0.14	0.268	0.246	0.247	0.277
	RENTOKIL GROUP PLC	46.1%	0.73	0.496	0.158	0.342	0.181
	YULE CATTO & CO PLC	37.5%	0.44	0.346	0.191	-0.054	0.216

Table 36: The list of firms used in this study by industry. First table. The commonality, the size and the four factor loadings obtained is also displayed.

ind	Name	Commonality	Size	Factor 1	Factor 2	Factor 3	Factor 4
ELEC	BICC PLC	64.2%	1.73	0.603	0.274	0.361	0.075
	CHLORIDE GROUP PLC	46.8%	1.07	0.520	0.353	0.164	-0.023
	DOWDING & MILLS PLC	47.9%	-0.14	0.382	0.272	0.021	0.192
	SCHOLES GROUP PLC	44.2%	-0.04	0.131	0.110	0.132	0.039
	VOLEX GROUP PLC	44.6%	0.27	0.385	0.395	0.149	0.262
	WHOLESALE FITTINGS PLC	40.1%	-0.05	0.121	0.106	0.500	0.182
ELTN	A.B.ELECTRONIC PRODUCTS	46.9%	0.45	0.150	0.176	0.572	0.297
	BOWTHORPE HOLDINGS PLC	51.2%	0.58	0.372	0.223	0.499	0.138
	CRAY ELECTRONICS HOLD.	47.9%	0.00	0.164	0.030	0.169	0.641
	CRYSTALATE HOLDINGS PLC	24.8%	0.37	0.240	0.082	0.382	0.190
	DIPLOMA PLC	64.6%	0.38	0.370	0.217	0.637	0.109
	ELECTROCOMPONENTS PLC	61.2%	0.52	0.436	0.182	0.608	0.027
	FARNELL ELECTRONICS PLC	73.6%	0.28	0.462	0.205	0.666	0.048
	PLESSEY CO PLC(THE)	51.2%	1.66	0.516	0.036	0.445	0.146
	UNITECH PLC	59.5%	0.64	0.368	0.268	0.599	0.133
TEXT	BAIRD(WILLIAM)PLC	56.5%	0.88	0.413	0.282	0.453	0.179
	DEWHIRST(I.J.)HOLDINGS	34.7%	-0.05	0.085	0.226	0.193	0.464
	ALLIED TEXTILE COMPAN.	57.8%	0.23	0.456	0.272	0.265	0.195
	DAWSON INTERNATIONAL	49.0%	0.81	0.199	0.427	0.417	0.281
	ILLINGWORTH,MORRIS PLC	31.3%	0.40	0.294	0.332	0.111	0.177
	COATS VIYELLA PLC	48.2%	1.32	0.324	0.458	0.268	0.191
	LAMONT HOLDINGS PLC	16.2%	0.12	0.067	0.293	0.180	0.100
	READICUT INTERNATIONAL	41.6%	0.47	0.403	0.458	0.174	0.036
	SCAPA GROUP PLC	52.9%	0.65	0.485	0.287	0.276	0.303
	TOOTAL GROUP PLC	57.9%	1.09	0.557	0.362	0.288	0.155
FOOD	ASSOCIATED BRIT. FOODS	62.1%	1.75	0.639	0.188	0.301	0.189
	BERISFORD INTERNATIONAL	50.9%	1.67	0.626	0.135	0.271	0.047
	BOOKER PLC	47.8%	1.24	0.561	0.243	0.166	0.130
	CADBURY SCHWEPPES PLC	64.4%	1.71	0.699	0.197	0.177	0.271
	DALGETY PLC	61.6%	1.61	0.670	0.207	0.183	0.142
	FITCH LOVELL PLC	57.5%	0.88	0.681	0.060	0.084	0.141
	HAZLEWOOD FOODS PLC	41.1%	0.01	0.083	0.070	0.105	0.552
	MATTHEWS(BERNARD)PLC	49.0%	0.39	0.405	0.251	0.103	0.276
	NORTHERN FOODS PLC	71.9%	1.36	0.791	0.080	0.103	0.213
	RANKS HOVIS MCDOUGALL	56.3%	1.48	0.633	0.239	0.148	0.194
	TATE & LYLE PLC	53.0%	1.39	0.610	0.192	0.186	0.197
	UNIGATE PLC	62.6%	1.54	0.741	0.156	0.160	0.059
	UNILEVER PLC	64.1%	2.20	0.681	0.259	0.276	0.143
	UNITED BISCUITS	65.7%	1.57	0.686	0.155	0.278	0.173

Table 37: The list of firms used in this study by industry. Second table. The commonality, the size and the four factor loadings obtained is also displayed.

Chapter 10

Conclusions

This study attempted a development of concepts and tools for the extraction of knowledge from past experience contained in accounting and financial data. Its first part described the statistical characteristics of accounting data. Neural Networks were then used to solve three characteristic problems.

Main achievements: Our programmatic statement for the assessment of the statistical characteristics of accounting data was to study items first and then ratios. Items prove to be much more regular than ratios. The observed ones — extracted from the reports of industrial firms during the period 1983-1987 — were two or three-parametric lognormal. McLeay [86] observed lognormality in large samples of items which are sums of similar transactions with the same sign. We extended this empirical finding: Lognormality cannot be rejected also for stocks like Fixed and Total Assets or Net Worth and non-accounting items related to size like the number of employees. Positive values of accounting items having both positive and negative cases as well as the absolute value of the negative ones are lognormal too. We also gathered detailed evidence on the lognormality of homogeneous samples formed with one industry at a time.

Lognormality allowed us to explain the existence of outliers and the heteroscedasticity of accounting data often referred to in the literature. We have shown that regressions should not be used to model relations between lognormal variables and that weighting is not an adequate recipe since it simply transfers the influence from the largest to the smallest cases in the sample. Finally, we pointed out that the trimming of outliers is useless for two-variate lognormal data.

Another important empirical finding of this study is the existence of a common source of variability in the observed log items. In log space these variables are the addition of two processes. The first one is common to all items and seems to reflect the relative size of firms. The second one, particular to each item, reflects its uniqueness. Hence, items should be explained in terms of size and deviations from size. Instead of viewing each item individually — eventually correlated with other few ones — we should first account for an effect common to them all and then take the residual variability as the contribution of that item.

We also studied the problems posed to cross-sectional models by items having both positive and negative cases. There is no continuity between the positive cases and the negative ones. We remarked that negative cases should be viewed as a different group.

Based on these findings we extended ratios so as to cope with non-proportionality and non-linearity. We based our approach on the existence of a common effect and on the three-parametric lognormality observed in a few items. Three extensions of the ratio concept were developed: Firstly, ratios can have more than two components. The sole requirement for the statistical validity of such ratios is the use of multiplicative residuals. Next, ratios can also be viewed in log space as a regression. Such free-slope ratios preserve proportionality. Finally, the three-parametric lognormality present in a few samples leads to base-line ratios. Base-line ratios account for the non-proportionality often mentioned in the literature. They seem promising for ratio analysis and statistical manipulation. They are robust, easy to estimate and it is likely that they will be able to gather in one unique relation financial features of firms with very different sizes.

Finally, we remarked that the reasons often invoked in the literature for expecting significant intercept terms in cross-sectional samples don't lead to non-proportional relations. Only an overall cost or income impinging upon the whole of the sample is able to yield non-proportionality. This overall base-line couldn't be very far away from the smallest case in the sample. And the effect of such a translation would not be noticeable except for small firms.

We also studied the distribution of ratios. We found a clear trend towards positive skewness, as expected. However, a few factors affect the distribution particular ratios assume. Firstly, accounting identities and other external forces act as constraints, hiding its skewed distribution. This explains why ratios like NW/TA and TD/TA are so often reported in the literature as being near normality. Secondly, when observing the multiplicative residuals in log space, leptokurtosis becomes visible. We identified the particular variability of each item as the source of leptokurtosis in accounting data. The source of their Gaussian behaviour is the strong effect common to all items. Next we argued that ratios are ordinal and ratio standards are not affected by any anomalies in their distribution because they only use one degree of freedom. No consideration of the spread of items is required to model with ratios. Conversely, no disturbances in their spread can affect ratio standards.

Before finishing this part we studied the problem of building an estimator of the common effect. Such a general deflator can enhance the interpretability of results in statistical models. We have shown that simple case-averages of selected items approach the common effect. We also discussed the reduction of the dimension of the input space in statistical models. We suggested the use of the Hadamard rotation, able to isolate the common effect and re-distribute the remaining variability by a number of factors. Finally, we studied the importance and effect of the SEIC industrial grouping using intra-class correlations. Both the spread of size and the one of financial features of firms are dependent on the group to which each case belongs. We identified higher order effects in the space of firm's features, demanding the use of algorithms able to model them.

In the second part of this study we have shown that Neural Networks are self-explanatory tools when extracting knowledge from accounting data. Based on a known problem, the discrimination between industries using accounting numbers, we enhanced a Multi-Layer Perceptron so as to form in its hidden units ratios appropriate to model that relation. The MLP also proved able to outperform the classification of a traditional discriminant analysis approach. This performance was achieved with half the number of inputs and within a much simpler framework. Namely, the search for appropriate ratios, the pruning of outliers and the extraction of a somehow arbitrary number of factors were avoided.

Next, Kohonen's Maps were used to automate financial diagnosis. Firstly, we explored the graphical possibilities offered by the homogeneity of two-variate relations in log space. We outlined the two kinds of complementary questions ratios are called upon to answer and we described graphical tools able to give joint answers to those questions. Secondly, Kohonen's Maps performed quantization of these maps, allowing the assigning of a diagnostic to each region in such graphical tools. The devised plots can be used as direct tools for diagnosis in the same way financial ratios are. But they yield a richer information, incorporating relative size — or, alternatively, deviations from the expected for a given size — and allowing the study of trajectories. The examination of trajectories instead of simple trends is potentially revealing for financial analysts. Size is important in specific problems like the prediction of firm failure. The set of rules automatically generated in this fashion can be seen as the result of a pre-processing for symbol-based expert systems.

The last experiment of this study was meant to test the ability of Neural Networks to model complex maps like the relation between the expectations of investors and the characteristics of traded assets. Using a particular kind of asset, the large and frequently traded industrial firms in the U.K. we extracted sensitivities from the time-history of 121 returns. Then, we modelled the relation between these sensitivities and accounting information using Neural Networks. Finally, we studied the behaviour of the obtained models. We found a clear relation between features of accounting reports like size, industrial group, work force, payment pattern and a firm's appraisal by the market. It turns out that a very significant portion of the cross-sectional variability of these sensitivities can be explained by accounting numbers. The developed method opens the possibility of building a taxonomy of risk. Also, the knowledge about the way each market force impinges upon specific features of the firm is potentially valuable in the identification of the nature of such forces.

Limitations and other negative aspects: The second part of this study would require further research. Namely, it would be interesting to replicate the building of ratios by MLPs using other problems. Apart from this, we didn't succeed in incorporating base-lines into the Back-Propagation algorithm and in further improving the interpretability of the ratios discovered by the MLP. Also the study of two-dimensional tools and the automatic extraction of rules would have benefitted from a diversification of the features explored.

Directions for future research: This study opens up a large field of rich possibilities for future research. In the first place, it is important to find out to what extent the promising regularities observed in accounting items — lognormality and the strong, common, effect — can be extrapolated to different samples: Small or non-industrial firms, items other than the observed ones.

Base-lines should be searched and related to other characteristics of the firm. Their internal mechanism, if any, should be sorted out. This should be done using samples containing firms covering very different sizes. The effectiveness of base-line ratios for setting standards could then be established. The leptokurtosis observed in ratios would be, in our opinion, the next subject of investigation.

It seems also promising to try and build a proxy for the common effect with case-averages containing many items. A reliable estimation of the common effect would allow the exploring of residuals, that is, the particular contribution of each item to the overall variability.

The logarithmic nature of accounting data allows a systematic study of ratios: The way they behave inside industries, the stability of their expected values, the search for the most promising ones for specific tasks, the way their information can be complemented by other ratios, the ones more affected by base-lines and in what industries, their range of application to firms of different sizes and so on. In the same line, the behaviour of industrial groups should be studied one by one.

Neural Networks deserve further research. Aspects related to ours are the direct modelling of base-lines by the Back-Propagation algorithm, the study of learning procedures aimed at enhancing the interpretability of the ratios formed in the hidden layers, the testing of more effective learning techniques and distance measures in Kohonen maps. The possibility of drawing a taxonomy of risk as shown in the last chapter of this study should, of course, be explored. Larger learning sets would allow the building of more detailed and reliable models.

Part III

Appendices

Appendix A

The Statistical Description of Accounting Items

This appendix contains detailed results or developments related to the first part of our study. They were relegated to an appendix because they would break the sequence of the presentation should they remain in the main text.

Some of the sections presented here are self-contained and eventually important for the understanding of the main body of research. Therefore, this appendix should not be regarded as a simple storage place for large tables.

However, large tables do exist as well. And owing to the need for printing a huge amount of such tables the organization of this appendix doesn't follow the sequential order of the main text. Subjects requiring the displaying of many tables alternate with others not demanding it. Even so, some sets of tables had to be placed at the end.

A.1 Persistency of Deviations from Two-Parametric Lognormality

We measured the number of times a significant departure from a two-parametric lognormal distribution was observed during the whole period of five years for a given sample. Such a measure can give us an idea of the persistency of two-parametric lognormality.

Tables 38 on page 246, and the one on next page show, by items and by industrial group, the number of times a significant departure from a two-parametric lognormal distribution was observed during the five-years period.

As an example, Sales had 3 groups which exhibited departures once in five years, another group had three departures in the same period and finally there were two industrial groups which were

Item:	ONCE	TWICE	THREE TIMES	FOUR TIMES	FIVE TIMES
Sales	3		1	2	
Net Worth	2	3			
Wages	2	1			1
Inventory	3		1		
Debtors	1	3			
Creditors	2	1		1	
Current Assets	3	3			
Fixed Assets	1	2			
Total Assets		1	3		
Current Liabilities		1	3		
Number of employees	1	1	1	1	
Expenses	2		1	1	
Tot. Capital Empl.		3		1	
EBIT	2	2			
Operating Profit	2	2			
Long Term Debt	3	1		1	
Funds Flow Fr. Ops.	6	1	1		
Working Capital	3	2			

Table 38: Persistency of departures from the two-parameters model. By item.

non-significantly two-parametric in four of the five years considered. And the industrial group Wool had two variables which departed from the two-parametric hypothesis once in five years. Working Capital — in three industrial groups — departed once in five years. Two other groups departed twice. And when considering groups instead of items, metallurgy had no departures at all, Building Materials had three items which departed once in five years and one item which departed twice. And so on.

As we can see departures are sporadic. They hardly occur in more than one or two years. Only Wages (Electronics) is persistently three-parametric lognormal.

Apart from the 20 cases known as "bad cases", all the departures from the two-parametric model are three-parametric lognormal: There is always a small δ for which the Shapiro-Wilk W becomes non-significantly different from 1.

A.2 Simulation of Working Capital

The distribution of x, a desired p-variate normal deviate, can be represented as a linear transformation of p independent normal variates $g = (g_1, \cdots, g_p)^{'}$ as $x = Ag + \mu$.

A is any $p \times p$ matrix for which $A\ A^{'} = \Sigma$. This matrix is not unique. If, for instance

$$\Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$$

a 2×2 correlation matrix, then a particularly simple A yielding $A\ A^{'} = \Sigma$ is

$$\begin{bmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{bmatrix}$$

Industrial Groups	ONCE	TWICE	THREE TIMES	FOUR TIMES	FIVE TIMES
Building Materials	3	1			
Metallurgy					
Paper and Packing	3	1	1		
Chemicals	4	1	1		
Electricity	3	3	2	1	
Industrial Plants	7	9	1	1	
Machine Tools	4				
Electronics	2	7	3	1	1
Motor Components	2				
Clothing	3	1			
Wool	2				
Textiles Mix.					
Leather					
Food Manufacturers	3	4	3	4	

Table 39: Persistency of departures from the two-parameters model. By industry.

In general, a possible choice for A is provided by the Cholesky factorization, that is, the lower triangular matrix A for which $A\,A^{'} = \Sigma$. In multi-variate simulations we used such procedure.

In the case of a two-variate distribution of x and y it yields a simple expression. For a desired ρ and using g_1, g_2 independent random deviates,

$$\begin{cases} x = & E(x) & +g_1 \\ y = & E(y) & +g_1 \times \rho + g_2 \times \sqrt{1-\rho^2} \end{cases}$$

$E(x), E(y)$ are the expected values of the variates we want to simulate. This simple manipulation can be carried out easily.

In the first mentioned simulation we used these starting values

$$\begin{aligned} \rho &= 0.98043 \\ \overline{\log CA} &= 4.343 \\ \overline{\log CL} &= 4.192 \end{aligned}$$

inspired in the parameters observed in the Electronics industry (1987). We simulated 2000 values and observed the lognormality of the resulting $WC = CA - CL$ in two separate cases, positive deviates and absolute values of negative deviates. The results are displayed in next table.

Statistic	CA for $WC < 0$	CL for $WC < 0$	Positive WC	Negative WC
SKEW	-0.217	-0.242	-0.055	-0.471
KURT	0.592	0.753	0.301	1.106
W	0.9837	0.9839	0.98	0.9838
sig W	non-sig.	non-sig.	non-sig.	non-sig.
N. Cases	463	463	1537	463

Therefore simulated results agree with empirical observations that absolute values of negative accounting items are lognormally distributed.

The following is an example of multi-variate simulation.

We based the starting values in the building materials group (1983) and we selected the four variables displayed next.

Item	Mean	St. deviation	Item	Mean	St. deviation
Sales	4.6253	0.605	Current Assets	4.2339	0.583
Fixed Assets	4.0532	0.599	Current Liabilities	4.0273	0.607

The Σ used in this case was the variance and co-variance matrix: $\begin{bmatrix} S & CA & FA & CL \\ 0.366 & & & \\ 0.344 & 0.339 & & \\ 0.341 & 0.324 & 0.359 & \\ 0.353 & 0.339 & 0.327 & 0.369 \end{bmatrix}$

After generating 2,000 cases with multivariate lognormality we obtained 211 of them with negative Working Capital. The mean values and standard deviations for such item, as well as the skewness and kurtosis were

	Mean	St.d.	N. cases	Skewness	Kurtosis
for $WC > 0$	3.754	0.67	1789	-0.162	-0.023
for $WC < 0$	3.272	0.85	211	-0.442	0.551

As seen, both positive and negative cases were lognormal. The resulting co-variances are

```
  S     FA    CL    CA    WC      S     FA    CL    CA    WC

0.367             for WC > 0  0.384            for WC < 0    S
0.346 0.369       1789 cases  0.348 0.355       211 cases   FA
0.348 0.327 0.354             0.371 0.335 0.376             CL
0.344 0.327 0.335 0.338       0.370 0.335 0.370 0.369       CA
0.336 0.325 0.300 0.341 0.450 0.396 0.350 0.419 0.386 0.722 WC
```

The results of simulation agree with the few observed real cases. Negative cases also exhibit the common effect but they spread more than the positive cases.

A.3 Description of the Extreme Departures From Lognormality

We pointed out in chapter 1 that a few tests of lognormality yielded values of P (Shapiro-Wilk's W significance) which were so small that it would not be possible to apply logits.

We also noticed that such a set of bad cases behave differently from the other ones. The significance (P values) obtained from all other tests form in logit space a normal distribution. Bad cases do not fit well in such a distribution. They are more numerous than the expected and they form a cluster sticking out well below the lower normal values of *Logit* P.

They are also insensitive to a three-parametric transformation. No δ exists able to turn them lognormal. This feature is a very particular one since in all the other cases yielding significant departures a δ exists able to bring W to non-significant values.

We examined each one of such bad cases in order to find out the reasons for their erratic behaviour. Here we present the results.

1983: In 1983 there were no bad cases.

1984: There are bad cases in two industries.

- **CLOTHING:** Sales and Expenses. The firm STORMGARD PLC sold 54 (units are thousands of pounds) and had also very small expenses. The sample has only 46 cases and the next smallest value of sales is 2402. STORMGARD turns out to be a strong outlier in a very small sample. We further notice that this firm has sales which are larger than earnings.

- **ELECTRONICS:** Wages and Number of Employees. There are three clear clusters. A cluster of eight large firms is clearly detached from the rest of the distribution. It is interesting to notice that neither the skewness nor the kurtosis exhibit values far from the acceptable.

1985: There are bad cases in three industries.

- **PAPER AND PACKING:** Current Assets and Expenses. In 57 cases there is one firm, EAST LANCASHIRE PAPER GRO, with $CA = 1$. The next smallest value in the sample is 1265. The same as Expenses. EAST LANCASHIRE is also one of the 5 firms exhibiting EBIT larger than Sales during one or two years of the period.

- **FOOD:** Current Assets. There are three clear clusters. Again, Skewness and Kurtosis are unable to trace the irregular shape of this distribution.

- **CLOTHING:** Operating Profit. The firm UNIGROUP PLC appears in the database with $OPP = 1$, and such a profit turns out to be a strong outlier in a small group. The next smallest case is $OPP = 52$. The sample has 44 cases.

1986: There are bad cases in three industries.

- **ELECTRONICS:** Wages, Number of Employees and Current Liabilities. Again three clusters, large firms well separated from the distribution. Skewness and kurtosis are normal.

- **INDUSTRIAL PLANTS:** Inventory. BIMEC PLC has $I = 1$. Next smallest value, 278. In a sample of 22 cases this is enough to influence normality tests.

- **FOOD:** Long term Debt. Very clear three-modal distribution.

1987: There are bad cases in three industries.

- **FOOD:** Sales, Expenses and Earnings. Again, three very clear clusters but in this case the cluster of small firms is detached from the others. Then, there is also a peaking central cluster. Skewness and kurtosis again fail to trace the lack of normality. In this group there are four firms with EBIT larger than Sales.

ELECTRONICS: Wages, Number of Employees and Total Capital Employed. There are three clusters as in two previous years. But the group of eight very large firms is now less detached from the others than it was in previous years.

CLOTHING: Funds Flow From Operations. The firm GOODMAN GROUP PLC displays a $FL = 9$ which is a clear outlier. The next smallest cases have $FL = 339$. The sample has 45 cases.

In short: The causes for the existence of the described bad cases seem to be twofold:

Anomalous Cases. Errors, extreme outliers, very particular situations.

Non-Homogeneous Groups. The existence of clusters of firms well detached inside the same industrial group is perhaps a result of a temporary expansion of the sector. Or it can be a consequence of an intrinsic non-homogeneity of an industrial group. It happens from 1984 to 1987 in the Electronics and Food industries, affecting Sales, Number of Employees and Wages mainly.

We conclude that there seems to be an explanation or cause for each one of the observed strong departures from the lognormal hypothesis. These causes should be considered as external to the generative mechanism governing the cross-sectional characteristics of accounting items since their frequency, 20 cases in 1260 different samples examined, makes them exceptional.

A.4 Results of Lognormality Tests

A.4.1 All Groups Together

In this section we display and comment the detailed results obtained when measuring the kurtosis, the skewness and the Shapiro-Wilk's W of each of the 13 positive-valued accounting items and the positive values of the 4 items having both positive and negative cases, during a period of five years.

We also include Long Term Debt for which only the non-zero cases were selected.

Such results are contained in three tables: Table 40 on page 251, for the items Sales, Net Worth, Wages, Inventory, Debtors and Creditors. In the one on page next to this one, for Fixed Assets, Total Assets, Current Assets, Current Liabilities, the number of employees and Expenses. Finally, in the next one for Total Capital Employed, Earnings, Operating Profit, Long Term Debt, Gross Funds From Operations and Working Capital. The items selected for testing represent very different situations and manipulations of data.

Item	Year	1983	1984	1985	1986	1987
	N. Cases	555	649	677	702	688
Sales	SKEW	0.058	-0.035	0.126	0.107	-0.08
	KURT	0.765	0.842	0.604	0.39	0.854
	W	0.982	0.983	0.984	0.983	0.982
	sig W					
	best W					
Net Worth	SKEW	0.174	0.163	0.253	0.289	0.289
	KURT	0.468	0.402	0.255	0.263	0.353
	W	0.987	0.985	0.984	0.983	0.983
	sig W					
	best W					
Wages	SKEW	0.322	0.246	0.384	0.35	0.239
	KURT	0.177	0.252	0.067	0.091	0.266
	W	0.978	0.984	0.975	0.977	0.98
	sig W	0.015		0.001	0.01	0.02
	best W	0.988		0.988	0.989	0.981
Inventory	SKEW	-0.089	-0.117	0.019	-0.202	-0.302
	KURT	0.69	0.577	0.572	1.217	1.331
	W	0.985	0.986	0.989	0.986	0.985
	sig W					
	best W					
Debtors	SKEW	0.052	-0.003	0.066	0.126	-0.036
	KURT	0.309	0.411	0.386	0.318	0.76
	W	0.984	0.987	0.986	0.987	0.991
	sig W					
	best W					
Creditors	SKEW	0.285	0.176	0.236	0.242	0.196
	KURT	0.307	0.331	0.356	0.289	0.369
	W	0.978	0.983	0.981	0.979	0.982
	sig W	0.014		0.06	0.006	
	best W	0.985			0.985	

Table 40: Lognormality of all groups together. First table.

Samples were drawn directly from the Micro-EXSTAT data-base. No pre-conditions were established, apart from the one of firms being in the U.K. None of the cases was considered as an outlier. Therefore, samples are quite representative.

The number of cases in each sample is displayed in the tables referred to along with the other statistics.

The results show that 11 of the 18 items — Sales, Net Worth, Debtors, Fixed Assets, Expenses, Inventory and Total Capital Employed, along with the positive values of Earnings, Operating Profit, Long Term Debt and Working Capital — are two-parametric lognormal in the whole period of 1983 to 1987.

The remaining 7 variables are either two-parametric or three-parametric lognormal depending on the year. None is persistently three-parametric during the five years. In at least one year all variables achieved lognormality with just a simple log transformation.

The most three-parametric cases are Total Assets and Wages with four years in five requiring a three-parametric transformation. In general, the positive values of McLeay's Δ variables are more near two-parametric lognormality than the Σ ones. Only Gross Funds From Operations exhibit one departure from a two-parameters distribution, in 1987.

Only when $P < 0.05$ is very near this value we display its value (sig W). In such cases we also show the value of W obtained by introducing an optimal δ in the log transformation (best W).

The algorithm for assessing skewness and kurtosis are those adapted by the SPSS-X package.

(cont.)		1983	1984	1985	1986	1987
Fixed	SKEW	0.097	0.177	0.119	0.124	0.159
Assets	KURT	0.421	0.11	0.114	0.113	-0.008
	W	0.988	0.983	0.984	0.987	0.984
	sig W					
	best W					
Total	SKEW	0.301	0.351	0.404	0.343	0.425
Assets	KURT	0.546	0.276	0.228	0.349	0.309
	W	0.983	0.979	0.978	0.979	0.978
	sig W		0.01	0.005	0.01	0.005
	best W		0.985	0.987	0.983	0.984
Current	SKEW	0.237	0.349	0.056	0.295	0.345
Assets	KURT	0.372	0.374	1.84	0.345	0.48
	W	0.982	0.984	0.985	0.98	0.979
	sig W				0.03	0.02
	best W				0.985	0.985
Current	SKEW	0.26	0.155	0.21	0.273	0.262
Liabilities	KURT	0.366	0.446	0.462	0.36	0.417
	W	0.979	0.983	0.984	0.981	0.982
	sig W	0.026			0.04	
	best W	0.986			0.988	
Number	SKEW	0.171	0.191	0.283	0.221	0.159
of	KURT	0.282	0.181	0.187	0.251	0.346
Employees	W	0.981	0.982	0.98	0.981	0.981
	sig W			0.02	0.04	0.05
	best W			0.985	0.985	0.983
Expenses	SKEW	0.093	-0.108	-0.043	0.012	-0.124
	KURT	0.327	0.738	0.742	0.29	0.641
	W	0.981	0.986	0.988	0.985	0.984
	sig W					
	best W					

Table 41: Lognormality of all groups together. Second table.

One interesting feature of the distribution of the observed items is the positive nature of its kurtosis. With very few exceptions the values obtained for the kurtosis are positive. We saw in section 4.1 (page 90) that the distribution of the ratio residuals is characterized by the same feature but magnified.

A.4.2 By Industrial Group

Tables 44 and the next five tables contain the detailed results of the tests of lognormality of the observed eighteen items by industrial group and by year. They can be found on pages 259 and the next five pages towards the end of this appendix.

Each table displays, for a particular item, the number of cases, the Shapiro-Wilk's W, and the associated probability P. This P should be interpreted as the likelihood of obtaining a W as small as that observed, when, in the population from which the sample was drawn, W would have the value of 1. Since a W of 1 means lognormality, any small P will denote a departure from the lognormal hypothesis. It is usual to reject the null hypothesis of no departure from lognormality for $P < 0.05$. In this case we know that there are less than five chances in one hundred that some hazardous circumstances of sampling would lead to a value of W as small as the observed one despite our sample being drawn from a lognormal population.

When P is displayed as having a value of 0 it means that P is smaller than the precision of the algorithm. Values of 0.00 mean a $P < 0.005$. The values of $P = 0$ are denoted in our study as the

(cont.)		1983	1984	1985	1986	1987
Total	SKEW	0.28	0.156	0.322	0.338	0.34
Capital	KURT	0.35	0.449	0.126	0.18	0.282
Employed	W	0.9828	0.9854	0.9819	0.9823	0.9815
	sig W					
	best W					
EBIT	SKEW	-0.061	0.094	0.165	0.232	0.305
	KURT	0.678	0.244	0.409	0.402	0.451
	W	0.9846	0.9887	0.9841	0.9854	0.9823
	sig W					
	N. Cases	514	606	629	645	641
Operating	SKEW	-0.097	-0.053	-0.13	0.128	0.176
Profit	KURT	0.68	0.36	0.81	0.215	0.469
	W	0.9898	0.9918	0.9854	0.9843	0.9848
	sig W					
	N. Cases	497	589	615	627	619
Long Term	SKEW	-0.106	-0.049	0.029	-0.01	-0.059
Debt	KURT	0.095	-0.023	-0.101	-0.15	-0.016
	W	0.9868	0.9842	0.9839	0.9857	0.985
	sig W					
	N. Cases	358	439	479	518	510
Gross Funds	SKEW	0.084	0.049	0.228	0.176	0.1
From	KURT	0.492	0.448	0.286	0.4	0.938
Operations	W	0.9867	0.9872	0.9842	0.9833	0.98
	sig W					0.026
	N. Cases	527	625	647	666	650
Working	SKEW	-0.093	0.215	0.103	0.061	0.288
Capital	KURT	0.487	-0.062	0.532	0.269	0.311
	W	0.9926	0.9839	0.9881	0.9850	0.9807
	sig W					0.052
	N. Cases	505	587	610	641	626

Table 42: Lognormality of all groups together. Third table.

bad cases. Each of these samples have been observed and discussed elsewhere.

The minimum sample size is 6. The maximum is 145. Most of the samples have sizes between 20 and 60.

A.5 The Estimation And The Significance of a Base-Line

We suggest two methods for estimating base-lines. Firstly, base-lines can be estimated for each item individually using the method described by Royston [105] and explained at the beginning of chapter 1: The estimated δ is the one which maximizes any statistic linked with a test of normality. Secondly, base-lines can be estimated in ratios by building models in log space and then using an iterative Least-Squares algorithm for finding both the expected value of the ratio and the base-line in the denominator.

The first method is the only one available for multi-variate modelling. Its drawback is the over-estimation introduced by the Shapiro-Wilk test. The second one should be tried when working with simple ratios since it is easier to carry out and more robust to problems of over-precision. But it has a few problems of its own.

To illustrate both methods we are going to use five pairs of samples obtained from simulated data as explained in section 3.4.3. Figures 27 on page 78 and figures 28 on page 79 are a graphical representation of such a set when different base-lines are simulated.

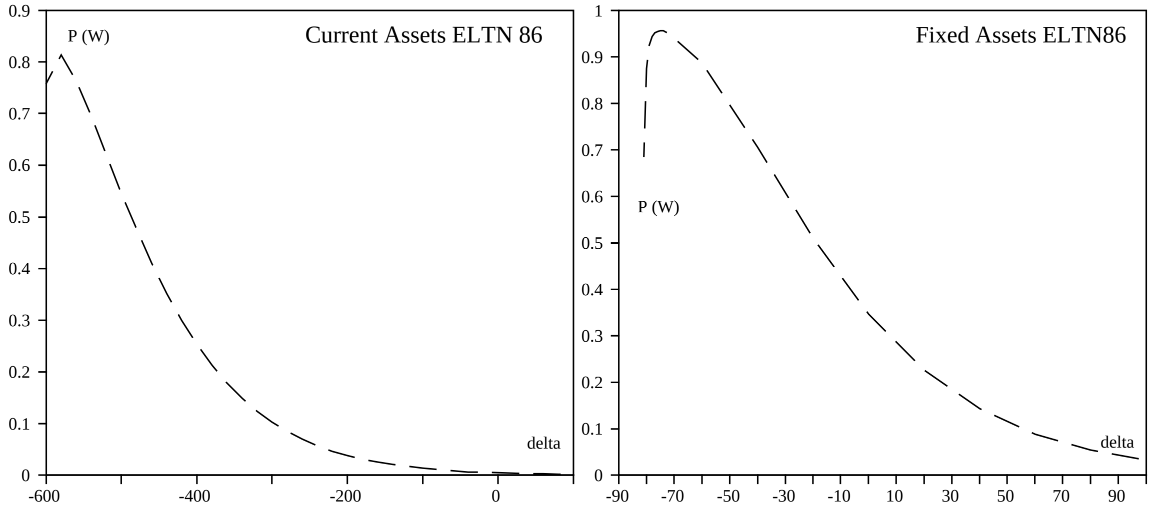

Figure 82: Estimation of δ using the method based on tests of lognormality. This figure shows the evolution of $P(W)$ for several δ in the Electronics industry.

A.5.1 The Method Based On Tests of Normality

This method consists of detecting the δ which makes the sample closer to normality. When using the Shapiro-Wilk test the statistic to optimize is W. For estimating δ we find the maximum W obtained when performing normality tests of $\log(x+\delta)$. The use of W yields sharp, easy to obtain δ (see figure 82).

As an example, we used the simulated five sets of lognormal data mentioned above, each one with a known base-line. Then, we estimated the δ using W. Next table compares the simulated with the estimated base-lines. Notice that there are two samples in each set, the one named "Numerator" and the one named "Denominator". This is because we are going to use these sets later on as the numerator and the denominator of ratios.

Sample number	Simulated base-lines		W estimated δ	
	Numerator	Denominator	Numerator	Denominator
1: No base-lines	0	0	140	120
2: + in Numerator	300	0	-160	120
3: + in Denominator	0	300	140	-180
4: - in Numerator	-50	0	200	120
5: - in Denominator	0	-30	140	140

Firstly, we notice that the samples with no simulated base-lines optimized W for values of δ of 140 and 120. They were lognormal even with no δ at all but an optimum W corresponds to these values, not the zero ones. Notice also that a δ of 140, as detected by the Shapiro-Wilk test, means that we have to add 140 to all the cases in the sample in order optimize W. Therefore, if we were to simulate such a base-line using a perfect lognormal sample, we should *subtract* — not add — 140 from it. The conclusion is that in order to estimate the real base-line we invert the sign of the δ

obtained from optimizing W.

Secondly, we observe that when a simulated base-line of +300 was introduced, the δ detected by the Shapiro-Wilk test was -160. The meaning of this is straightforward. Since the original sample had already a supposed base-line of -140, by adding to it the simulated +300 we obtain a sample with a base-line of +160. The estimated δ is then expected to be -160.

The δ obtained for all the other samples are explained in the same way.

Over-estimation of δ: The δ estimated by the Royston method are in general too large. To avoid over-estimations it seems a good practice to follow these rules:

1. Any item achieving two-parametric lognormality should be considered as not having a significant base-line as well. In figure 82 (right) the probability of the sample being drawn from a two-parametric population is about 20%. Therefore we conclude that there is no significant base-line despite the fact that the introduction of one would make the significance of W attain almost 100%.

2. Only those items not achieving a two-parametric lognormality should be candidates for having significant base-lines. In that case, instead of finding the maximum W, we just find the smallest δ able to render the sample two-parametric. In the same figure (left) a base-line of about 300 is required to bring the transformed sample to a state of non-significant departure from normality. We accept this value of $\delta \approx -300$ despite being possible to attain a much more significant W with larger absolute values of δ.

These rules ensure a better estimation of base-lines. They are justified by our experience and by the theory of significance of estimators.

A.5.2 The Method Based On Iterative Least-Squares Estimation

We now use the previous samples to form base-line ratios. Then we estimate jointly the expected value of such ratios and the base-line impinging upon the denominator. The appropriate tool is an iterative Least-Squares algorithm. If N stands for the numerator and D stands for the denominator, our problem consists of finding the δ and $\Delta\mu$ such that, in

$$\log N_j = \Delta\mu_{N/D} + \log(D_j + \delta_D) + \varepsilon_j^{N/D} \quad \text{or in} \quad \log D_j = \Delta\mu_{D/N} + \log(N_j + \delta_N) + \varepsilon_j^{D/N}$$

the sum of squared ε_j is minimized.

For the above samples we obtained the following estimated models:

$$
\begin{array}{llcrcl}
\text{1: No base-lines:} & \log N & = & 0.131 & + & \log(D+9) \\
\text{2: +300 in Numerator:} & \log D & = & -0.12 & + & \log(N-373) \\
\text{3: +300 in Denominator:} & \log N & = & 0.131 & + & \log(D+291) \\
\text{4: -30 in Denominator:} & \log N & = & 0.131 & + & \log(D+39) \\
\text{5: -50 in Numerator:} & \log D & = & -0.13 & + & \log(N+51)
\end{array}
$$

These estimated parameters seem to comply with the simulated base-lines, thus ignoring the over-estimated base-lines detected by the Shapiro-Wilk test in the non-displaced samples.

Two drawbacks specific to this method are:

- The algorithm is not guaranteed to converge, or to converge in an acceptable way. For example, the algorithm frequently generates δ which, when added to the observations in the sample, yield negative values. Hence, a few or even a lot of cases are thrown away from the sample during the iterative computation. The resulting parameters are useless.

- In the previous example we knew in advance where each base-line was. But in the real cases we don't know how to build the model so that the base-line appears in the denominator. The resulting models — when we try to estimate base-lines by putting them in the denominator when they should be in the numerator — yield ratios which are formally correct but of not so easy interpretation. Since a base-line, when accounted for in the wrong member, has a value which can be very far away from the real one, these ratios suggest the existence of exaggerated or smoothed base-lines.

We think that both methods should be used for building correct base-line ratios. They complement each other. First, the Shapiro-Wilk test or any similar one can determine where the strong base-lines are and their approximated magnitude. Then, a Least-Squares modelling in log space will hopefully yield the estimated parameters for ratio analysis. When the problem is the estimation of base-lines for multi-variate models, the former technique is the only one available.

A.5.3 The Significance of a Base-Line

The examination of two-variate scatter-plots of accounting variables in logarithmic space can detect departures from strict proportionality when they turn out to be significant. In fact, the log transformation — and also the ratio one — produces a trade-off between non-proportionality and non-linearity so that even small departures from proportionality result in clear departures from linearity.

Given this, any of the usual methods for tracing non-independence of residuals could be used in log space to detect non-negligible δ. The plots developed in chapter 3, page 73 are a simple application of these principles.

However, some advantages would arise from applying methods able to explore the particular nature of the distortion. Next we suggest two procedures.

For a given two-variate sample $\{y, x\}$, we could consider any constant term a as significant regarding the introduction of non-proportionality in a relation $y = a+x$ if it would produce significant non-linearity in log or ratio space. The significance of such a non-linearity could be assessed by comparing the variability explained by a linear model with the one the introduction of a quadratic term would account for.

Such a method is formally correct but it is open to misleading influences. A few cases could dictate the final result.

Another possible way could be the following procedure:

- Create two new variables each containing the ranking of the ratio by ascending order of the numerator and denominator. For each of these ranks,
- identify the samples containing the first N percentiles (typically, the first 5 percentiles) and the central ones (the 3^{th} quintile).
- Compare the log mean-values of the ratio in the centre with the sample containing the first percentiles. A simple t-test can be used for this.

Significant mean differences in any of the cases indicate a perceptible base-line. This procedure is effective because, as we saw, base-lines only affect the smallest cases in the sample. Notice that the ratios should be handled in log space.

This method is more robust than the first one but it is more empirical too. For example, the ambiguity in the size of the samples to be compared introduces arbitrariness.

As an example, we assessed the non-proportionality in the relation between Earnings and Sales for the Electronics industry in 1987. The mean of the log ratio in the first decile was -0.73. The mean for the central quintile was -1.10. The difference between these means is significant ($P < 0.001$). After computing the model

$$\log EBIT_j = -2.02 + 1.12 \times \log(S_j + 5510) + \varepsilon_j$$

we observed the mean values of the residuals, ε_j in the same places as before. Both mean values yielded the same value of approximately zero.

In general, the simple examination of scatter-plots of any two accounting items in log space is enough to detect base-lines. They draw very homogeneous, highly correlated, linear scatters, always with a slope of 45^o corresponding to the common, strong effect. A typical example is displayed in figure 23 on page 58. In most of this relations traces of non-linearity cannot be observed. However in a few cases a real convexity affecting small values is clearly visible. This convexity is consistent with figure 26 (left) on page 77.

Minimum	1983	1984	1985	1986	1987
Building Materials	0.29	0.40	0.36	0.35	0.34
Metallurgy	0.25	0.20	0.36	0.34	0.29
Paper and Packing	0.40	0.34	0.32	0.34	0.35
Chemicals	0.32	0.30	0.29	0.27	0.29
Electrical	0.39	0.35	0.30	0.28	0.28
Industrial Plants	0.21	0.23	0.28	0.37	0.33
Machine Tools	0.16	0.23	0.18	0.19	0.16
Electronics	0.41	0.46	0.43	0.43	0.39
Motor Components	0.49	0.49	0.46	0.44	0.42
Clothing	0.16	0.15	0.14	0.13	0.14
Miscellaneous Textiles	0.11	0.20	0.21	0.21	0.21
Wool	0.40	0.50	0.61	0.58	0.56
Leather	0.30	0.49	0.49	0.44	0.40
Food Manufacturers	0.56	0.58	0.52	0.56	0.56

Table 43: Minimum variance and co-variance obtained from Σ matrices by industry and year.

A.6 The Common Effect

In this section we replicate the experiment carried out by Fieldsend *et al.* [42] but always using the same independent log variable, the proxy for size, s, developed in section 5.1. For 14 industrial groups during a period of five years (1983-1986) we observe the slopes, b, and the proportion of explained variability, R^2, of regressions in which $\log s$ explains individual log items. For models like

$$\log x_j = a + b \times \log s_j + \varepsilon_j$$

in which x_j is an accounting item, the estimated values of b should scatter around 1. In tables 50 and next, on pages 265 and the next one towards the end of this appendix we display the resulting slopes. Tables 52 and next, on pages 267 and the next one towards the end of this chapter display the obtained R^2. Long Term Debt is the item with largest departures from the simple ratio model ($b = 1$). Metallurgy is the industry with the same quality.

On the whole, the displayed results are an argument in favour of a unique, strong effect representing the relative growth of accounting items. The slope emerges as a non-important parameter. Its value is predictable and departures from such a prediction are very small. They can be explained by the bias resulting from using regressions instead of algorithms able to deal with this errors-in-both-variables model.

A.7 Variance and Co-variance Matrices

In this section we display a few more variance and co-variance matrices of accounting items in log space illustrating their particular features. Very typical shapes are displayed in figures 83 and next on pages 269 and the next one towards the end of this appendix.

It is also interesting to observe the minimum variance or co-variance in Σ matrices belonging to the 14 industrial groups during the five-year period. Table 43 shows these values.

		1983			1984			1985			1986			1987		
V	ind	N	W	P	N	W	P	N	W	P	N	W	P	N	W	P
S	BUIL	29	0.94	0.18	34	0.96	0.43	36	0.97	0.66	39	0.96	0.43	38	0.97	0.56
	METL	29	0.98	0.95	32	0.98	0.97	33	0.97	0.77	34	0.96	0.50	33	0.97	0.64
	PAPR	50	0.95	0.06	56	0.95	0.08	57	0.94	0.02	60	0.96	0.12	58	0.96	0.17
	CHEM	53	0.95	0.09	55	0.95	0.12	56	0.96	0.20	55	0.96	0.19	55	0.96	0.30
	ELEC	37	0.95	0.16	44	0.93	0.02	48	0.96	0.22	49	0.97	0.40	46	0.96	0.28
	I.PL	19	0.91	0.10	20	0.84	0.00	22	0.83	0.00	23	0.89	0.02	23	0.90	0.02
	TOOL	22	0.91	0.05	24	0.95	0.29	24	0.97	0.68	26	0.97	0.84	25	0.95	0.29
	ELTN	98	0.97	0.33	130	0.96	0.03	138	0.96	0.04	145	0.96	0.02	143	0.97	0.07
	MOTR	25	0.92	0.08	30	0.95	0.26	31	0.96	0.36	30	0.96	0.42	29	0.96	0.46
	CLOT	39	0.98	0.92	46	0.89	0	48	0.98	0.94	52	0.97	0.70	50	0.99	0.98
	WOOL	15	0.93	0.36	21	0.97	0.82	20	0.96	0.62	20	0.97	0.84	20	0.96	0.55
	TX.M	32	0.95	0.31	36	0.97	0.75	36	0.96	0.37	37	0.97	0.66	38	0.97	0.67
	LEAT	13	0.95	0.65	16	0.95	0.52	16	0.95	0.50	16	0.94	0.42	16	0.96	0.79
	FOOD	94	0.96	0.03	105	0.96	0.03	112	0.95	0.00	116	0.96	0.05	114	0.93	0
NW	BUIL	29	0.93	0.08	34	0.94	0.11	35	0.93	0.04	38	0.96	0.29	38	0.94	0.07
	METL	29	0.96	0.49	32	0.97	0.65	32	0.98	0.87	33	0.97	0.78	31	0.97	0.77
	PAPR	48	0.98	0.90	55	0.98	0.96	56	0.98	0.84	60	0.98	0.80	58	0.97	0.51
	CHEM	52	0.97	0.70	54	0.98	0.74	55	0.98	0.79	54	0.98	0.91	54	0.97	0.61
	ELEC	37	0.93	0.03	44	0.94	0.05	48	0.94	0.04	49	0.95	0.09	46	0.94	0.05
	I.PL	19	0.91	0.10	20	0.90	0.05	22	0.89	0.02	23	0.91	0.04	23	0.93	0.13
	TOOL	22	0.95	0.36	24	0.96	0.47	24	0.96	0.6	26	0.97	0.71	25	0.96	0.60
	ELTN	97	0.98	0.58	130	0.98	0.5	134	0.96	0.04	142	0.98	0.45	138	0.96	0.00
	MOTR	25	0.96	0.61	30	0.97	0.81	31	0.97	0.64	30	0.97	0.77	29	0.96	0.42
	CLOT	39	0.97	0.75	46	0.97	0.49	48	0.98	0.88	52	0.98	0.94	50	0.98	0.95
	WOOL	15	0.94	0.37	21	0.97	0.76	20	0.97	0.75	20	0.95	0.53	20	0.96	0.64
	TX.M	32	0.94	0.14	36	0.97	0.61	36	0.97	0.73	36	0.97	0.79	37	0.97	0.64
	LEAT	13	0.87	0.05	16	0.90	0.08	16	0.93	0.29	16	0.94	0.38	16	0.95	0.52
	FOOD	94	0.98	0.81	104	0.97	0.24	109	0.98	0.65	111	0.96	0.04	113	0.97	0.24
W	BUIL	29	0.97	0.72	34	0.97	0.64	36	0.97	0.64	39	0.97	0.53	38	0.97	0.57
	METL	29	0.97	0.65	32	0.95	0.26	33	0.95	0.23	34	0.96	0.49	33	0.95	0.25
	PAPR	50	0.97	0.58	55	0.97	0.38	56	0.97	0.42	59	0.96	0.23	57	0.96	0.31
	CHEM	52	0.95	0.09	54	0.96	0.16	55	0.96	0.20	54	0.96	0.33	54	0.97	0.40
	ELEC	37	0.94	0.10	44	0.94	0.05	48	0.94	0.04	49	0.95	0.06	46	0.94	0.05
	I.PL	19	0.94	0.32	20	0.89	0.04	22	0.93	0.12	23	0.93	0.12	23	0.95	0.45
	TOOL	22	0.93	0.15	24	0.93	0.14	24	0.93	0.10	26	0.94	0.16	25	0.94	0.18
	ELTN	97	0.95	0.02	130	0.96	0.00	138	0.93	0	144	0.94	0	143	0.94	0
	MOTR	25	0.95	0.33	30	0.95	0.28	31	0.95	0.33	30	0.95	0.34	29	0.96	0.43
	CLOT	39	0.97	0.75	45	0.99	0.99	48	0.97	0.66	52	0.97	0.45	50	0.98	0.73
	WOOL	15	0.94	0.47	21	0.98	0.93	20	0.97	0.89	20	0.98	0.96	20	0.97	0.88
	TX.M	31	0.94	0.16	35	0.96	0.29	35	0.95	0.15	36	0.94	0.09	37	0.95	0.18
	LEAT	13	0.95	0.68	16	0.91	0.15	16	0.92	0.22	16	0.93	0.26	16	0.93	0.26
	FOOD	92	0.97	0.40	105	0.96	0.02	111	0.96	0.02	116	0.96	0.05	114	0.96	0.08

Table 44: The Shapiro-Wilk test of lognormality. Items by industrial group and by year. N is the number of cases. First table.

		1983			1984			1985			1986			1987		
V	ind	N	W	P	N	W	P	N	W	P	N	W	P	N	W	P
I	BUIL	29	0.96	0.45	34	0.96	0.40	36	0.96	0.28	39	0.95	0.21	38	0.94	0.11
	METL	29	0.95	0.36	32	0.98	0.86	33	0.97	0.7	34	0.94	0.13	32	0.93	0.05
	PAPR	50	0.97	0.71	54	0.97	0.54	55	0.97	0.65	59	0.96	0.28	55	0.96	0.26
	CHEM	53	0.95	0.13	55	0.94	0.04	56	0.95	0.08	55	0.97	0.44	55	0.95	0.10
	ELEC	37	0.93	0.04	44	0.97	0.61	48	0.94	0.02	49	0.94	0.05	45	0.94	0.04
	I.PL	19	0.93	0.22	19	0.94	0.29	21	0.92	0.08	22	0.78	0	23	0.95	0.46
	TOOL	22	0.89	0.02	24	0.94	0.24	24	0.97	0.66	26	0.97	0.77	25	0.96	0.57
	ELTN	94	0.97	0.18	127	0.97	0.32	132	0.98	0.86	139	0.98	0.66	136	0.97	0.33
	MOTR	25	0.94	0.16	30	0.95	0.34	30	0.97	0.70	30	0.97	0.82	29	0.96	0.51
	CLOT	39	0.97	0.68	45	0.97	0.59	48	0.97	0.63	52	0.97	0.69	50	0.98	0.90
	WOOL	15	0.96	0.77	21	0.98	0.96	20	0.97	0.86	20	0.98	0.92	20	0.96	0.72
	TX.M	32	0.96	0.37	36	0.98	0.88	36	0.97	0.80	37	0.97	0.61	38	0.96	0.38
	LEAT	13	0.95	0.64	16	0.93	0.3	16	0.92	0.21	16	0.94	0.36	16	0.94	0.42
	FOOD	93	0.98	0.65	105	0.98	0.53	112	0.97	0.37	114	0.98	0.51	111	0.98	0.79
D	BUIL	29	0.96	0.52	34	0.98	0.82	36	0.98	0.81	39	0.98	0.93	38	0.98	0.90
	METL	29	0.96	0.47	32	0.95	0.21	33	0.97	0.67	34	0.97	0.55	33	0.97	0.65
	PAPR	50	0.95	0.07	55	0.94	0.03	56	0.94	0.02	59	0.95	0.07	56	0.95	0.06
	CHEM	53	0.96	0.19	55	0.96	0.14	56	0.95	0.10	55	0.96	0.14	55	0.96	0.18
	ELEC	37	0.95	0.14	44	0.97	0.58	48	0.96	0.20	49	0.97	0.45	46	0.95	0.11
	I.PL	19	0.97	0.78	20	0.89	0.02	22	0.88	0.01	23	0.94	0.2	23	0.92	0.07
	TOOL	22	0.94	0.30	24	0.89	0.01	24	0.95	0.38	26	0.96	0.43	25	0.97	0.68
	ELTN	98	0.98	0.70	130	0.97	0.31	138	0.96	0.06	145	0.96	0.04	143	0.96	0.03
	MOTR	25	0.95	0.37	30	0.95	0.21	31	0.94	0.17	30	0.97	0.59	29	0.96	0.52
	CLOT	39	0.97	0.71	46	0.97	0.57	48	0.97	0.56	52	0.97	0.41	50	0.98	0.97
	WOOL	15	0.93	0.30	21	0.97	0.79	20	0.97	0.89	20	0.98	0.96	20	0.94	0.36
	TX.M	32	0.96	0.52	36	0.98	0.91	36	0.97	0.68	37	0.97	0.59	38	0.97	0.79
	LEAT	13	0.90	0.14	16	0.92	0.18	16	0.92	0.23	16	0.92	0.21	16	0.91	0.12
	FOOD	93	0.97	0.46	104	0.98	0.62	112	0.97	0.43	116	0.97	0.44	113	0.97	0.12
C	BUIL	29	0.95	0.22	34	0.97	0.7	36	0.97	0.61	39	0.99	0.99	38	0.96	0.3
	METL	29	0.98	0.86	32	0.98	0.83	33	0.98	0.83	34	0.96	0.37	33	0.97	0.76
	PAPR	50	0.96	0.32	55	0.96	0.32	56	0.96	0.2	59	0.95	0.07	55	0.97	0.37
	CHEM	53	0.96	0.16	55	0.94	0.02	56	0.95	0.09	55	0.96	0.14	55	0.96	0.24
	ELEC	37	0.95	0.13	44	0.97	0.60	48	0.94	0.05	49	0.96	0.31	46	0.96	0.27
	I.PL	19	0.95	0.40	20	0.90	0.04	22	0.90	0.03	23	0.95	0.36	23	0.98	0.90
	TOOL	22	0.95	0.41	24	0.94	0.21	24	0.96	0.56	26	0.95	0.34	25	0.95	0.36
	ELTN	98	0.98	0.94	130	0.97	0.30	138	0.97	0.27	145	0.96	0.02	143	0.97	0.13
	MOTR	25	0.95	0.29	30	0.96	0.45	31	0.96	0.50	30	0.96	0.55	29	0.96	0.47
	CLOT	39	0.97	0.66	46	0.96	0.34	48	0.98	0.92	52	0.97	0.63	50	0.98	0.91
	WOOL	15	0.90	0.12	21	0.96	0.54	20	0.97	0.76	20	0.97	0.90	20	0.96	0.63
	TX.M	32	0.96	0.42	36	0.97	0.69	36	0.98	0.83	37	0.98	0.80	38	0.97	0.76
	LEAT	13	0.94	0.53	16	0.92	0.19	16	0.92	0.19	16	0.92	0.18	16	0.94	0.44
	FOOD	93	0.95	0.01	105	0.96	0.03	112	0.96	0.02	116	0.96	0.06	113	0.96	0.02

Table 45: The Shapiro-Wilk test of lognormality. Items by industrial group and by year. N is the number of cases. Second table.

		1983			1984			1985			1986			1987		
V	ind	N	W	P	N	W	P	N	W	P	N	W	P	N	W	P
CA	BUIL	29	0.95	0.36	34	0.97	0.62	36	0.96	0.41	38	0.96	0.33	38	0.95	0.13
	METL	29	0.97	0.67	32	0.97	0.75	33	0.97	0.68	34	0.96	0.46	33	0.96	0.52
	PAPR	50	0.96	0.30	54	0.95	0.11	57	0.83	0	60	0.97	0.35	58	0.96	0.29
	CHEM	52	0.96	0.18	55	0.95	0.05	56	0.95	0.07	54	0.97	0.36	55	0.95	0.07
	ELEC	37	0.94	0.07	44	0.94	0.05	48	0.93	0.01	49	0.94	0.05	46	0.94	0.02
	I.PL	19	0.91	0.09	20	0.90	0.05	22	0.89	0.02	23	0.92	0.06	23	0.96	0.53
	TOOL	21	0.90	0.03	24	0.93	0.13	24	0.96	0.47	26	0.98	0.88	25	0.97	0.80
	ELTN	97	0.98	0.82	127	0.98	0.71	133	0.96	0.05	142	0.96	0.00	141	0.96	0.01
	MOTR	25	0.94	0.19	30	0.96	0.39	31	0.97	0.58	30	0.97	0.67	29	0.97	0.69
	CLOT	39	0.98	0.83	46	0.97	0.68	48	0.98	0.77	52	0.98	0.71	50	0.98	0.73
	WOOL	15	0.95	0.65	21	0.98	0.93	20	0.98	0.93	20	0.97	0.82	20	0.96	0.71
	TX.M	32	0.93	0.08	36	0.96	0.40	36	0.97	0.68	37	0.97	0.73	37	0.97	0.51
	LEAT	13	0.94	0.49	16	0.94	0.47	16	0.94	0.34	16	0.94	0.41	16	0.94	0.37
	FOOD	93	0.96	0.11	105	0.97	0.13	112	0.94	0	115	0.97	0.20	112	0.96	0.02
FA	BUIL	29	0.92	0.04	34	0.95	0.28	36	0.95	0.15	39	0.93	0.04	38	0.94	0.11
	METL	29	0.96	0.50	32	0.97	0.74	33	0.96	0.37	34	0.97	0.53	32	0.96	0.42
	PAPR	50	0.97	0.37	56	0.97	0.61	57	0.98	0.95	60	0.98	0.94	57	0.98	0.80
	CHEM	53	0.95	0.09	55	0.95	0.09	56	0.95	0.08	55	0.94	0.03	55	0.95	0.09
	ELEC	37	0.95	0.17	44	0.95	0.15	48	0.95	0.16	49	0.96	0.19	46	0.94	0.05
	I.PL	19	0.93	0.24	20	0.90	0.06	22	0.90	0.03	22	0.90	0.04	23	0.93	0.17
	TOOL	22	0.95	0.43	24	0.95	0.38	24	0.95	0.40	26	0.94	0.18	25	0.95	0.36
	ELTN	98	0.98	0.84	130	0.98	0.54	138	0.98	0.78	145	0.97	0.35	143	0.97	0.08
	MOTR	25	0.95	0.32	30	0.98	0.89	30	0.97	0.73	30	0.98	0.86	29	0.97	0.78
	CLOT	39	0.97	0.74	46	0.96	0.31	48	0.97	0.45	52	0.96	0.27	50	0.97	0.41
	WOOL	15	0.95	0.61	21	0.97	0.90	20	0.98	0.99	20	0.99	0.99	20	0.98	0.93
	TX.M	32	0.96	0.53	36	0.96	0.34	36	0.95	0.26	36	0.96	0.36	37	0.96	0.32
	LEAT	13	0.90	0.13	16	0.93	0.28	16	0.96	0.67	16	0.95	0.51	16	0.96	0.72
	FOOD	94	0.98	0.69	105	0.97	0.32	112	0.97	0.24	116	0.97	0.45	114	0.97	0.23
TA	BUIL	29	0.94	0.12	34	0.96	0.39	36	0.95	0.24	39	0.94	0.07	38	0.94	0.06
	METL	29	0.97	0.76	32	0.98	0.88	33	0.97	0.74	34	0.97	0.71	33	0.97	0.55
	PAPR	50	0.96	0.36	56	0.96	0.15	57	0.96	0.24	60	0.96	0.10	58	0.97	0.46
	CHEM	53	0.96	0.16	55	0.95	0.09	56	0.95	0.10	55	0.96	0.21	55	0.96	0.19
	ELEC	37	0.93	0.04	44	0.94	0.06	48	0.93	0.02	49	0.95	0.09	46	0.94	0.04
	I.PL	19	0.91	0.10	20	0.91	0.08	22	0.90	0.03	23	0.90	0.03	23	0.96	0.59
	TOOL	22	0.93	0.15	24	0.94	0.25	24	0.97	0.78	26	0.96	0.45	25	0.96	0.59
	ELTN	98	0.99	0.99	130	0.97	0.26	137	0.96	0.02	144	0.96	0.00	142	0.95	0.00
	MOTR	25	0.94	0.21	30	0.97	0.61	31	0.97	0.74	30	0.96	0.44	29	0.95	0.33
	CLOT	39	0.97	0.74	46	0.98	0.93	48	0.96	0.31	52	0.98	0.81	50	0.98	0.83
	WOOL	15	0.94	0.4	21	0.98	0.93	20	0.98	0.96	20	0.98	0.92	20	0.97	0.77
	TX.M	32	0.93	0.08	36	0.96	0.47	36	0.97	0.51	37	0.96	0.46	37	0.96	0.42
	LEAT	13	0.91	0.19	16	0.92	0.17	16	0.93	0.29	16	0.94	0.36	16	0.96	0.75
	FOOD	94	0.96	0.05	105	0.96	0.03	112	0.94	0.00	116	0.96	0.10	113	0.96	0.03

Table 46: The Shapiro-Wilk test of lognormality. Items by industrial group and by year. N is the number of cases. Third table.

		1983			1984			1985			1986			1987		
V	ind	N	W	P	N	W	P	N	W	P	N	W	P	N	W	P
CL	BUIL	29	0.95	0.29	34	0.97	0.66	36	0.97	0.77	39	0.97	0.61	38	0.97	0.56
	METL	29	0.97	0.83	32	0.98	0.97	33	0.97	0.81	34	0.96	0.51	33	0.97	0.61
	PAPR	50	0.96	0.20	56	0.96	0.13	57	0.96	0.11	60	0.96	0.29	57	0.96	0.31
	CHEM	53	0.94	0.04	55	0.92	0.00	56	0.93	0.00	55	0.95	0.05	55	0.95	0.05
	ELEC	37	0.93	0.05	44	0.97	0.66	48	0.95	0.13	49	0.97	0.40	46	0.94	0.07
	I.PL	19	0.96	0.57	20	0.86	0.01	22	0.86	0.00	23	0.93	0.17	23	0.97	0.71
	TOOL	22	0.94	0.22	24	0.95	0.38	24	0.97	0.84	26	0.95	0.39	25	0.96	0.64
	ELTN	98	0.97	0.35	130	0.97	0.21	138	0.96	0.02	145	0.95	0	143	0.95	0.00
	MOTR	25	0.95	0.42	30	0.95	0.30	31	0.97	0.62	30	0.95	0.27	29	0.96	0.48
	CLOT	39	0.98	0.88	46	0.97	0.72	48	0.97	0.57	52	0.97	0.61	50	0.98	0.96
	WOOL	15	0.90	0.09	21	0.96	0.70	20	0.98	0.97	20	0.97	0.83	20	0.96	0.66
	TX.M	32	0.97	0.56	36	0.97	0.8	36	0.97	0.79	37	0.98	0.93	38	0.98	0.92
	LEAT	13	0.95	0.67	16	0.94	0.43	16	0.92	0.23	16	0.94	0.41	16	0.96	0.64
	FOOD	94	0.95	0.01	105	0.96	0.02	112	0.95	0.01	116	0.96	0.06	113	0.96	0.05
N	BUIL	28	0.97	0.64	34	0.96	0.5	36	0.97	0.61	39	0.97	0.50	38	0.96	0.44
	METL	29	0.97	0.73	32	0.95	0.19	33	0.94	0.13	34	0.95	0.20	33	0.94	0.15
	PAPR	50	0.96	0.35	55	0.98	0.82	56	0.97	0.62	59	0.97	0.52	57	0.97	0.59
	CHEM	52	0.97	0.70	54	0.97	0.36	55	0.97	0.46	54	0.97	0.61	53	0.97	0.53
	ELEC	36	0.94	0.11	43	0.96	0.20	47	0.95	0.10	49	0.94	0.03	46	0.94	0.02
	I.PL	19	0.92	0.14	20	0.87	0.01	22	0.90	0.03	23	0.91	0.04	23	0.94	0.18
	TOOL	22	0.91	0.07	24	0.94	0.19	24	0.92	0.06	26	0.91	0.03	25	0.92	0.05
	ELTN	97	0.96	0.10	130	0.95	0.00	138	0.94	0	145	0.94	0	143	0.94	0
	MOTR	25	0.95	0.42	30	0.95	0.31	30	0.96	0.53	30	0.96	0.52	29	0.96	0.55
	CLOT	39	0.98	0.88	45	0.99	0.99	48	0.98	0.77	52	0.98	0.78	50	0.98	0.86
	WOOL	15	0.95	0.57	21	0.98	0.99	20	0.99	0.99	20	0.99	0.99	20	0.98	0.99
	TX.M	32	0.95	0.19	36	0.97	0.73	36	0.96	0.45	37	0.96	0.31	38	0.97	0.56
	LEAT	13	0.92	0.29	16	0.91	0.12	16	0.91	0.16	16	0.92	0.17	16	0.92	0.21
	FOOD	92	0.97	0.50	105	0.97	0.26	111	0.97	0.29	116	0.98	0.53	113	0.98	0.50
EX	BUIL	29	0.93	0.08	34	0.96	0.36	36	0.97	0.58	39	0.97	0.60	38	0.97	0.77
	METL	29	0.98	0.98	32	0.98	0.97	33	0.95	0.23	34	0.96	0.47	33	0.97	0.80
	PAPR	49	0.96	0.18	56	0.94	0.01	57	0.90	0	60	0.95	0.06	58	0.94	0.02
	CHEM	53	0.96	0.20	55	0.96	0.21	56	0.96	0.14	55	0.96	0.25	55	0.97	0.44
	ELEC	37	0.95	0.23	44	0.97	0.65	47	0.96	0.27	49	0.96	0.35	46	0.96	0.29
	I.PL	19	0.92	0.13	20	0.87	0.01	22	0.91	0.05	23	0.92	0.10	23	0.93	0.12
	TOOL	22	0.93	0.14	24	0.95	0.32	24	0.97	0.74	26	0.98	0.92	25	0.95	0.33
	ELTN	98	0.98	0.84	130	0.98	0.75	138	0.98	0.80	145	0.98	0.67	143	0.98	0.87
	MOTR	25	0.92	0.07	30	0.95	0.28	31	0.95	0.22	30	0.96	0.43	29	0.95	0.25
	CLOT	39	0.98	0.86	46	0.84	0	48	0.97	0.69	52	0.97	0.69	50	0.98	0.97
	WOOL	15	0.96	0.70	21	0.98	0.95	20	0.98	0.92	20	0.97	0.88	20	0.97	0.78
	TX.M	32	0.97	0.80	36	0.98	0.87	36	0.96	0.49	37	0.98	0.92	38	0.98	0.84
	LEAT	13	0.95	0.63	16	0.96	0.73	16	0.95	0.54	16	0.93	0.31	16	0.95	0.61
	FOOD	94	0.96	0.06	105	0.96	0.02	112	0.95	0.00	116	0.96	0.04	114	0.94	0

Table 47: The Shapiro-Wilk test of lognormality. Items by industrial group and by year. N is the number of cases. Fourth table.

		1983			1984			1985			1986			1987		
V	ind	N	W	P	N	W	P	N	W	P	N	W	P	N	W	P
TC	BUIL	29	0.94	0.15	34	0.94	0.12	35	0.93	0.06	38	0.95	0.13	38	0.94	0.08
	METL	29	0.96	0.47	32	0.96	0.34	33	0.98	0.88	34	0.98	0.86	32	0.98	0.92
	PAPR	49	0.98	0.91	55	0.98	0.74	57	0.97	0.63	60	0.97	0.52	58	0.97	0.61
	CHEM	53	0.97	0.57	55	0.97	0.46	56	0.97	0.38	55	0.97	0.62	55	0.97	0.35
	ELEC	37	0.93	0.02	44	0.94	0.04	48	0.94	0.04	49	0.95	0.06	46	0.94	0.04
	I.PL	19	0.90	0.07	20	0.90	0.06	22	0.88	0.01	23	0.90	0.03	23	0.92	0.10
	TOOL	22	0.93	0.19	24	0.95	0.28	24	0.95	0.40	26	0.97	0.72	25	0.95	0.41
	ELTN	97	0.98	0.72	130	0.98	0.62	135	0.96	0.01	143	0.97	0.10	138	0.95	0
	MOTR	25	0.95	0.30	30	0.97	0.81	31	0.97	0.75	30	0.96	0.51	29	0.95	0.24
	CLOT	39	0.98	0.93	46	0.97	0.42	48	0.98	0.88	52	0.98	0.94	50	0.98	0.77
	WOOL	15	0.95	0.63	21	0.96	0.66	20	0.96	0.69	20	0.95	0.53	20	0.95	0.50
	TX.M	32	0.93	0.09	36	0.97	0.58	36	0.97	0.61	36	0.97	0.59	37	0.97	0.57
	LEAT	13	0.88	0.09	16	0.90	0.11	16	0.93	0.26	16	0.94	0.41	16	0.95	0.60
	FOOD	94	0.98	0.56	105	0.97	0.48	111	0.96	0.04	112	0.96	0.02	113	0.96	0.05
EB	BUIL	29	0.96	0.43	33	0.88	0.00	36	0.97	0.59	38	0.97	0.49	38	0.94	0.07
	METL	25	0.96	0.56	29	0.97	0.82	27	0.95	0.28	31	0.98	0.84	29	0.95	0.30
	PAPR	45	0.98	0.84	53	0.99	0.99	57	0.97	0.50	59	0.97	0.46	57	0.97	0.56
	CHEM	51	0.96	0.18	54	0.97	0.47	54	0.98	0.80	53	0.97	0.69	52	0.95	0.11
	ELEC	37	0.98	0.84	41	0.98	0.91	47	0.96	0.26	46	0.98	0.92	43	0.96	0.33
	I.PL	17	0.85	0.01	18	0.90	0.05	21	0.92	0.12	19	0.94	0.27	19	0.92	0.13
	TOOL	17	0.91	0.14	23	0.98	0.95	22	0.97	0.72	24	0.97	0.71	25	0.96	0.59
	ELTN	91	0.98	0.76	122	0.97	0.28	121	0.96	0.05	121	0.95	0.00	128	0.95	0.00
	MOTR	22	0.95	0.38	28	0.98	0.87	27	0.97	0.67	27	0.96	0.55	27	0.97	0.80
	CLOT	36	0.96	0.41	39	0.97	0.69	44	0.97	0.50	50	0.98	0.90	45	0.98	0.93
	WOOL	15	0.95	0.57	21	0.96	0.69	20	0.96	0.60	20	0.97	0.88	20	0.98	0.97
	TX.M	29	0.93	0.07	33	0.96	0.45	32	0.95	0.26	34	0.95	0.26	36	0.97	0.79
	LEAT	13	0.91	0.20	15	0.95	0.60	15	0.96	0.73	16	0.95	0.62	16	0.95	0.62
	FOOD	87	0.96	0.10	97	0.96	0.02	106	0.97	0.16	107	0.97	0.12	106	0.93	0
OP	BUIL	29	0.96	0.53	33	0.96	0.54	34	0.97	0.68	36	0.97	0.75	37	0.94	0.06
	METL	24	0.95	0.38	28	0.96	0.50	26	0.97	0.73	31	0.98	0.89	28	0.97	0.67
	PAPR	44	0.98	0.86	51	0.95	0.14	55	0.96	0.21	58	0.97	0.55	57	0.97	0.43
	CHEM	49	0.97	0.58	54	0.97	0.59	53	0.97	0.65	51	0.97	0.63	52	0.97	0.37
	ELEC	33	0.97	0.79	40	0.98	0.80	45	0.97	0.49	43	0.97	0.58	41	0.96	0.39
	I.PL	17	0.86	0.02	18	0.88	0.03	19	0.92	0.14	18	0.91	0.11	16	0.94	0.46
	TOOL	16	0.95	0.60	21	0.97	0.85	20	0.97	0.86	23	0.97	0.70	24	0.98	0.90
	ELTN	87	0.98	0.81	118	0.97	0.47	121	0.96	0.09	120	0.96	0.04	126	0.96	0.01
	MOTR	22	0.95	0.36	26	0.97	0.65	27	0.96	0.61	25	0.97	0.75	26	0.96	0.47
	CLOT	35	0.95	0.14	38	0.97	0.51	44	0.86	0	50	0.97	0.70	43	0.98	0.8
	WOOL	14	0.96	0.72	21	0.96	0.65	20	0.97	0.87	20	0.98	0.95	20	0.96	0.56
	TX.M	27	0.95	0.25	31	0.96	0.50	32	0.96	0.55	32	0.96	0.46	35	0.98	0.93
	LEAT	13	0.91	0.18	15	0.95	0.57	15	0.95	0.64	16	0.97	0.86	16	0.96	0.78
	FOOD	87	0.98	0.64	95	0.96	0.05	104	0.97	0.26	104	0.97	0.27	98	0.95	0.00

Table 48: The Shapiro-Wilk test of lognormality. Items by industrial group and by year. N is the number of cases. Fifth table.

		1983			1984			1985			1986			1987		
V	ind	N	W	P	N	W	P	N	W	P	N	W	P	N	W	P
DB	BUIL	22	0.97	0.78	28	0.97	0.75	29	0.97	0.83	34	0.94	0.13	34	0.95	0.23
	METL	19	0.96	0.66	21	0.97	0.77	20	0.97	0.82	21	0.93	0.15	19	0.97	0.80
	PAPR	37	0.97	0.75	43	0.96	0.29	42	0.98	0.80	46	0.97	0.56	47	0.95	0.17
	CHEM	36	0.94	0.11	39	0.96	0.42	42	0.96	0.26	43	0.93	0.01	43	0.93	0.02
	ELEC	19	0.93	0.25	27	0.95	0.23	29	0.95	0.35	33	0.97	0.57	33	0.97	0.63
	I.PL	12	0.97	0.87	14	0.85	0.02	16	0.90	0.11	17	0.93	0.31	18	0.95	0.52
	TOOL	15	0.88	0.05	19	0.96	0.60	20	0.96	0.60	21	0.96	0.57	21	0.91	0.06
	ELTN	55	0.97	0.42	78	0.97	0.22	94	0.98	0.82	101	0.97	0.22	100	0.98	0.55
	MOTR	21	0.89	0.02	26	0.95	0.29	25	0.95	0.35	28	0.94	0.21	27	0.96	0.43
	CLOT	25	0.94	0.22	29	0.93	0.07	33	0.94	0.13	37	0.95	0.25	35	0.98	0.86
	WOOL	7	0.78	0.03	10	0.95	0.72	11	0.95	0.72	12	0.97	0.87	12	0.97	0.93
	TX.M	20	0.96	0.56	23	0.97	0.77	23	0.95	0.40	26	0.96	0.52	26	0.95	0.24
	LEAT	6	0.93	0.58	8	0.97	0.91	9	0.90	0.28	9	0.95	0.77	11	0.94	0.57
	FOOD	64	0.95	0.04	74	0.95	0.03	86	0.95	0.01	90	0.93	0	84	0.97	0.26
FL	BUIL	29	0.96	0.43	34	0.91	0.01	36	0.96	0.27	39	0.95	0.23	38	0.95	0.13
	METL	28	0.96	0.59	31	0.97	0.80	31	0.98	0.93	31	0.98	0.94	29	0.96	0.37
	PAPR	48	0.97	0.70	56	0.98	0.92	57	0.97	0.36	60	0.97	0.45	57	0.97	0.44
	CHEM	51	0.95	0.09	55	0.96	0.21	54	0.96	0.16	53	0.96	0.19	54	0.94	0.03
	ELEC	37	0.96	0.37	42	0.98	0.78	48	0.95	0.10	48	0.93	0.02	45	0.95	0.11
	I.PL	17	0.86	0.02	18	0.90	0.06	21	0.91	0.06	19	0.93	0.23	20	0.98	0.98
	TOOL	19	0.95	0.52	24	0.96	0.65	24	0.99	0.99	25	0.94	0.25	24	0.96	0.50
	ELTN	91	0.97	0.23	123	0.97	0.12	126	0.96	0.07	130	0.97	0.22	130	0.95	0.00
	MOTR	22	0.96	0.60	30	0.97	0.82	29	0.97	0.64	28	0.96	0.40	28	0.92	0.04
	CLOT	36	0.97	0.78	40	0.93	0.04	44	0.99	0.98	50	0.97	0.68	45	0.89	0
	WOOL	15	0.94	0.44	21	0.96	0.65	20	0.95	0.53	20	0.96	0.74	20	0.98	0.94
	TX.M	29	0.94	0.12	34	0.96	0.38	33	0.95	0.27	34	0.95	0.17	36	0.97	0.63
	LEAT	13	0.92	0.25	16	0.97	0.93	15	0.95	0.54	16	0.96	0.73	16	0.96	0.70
	FOOD	92	0.97	0.52	101	0.96	0.03	109	0.96	0.03	113	0.97	0.25	108	0.95	0.00
WC	BUIL	28	0.95	0.24	30	0.98	0.89	32	0.96	0.36	37	0.97	0.62	35	0.97	0.75
	METL	27	0.96	0.54	30	0.94	0.17	32	0.97	0.60	33	0.96	0.34	31	0.95	0.19
	PAPR	43	0.93	0.01	48	0.97	0.51	50	0.96	0.16	54	0.97	0.51	54	0.97	0.41
	CHEM	47	0.98	0.90	50	0.98	0.83	53	0.98	0.75	49	0.95	0.10	50	0.97	0.40
	ELEC	36	0.97	0.55	42	0.97	0.57	47	0.98	0.85	46	0.98	0.97	42	0.95	0.12
	I.PL	19	0.89	0.03	19	0.90	0.07	21	0.89	0.02	23	0.94	0.23	21	0.93	0.20
	TOOL	21	0.95	0.49	24	0.95	0.28	24	0.94	0.27	26	0.97	0.68	24	0.96	0.61
	ELTN	89	0.98	0.56	117	0.96	0.09	113	0.96	0.02	129	0.97	0.35	127	0.95	0.00
	MOTR	25	0.95	0.39	29	0.96	0.49	30	0.96	0.49	27	0.95	0.32	28	0.97	0.62
	CLOT	37	0.96	0.27	45	0.96	0.21	44	0.98	0.86	51	0.97	0.62	49	0.98	0.84
	WOOL	13	0.95	0.64	19	0.95	0.40	18	0.94	0.35	18	0.91	0.11	18	0.89	0.04
	TX.M	29	0.92	0.05	34	0.95	0.26	33	0.95	0.25	34	0.94	0.08	35	0.95	0.17
	LEAT	12	0.96	0.73	15	0.95	0.60	15	0.95	0.58	14	0.97	0.84	15	0.89	0.06
	FOOD	79	0.97	0.54	85	0.96	0.12	98	0.98	0.84	100	0.98	0.68	97	0.95	0.00

Table 49: The Shapiro-Wilk test of lognormality. Items by industrial group and by year. N is the number of cases. Sixth table.

item	ind	1983	1984	1985	1986	1987	item	ind	1983	1984	1985	1986	1987
C	1	1.00	1.00	1.01	1.00	1.02	FA	1	0.97	1.05	1.05	1.05	1.11
	2	1.03	0.97	1.03	1.01	1.00		2	0.97	0.96	1.18	1.20	1.07
	3	0.96	1.06	1.04	1.03	1.04		3	0.94	0.94	0.87	0.99	0.99
	4	0.97	0.96	0.99	0.98	0.96		4	1.28	1.25	1.24	1.25	1.21
	5	1.11	1.12	1.11	1.12	1.14		5	0.98	0.89	0.93	0.93	0.89
	6	1.08	1.06	1.04	1.06	0.99		6	0.95	0.81	0.87	0.98	0.89
	7	0.87	1.10	1.05	1.03	1.06		7	0.70	0.83	0.93	0.98	1.16
	8	1.00	0.98	1.01	0.98	1.02		8	1.03	1.00	1.07	1.04	1.09
	9	1.00	0.93	0.97	0.96	1.00		9	1.12	1.02	1.10	1.08	1.08
	10	1.01	1.07	1.03	1.02	0.98		10	1.03	1.13	1.06	1.11	1.04
	11	1.03	0.98	0.94	1.03	0.97		11	0.89	0.93	1.00	1.04	0.97
	12	1.02	0.99	0.99	0.99	0.98		12	1.03	1.04	1.04	1.04	1.04
	13	1.05	1.01	0.96	0.96	0.97		13	1.07	1.19	1.20	1.19	1.22
	14	0.97	0.98	0.99	1.00	0.98		14	1.12	1.11	1.16	1.10	1.09
D	1	1.00	0.98	0.97	0.97	0.97	FL	1	0.97	1.13	1.07	1.06	1.08
	2	1.02	0.92	0.92	0.91	0.94		2	1.13	1.02	0.92	1.07	1.09
	3	1.08	1.01	1.04	1.04	1.06		3	1.13	1.07	1.02	1.02	0.98
	4	0.90	0.90	0.90	0.89	0.91		4	1.11	1.14	1.12	1.08	1.12
	5	0.97	0.93	0.98	1.00	1.01		5	0.97	1.00	0.94	0.93	0.89
	6	0.94	1.01	0.97	0.95	1.01		6	1.05	1.04	0.95	1.01	0.78
	7	1.07	1.11	1.00	0.90	0.85		7	0.90	0.98	1.08	1.27	0.99
	8	0.95	0.93	0.97	0.96	0.98		8	0.91	0.96	0.93	0.99	0.98
	9	0.96	1.00	1.09	0.98	1.00		9	0.95	1.03	1.00	0.95	1.13
	10	1.03	0.94	0.87	0.83	0.86		10	0.98	1.06	0.97	0.89	1.06
	11	0.95	0.95	0.93	0.96	1.01		11	1.16	1.14	1.15	1.06	1.04
	12	0.96	0.97	0.96	0.95	0.99		12	1.14	1.19	1.07	1.10	1.10
	13	0.99	0.97	0.93	0.91	0.94		13	1.00	1.11	1.10	1.09	1.03
	14	0.94	0.94	0.93	0.96	0.94		14	1.12	1.06	1.12	1.08	1.03
I	1	1.13	1.06	1.04	1.08	1.05	NW	1	0.94	0.98	0.98	1.03	1.02
	2	1.19	1.18	1.07	0.99	1.18		2	1.01	1.01	1.03	0.96	0.94
	3	1.03	1.02	1.00	1.05	1.05		3	0.92	0.89	0.80	0.97	0.90
	4	1.00	1.02	1.00	0.99	0.96		4	1.18	1.16	1.14	1.12	1.07
	5	0.99	1.03	1.00	1.02	1.02		5	0.89	0.84	0.88	0.87	0.85
	6	1.07	1.13	1.11	1.15	1.12		6	0.95	0.78	0.87	0.87	0.88
	7	1.02	0.98	1.00	1.02	1.06		7	0.93	0.80	0.94	1.01	0.98
	8	1.07	1.10	1.05	1.06	1.07		8	1.03	1.05	1.04	1.05	1.04
	9	1.07	0.99	1.00	1.01	1.03		9	1.06	0.99	1.03	0.98	0.96
	10	1.16	1.15	1.08	1.13	1.13		10	1.01	0.94	0.96	0.98	0.90
	11	0.92	0.96	0.98	0.95	0.93		11	1.05	1.02	1.06	1.03	1.00
	12	1.08	1.06	1.07	1.16	1.10		12	0.98	0.98	0.99	0.99	0.98
	13	1.09	1.08	1.06	1.08	1.07		13	0.95	1.06	1.08	1.10	1.11
	14	1.09	1.08	1.07	1.07	1.04		14	1.06	1.08	1.07	1.01	0.99

Table 50: Slopes of the regression in which size explains twelve accounting items by industry and by year. First table.

item	ind	1983	1984	1985	1986	1987	item	ind	1983	1984	1985	1986	1987
S	1	1.01	0.99	1.01	0.98	1.00	WC	1	1.14	0.95	0.91	1.16	1.06
	2	0.91	0.97	0.93	1.01	0.93		2	1.02	0.95	0.97	0.92	0.81
	3	1.10	1.07	1.05	1.04	1.05		3	0.90	0.95	0.94	1.04	0.99
	4	0.91	0.92	0.94	0.89	0.88		4	1.07	1.09	1.07	1.06	0.96
	5	0.98	1.11	1.01	1.02	1.04		5	0.91	0.90	1.02	0.91	0.92
	6	1.03	1.13	1.14	1.11	1.09		6	1.19	0.79	0.93	0.82	0.85
	7	1.17	1.14	1.13	1.07	1.05		7	1.22	0.98	1.14	1.27	1.09
	8	0.97	0.97	1.00	1.01	1.01		8	1.13	1.09	1.04	1.08	1.07
	9	1.01	1.02	0.95	0.96	0.98		9	1.17	1.09	0.99	0.92	0.92
	10	0.97	1.11	0.97	0.93	0.96		10	0.86	0.82	0.83	0.91	0.75
	11	0.92	0.94	0.93	0.91	0.94		11	1.12	1.01	1.02	0.99	0.94
	12	0.99	0.97	0.96	0.98	1.00		12	1.01	0.97	0.94	0.90	0.89
	13	1.01	0.99	1.03	1.05	1.05		13	0.76	0.87	0.95	0.80	0.94
	14	0.96	0.92	0.92	0.92	0.99		14	0.97	0.95	0.95	0.99	0.87
W	1	1.01	1.04	1.05	1.02	1.02	DEBT	1	1.25	1.23	1.35	1.15	1.42
	2	1.04	1.09	1.10	1.11	1.12		2	1.02	0.86	1.06	1.28	1.32
	3	1.03	1.05	1.04	1.03	1.04		3	1.26	0.95	0.90	1.14	1.18
	4	1.03	1.06	1.07	1.10	1.06		4	1.29	1.26	1.19	1.19	1.09
	5	1.02	1.14	1.06	1.03	1.01		5	1.00	0.93	1.05	0.96	0.91
	6	1.03	1.06	1.00	1.04	1.05		6	0.96	1.27	1.35	1.04	0.72
	7	0.98	1.00	1.01	1.07	1.06		7	0.42	0.94	0.90	1.00	1.41
	8	1.02	1.04	1.00	0.99	0.97		8	0.88	0.91	0.94	0.99	1.03
	9	1.01	1.08	1.03	1.04	1.06		9	1.41	1.27	1.09	1.25	1.18
	10	0.98	1.05	1.09	1.10	1.12		10	0.99	0.92	1.07	1.07	0.82
	11	0.92	0.99	0.99	1.00	0.98		11	0.63	0.59	1.17	1.04	1.34
	12	1.03	1.03	1.06	1.04	1.01		12	1.13	1.17	1.11	1.20	1.11
	13	0.87	0.92	0.93	0.92	0.92		13	0.66	0.83	1.21	0.79	0.87
	14	1.05	1.05	1.07	1.04	1.07		14	1.17	1.14	1.13	1.05	1.11
CA	1	0.98	0.97	0.97	1.01	0.99	EBIT	1	0.94	1.09	1.03	1.06	1.08
	2	0.99	0.96	0.93	0.88	0.89		2	1.04	0.98	0.83	1.02	0.99
	3	0.93	0.97	0.96	1.00	1.00		3	1.17	0.99	1.03	0.96	0.98
	4	0.99	0.98	0.98	0.98	1.01		4	1.10	1.17	1.14	1.09	1.05
	5	1.01	0.96	1.02	1.02	1.02		5	1.04	1.00	0.95	0.95	0.93
	6	1.04	0.89	0.95	0.92	0.91		6	1.14	1.10	0.94	1.04	0.88
	7	1.03	0.97	1.01	1.01	1.02		7	1.00	1.09	1.06	1.22	0.94
	8	1.03	1.02	1.00	1.02	1.03		8	0.88	0.93	0.92	0.94	0.96
	9	1.00	0.95	0.97	1.02	0.96		9	0.97	0.98	0.99	0.93	1.00
	10	1.00	0.95	0.97	0.98	0.92		10	0.82	0.91	0.92	0.79	0.80
	11	1.20	1.06	1.06	1.03	1.02		11	1.26	1.19	1.20	1.07	1.06
	12	1.01	0.99	1.00	0.99	0.98		12	1.27	1.20	1.08	1.13	1.10
	13	1.02	1.00	0.99	1.00	0.98		13	1.03	1.05	1.17	1.22	1.11
	14	0.97	0.99	0.96	1.00	0.94		14	1.08	1.05	1.14	1.07	0.98

Table 51: Slopes of the regression in which size explains twelve accounting items by industry and by year. Second Table.

item	ind	1983	1984	1985	1986	1987	item	ind	1983	1984	1985	1986	1987
C	1	0.89	0.9	0.91	0.92	0.91	FA	1	0.9	0.89	0.85	0.86	0.88
	2	0.8	0.86	0.81	0.74	0.77		2	0.62	0.5	0.6	0.63	0.6
	3	0.89	0.89	0.92	0.93	0.94		3	0.76	0.76	0.72	0.83	0.83
	4	0.91	0.92	0.9	0.9	0.88		4	0.89	0.86	0.86	0.81	0.8
	5	0.95	0.96	0.96	0.95	0.94		5	0.81	0.72	0.74	0.74	0.71
	6	0.86	0.91	0.93	0.95	0.87		6	0.82	0.81	0.83	0.87	0.8
	7	0.92	0.96	0.92	0.9	0.91		7	0.82	0.83	0.85	0.88	0.81
	8	0.9	0.9	0.9	0.9	0.91		8	0.75	0.83	0.85	0.84	0.81
	9	0.94	0.86	0.93	0.92	0.92		9	0.94	0.92	0.96	0.93	0.91
	10	0.88	0.9	0.88	0.84	0.8		10	0.76	0.82	0.75	0.73	0.76
	11	0.8	0.87	0.86	0.85	0.89		11	0.67	0.82	0.83	0.87	0.87
	12	0.91	0.93	0.96	0.95	0.94		12	0.78	0.81	0.83	0.85	0.83
	13	0.92	0.94	0.96	0.95	0.95		13	0.82	0.88	0.9	0.87	0.88
	14	0.93	0.94	0.93	0.94	0.94		14	0.84	0.9	0.89	0.87	0.84
D	1	0.96	0.97	0.96	0.97	0.97	FL	1	0.86	0.79	0.9	0.87	0.9
	2	0.85	0.89	0.84	0.79	0.78		2	0.73	0.81	0.81	0.83	0.75
	3	0.9	0.84	0.83	0.87	0.87		3	0.85	0.82	0.9	0.83	0.91
	4	0.9	0.89	0.91	0.85	0.87		4	0.91	0.89	0.9	0.88	0.67
	5	0.93	0.92	0.91	0.92	0.92		5	0.88	0.88	0.86	0.82	0.8
	6	0.92	0.95	0.95	0.95	0.89		6	0.81	0.82	0.82	0.84	0.6
	7	0.86	0.88	0.87	0.83	0.71		7	0.52	0.75	0.71	0.58	0.91
	8	0.91	0.92	0.95	0.95	0.95		8	0.87	0.89	0.84	0.82	0.91
	9	0.98	0.97	0.93	0.97	0.96		9	0.95	0.84	0.87	0.91	0.85
	10	0.85	0.84	0.78	0.72	0.78		10	0.76	0.6	0.78	0.6	0.63
	11	0.85	0.89	0.91	0.91	0.9		11	0.78	0.88	0.89	0.87	0.89
	12	0.93	0.95	0.95	0.95	0.94		12	0.92	0.87	0.95	0.94	0.94
	13	0.8	0.86	0.89	0.88	0.87		13	0.96	0.9	0.96	0.96	0.96
	14	0.95	0.93	0.94	0.94	0.94		14	0.89	0.92	0.89	0.91	0.91
I	1	0.93	0.94	0.92	0.92	0.92	NW	1	0.85	0.9	0.89	0.9	0.9
	2	0.84	0.87	0.79	0.87	0.69		2	0.7	0.82	0.8	0.8	0.77
	3	0.95	0.95	0.94	0.79	0.8		3	0.89	0.81	0.79	0.8	0.89
	4	0.82	0.8	0.79	0.83	0.78		4	0.88	0.9	0.9	0.89	0.88
	5	0.93	0.91	0.89	0.91	0.91		5	0.9	0.85	0.85	0.84	0.83
	6	0.83	0.85	0.91	0.91	0.9		6	0.75	0.77	0.84	0.86	0.79
	7	0.87	0.88	0.87	0.9	0.92		7	0.84	0.75	0.71	0.82	0.82
	8	0.82	0.81	0.78	0.82	0.73		8	0.94	0.91	0.93	0.93	0.92
	9	0.98	0.92	0.94	0.94	0.92		9	0.94	0.95	0.91	0.92	0.9
	10	0.91	0.9	0.88	0.87	0.89		10	0.81	0.84	0.84	0.85	0.87
	11	0.75	0.84	0.78	0.8	0.73		11	0.77	0.85	0.89	0.89	0.89
	12	0.94	0.96	0.95	0.93	0.93		12	0.91	0.93	0.94	0.94	0.94
	13	0.95	0.98	0.98	0.98	0.97		13	0.93	0.96	0.97	0.97	0.96
	14	0.91	0.93	0.92	0.88	0.87		14	0.89	0.88	0.88	0.92	0.88

Table 52: Proportion of explained variability when a proxy for size explains twelve accounting items by industry and by year. First Table.

item	ind	1983	1984	1985	1986	1987	item	ind	1983	1984	1985	1986	1987
S	1	0.97	0.98	0.98	0.98	0.98	WC	1	0.64	0.81	0.75	0.78	0.75
	2	0.64	0.66	0.65	0.68	0.64		2	0.87	0.87	0.77	0.75	0.82
	3	0.95	0.88	0.94	0.9	0.91		3	0.52	0.75	0.74	0.77	0.81
	4	0.9	0.88	0.88	0.86	0.86		4	0.87	0.83	0.81	0.61	0.71
	5	0.96	0.9	0.94	0.93	0.94		5	0.61	0.81	0.74	0.62	0.82
	6	0.97	0.98	0.96	0.97	0.94		6	0.68	0.69	0.7	0.56	0.76
	7	0.94	0.97	0.96	0.95	0.92		7	0.72	0.7	0.64	0.84	0.78
	8	0.94	0.94	0.93	0.93	0.93		8	0.85	0.87	0.88	0.8	0.84
	9	0.99	0.95	0.97	0.96	0.95		9	0.75	0.86	0.86	0.78	0.79
	10	0.96	0.9	0.96	0.91	0.92		10	0.72	0.49	0.62	0.56	0.41
	11	0.86	0.93	0.93	0.92	0.94		11	0.68	0.84	0.83	0.83	0.73
	12	0.95	0.96	0.97	0.97	0.97		12	0.88	0.79	0.58	0.87	0.81
	13	0.99	0.99	0.98	0.98	0.98		13	0.91	0.96	0.92	0.92	0.82
	14	0.91	0.93	0.92	0.92	0.89		14	0.8	0.8	0.74	0.78	0.74
W	1	0.98	0.98	0.98	0.98	0.97	DEBT	1	0.55	0.68	0.69	0.55	0.65
	2	0.77	0.78	0.83	0.85	0.86		2	0.38	0.31	0.38	0.43	0.5
	3	0.94	0.93	0.92	0.92	0.93		3	0.51	0.27	0.4	0.5	0.63
	4	0.95	0.95	0.95	0.96	0.95		4	0.73	0.73	0.63	0.52	0.53
	5	0.95	0.92	0.88	0.95	0.94		5	0.73	0.49	0.46	0.41	0.36
	6	0.96	0.98	0.96	0.97	0.94		6	0.23	0.86	0.86	0.6	0.33
	7	0.93	0.97	0.96	0.96	0.94		7	0.04	0.34	0.24	0.37	0.58
	8	0.88	0.9	0.91	0.92	0.91		8	0.54	0.62	0.55	0.49	0.49
	9	0.98	0.95	0.97	0.95	0.97		9	0.72	0.92	0.83	0.68	0.66
	10	0.88	0.88	0.87	0.88	0.88		10	0.34	0.26	0.39	0.32	0.29
	11	0.77	0.87	0.89	0.91	0.9		11	0.12	0.31	0.54	0.52	0.59
	12	0.88	0.9	0.92	0.93	0.92		12	0.71	0.74	0.83	0.84	0.76
	13	0.93	0.95	0.94	0.93	0.91		13	0.65	0.85	0.72	0.66	0.54
	14	0.93	0.94	0.95	0.95	0.95		14	0.82	0.77	0.73	0.68	0.66
CA	1	0.98	0.99	0.95	0.98	0.99	EBIT	1	0.82	0.78	0.82	0.82	0.88
	2	0.88	0.91	0.9	0.91	0.93		2	0.75	0.83	0.78	0.83	0.84
	3	0.95	0.97	0.96	0.96	0.97		3	0.89	0.77	0.83	0.86	0.88
	4	0.97	0.93	0.94	0.91	0.87		4	0.86	0.89	0.83	0.85	0.81
	5	0.98	0.96	0.97	0.97	0.97		5	0.75	0.82	0.82	0.7	0.88
	6	0.97	0.95	0.97	0.96	0.92		6	0.66	0.76	0.73	0.75	0.48
	7	0.95	0.96	0.96	0.96	0.95		7	0.4	0.75	0.8	0.77	0.84
	8	0.9	0.95	0.95	0.96	0.94		8	0.82	0.8	0.77	0.84	0.82
	9	0.99	0.94	0.97	0.84	0.81		9	0.89	0.79	0.92	0.75	0.77
	10	0.94	0.93	0.92	0.92	0.92		10	0.35	0.6	0.72	0.58	0.45
	11	0.87	0.93	0.92	0.93	0.91		11	0.68	0.83	0.86	0.83	0.84
	12	0.95	0.96	0.97	0.97	0.94		12	0.84	0.89	0.95	0.91	0.92
	13	0.97	0.98	0.98	0.97	0.96		13	0.97	0.96	0.92	0.9	0.93
	14	0.95	0.96	0.96	0.92	0.96		14	0.87	0.89	0.86	0.89	0.85

Table 53: Proportion of explained variability when a proxy for size explains twelve accounting items by industry and by year. Second Table.

1983, Building Materials

S	NW	W	I	D	C	CA	FA	TA	CL	TC	
0.366											S
0.325	0.357										NW
0.350	0.331	0.360									W
0.388	0.359	0.398	0.472								I
0.350	0.304	0.345	0.383	0.358							D
0.358	0.296	0.340	0.380	0.358	0.387						C
0.344	0.314	0.342	0.386	0.338	0.335	0.339					CA
0.341	0.337	0.335	0.369	0.324	0.327	0.324	0.359				FA
0.344	0.323	0.339	0.379	0.334	0.333	0.334	0.338	0.336			TA
0.353	0.303	0.341	0.383	0.356	0.370	0.339	0.327	0.336	0.369		CL
0.345	0.350	0.348	0.382	0.327	0.317	0.336	0.356	0.343	0.321	0.369	TC

1983, Clothing

S	NW	W	I	D	C	CA	FA	TA	CL	TC	
0.186											S
0.182	0.236										NW
0.176	0.172	0.205									W
0.214	0.221	0.199	0.277								I
0.188	0.194	0.168	0.238	0.236							D
0.189	0.181	0.178	0.226	0.206	0.220						C
0.184	0.199	0.170	0.225	0.202	0.195	0.200					CA
0.182	0.196	0.197	0.214	0.193	0.188	0.184	0.262				FA
0.180	0.193	0.174	0.216	0.195	0.189	0.191	0.202	0.190			TA
0.191	0.175	0.186	0.233	0.213	0.215	0.198	0.209	0.199	0.235		CL
0.181	0.227	0.171	0.220	0.194	0.182	0.199	0.197	0.193	0.176	0.226	TC

1983, Food

S	NW	W	I	D	C	CA	FA	TA	CL	TC	
0.620											S
0.578	0.745										NW
0.588	0.682	0.724									W
0.656	0.675	0.660	0.777								I
0.575	0.585	0.578	0.636	0.570							D
0.587	0.589	0.595	0.651	0.571	0.608						C
0.576	0.620	0.591	0.660	0.569	0.584	0.597					CA
0.612	0.764	0.769	0.699	0.612	0.629	0.634	0.904				FA
0.579	0.647	0.626	0.662	0.570	0.587	0.596	0.692	0.615			TA
0.579	0.586	0.588	0.652	0.563	0.591	0.579	0.633	0.586	0.589		CL
0.592	0.736	0.697	0.688	0.594	0.602	0.628	0.785	0.660	0.601	0.745	TC

Figure 83: Typical Σ matrices. Items having only positive cases.

1985, Building Materials

	S	W	I	D	C	CA	FA	DEBT	WC	FL	EBIT	N
S	0.38											
W	0.39	0.42										
I	0.40	0.42	0.46									
D	0.33	0.35	0.35	0.31								
C	0.35	0.36	0.37	0.32	0.35							
CA	0.35	0.37	0.38	0.32	0.33	0.34						
FA	0.43	0.46	0.48	0.39	0.40	0.42	0.60					
DEBT	0.52	0.56	0.58	0.46	0.50	0.50	0.68	1.01				
WC	0.35	0.36	0.40	0.31	0.31	0.36	0.44	0.52	0.48			
FL	0.38	0.39	0.42	0.34	0.35	0.37	0.49	0.56	0.40	0.42		
EBIT	0.36	0.37	0.41	0.33	0.34	0.36	0.48	0.53	0.42	0.42	0.44	
N	0.37	0.40	0.40	0.33	0.34	0.35	0.45	0.54	0.34	0.38	0.36	0.39

1985, Electronics

	S	W	I	D	C	CA	FA	DEBT	WC	FL	EBIT	N
S	0.63											
W	0.63	0.68										
I	0.66	0.66	0.77									
D	0.58	0.61	0.62	0.57								
C	0.62	0.64	0.66	0.59	0.64							
CA	0.61	0.63	0.66	0.59	0.62	0.63						
FA	0.64	0.67	0.67	0.61	0.64	0.64	0.72					
DEBT	0.54	0.59	0.63	0.55	0.59	0.58	0.63	0.93				
WC	0.65	0.68	0.71	0.63	0.66	0.68	0.68	0.62	0.78			
FL	0.60	0.62	0.63	0.58	0.61	0.61	0.63	0.53	0.65	0.63		
EBIT	0.61	0.62	0.66	0.58	0.62	0.61	0.62	0.50	0.66	0.65	0.71	
N	0.61	0.65	0.65	0.58	0.62	0.61	0.63	0.57	0.65	0.59	0.60	0.64

1985, Clothing

	S	W	I	D	C	CA	FA	DEBT	WC	FL	EBIT	N
S	0.18											
W	0.20	0.27										
I	0.19	0.23	0.22									
D	0.16	0.16	0.17	0.18								
C	0.18	0.20	0.18	0.16	0.19							
CA	0.18	0.19	0.20	0.17	0.18	0.19						
FA	0.19	0.24	0.21	0.16	0.20	0.19	0.26					
DEBT	0.22	0.24	0.24	0.23	0.23	0.24	0.23	0.53				
WC	0.16	0.17	0.19	0.18	0.16	0.19	0.17	0.27	0.25			
FL	0.21	0.23	0.22	0.19	0.21	0.21	0.23	0.30	0.21	0.28		
EBIT	0.21	0.21	0.21	0.20	0.21	0.21	0.21	0.29	0.21	0.27	0.28	
N	0.20	0.28	0.22	0.14	0.19	0.18	0.24	0.22	0.16	0.21	0.20	0.29

1985, Food

	S	W	I	D	C	CA	FA	DEBT	WC	FL	EBIT	N
S	0.55											
W	0.55	0.62										
I	0.56	0.58	0.63									
D	0.51	0.51	0.54	0.50								
C	0.56	0.57	0.60	0.54	0.61							
CA	0.54	0.55	0.59	0.52	0.57	0.56						
FA	0.58	0.65	0.63	0.55	0.60	0.60	0.72					
DEBT	0.58	0.60	0.61	0.56	0.60	0.58	0.65	0.87				
WC	0.55	0.57	0.61	0.54	0.57	0.59	0.62	0.59	0.78			
FL	0.58	0.63	0.64	0.56	0.62	0.61	0.68	0.64	0.67	0.71		
EBIT	0.62	0.66	0.70	0.62	0.67	0.67	0.72	0.69	0.74	0.78	0.88	
N	0.52	0.61	0.56	0.48	0.56	0.53	0.63	0.57	0.54	0.60	0.63	0.61

Figure 84: Typical Σ matrices. Only cases with positive items were accepted.

Appendix B

Classification Results Using the Multi-Layer Perceptron

In this appendix we gather information concerning the experiment described in section 7.3 about the Multi-Layer Perceptron (MLP) as a modelling tool for accounting relations. But here we examine the MLP as a classifier, intended to be used instead of Multiple Discriminant Analysis (MDA). Therefore, we focus on the classification performance rather than on the acquisition of knowledge.

Each section of this appendix contains the description of a particular test. Firstly, the technique usual in accounting research, involving 18 ratios as input variables, is described. The results of using MDA are compared with those of using MLP. Next we apply the framework developed in the first part of this study instead of the usual one, both with MDA and MLP modelling. It uses eight log items as inputs. Finally, we show the classification obtained with the new ratios devised by the MLP when used as inputs for MDA.

This appendix is intended to show the importance of implementing our framework in a particular, well known, problem. Also, the circumstances leading the MLP to outperform the linear tools can be devised.

B.1 Results: The Usual Technique

In this section we describe the procedures and results obtained when applying to the classification problem the techniques which are usual in accounting research, that is,

- Input variables are ratios selected so as to reflect desired features.
- Ratios suffer ad-hoc transformations. The goal is to achieve improvements in their normality.
- Factor Analysis is used to extract a few variables from the set of transformed ratios.

Ratio	Skewness	Kurtosis	Ratio	Skewness	Kurtosis
log NW	0.42	0.01	log S	0.38	0.09
DD	0.59	1.60	FA/TA	0.33	-0.15
S/FA	4.5	28.5	W/TA	1.09	2.17
VA/TCE	2.5	13.8	OPP/S	0.17	6.05
EBIT/S	0.53	5.95	OPP/TCE	1.85	34.8
EBIT/TCE	1.25	24.5	S/TA	2.01	6.85
S/I	2.94	12.1	D/CA	1.45	9.70
D/I	2.41	11.2	D/C	1.78	5.74
DEBT/NW	3.31	18.1	DEBT/TCE	3.31	18.1

Table 54: Skewness and kurtosis of ratios used in the replica of the traditional study. The displayed values were obtained after applying transformations. DD is the Days Debtors ratio.

- Multiple Discriminant Analysis uses such factors as input variables. In this case, the industrial grouping according to the SEIC is the outcome.

Our study reproduces a reputed one, carried out in 1984 by Sudarsanam and Taffler [124] and quoted by Foster. The ratios used and their transformations are displayed elsewhere (see page 174).

Normality of transformed ratios: We obtained a broad set of values for the skewness and kurtosis of the ratios used in the replication of the study referred to. Such values are displayed in table 54.

DEBT has a large number of zero cases corresponding to non-leveraged firms. It will not yield homogeneous distributions with any transformation. The factor extracted from DEBT ratios exhibit a very strong two-modality.

Extraction of factors from ratios: After obtaining the transformed ratios we extract the eight largest components of their variability. Next we display the differences between our study and the original one concerning the affinity of input variables with the resulting factors.

1. Operating Scale: We obtained the same groups.
2. Fixed Capital Intensity, the same groups.
3. Labour Capital Intensity, the same groups.
4. Profitability, the same groups.
5. Asset Turnover: This factor was formed with variability from S/TA and S/I mainly.
6. Short Term Asset Intensity, DD, D/C, D/CA and D/I.
7. Net Trade Credit, DD, D/C, D/CA and D/I.
8. Leverage, the same groups.

Therefore, our study found differences in the interpretation of the factors related to Short Term. In our data the three factors representing short-term features have their variables mixed up.

The co-variance matrix was almost singular. The main correlations were observed between

- $\log S$ and $\log NW$ (0.995),
- $\sqrt{OPP/S}$ and $\log(EBIT/S)$ (0.971),
- $\sqrt{OPP/TCE}$ and $\sqrt{EBIT/TCE}$ (0.980),
- $\sqrt{FA/TA}$ and $\log(S/TA)$ (0.970),
- $\sqrt{S/TA}$ and $\log(S/I)$ (0.922) and finally between
- $\sqrt{DEBT/NW}$ and $DEBT/TCE$ (0.970).

The eigenvalue sequence doesn't exhibit the smallest trace of a break in the rate of decay. It decays smoothly in an exponential way. The factors are, more or less, replicating the original variables. Hence, there is no clear distinction between the selected factors and the rejected ones. There is no real commonality or real uniqueness and each factor contains a good portion of the information others contain. Since the purpose is the reduction in the number of dimensions, not the discovering of features, this is just as well.

An eigenvalue sequence can have values like these:

$$20\%,\ 19\%,\ 15\%,\ 15\%,\ 13\%,\ 9\%,\ 8\%,\ 6\%$$

Typically, eight factors will account for more than 90% of the variability.

Transformations: We must remark that the effect of using different transformations inside the same set of input variables introduces a non-negligible amount of non-linearity in input space. If two linearly related variables are exposed to different transformations, say, one square root and the other logs, the resulting relation between them is no longer linear. Afterwards, when factor analysis is used to extract new variables from these non-linear ones, the clear result will be that most of the variability associated with the extreme values — the ones which are most curled by the non-linearity — is flattened away. Factor analysis extract linear patterns.

Hence, the final result of this interaction between artificial non-linearity and linear factor analysis is that the extreme values of the distribution will be pushed towards the centre of the distribution.

Multiple Discriminant Analysis: A diversion from the original study consisted of dividing the set of examples randomly in two approximately equal sized samples. The MDA model was built with one of the samples but its performance was checked with the other one. In general, the size of each group in one set and in the other are not very similar. This fact introduces a distortion in the

N.	SEIC Code	Group Name	N. Cases	Correct	N. Cases	Correct
1	14	Building Materials	8	3	23	1
2	32	Metallurgy	11	1	8	2
3	54	Paper and Pack	25	5	21	1
4	68	Chemicals	22	4	23	7
5	19	Electrical	16	3	18	4
6	22	Industrial Plants	8	2	9	1
7	28	Machine Tools	11	2	10	1
8	35	Electronics	49	11	35	14
9	41	Motor Components	17	4	6	5
10	59	Clothing	19	10	23	9
11	61	Wool	7	1	12	1
12	62	Miscellaneous Textiles	11	1	19	1
13	64	Leather	8	1	8	3
14	49	Food Manufacturers	43	25	37	23

Table 55: Classification results with MDA and 8 factors.

classification results since the likelihood of each group in the test set is different from the likelihood in the training set. However, by imposing equal prior probabilities across groups this distortion is minimized.

We are mainly interested in comparing the performance of MDA with that of the Multi-Layer Perceptron. Provide the samples are the same and the prior assumptions coincide, this comparison can be carried out.

Table 55 shows the classification results. *N. Cases* displays the number of cases in a group after split in two random samples. *Correct* shows the number of correct classifications when that group was used to model and the other group was used to test.

The displayed results and all the other results reported were obtained under the supposition of equal prior likelihood of any firm to belong to this group or the other. There is no special reason why a prior knowledge about relative size of groups should be included in this study.

For small groups the classification is very poor. It increases dramatically with the size of the group. An overall 29% of success in both cases is attained almost because of very good classification of groups like Food and Electronics.

B.2 MLP With 8 Factors as Input Variables

The same eight factors which were used as input variables for MDA were also tested as input for an MLP. After several experiments we found that the best results would be achieved with an MLP with one hidden layer and six nodes on it. Table 56 contains the number of correct classifications in the test set, by group.

These results concern an MLP with six nodes in a unique hidden layer and 14 output nodes (one per group). Outputs were post-processed as described in section 7.4.3, on page 180 but no random penalization of weights were applied. The criterion used for convergence was the maximization of the likelihood input-outcomes.

N.	SEIC Code	Group Name	N. Cases	Correct	N. Cases	Correct
1	14	Building Materials	8	3	23	0
2	32	Metallurgy	11	0	8	1
3	54	Paper and Packing	25	6	21	1
4	68	Chemicals	22	4	23	7
5	19	Electrical	16	4	18	6
6	22	Industrial Plants	8	2	9	0
7	28	Machine Tools	11	1	10	0
8	35	Electronics	44	12	35	16
9	41	Motor Components	17	3	6	5
10	59	Clothing	19	12	23	10
11	61	Wool	7	0	12	0
12	62	Miscellaneous Textiles	11	2	19	1
13	64	Leather	8	0	8	3
14	49	Food Manufacturers	43	28	37	24

Table 56: Classification results with MLP and 8 factors.

The training was interrupted when the likelihood, measured in the training set, reached a maximum. This procedure is therefore different from the one referred to in section 7.4.1, page 176. It allows a direct comparison with the results of the MDA modelling.

Under the displayed conditions, the MLP shows a performance which is similar to the one of MDA (about 30% of correct classifications), a linear, analytic tool. We believe that the improvements in performance achieved in later experiments stem from the interruption of training before its completion and also from the more robust pre-processing of the input data.

B.3 MLP and MDA With Eight Log Items as Input Variables

We now describe our procedure for modelling the relation between accounting information and industry grouping.

We recall that the new approach consisted of using eight accounting variables directly, not in the form of ratios. A simple two-parameter log transformation and a mean-adjustment was all the manipulation suffered by the items before being used as input variables for classification. The logs used were the decimal ones. Notice that there is a more subtle difference between the MLP and the MDA procedures in what concerns the pre-processing of data. The MDA standardizes the input variables one by one. The MLP uses all the information contained in the differences of spread.

The selected items were Fixed Assets, Inventory, Debtors, Creditors, Long Term Debt, Net Worth, Wages and Sales less Operating Expenses. All these variables were present in the original 18 ratios, along with others like Earnings, Value Added, Total Capital Employed and Total Assets which we didn't use in the new approach.

All the log items were mean-adjusted before being presented as input. The overall mean, not the industry-specific one, was used for this. Therefore, the input variables are not just log items but what we call relative positions (see equation 4 on page 59). No correction for δ was introduced.

When using the analytical tool for modelling with these eight positions we obtained about 33-34%

N.	SEIC Code	Group Name	N. Cases	Correct	N. Cases	Correct
1	14	Building Materials	8	2	23	1
2	32	Metallurgy	11	2	8	2
3	54	Paper and Packing	25	5	21	6
4	68	Chemicals	22	4	23	7
5	19	Electrical	16	5	18	4
6	22	Industrial Plants	8	1	9	2
7	28	Machine Tools	11	2	10	2
8	35	Electronics	44	21	35	14
9	41	Motor Components	17	4	6	5
10	59	Clothing	19	10	23	9
11	61	Wool	7	1	12	4
12	62	Miscellaneous Textiles	11	2	19	4
13	64	Leather	8	1	8	1
14	49	Food Manufacturers	43	26	37	23

Table 57: Classification results with MDA and 8 log items.

N.	SEIC Code	Group Name	N. Cases	Correct	N. Cases	Correct
1	14	Building Materials	8	4	23	10
2	32	Metallurgy	11	1	8	1
3	54	Paper and Packing	25	5	21	2
4	68	Chemicals	22	4	23	9
5	19	Electrical	16	5	18	6
6	22	Industrial Plants	8	2	9	0
7	28	Machine Tools	11	5	10	1
8	35	Electronics	44	17	35	19
9	41	Motor Components	17	5	6	5
10	59	Clothing	19	10	23	11
11	61	Wool	7	2	12	0
12	62	Miscellaneous Textiles	11	2	19	2
13	64	Leather	8	1	8	3
14	49	Food Manufacturers	43	32	37	25

Table 58: The best classification results with MLP and 8 log items.

of correct classifications in the test set. The detailed results are gathered in table 57. It seems clear that, just by avoiding all the entangling pre-processing of data traditional in accounting research and using the log space instead, some improvements in performance can be observed.

Table 58 shows the best classification results the MLP is able to achieve. The improvement, from 33%-34% to 37%-38%, is due to the interruption of training in the optimum for the test set rather than in the optimum for the training set. It is also a consequence of the better generalisation introduced by forcing a reduction in the number of free parameters in the net.

B.4 Using the Devised Set of Ratios With MDA

We now show the results obtained when using a devised set of ratios for modelling with analytic tools. These ratios are a free interpretation of the best topology the MLP builds after learning the

N.	SEIC Code	Group Name	N. Cases	Correct	N. Cases	Correct
1	14	Building Materials	8	2	23	6
2	32	Metallurgy	11	2	8	2
3	54	Paper and Packing	25	4	21	5
4	68	Chemicals	22	4	23	4
5	19	Electrical	16	5	18	2
6	22	Industrial Plants	8	1	9	0
7	28	Machine Tools	11	0	10	2
8	35	Electronics	44	14	35	14
9	41	Motor Components	17	3	6	1
10	59	Clothing	19	10	23	7
11	61	Wool	7	0	12	3
12	62	Miscellaneous Textiles	11	1	19	3
13	64	Leather	8	1	8	1
14	49	Food Manufacturers	43	25	37	19

Table 59: Classification results with MDA and the five discovered ratios plus a proxy for size.

relation. We recall that the ratios used were five,

$$\begin{cases} \text{In the } 2^{st} \text{ node:} & \dfrac{NW \times I}{FA \times EX} \\ \text{In the } 3^{st} \text{ node:} & \dfrac{EX}{FA} \\ \text{In the } 4^{st} \text{ node:} & \dfrac{EX}{DB} \\ \text{In the } 5^{st} \text{ node:} & \dfrac{FA \times C}{W \times EX} \\ \text{In the } 6^{st} \text{ node:} & \dfrac{EX}{\sqrt{FA \times D}} \end{cases}$$

along with a proxy for the strong, common effect.

Table 59 shows the best generalisation achieved. It is around 29%.

Though the results are not impressive by themselves, we must remember that they approach those obtained with 18 ratios. Anyway, the important point here is to notice that the MLP was able to point out the items which are important in the modelling of the relation.

B.5 Conclusions

Generalisation results show that under similar conditions little difference exists between the MDA and the MLP results for the particular problem we studied. Clearly, the relation to be modelled must be near linearity. This is a fortunate circumstance. It allows the strict comparing of these tools in a problem for which the linear, analytical, procedure has not been put in a position of disadvantage.

An interesting achievement is the ability displayed by the MLP to deal with simple, linear, relations with no losses in generalisation. Algorithms like polynomial fitting would perform badly if required to model a straight line. The MLP did it easily. Hence, the Multi-layer Perceptron emerges as a general-purpose tool, to which we can trust the task of modelling a broad class of relations, ranging from the simple, linear, ones to the most complex ones.

When using 18 ratios and the procedures typical in accounting research — including the extraction of eight factors — both the MDA and MLP generalisation results range from 29% to 30%. The use of eight log items instead of the eighteen transformed and rotated ratios introduces an expected improvement in the generalisation achieved. Both the MLP and the MDA now range from 33% to 34% of correct classifications in the test set. This clearly shows the disadvantages of such techniques based on standard recipes.

By stopping the learning process in the optimal classification for the test set rather than for the learning one a considerable improvement is added to the experiment with eight items. The generalisation is up to 37% - 38%. Naturally, analytic tools like the MDA cannot replicate this experiment. The classification results are summarized in next table.

INPUT	MDA	MLP
18 ratios	29%	30%
8 variables	34%	38%

Finally, the five ratios inspired by the ones formed inside the MLP plus the estimated size, are able to achieve 28% - 29% of correct classification in the test set, which is similar to the performance of the original 18 ratios.

All the results suffer from the same problem, the virtual disappearance of the small groups — the overwhelming dominance of the large ones. The proportion of correct classification is related to the proportion of cases in the learning set in the sense that large groups are correctly recognized whilst the small ones are ignored.

Appendix C

Tools And Algorithms Used In This Study

This appendix enumerates the main tools and algorithms used during this study to perform data manipulation.

Statistical computation was performed with the SPSS-X package. The values of skewness and kurtosis, whenever displayed, refer to the corresponding algorithm of this package. Notice that the normal kurtosis is zero. Intra-Class correlations (chapter 5) were computed from the mean-squares obtained with SPSS-X. ALSCAL routines were used to build two-dimensional maps from standard deviations (chapter 5).

The algorithms for simulation of Neural Networks were built in FORTRAN 77. All the main manipulation of data and calculations were also programed in FORTRAN 77. No external subroutines were used. The code was written by the author.

The simulation of accounting variables and the testing of normality (chapter 1) was made also in the same environment. In this case we used the NAG Mark 11 library of subroutines. For example, the generation of random normal deviates used the NAG subroutine G05DDF and the Shapiro-Wilk tests were performed with the algorithm G01DDF.

Bibliography

[1] J. Aitchison and J. Brown. *The Lognormal Distribution.* Cambridge University press, 1957.

[2] L. Almeida. A learning rule for asynchronous perceptrons with feedback in a combinational environment. In M. Caudil and C. Butler, editors, *Proceedings of the First IEEE International Conference on Neural Networks*, 1987. San Diego, California, June.

[3] E. Altman. Financial ratios, discriminant analysis and the prediction of corporate bankrupcy. *The Journal of Finance*, pages 589–609, September 1968.

[4] E. Altman, R. Haldeman, and P. Narayanan. Zeta analysis: A new model to identify bankrupcy risk of corporations. *Journal of Banking and Finance*, pages 29–54, June 1977.

[5] P. Barnes. Methodological implications of non-normally distributed financial ratios. *Journal of Business, Finance and Accounting*, pages 51–62, Spring 1982.

[6] E. Baum and F. Wilkzek. Supervised learning of probability distributions by neural networks. In American Institute of Physics, editor, *Neural Information Processing Systems - Natural and Synthetic*, pages 52–61, 1987. Denver, Colorado.

[7] W. Beaver. Financial ratios as predictors of failure. *Journal of Accounting Research*, pages 71–111, 1966. Suplement.

[8] W. Beaver. Market prices, financial ratios and the prediction of failure. *Journal of Accounting Research*, Autumn:179–192, 1968.

[9] W. Beaver, H. Kettler, and M. Sholes. The association between market-determined and accounting-determined risk measures. *The Accounting Review*, pages 654–682, October 1970.

[10] W. Beaver and J. Manegold. The association between market-determined and accounting-determined measures of systematic risk: Some further evidence. *Journal of Financial and Quantitative Analysis*, pages 231–284, June 1975.

[11] R. Berry and S. Nix. Regression analysis vs. ratios in the cross-sectional analysis of financial statements. *Accounting and Business Research*, 21(82):107–117, 1991.

[12] R. Berry and D. Trigueiros. Using the id3 algorithm to interpret the results of financial models. Technical Report SYS, University of East Anglia — Presented in the British Accounting Association Annual Meeting, April 1990, Dundee, Scotland, 1990.

[13] J. Bildersee. The association between a market-determined measure of risk and alternative measures of risk. *The Accounting Review*, pages 81–88, January 1975.

[14] R. Bird and A. McHugh. Financial ratios - an empirical study. *Journal of Business Finance and Accounting*, pages 29–45, Spring 1977.

[15] M. Blume. Betas and thair regression tendencies. *The Journal of Finance*, pages 785–795, June 1975.

[16] R. Booton. Nonlinear control systems with stochastic inputs. Technical Report 61, MIT: Dyn. An. and Contr. Lab., 1952.

[17] P. D. Bougen and J. Drury. U.k. statistical distributions of financial ratios, 1975. *Journal of Business, Finance and Accounting*, 1(7):39–47, 1980.

[18] R. Bowman. The theoretical relationship between systematic risk and financial (accounting) variables. *The Journal of Finance*, pages 617–630, June 1979.

[19] R. Bowman. The importance of market value measurement of debt in assessing leverage. *Journal of Accounting Research*, Spring:242–254, 1980.

[20] P. Brown and R. Ball. *Some Perliminary Findings on the Association Between the Earnings of a Firm, its Industry, and the Economy.* Empirical Research in Accounting: Selected Studies, 1967. Suplement to Journal of Accounting Research.

[21] W. Buijink and M. Jegers. Cross-sectional distributional properties of financial ratios in belgian manufacturing industries: Some empirical evidence. Technical report, University of Antwerp, Belgium, 1984.

[22] K. Chen and A. Shimerda. An empirical analysis of useful financial ratios. *Financial Management*, pages 11–26, Spring 1981.

[23] N. Chen. Some empirical tests of the theory of arbitrage pricing. *The Journal of Finance*, pages 1393–1414, December 1983.

[24] G. Connor. A unified beta pricing theory. *Journal of Economic Theory*, 34:13–31, 1984.

[25] R. Cook. Detection of influential observations in linear regression. *Technometrics*, 19:15–18, 1977.

[26] T. Copeland and J. Weston. *Financial Theory and Corporate Policy*. Addison-Wesley Publishing Co., second edition, 1983.

[27] E. Deakin. Distributions of financial accounting ratios: Some empirical evidence. *The Accounting Review*, pages 90–96, January 1977.

[28] G. Diacogiannis. Arbitrage pricing model: A critical examination of its empirical applicability for the london stock exchange. *Journal of Business, Finance and Accounting*, pages 489–503, Winter 1986.

[29] E. Dimson. Risk measurement when shares are subject to infrequent trading. *Journal of Financial Economics*, 7:197–226, 1979.

[30] T. Dimson, E. Marsh. Risk, return and company size effect: The thirty year record. Working Paper, London Business School, 1985.

[31] S. Dudda and S. Shekhar. A non-conservative application of neural networks. In R. Trippi and E. Turban, editors, *Investment Management, Decision Support and Expert Systems*, pages 271–282. Boyd and Fraser, 1990.

[32] P. Dybvig. An explicit bound on individual asset's deviations from apt pricing in a finite economy. *Journal of Financial Economics*, 12:483–497, 1983.

[33] P. Dybvig and S. Ross. Yes. the apt is testable. *The Journal of Finance*, pages 1173–1188, September 1985.

[34] R. Eisenbeis. Pitfalls in the application of discriminant analysis in business, finance and economics. *The Journal of Finance*, 32(3):875–899, June 1977.

[35] E. Elton and M. Gruber. Homogeneous groups and the test of economic hypothesis. *Journal of Financial and Quantitative Analysis*, 4:581–602, 1970.

[36] E. Elton and M. Gruber. Improved forecasting through the design of homogeneous groups. *The Journal of Business*, 44:432–449, 1971.

[37] R. Eubank. *Spline Smoothing and Non-Parametric Regression*. Marcel Dekker, Inc., 1988.

[38] M. Ezzamel and C. Mar-Molinero. The distributional proprieties of financial ratios in u.k. manufacturing companies. *Journal of Business, Finance and Accounting*, 17(1):1–29, 1990.

[39] M. Ezzamel, C. Mar-Molinero, and A. Beecher. On the distributional properties of financial ratios. *Journal of Business, Finance and Accounting*, 14(4):463–481, 1987.

[40] E. Fama and R. French. Dividend yields and expected stock returns. *Journal of Financial Economics*, 22:3–25, 1988.

[41] E. Fama and G. Schwert. Asset returns and inflation. *Journal of Financial Economics*, 5:115–146, 1977.

[42] S. Fieldsend, N. Longford, and S. McLeay. Industry effects and the proportionality assumption in ratio analysis: A variance cmponent analysis. *Journal of Business Finance and Accounting*, 14(4):497–517, Winter 1987.

[43] R. Forsyth. Ai: East / west conference report. *AISB Quarterly*, 66:14–17, 1988.

[44] G. Foster. *Financial Statement Analysis*. Prentice-Hall, 1986.

[45] T. Frecka and W. Hopwood. The effect of outliers on the cross-sectional distributional properties of financial ratios. *The Accounting Review*, pages 115–128, January 1983.

[46] D. Gabor, W. Wilby, and R. Woodcock. A universal non-linear filter, predictor and simulator which optimizes itself by a learning process. *Journal of the Institution of Electrical Engineers*, pages 422–435, 1960.

[47] M. Gahlon and J. Gentry. On the relationship between systematic risk and the degrees of operating and financial leverage. *Financial Management*, Summer:15–23, 1982.

[48] R. Gibrat. *Les Inegalites Economiques*. Librairie du Recueil Sirey, 1931.

[49] N. Gonedes. A test of the equivalent risk class hypothesis. *Journal of Financial and Quantitative Analysis*, 4:159–177, June 1969.

[50] N. Gonedes. Properties of accounting numbers: Models and tests. *Journal of Accounting Research*, Spring:212–237, 1972.

[51] Z. Griliches. Cost allocation in railroad regulation. *Bell Journal of economics and Management Science*, pages 26–41, Spring 1972.

[52] M. Grinblat and S. Titman. Factor pricing in a finite economy. *Journal of Financial Economics*, 12:497–507, 1983.

[53] M. Gultekin and N. Gultekin. Stock return anomalies and the tests of the apt. *The Journal of Finance*, pages 1213–1224, December 1987.

[54] M. Gupta and R. Huefner. A cluster analysis study of financial ratios and industry characteristics. *Journal of Accounting Research*, Spring:77–95, 1972.

[55] R. Hamada. The effect of firm's capital structure on the systematic risk of common stock. *The Journal of Finance*, pages 435–452, March 1972.

[56] J. Hebb. *The Organization of Behavior*. Wiley, 1949.

[57] R. Hecht-Nielsen. Neurocomputer applications. In *Neural Computers*, pages 445–453. NATO ASI Series, Serie F, Number 41, 1987. Edited by R. Eckmiller and C. Malsburg in Springer Verlag, Berlin.

[58] N. Hill and B. Stone. Accounting betas, systematic operating risk and financial leverage: A risk composition approach to the determinants of systematic risk. *Journal of Financial and Quantitative Analysis*, pages 595–637, September 1980.

[59] G. Hinton and T. Sejnowski. Optimal perceptual inference. In *Conference on Computer Vision and Pattern Recognition*, pages 448–453. IEEE, Washington, DC, 1983.

[60] J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. In *Proceedings of the National Academy of Science USA*, page 2554. USA Academy of Science, 1982. Volume 79.

[61] J. Hopfield. Learning algorithms and probability distributions in feed-forward and feed-back networks. In *Proceedings of the National Academy of Science USA*, pages 8429–8433. USA Academy of Science, 1987. Volume 84.

[62] J. Horrigan. Some empirical bases of financial ratio analysis. *The Accounting Review*, pages 558–568, July 1965.

[63] J. Horrigan. The determination of long-term credit standing with financial ratios. *Journal of Accounting Research, Suplement. Empirical Research in Accounting: Select Studies*, pages 44–68, 1966.

[64] J. Horrigan. Methodological implications of non-normally distributed financial ratios: A comment. *Journal of Business Finance and Accounting*, pages 683–689, Winter 1983.

[65] Y. Ijiri and H. Simon. *Skewed Distributions and the Size of Business Firms*. North-Holland, Amsterdam, 1977.

[66] R. Jarrow. *Finance Theory*. Prenice Hall, 1988.

[67] R. Jensen. A cluster analysis study of financial performance of selected business firms. *The Accounting Review*, pages 36–55, January 1971.

[68] I. Kasakov. Approximate probability analysis of the operational precision of essentially non-linear feedback control systems. *Automation and Remote Control*, 17, 1956.

[69] D. Keim. Size-related anomalies and stock return seasonality. further empirical evidence. *Journal of Financial Economics*, pages 13–32, June 1983.

[70] D. Keim and R. Stambaugh. Predicting returns in the stock and bond markets. *Journal of Financial Economics*, 17:357–390, 1986.

[71] R. Keim, D. Stambaugh. A further investigation on the weekend effect in stock returns. *The Journal of Finance*, pages 819–835, July 1984.

[72] T. Kohonen. Clustering, taxonomy and topological maps of patterns. In *Proc Int Conf on Pattern Recognition*. IEEE, 1982.

[73] T. Kohonen. *Self-Organization and Associative Memory*. Springer Verlag,Berlin, 1984.

[74] A. Kolmogoroff. *Interpolation and Extrapolation of Stationary Series*. Academy of Sciences of the URSS, 1942.

[75] A. Lapedes and T. Farber. Nonlinear signal processing using neural networks. Technical Report LA UR 87 2662, Los Alamos National Laboratory, 1987.

[76] Y. LeCun. Optimal brain damage. In *Neural Information Processing Systems*, volume 1. Morgan Kaufmann, Denver 1990.

[77] Y. W. Lee. Contributions of norbert wiener to lineasr and non-linear theory in engineering. *Selected Papers of Norbert Wiener*, 1:17–34, 1964.

[78] B. Lev. On the use of index models in analitical review by auditors. *Journal of Accounting Research*, Autumn:524–550, 1980.

[79] B. Lev and S. Sunder. Methodological issues in the use of financial ratios. *Journal of Accounting and Economics*, pages 187–210, December 1979.

[80] R. Lippmann. An introduction to computing with neural nets. *IEEE ASSP Magazine*, 4:4–22, 1987.

[81] R. Magee. Industry wide commodities in earnings. *Journal of Accounting Research*, Autumn:270–287, 1974.

[82] B. Mandelbrot. The variation of certain speculative prices. *Journal of Business*, 36:394–419, October 1963.

[83] D. Martin, J. Scott and R. Vandell. Equivalent risk classes: A multidimensional examination. *Journal of Financial and Quantitative Analysis*, XIV:101–118, March 1979.

[84] W. McCulloch and W. Pitts. A logical calculus of the ideas immanent in nervous activity. *Boulletin of Mathematical Biophysics*, 5:115–133, 1943.

[85] W.S. McCulloch. *Embodiments of Mind*. The MIT Press, 1965.

[86] S. Mcleay. The ratio of means, the mean of ratios and other benchmarks. *Finance, Journal of the French Finance Society*, 7(1):75–93, 1986.

[87] S. Mcleay. Student's t and the distribution of financial ratios. *Journal of Business, Finance and Accounting*, 13(2):209–222, 1986.

[88] J. Mingers. Rule induction with statistical data — a comparison with multiple regression. *Journal of the Operational Research Society*, 38(4):347–351, 1986.

[89] M. Minsky and S. Papert. *Perceptrons: An Introduction to Computational Geometry*. MIT Press, 1969.

[90] M. Mozer and P. Smolensky. Skeletonization: A technique for trimming the fat from a network via relevance assessment. In *Neural Information Processing Systems*, volume 1. Morgan Kaufmann, Denver 1988.

[91] M. O'Connor. On the usefulness of financial ratios to the investor in common stock. *The Accounting review*, pages 339–352, April 1973.

[92] Y. Pao. *Adaptive Pattern Recognition and Neural Networks*. Addison Wesley, Reading, Massashusets, 1989.

[93] E. Peters. Fractal structure in the capital markets. *The Financial Analysts Journal*, pages 32–37, July 1989.

[94] W. Pitts and W.S. McCulloch. How we know universals: The perception of auditory and visual forms. *Bulletin of Mathematical Biophysiology*, 9:127–147, 1947.

[95] J. Plymen. Classification of stock exchange securities by industry. *Journal of the Institute of Actuaries*, 97(406), 1971.

[96] J. Quinlan. Discovering rules from large collections of samples — a case study. In D. Michie, editor, *Expert Systems in the Micro Electronic Age*. Edimburgh University Press, 1979.

[97] Brealey R. and Myers S. *Principles of Corporate Finance*. McGraw-Hill Book Co., fourth edition, 1981.

[98] P. Race and R. Thomas. Rule induction in investment appraisal. *Journal of the Operational Research Society*, 39(12):1113–1123, 1988.

[99] N. Rao. Equivalent-risk class hypothesis: An empirical study. *Journal of Financial and Quantitative Analysis*, pages 1763–1771, June 1972.

[100] R. Roll. r^2. *The Journal of Finance*, XLIII(2):541–567, July 1988.

[101] R. Roll and S. Ross. An empirical investigation of the arbitrage pricing theory. *The Journal of Finance*, pages 1073–1103, December 1980.

[102] B. Rosenberg and J. Guy. Prediction of beta from investment fundamentals. *The Financial Analysts journal*, I:3–15, May - June 1976.

[103] F. Rosenblatt. *Principles of Neurodynamics.* Spartan Books, Washington, 1961.

[104] S. Ross. The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, 13:341–360, 1976.

[105] J. P. Royston. An extension of the shapiro and wilk's test for normality to large samples. *Applied Statistics*, 31(2):115–124, 1982.

[106] D. Rumelhart, G. Hinton, and R. Williams. Learning internal representations by error propagation. In *Parallel Distributed Processing*, volume 1. MIT Press, 1986.

[107] H. Ryser. *Combinatorial Mathematics.* Wiley, 1963.

[108] M. Schetzen. *The Volterra and Wiener Theories of Non-Linear Systems.* Wiley, 1980.

[109] Scholes and J. Williams. Estimating betas from nonsynchronous data. *Journal of Financial Economics*, 5:309–327, 1977.

[110] T. Sejnowski and C. Rosenberg. Nettalk: A parallel network that learns to read aloud. In *Electr. Engin. Dept.* The John Hopkins University, 1986. JHU/EECS-86/01.

[111] J. Shanken. The arbitrage pricing theory: Is it testable? *The Journal of Finance*, pages 341–360, December 1982.

[112] J. Shanken. Multi-beta capm or equilibrium apt: A reply. *The Journal of Finance*, pages 1189–1196, September 1985.

[113] J. Shanken. Nonsynchronous data and the covariance-factor structure of returns. *The Journal of Finance*, 42:221–231, June 1987.

[114] D. Shen and A. Rosenberg. A nonlinear stochastic learning model. In *Trans. of the third Prague Conf. on Information Theory, Statistical Decision Functions and Random Processes*, pages 629–636. Czekoslovak Academy of Sciences, 1964.

[115] W. Shen. Nonlinear amplitude sensitive control systems with stochastic inputs. In *WESCON Convention Record.* IRE, 1957. part 4.

[116] F. M. Silva and L. B. Almeida. Acceleration techniques for the backpropagation algorithm. In *Neural Networks.* EURASIP Workshop, Sesimbra, Portugal, 1990. L. B. Almeida and C. J. Wellekens (Eds.), Springer-Verlag.

[117] N. Sinclair. Aspects of the factor structure implicit in the australian industrial equity market. *Australian Journal of Management*, pages 23–36, June 1984.

[118] A. Singh and G. Whittington. *Growth Profitability & Valuation*. Cambridge University Press, 1968.

[119] G. Snedecor and W. Cochran. *Statistical Methods*. Iowa State University Press, nineth edition, 1965.

[120] S. Solla, E. Levin, and M. Fleisher. Accelerated learning in layered neural networks. *Complex Systems*, 2:625–640, 1988.

[121] J. Steindl. *Random Process and the Growth of Firms: A Study of the Pareto Law*. Hafner, New York, 1965.

[122] M. Stone. Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society*, B-36, 1974.

[123] M. Stone. Cross-validation: A review. *Math. Operationsforsch. Statist., Ser. Statistics*, 9(1), 1978.

[124] P. Sudarsanam and R. Taffler. Industrial classification in u.k. capital markets: A test of economic homogeneity. *Applied Economics*, 17:291–308, 1985.

[125] R. Taffler. Forecasting company failure in the u.k. using discriminant analysis and financial ratios data. *Journal of the Royal Statistical Society*, 145:342–358, 1982.

[126] G. Tattersall, P. Linford, and R. Linggard. Neural arrays for speech recognition. *Br. Telecom Technol. J.*, 6:140–163, 1988.

[127] D. Thompson. Sources of systematic risk in common stock. *The Journal of Business*, pages 173–188, April 1976.

[128] M. Tippett. An induced theory of financial ratios. *Accounting and Business Research*, pages 77–85, January 1991.

[129] D. Trigueiros and R. Berry. The application of neural network based methods to the extraction of knowledge from accounting reports. In *Proceedings of the* 24^{th} *Hawaii International Conference on System Sciences*, 1991. Honolulu, Hawaii.

[130] J. Utans and J. Moody. Selecting neural network architectures via the prediction risk: Application to corporate bond rating prediction. In IEEE Computer Society Press, editor, *Proceedings of the* 1^{st} *International Conference on Artificial Intelligence Applications on Wall Street*, 1991. Los Alamitos, California.

[131] V. Volterra. *Theory of Functionals*. Blackie, 1930.

[132] A. Weigend, B. Huberman, and Rumelhart D. Predicting the future. a connectionist approach. Technical Report PARC-SSL-90-20, Stanford University, PDP Research Group, 1990.

[133] H. White. Economic prediction using neural networks: The case of ibm daily returns. In R. Trippi and E. Turban, editors, *Investment Management, Decision Support and Expert Systems*, pages 283–292. Boyd and Fraser, 1990.

[134] G. Whittington. Some basic properties of accounting ratios. *Journal of Business, Finance and Accounting*, 7(2):219–232, 1980.

[135] B. Widrow and E. Hoff. Adaptive switching circuits. In *WESCON Convention Record*. IRE, 1960. part 4.

[136] N. Wiener. *The extrapolation, intrapolation and Smoothing of stationary time-series*. Wiley, 1949.

[137] N. Wiener. *Nonlinear Problems in Random Theory*. Wiley, 1958.

[138] N. Wiener. *Cybernetics*. The MIT Press, 1961.

[139] R. Wippern. A note on the equivalent risk class assumption. *Engineering Economist*, Spring:13–22, 1966.