UNIVERSITY OF CAPE TOWN

DEPARTMENT OF MATHEMATICAL STATISTICS

****

THE APPLICATION OF MULTIVARIATE STATISTICAL TECHNIQUES IN THE ANALYSIS OF STOCK MARKET DATA

BY

J.F. AFFLECK-GRAVES

****

A thesis prepared under the supervision of Professor C.G. Troskie and Associate Professor A.H. Money in fulfilment of the requirements for the degree of Doctor of Philosophy in Mathematical Statistics

****

TO RITA

CONTENTS

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to Professor C.G. Troskie and Associate Professor A.H. Money for the advice and encouragement they have given me, not only during the writing of this thesis, but since my undergraduate days. I thank them for guiding me to this extremely interesting field of research, for their many suggestions of worthwhile topics, and for the numerous hours they have sacrificed in discussing the results obtained.

I must also thank the members of the staff of the Department of Mathematical Statistics at the University of Cape Town for the encouragement they have given me. In particular, I must mention Drs. Underhill, Juritz, and Hart whose doors were always open to me. It is indeed an honour to work amongst such people. A special thanks is also due to Mrs. M.I. Cousins for the conscientious and able manner in which she has typed this thesis and for her constant demands for "more work" which ensured its early completion.

I also thank Associate Professor F. Jackson of the Department of Applied Mathematics at the University of Cape Town for his advice and suggestions on nonlinear optimization techniques and for providing the computer program used to solve the Flexible Tolerance Method in Chapter 4.

My wife, Rita, drew all the graphs and sketches presented in this thesis and I thank her for the competent and painstaking manner in which she accomplished this task.

Finally, all computations discussed in this thesis were performed on the Univac 1106 of the University of Cape Town and I must thank the staff of the Computer Centre for their help in solving numerous technical problems.

CHAPTER ONE

INTRODUCTION AND SUMMARY

Securities were first traded in the sixteenth century and it is reasonable to assume that since that time investors have always devoted attention to research into the relative merits and demerits of individual stock market ventures. In the nineteenth and twentieth centuries, man has placed increasing emphasis on numerical and statistical analysis, and it was only a matter of time before these two related approaches focussed their attention on the problems of stock market research.

As early as 1900, a Frenchman, Louis Bachelier, published a paper (Bachelier (1900)) which was to be the forerunner of a vigorous, and as yet unresolved, debate on what has come to be called "the random walk model of stock market prices," or "the efficient market hypothesis." Unfortunately, in the year 1900 Bachelier's paper produced little response, and interest in the model waned until the late 1950's and early 1960's when it was rediscovered and empirically examined by researchers such as Osborne, Mandelbrot, Cootner and Fama. Their work, together with that of other researchers such as Markowitz and Sharpe, sowed the seed of interest and the last fifteen to twenty years have seen an ever increasing interest in what can possibly be called Mathematical and Statistical Stock Market research. In fact, as Latané, Tuttle and Jones (1975) conclude:

> "the modern computer, magnetic data banks, and other new quantitative tools are producing an information revolution in security analysis. New computer-based information systems provide a clear opportunity to make security analysis and portfolio management more of a science than an art."

In this thesis, some of the many aspects of stock market behaviour which can be investigated using statistical analysis, are examined. At the outset it must be stated that a comprehensive examination of all possible areas in which statistical analysis can be used is not attempted. Instead, the thesis comprises of a few different, independent studies, which arose from the consideration of some basic questions concerning the behaviour of stock market prices. For this reason, the reader should not place much emphasis on the order in which the chapters appear as each chapter can be considered a separate entity. However, for convenience the thesis may be divided into two halves. The first half (Chapters 2, 3 and 4) deals with the construction and analysis of stock market indices, while the second half (Chapters 5, 6, 7 and 8) deals with the analysis of stock market securities and their prices.

In Chapter 2, a problem which has received little attention in the literature is discussed, namely the appropriate selection of the constituents to be used in the construction of a stock market index. Probably the main reason for this lack of attention is that there is no universal agreement on exactly what constitutes a good or suitable index. Consequently it is difficult to evaluate alternative choices of constituents, even though they do have quite a marked effect on the index. This problem is discussed and two rules which, it is argued, will help to enable a fairly logical and scientific selection to be made, are presented. These two rules are empirically examined for different types of indices so as to determine their possible performance in practice. Finally, the relative pros and cons of these two rules are discussed.

In Chapter 3, five different methods of constructing stock market indices are examined to ascertain the effect that the method of construction has on the subsequent performance of the index. As is true of almost all stock market analysis, the results obtained do not indicate any single most suitable method. Nevertheless, they do confirm certain intuitive beliefs about the behaviour of certain types of indices, and do enable some general conclusions to be drawn. Moreover, even where statistical conclusions cannot be made with suitable confidence, the results still provide some interesting possibilities which are discussed in the text.

The use of principal component analysis in the construction of stock market indices is not new, but has floundered on two important disadvantages. The first is that the weight assigned to a security can be negative which is unacceptable to most investors, while the second is that no limit is imposed on the weight which can be assigned to any individual security and this can result in a single security dominating the index. These two problems are examined in Chapter 4 where it is shown that by a simple restatement of the problem of principal components, both of the above disadvantages can be overcome. This restatement results in the formulation of a quadratic programming problem subject to a number of constraints, at least one of which is nonlinear, and it is shown how this problem can be solved using a nonlinear optimization technique. Some examples are presented to indicate how this procedure could be used to construct stock market indices, and their performance is contrasted with that of the original principal component indices. It is indicated that this method is not restricted to stock market problems, but can be used to solve any principal component analysis problem where it is desirable to impose some form of constraint on the weights of the components.

In Chapter 5, the well known random walk model is discussed. A previous study has shown that securities quoted on the Johannesburg Stock Exchange appear to obey this model to the same extent as similar securities quoted on the New York Stock Exchange and some of the major European exchanges.

The major reason for the contentiousness of this model, is that it has been claimed that acceptance of the random walk model proves that the charting of stock market prices for prediction purposes is worthless. In this chapter it is argued that, while the random walk model does almost certainly hold, it does not preclude success for the chartist. In practice, it is argued, the chartist looks at more information than merely the past price, and it is shown that by considering only a few additional variables, the ability to predict future prices is considerably increased. Although the results obtained do not indicate that charting definitely can be successful, they do indicate that there is far more hope for the chartist than has been suggested previously by proponents of the random walk model.

The fairly well established theory of beta coefficients is discussed in Chapter 6, with particular reference to the Johannesburg Stock Exchange. Various tests which have been performed on the New York Stock Exchange and other exchanges are repeated for securities quoted on the J.S.E., and the behaviour of these beta coefficients are compared. In addition, an important aspect of these coefficients, namely the suitability of the regression line, is examined and some interesting results obtained which indicate that perhaps the degree of fit is more important in determining the beta coefficient, than is the relative risk of the security. This leads to the proposal of an alternative measurement of risk which, it is argued, provides a better estimate of the relative riskiness of the security. Estimates of the proposed measure are found for a number of securities quoted on the Johannesburg Stock Exchange, and these are compared to the estimates of the corresponding beta coefficients.

Two popular portfolio selection models, the original Markowitz Model and Sharpe's Index Model, are examined in Chapter 7, where an attempt is made to determine the differences in the portfolios selected by these two models. Both models were used to select portfolios from five different random samples of fifty securities each, for differing values of the upper bound of the funds which can be invested in any one security. From the results presented, an indication of the extent to which the portfolios selected by the two models are likely to differ in practice, is given for various situations. In addition, the possibility of using principal component indices for Sharpe's Model is also investigated.

Finally, Chapter 8 presents a discussion of how trapezoidal data can be handled when performing a regression analysis. A number of alternative methods are proposed, and the effect of each of these is evaluated using a simulation model and a "real world" example. Although the results obtained are not theoretically derived, they do indicate that certain methods are clearly superior to others as well as giving some indication as to which is the most appropriate method to use in particular situations.

CHAPTER TWO

THE SELECTION OF CONSTITUENTS FOR A STOCK MARKET INDEX

2.1 INTRODUCTION

Over the past twenty years there has been a considerable amount of academic research into various aspects of Stock Markets and their behaviour. One of the topics which has received surprisingly little attention has been that of Stock Market Indices. The relatively few studies in the literature (compared with those on the random walk model for instance) have discussed the different methods of constructing stock market indices; that is, arithmetic averages, weighted averages, geometric averages etc., and have contrasted the difference in the performance of an index based on these alternative methods. However, a factor which has considerable bearing on the performance of an index has been almost completely ignored, namely which securities should be used in the construction of the index.

Obviously, the appropriate selection criteria will be dependent, at least to some extent, on the type of index to be constructed. In this chapter an attempt is made to give some scientific guidance to a researcher or stock market analyst faced with the following problem : Given that a particular type of stock market index is to be constructed, which securities should be chosen as constituents.

Section 2 contains a brief description of some of the main methods for constructing stock market indices while Section 3 examines a fairly elementary selection method and develops certain rules and proposals for use in practice. Section 4 introduces a proposal for the selection of securities based on the technique of Cluster Analysis, and discusses how this method could be implemented. Finally, Section 5 presents some overall conclusions which can be drawn from the results presented in this section.

2.2 METHODS OF CONSTRUCTING STOCK MARKET INDICES

The method used in the construction of a stock market index will clearly have an impact on the performance of that index (cf for example, Latané, Tuttle and Young (1971)). Moreover, it could affect the choice of constituents of the index. Therefore, in this section, some of the more important types of stock market indices will be discussed.

For the purposes of this chapter, four main types of stock market indices will be defined.

A. ARITHMETIC AVERAGE OF PRICE: This type of index is constructed as follows:

$$I_t = 1/N \sum_{j=1}^{N} P_{j;t}$$

where I_t is the level of the index at period t;

N is the number of securities included in the index;

$P_{j;t}$ is the price of the j^{th} security in period t.

This is perhaps the most intuitive type of Stock Market index and is the methodology upon which the Dow Jones Averages are based. As far as investment is concerned, this type of index corresponds to an investor who buys one share in each of the N constituent companies.

B. WEIGHTED AVERAGE OF PRICE: This type of index is computed as follows:

$$I_t = \sum_{j=1}^{N} w_{j;t} P_{j;t}$$

where $w_{j;t}$ is the weight assigned to the j^{th} security in period t; and I_t, N, and $P_{j;t}$ are as defined previously. By far the most popular method of weighting is by Market Capitalization, which is defined as the price of the security times the number of shares issued in that security. The weights for a market capitalization type index are usually computed relative to some base period; that is, the weight assigned to the j^{th} security in period t is the market capitalization of the j^{th} security in period t divided by the total market capitalization of all securities included in the index, at some base period t_0. Thus, the formula becomes

$$I_t = \left(\sum_{j=1}^{N} Q_{j;t}*P_{j;t}/\sum_{j=1}^{N} Q_{j;t_0}*P_{j;t_0}\right) * L.F.$$

where $Q_{j;t}$ is the number of issued shares in the j^{th} security at period t; and

L.F. is a linking factor to initially scale the index to some desired value.

This method of construction, which gives greater weight to the larger companies (those with bigger market capitalizations), has become extremely popular, especially with large investors such as mutual funds, since it takes the "availability" of the security into account. As far as investment is concerned, it is equivalent to an investor who spreads his money among the N constituents in proportion to the market capitalization of each security relative to the total market capitalization of all N companies. Some of the better known indices which use this approach are the Standard and Poor's Indices, the New York Stock Exchange Indices, the Financial Times Actuaries Index (London Stock Exchange) and the Rand Daily Mail Indices (Johannesburg Stock Exchange).

C. ARITHMETIC AVERAGE OF RETURN:

It has been argued by Cohen and Fitch (1966), that since investors are generally interested in return and not usually in price per se, stock market indices should be based on return and not price. Most of the empirical work pertaining to stock market indices based on return have not used return in the traditional sense of the word (difference in price over some period divided by price at the beginning of the period) but have used a related measure, the price relative $(P_{i;t}/P_{i;t-1})$. Thus, this type of index is usually constructed as follows:

$$I_t = 1/N \sum_{j=1}^{N} (P_{j;t}/P_{j;t-1}) * I_{t-1}$$

where the symbols are as previously defined. This index is equivalent to the performance of an investor who invests equal monetary amounts in each security and reallocates back to equal amounts at the start of each new period (whether a day, a week, a month or a year). Examples of stock market indices based on this methodology are the United Press International Market Indicator (New York Stock Exchange) and the ESE indices+ (Johannesburg Stock Exchange).

D. GEOMETRIC AVERAGE OF RETURN:

This method has received a fair amount of attention in the more recent literature and is also based on price relatives (as for C above). A Geometric Average index is constructed as follows:

$$I_t = \left(\Pi_{j=1}^{N} P_{j;t}/P_{j;t-1}\right)^{1/N} \times I_{t-1} .$$

This type of index has become known as the continuous reallocation type index since it reflects the behaviour (theoretical) of an investor who continuously reallocates his resources so as to maintain a portfolio with equal monetary

---

+The ESE indices are not computed strictly as defined above but are based on a very similar principal.

amounts in each security. The most famous index using this type of methodology is the Value Line Index.

Since the Geometric Average (of return) method is in fact a limiting case of the Arithmetic Average of Return method, it will not be discussed separately. For the remainder of this chapter, three types of indices will be examined, and they will be called:

(i) Price Indices (corresponding to A above);
(ii) Return Indices (corresponding to C above); and
(iii) Market Capitalization Indices (corresponding to B above).

It will be assumed that any results pertinent to Return indices will be equally justifiable for Geometric Averages (D above).

2.3 ELEMENTARY SELECTION METHODS

In this section some rather simple selection rules will be examined and certain proposals for selection methods to be used in practice will be presented. For simplicity of exposition, only market capitalization indices will be discussed initially, but results will be presented later in this section relevant to the other main types of indices It should be noted that the discussion will deal mainly with Sector Indices, but all remarks will be equally relevant to market indices. It should also be mentioned at this stage that market capitalization indicies were chosen for the initial investigation for two main reasons. Firstly, they are the best behaved and easiest to analyze as far as the selection procedures proposed are concerned, and secondly, they appear to be, perhaps, the most accepted type of index (Standard and Poor's for the New York Stock Exchange, Financial Times Actuaries Index for the London Stock Exchange, and the Rand Daily Mail Indices for the Johannesburg Stock Exchange).

As mentioned previously, very little attention has been paid to the problem of which securities to include when constructing a stock market index. There are, it is felt, two important reasons for this. Firstly, the number of different combinations of securities which can be used is so vast that a detailed numerical evaluation is virtually impossible. Secondly, no universal yardstick exists by which one can evaluate exactly what is meant by a "good index", as this will depend on the purpose for which the index has been constructed.

Now, there have been many attempts to define exactly what the purpose of a stock market index is. Some of these are:

(i) to reflect the performance of industry in a particular country; (ii) to enable a comparison of portfolio performance; and (iii) to indicate the movement of the "market".

However, while all of these have merit, it is felt that the most important function of a stock market index, as far as the average investor is concerned, is that it reflects the investment opportunities which presently exist, and which did exist in a particular sector or market.

Thus, in this chapter, it will be assumed that an index which is representative of the entire sector or market, and which summarizes most of the movement which takes place in that sector or market,is desired. It would seem reasonable then that the index should consist of all the shares in that sector. The general argument against this has been that it results in a somewhat sluggish index. However, it is felt that this is not a justifiable argument, and that in fact this is the ideal market capitalization index - the securities with very small market capitalizations have a very small effect (but nevertheless do contribute to the index) while those with large market capitalizations dominate the index (as they do the market or sector).

Unfortunately, there is considerable cost involved in the collection of data and the updating of such an index. Not only does information on all the securities quoted on the exchange have to be collected but, in addition, the more securities included in the index, the more frequently adjustments will have to be made to the base due to new listings, delistings, capitalization issues etc..

So, the situation arises where it might be advisable for some of the securities to be excluded from the index in order to cut down on data collection and computation costs. Yet, at the same time, the index should remain as representative as possible. But, if the above concept of the ideal market capitalization index is accepted, then a yardstick now exists by which to evaluate the alternative indices, namely, how closely does their movement resemble that of the index in which all securities in that sector have been included.

Now that a method of comparison exists the only question still to be answered is: How should the securities be chosen? If one merely chose at random, the alternatives available would be vast and comparisons virtually impossible. However, a rule of thumb for market capitalization indices has evolved in the literature. It is as follows:

> First select the security with the largest market capitalization. Then, select the security with the next largest market capitalization and continue in this manner until at least a certain percentage, $\underline{\alpha}$ (usually 75 or 80), of the market capitalization of the entire sector has been selected (regardless of whether this requires 2 securities or 50).

In this section, this rule will be empirically examined to investigate its suitability, and an attempt made to determine that percentage which appears to be most satisfactory for use in practice.

In order to do this, ten sectors of the Johannesburg Stock Exchange were chosen. They were the following: Coal; Gold - Witwatersrand; Financial Industrial; Insurance; Building; Food; Motors; Paper and Pulp; Stores; and Sugar. The securities for which data was available (weekly closing price for the period 5/1/73 to 20/2/76) in each of these sectors are listed in Tables 1 to 10 of Appendix A.

For each sector, the following computations were performed:

1) A market capitalization type index was constructed using all the securities available in that sector (henceforward referred to as the ALLINDEX).

2) Market capitalization type indices based on the selection rule above and using the following percentages (values of α) were constructed: 90%; 85%; 80%; 75%; 70%; 65%; 60%; 55%; 50%; and 45%. In addition it was required that each index comprise of at least two securities.

3) The correlation between the weekly return of each index thus constructed (step 2) and the weekly return on the ALLINDEX (step 1) was computed. The weekly return was used rather than the level of the index since it represents the amount of change in the market in the week and is independent of the level of the two indices.

These results are presented in Table 2.1 below.

In order to interpret the results presented in Table 2.1, a few basic decisions discussed above will be summarized. Firstly, it is assumed that the purpose of the index is to explain as much of the sector movement as possible. Secondly, the movement of the sector is assumed to be ideally described by an index comprising of all the securities quoted in that sector. Thirdly, it is assumed that data collection and administrative problems prevent the regular computation of an index involving all the securities in the sector. Finally, it is assumed that the selection process should be such that the resultant index is "similar" in movement to the ALLINDEX.

The problem which now arises is how to define "similar." For our purposes, it is argued that the square of the correlation coefficient between the weekly return of the selected index and the weekly return of the ALLINDEX (i.e. the square of the correlation presented in Table 2.1) is a suitable measure. In statistical terms, this is known as the coefficient of determination, and can be used as a measure of the proportion of the variation in the ALLINDEX explained by the selected index. Since a fairly popular rule in practice has been to choose an α of 75 to 80 per cent[+], it is argued that this is really the percentage of the variation in the entire sector that it was desired to explain. Now, a coefficient of determination of 0,8 (i.e. 80%) is equivalent to a correlation in Table 2.1 of 0,894. On examining Table 2.1, it can be seen that this is achieved when accounting for as little as 55% of the market capitalization since in nine of

---

[+] For example, Standard & Poors Composite Index comprises approximately 90% of the value of the NYSE common stock issues (Sprecher (1975)).

the ten sectors examined (the exception is the Sugar Sector+) application of the rule using an α of 55% resulted in an r greater than 0,894.

TABLE 2.1

SECTOR / α	COAL r	COAL Actual % M.Cap. Accted.	COAL No. of Shares	GOLD-WITWATERSRAND r	GOLD-WITWATERSRAND Actual % M.Cap. Accted.	GOLD-WITWATERSRAND No. of Shares
90%	0,9933	91	10	0,9983	92	11
85%	0,9895	86	8	0,9974	85	9
80%	0,9894	81	7	0,9944	82	8
75%	0,9803	75	6	0,9941	78	7
70%	0,9803	75	6	0,9927	73	6
65%	0,9647	69	5	0,9905	67	5
60%	0,9531	62	4	0,9905	67	5
55%	0,9123	55	3	0,9862	58	4
50%	0,9123	55	3	0,9862	58	4
45%	0,8676	48	2	0,9819	48	3
	FIN. INDUSTRIAL			INSURANCE		
90%	0,9971	90	13	0,9951	91	3
85%	0,9938	85	10	0,9951	91	3
80%	0,9875	81	8	0,9951	91	3
75%	0,9842	76	6	0,9681	78	2
70%	0,9807	72	5	0,9681	78	2
65%	0,9734	68	4	0,9681	78	2
60%	0,9620	63	3	0,9681	78	2
55%	0,9554	55	2	0,9681	78	2
50%	0,9554	55	2	0,9681	78	2
45%	0,9554	55	2	0,9681	78	2
	BUILDING			FOOD		
90%	0,9966	90	14	0,9919	91	8
85%	0,9924	85	12	0,9891	86	6
80%	0,9915	80	10	0,9862	82	5
75%	0,9851	75	8	0,9822	77	4
70%	0,9819	72	7	0,9744	72	3
65%	0,9770	66	6	0,9744	72	3
60%	0,9770	66	6	0,9744	72	3
55%	0,9421	59	5	0,9407	58	2
50%	0,9157	52	4	0,9407	58	2
45%	0,8965	44	3	0,9407	58	2

+The reason for this difference will be discussed fully in Section 4.

TABLE 2.1 (continued)

SECTOR / α	MOTORS r	MOTORS Actual % M.Cap. Accted.	MOTORS No. of Shares	PAPER r	PAPER Actual % M.Cap. Accted.	PAPER No. of Shares
90%	0,9971	91	11	0,9968	92	7
85%	0,9960	86	9	0,9951	86	5
80%	0,9909	80	7	0,9879	80	4
75%	0,9862	77	6	0,9879	80	4
70%	0,9663	71	5	0,9618	73	3
65%	0,9487	65	4	0,9618	73	3
60%	0,9487	65	4	0,9440	63	2
55%	0,9056	57	3	0,9440	63	2
50%	0,9056	57	3	0,9440	63	2
45%	0,8308	46	2	0,9440	63	2
	STORES			SUGAR		
90%	0,9972	90	7	0,9991	96	4
85%	0,9972	90	7	0,7325	88	3
80%	0,9923	84	6	0,7325	88	3
75%	0,9816	78	5	0,7236	77	2
70%	0,9714	72	4	0,7236	77	2
65%	0,9622	65	3	0,7236	77	2
60%	0,9622	65	3	0,7236	77	2
55%	0,9622	65	3	0,7236	77	2
50%	0,9622	65	3	0,7236	77	2
45%	0,9060	48	2	0,7236	77	2

Col. 1 : r ≡ Correlation between weekly return of selected index and the ALLINDEX

Col. 2 ≡ Actual % market capitalization accounted for by the selected index

Col. 3 ≡ No. of shares in the selected index

Table 2.2 below has been constructed using the above logic. Specifically, the proportion of the market capitalization (i.e. the α) needed for an index which is to explain a given proportion of the total variation in the ALLINDEX, was obtained by choosing the lowest α which satisfied that proportion of the total variation for at least nine of the ten sectors examined (Sugar being the exception in all cases).

Two additional columns are provided, the first giving the total number of shares which would be required to construct a sector index for each of the ten sectors satisfying the desired level of explanation, and the second expressing this as a percentage of the total number of securities available for the ten sectors (165 - see Appendix A).

TABLE 2.2

% Variation Explained (correlation)	Lowest α which would satisfy Column 1 for 9 of the 10 Sectors examined	No. of Shares from all 10 Sectors	% of Shares held
75 (0,8660)	55	28	17
80 (0,8944)	55	28	17
85 (0,9219)	60	34	21
90 (0,9487)	70	43	26
95 (0,9747)	75	50	30
97½ (0,9874)	85	72	44

As can be seen from Table 2.2, the choice of an α of 75% would result in an index which explained 95% of the variation in the ALLINDEX. While this is almost certainly acceptable, it is probably unnecessarily exact. An index which explained 80% of the variation in the ALLINDEX could be obtained by choosing an α of 55%.

Moreover, this would allow for a saving of approximately 44% in the amount of data required - for the ten sectors examined a total of only 28 securities would be required compared with 50 if one adhered to the original rule of an α of 75%. The considerable saving as compared with the ALLINDEX should also be noted. If the proposed rule (only account for 55% of the market capitalization) is accepted, only 17% of the data need be collected - certainly a considerable saving.

Of course, if it was necessary to be more accurate, Table 2.2 could be consulted, and an appropriate value of α chosen. In all cases, a considerable saving in data collection would result for a relatively small loss in accuracy (assuming ALLINDEX is the ideal).

The above results should prove very helpful to anyone wishing to construct a market capitalization type index. However, it must be remembered that the very nature of the rule used to select constituents (choosing those with largest market capitalization) should result in good performance for market capitalization weighted indices. But, how would this rule work if used on an unweighted method, namely on Arithmetic Price Averages (Price Indices in our notation of the previous section) and Arithmetic Return Averages (Return Indices)? In order to examine this aspect of the problem the above experiments were repeated for these two types of indices.

Firstly, the results for the Price Indices are presented in Tables 2.3 and 2.4 below which are analagous to Tables 2.1 and 2.2.

TABLE 2.3

SECTOR / α	COAL			GOLD : WITS & OTHERS			FIN INDUSTRIAL			INSURANCE		
	r	m	N	r	m	N	r	m	N	r	m	N
90%	0,9648	91	10	0,9960	92	11	0,9223	90	13	0,9303	91	3
85%	0,9490	86	8	0,9945	85	9	0,9089	85	10	0,9303	91	3
80%	0,9444	81	7	0,9898	82	8	0,8942	81	8	0,9303	91	3
75%	0,9185	75	6	0,9894	78	7	0,8574	76	6	0,8929	78	2
70%	0,9185	75	6	0,9865	73	6	0,8484	72	5	0,8929	78	2
65%	0,9046	69	5	0,9844	67	5	0,8335	68	4	0,8929	78	2
60%	0,8400	62	4	0,9844	67	5	0,7762	63	3	0,8929	78	2
55%	0,6943	55	3	0,9782	58	4	0,7630	55	2	0,8929	78	2
50%	0,6943	55	3	0,9782	58	4	0,7630	55	2	0,8929	78	2
45%	0,5408	48	2	0,9657	48	3	0,7630	55	2	0,8929	78	2
	BUILDING			FOOD			MOTORS			PAPER		
90%	0,9230	90	14	0,8519	91	8	0,9695	91	11	0,9314	92	7
85%	0,9206	85	12	0,8122	86	6	0,9675	86	9	0,8962	86	5
80%	0,9082	80	10	0,8112	82	5	0,9118	80	7	0,8383	80	4
75%	0,8744	75	8	0,7418	77	4	0,8482	77	6	0,8383	80	4
70%	0,8657	72	7	0,7392	72	3	0,8297	71	5	0,7366	73	3
65%	0,8547	66	6	0,7392	72	3	0,8080	65	4	0,7366	73	3
60%	0,8547	66	6	0,7392	72	3	0,8080	65	4	0,6531	63	2
55%	0,8563	59	5	0,7127	58	2	0,7811	57	3	0,6531	63	2
50%	0,8363	52	4	0,7127	58	2	0,7811	57	3	0,6531	63	2
45%	0,8341	44	3	0,7127	58	2	0,5604	46	2	0,6531	63	2
	STORES			SUGAR								
90%	0,9960	90	7	0,9896	96	4						
85%	0,9960	90	7	0,4576	88	3						
80%	0,9947	84	6	0,4576	88	3						
75%	0,9137	78	5	0,4348	77	2						
70%	0,8963	72	4	0,4348	77	2						
65%	0,8244	65	3	0,4348	77	2						
60%	0,8244	65	3	0,4348	77	2						
55%	0,8244	65	3	0,4348	77	2						
50%	0,8244	65	3	0,4348	77	2						
45%	0,6314	48	2	0,4348	77	2						

r ≡ Correlation between the weekly return of the selected index and the ALLINDEX

m ≡ Actual % market capitalization accounted for by the selected index

N ≡ Number of shares in the selected index

From Table 2.3 it is apparent that the correlations between the selected indices and the ALLINDEX (a new ALLINDEX, now based on the same methodology as the Price Indices - that is an arithmetic average of all the securities in the sector) are lower than those corresponding to the market capitalization indices in Table 2.1. This is to be expected since the Price indices (which are equivalent to the purchasing of one share in each of the constituent securities) are in effect price weighted - that is, the higher priced securities have more weight than the lower priced securities. Thus, choosing those securities with highest market capitalizations should not result in an index markedly closer to the ALLINDEX than an index chosen by random selection.

It can also be argued that the ALLINDEX does not really represent the "ideal" price average index since it does not favour the important securities, and their effect is usually swamped by the remaining securities. Unfortunately, no alternative "ideal" index exists, and hence the ALLINDEX has to be used in the following discussion. For these reasons, it was decided that in the case of Price indices a lower "percentage of the variation explained" than that used for the market capitalization indices would be acceptable, and a value of 70% was selected (compared with 80% for the market capitalization case).

It was with the above thoughts in mind that Table 2.4 below was constructed. In addition, whereas previously

(Table 2.2) the percentage of the market capitalization to be explained by the rule (i.e. the α) was given by that value which satisfied the desired correlation for at least nine of the ten sectors, in this instance the criterion was lowered to eight of the ten sectors.

Table 2.4 may be analysed in exactly the same manner as Table 2.2.

TABLE 2.4

% Variation Explained (correlation)	Lowest Possible α	No of Shares from all 10 Sectors	% of Shares Held
70 (0,8367)	75	50	30
75 (0,8660)	85	72	44
80 (0,8944)	85	72	44
85 (0,9219)	90	80	53
90 (0,9487)	Cannot be achieved unless more than 90% of m.cap. is accounted for		

Thus, for this type of index one could use an α of 75% for most cases, which would result in approximately thirty percent of the information being required as compared with the ALLINDEX. Once again, it must be stressed that if more accuracy is required one can consult Table 2.4 to determine a more suitable value of α.

Most of the remarks made above concerning the desired accuracy of Price Indices are equally relevant to Return Indices with the main difference being that Return Indices give exactly equal weight to all securities in the index (that is equal monetary amounts in each security so that those with higher market capitalizations or higher prices are not favoured). Thus, the conclusions arrived at for Price Indices as regards the percentage variation which it is desired to explain, are equally pertinent for return indices. Tables 2.5 and 2.6 below present the empirical results obtained from the ten sectors previously analysed, for the case of Return Indices.

Examination of Table 2.5 reveals that the correlations obtained between the selected indices and the ALLINDEX (now constructed as the arithmetic average of the weekly price relatives of all securities included in the sector) are lower than in the previous two cases. This is perhaps due to the fact that this method of constructing stock market indices gives each security exactly equal weight (i.e. equal monetary amounts are invested in each security) and hence one should expect to lose considerable accuracy by taking a sample of less than half the actual number of securities available in the sector. From Table 2.6 it is apparent that in order to explain 70% of the variation in the ALLINDEX, 90% of the market capitalization would have to be accounted for and data on 53% of all the securities in the sector or market would have to be obtained.

TABLE 2.5

SECTOR / α	COAL			GOLD : WITS & OTHERS			FIN INDUSTRIAL			INSURANCE		
	r	m	N	r	m	N	r	m	N	r	m	N
90%	0,9125	91	10	0,9793	92	11	0,8755	90	13	0,7534	91	3
85%	0,8997	86	8	0,9717	85	9	0,8385	85	10	0,7534	91	3
80%	0,8933	81	7	0,9613	82	8	0,7938	81	8	0,7534	91	3
75%	0,8679	75	6	0,9550	78	7	0,7768	76	6	0,6389	78	2
70%	0,8679	75	6	0,9554	73	6	0,7697	72	5	0,6389	78	2
65%	0,8432	69	5	0,9428	67	5	0,7430	68	4	0,6389	78	2
60%	0,8150	62	4	0,9428	67	5	0,6883	63	3	0,6389	78	2
55%	0,7662	55	3	0,9262	58	4	0,6964	55	2	0,6389	78	2
50%	0,7662	55	3	0,9262	58	4	0,6964	55	2	0,6389	78	2
45%	0,5875	48	2	0,9260	48	3	0,6964	55	2	0,6389	78	2

	BUILDING			FOOD			MOTORS			PAPER		
90%	0,8967	90	14	0,8098	91	8	0,9220	91	11	0,9539	92	7
85%	0,8858	85	12	0,7455	86	6	0,9134	86	9	0,8743	86	5
80%	0,8677	80	10	0,7256	82	5	0,8961	80	7	0,8710	80	4
75%	0,8450	75	8	0,6927	77	4	0,8514	77	6	0,8710	80	4
70%	0,8130	72	7	0,6596	72	3	0,8257	71	5	0,8135	73	3
65%	0,7886	66	6	0,6596	72	3	0,7962	65	4	0,8135	73	3
60%	0,7886	66	6	0,6596	72	3	0,7962	65	4	0,8360	63	2
55%	0,7754	59	5	0,5809	58	2	0,7166	57	3	0,8360	63	2
50%	0,7518	52	4	0,5809	58	2	0,7166	57	3	0,8360	63	2
45%	0,7493	44	3	0,5809	58	2	0,5832	46	2	0,8360	63	2

	STORES			SUGAR		
90%	0,9283	90	7	0,9354	96	4
85%	0,9283	90	7	0,6123	88	3
80%	0,9139	84	6	0,6123	88	3
75%	0,8921	78	5	0,5617	77	2
70%	0,8395	72	4	0,5617	77	2
65%	0,8159	65	3	0,5617	77	2
60%	0,8159	65	3	0,5617	77	2
55%	0,8159	65	3	0,5617	77	2
50%	0,8159	65	3	0,5617	77	2
45%	0,7787	48	2	0,5617	77	2

r ≡ Correlation between weekly return of the selected index and ALLINDEX

m ≡ Actual % market capitalization accounted for by the selected index

N ≡ No. of shares in the selected index

TABLE 2.6

% Variation Explained (correlation)	Lowest Possible α	No. of Shares from all 10 Sectors	% of Shares Held
40 (0,6325)	50	27	16
50 (0,7071)	80	61	37
60 (0,7746)	90	88	53
70 (0,8367)	90	88	53
80 (0,8944)	Cannot be obtained unless more than 90% of m.cap. is accounted for		

Before discussing an alternative (but similar) method which can be used to select the constituents of a stock market index in the following section, some conclusions on the empirical results presented in this section will be given.

The rule defined on page 2.9 for selecting constituents of a stock market index appears to be very satisfactory if a market capitalization type index is to be used. The researcher constructing the index can obtain an index very close to the ideal index in most cases (9 out of 10 Sectors examined), while saving considerable costs as far as data collection and adjustments are concerned (approximately as much as a 70 to 80% reduction in the amount of data required).

As far as the other two methods of constructing stock market indices are concerned, the results are not as satisfactory. However, it is felt that this is due to the fact that a suitable yardstick to evaluate the selected indices does not exist for these methods. While the ALLINDEX can justifiably (it is believed) be used as an ideal for a market capitalization index, the same is not true for a Price or a Return index. Thus, if an ideal Price index and an ideal Return index could be found then it is felt that the results presented in Tables 2.3 to 2.6 would be far more satisfactory (assuming that the securities with large market capitalizations are in general the more important securities). At present, all that the researcher can do is to set a lower criterion for his desired "closeness" to the ideal in these two cases. Even if the ALLINDEX remained the ideal, the above rule (page 2.9) does provide fairly satisfactory results (70% of the variation is explained when only approximately 50% of the available securities are used). In addition, it should be noted that the results obtained for both Price and Return indices were better than those indicated in Tables 2.4 and 2.6 for seven of the ten sectors examined (Insurance, Food and Sugar being the exceptions).

2.4 A CLUSTER ANALYSIS METHOD FOR SELECTING STOCK MARKET INDEX CONSTITUENTS

In the previous section an intuitive rule for the selection of securities to be included in a stock market index was empirically examined. The results obtained were fairly satisfactory for the ten sectors examined, with the exception of one sector, namely the Sugar Sector, which gave rather poor results. In this section the Sugar Sector is examined in detail so as to ascertain why the results were poorer than those for the other sectors.

In an attempt to overcome this difference, a modification of the rule defined in Section 3 is proposed. The results of some empirical tests on the suitability of this new proposal are also presented.

The Sugar Sector consists, for our purposes, of the six securities listed in Table A.10 of Appendix A (that is, all securities for which data was available over the entire period). Of these securities, C.G. Smith (155 526 000); Huletts (120 870 000); and Tongaat (40 694 200) had the largest market capitalizations on 5/1/73, and together accounted for approximately 88% of the total market capitalization of the sector (all 6 securities). The security with the next largest market capitalization was Swaziland Sugar (28 120 000), and this together with the previous three securities accounted for 96% of the total market capitalization of the sector. Now, it happens that over the period 1973 to 1976 for which data was available, while the general trend in the first mentioned three securities was up (an average increase over the 3 years of 168%), it was nothing as dramatic as the increase in Swaziland Sugar (1329%). Thus, once Swaziland Sugar was excluded from the securities comprising the selected index, then this selected index no longer had the considerable upward pull resulting from Swaziland Sugar (which the ALLINDEX had) and hence results in a fairly wide discrepancy between the selected index and the ALLINDEX as shown in Tables 2.1; 2.3; and 2.5 .

Clearly, this could occur in any sector analyzed, but the effect is likely to be more dominant in a small sector such as Sugar, and it is obviously not a desirable feature of the rule proposed in Section 3 above. Therefore, in an attempt to overcome this problem, the following rule is proposed.

> Firstly, divide the securities in the sector to be analyzed into a number of homogeneous groups (called clusters) which display similar price movements. It is suggested that this be done using the technique of cluster analysis.
>
> Secondly, select from each group or cluster that security with the largest market capitalization. Then, select from each cluster the security with the next largest market capitalization, and repeat this until at least α% of the market capitalization of EACH cluster has been accounted for.

The same tests which were performed in the previous section were performed for this rule in order to ascertain how successful such a rule might be in practice. However, since this rule results in far more securities being included in the index, the 90% and 85% α's were not computed. Instead an index called WST was computed, which consisted of choosing only the single security with largest market capitalization in each cluster and using these as constituents of the index.

The cluster analysis was performed on each sector using the weekly closing price of all securities in that sector for the period 22/3/68 to 29/12/72^{+}. The relevant dendrograms and subsequent clusterings for each sector are given in Figures B.1 to B.10 of Appendix B, which includes some relevant comments on the clustering algorithm used. It should be noted that the clustering procedure is fairly arbitrary as no firm basis exists for deciding exactly what constitutes a cluster and what does not. Thus, the clustering process is fairly subjective, but an attempt was made to be consistent by defining any combination with an average correlation of more than 0,8 as definitely constituting a cluster, and then attempting "to fit in" the remaining securities.

As in the previous section, the initial discussion will be limited to market capitalization indices, and then later extended to include both the Price indices and the Return indices. Once again, the ALLINDEX was assumed to be the ideal.

$^{+}$ The exception was the Insurance Sector for which the period 13/12/68 to 29/12/72 was used.

The results for the market capitalization indices are presented in Tables 2.7 and 2.8 below. As can be seen from Table 2.7, the Sugar Sector now performs as the others, which is reassuring. It should be noted that the percentage of the market capitalization accounted for (the columns under m) are not as meaningful in this instance as in the previous section. This follows since in this case the selection from each of the clusters must satisfy the desired percentage. Thus, if the α chosen is, say, 60% then it can occur, if for example there are three clusters, that two of them are satisfied at the 80% level and the third at the 60% level resulting in an overall percentage (m) of say 70. This situation would occur if the removal of one security from those selected from either of the clusters satisfied at the 80% level, results in that cluster having shares chosen from it that account for less than 60% of the market capitalization of the cluster.

From Table 2.8 it can be seen that once again, the results are most satisfactory for the market capitalization type indices. It should be noted that Table 2.8 is constructed by determining from Table 2.7, that α which must be accounted for per cluster to result in a correlation greater than that required (see parentheses in Table 2.8) in nine of the ten sectors examined.

In comparing these results with those from the previous section it is apparent that approximately 50% more securities

TABLE 2.7

SECTOR / α	COAL			GOLD : WITS & OTHERS			FIN INDUSTRIAL			INSURANCE		
α	r	m	N	r	m	N	r	m	N	r	m	N
80%	0,9907	87	10	0,9983	90	11	0,9937	85	15	0,9951	91	3
75%	0,9892	86	9	0,9983	90	11	0,9874	79	12	0,9951	91	3
70%	0,9726	79	8	0,9937	78	9	0,9874	79	12	0,9681	78	2
65%	0,9720	77	7	0,9937	78	9	0,9866	78	11	0,9681	78	2
60%	0,9545	69	6	0,9934	77	8	0,9866	78	11	0,9681	78	2
55%	0,9545	69	6	0,9934	77	8	0,9859	77	10	0,9681	78	2
50%	0,9545	69	6	0,9896	64	6	0,9416	57	7	0,9681	78	2
45%	0,9447	64	5	0,9867	58	5	0,9416	57	7	0,9681	78	2
WST	0,9295	60	4	0,9755	48	4	0,9416	57	7	0,9681	78	2
	BUILDING			FOOD			MOTORS			PAPER		
80%	0,9974	92	17	0,9991	97	12	0,9915	85	11	0,9973	94	8
75%	0,9959	90	16	0,9756	79	9	0,9757	79	9	0,9959	91	7
70%	0,9955	87	15	0,9756	79	9	0,9757	79	9	0,9714	84	6
65%	0,9852	76	14	0,9756	79	9	0,9757	79	9	0,9714	84	6
60%	0,9835	74	13	0,9735	77	8	0,9426	67	7	0,9701	82	5
55%	0,9835	74	13	0,9735	77	8	0,9426	67	7	0,9701	82	5
50%	0,9554	64	10	0,9735	77	8	0,9426	67	7	0,9619	76	4
45%	0,9121	55	8	0,9124	51	6	0,8813	48	5	0,9619	76	4
WST	0,9084	52	7	0,9091	49	5	0,8803	47	4	0,9619	76	4
	STORES			SUGAR								
80%	0,9927	85	8	0,9924	87	4						
75%	0,9927	85	8	0,9903	84	3						
70%	0,9820	79	6	0,9903	84	3						
65%	0,9820	79	6	0,9903	84	3						
60%	0,9717	72	5	0,9903	84	3						
55%	0,9717	72	5	0,9903	84	3						
50%	0,9625	66	4	0,9903	84	3						
45%	0,9625	66	4	0,9001	51	2						
WST	0,9078	49	3	0,9001	51	2						

r ≡ Correlation between the weekly return of the selected index and the ALLINDEX

m ≡ % Market capitalization actually accounted for by the selected index

N ≡ The number of securities used in the selected index

TABLE 2.8

% Variation Explained (Correlation)	Lowest Possible α	No. of shares Held	% of Shares Held
75 (0,8660)	WST*	42	25
80 (0,8944)	WST*	42	25
85 (0,9219)	50	57	34
90 (0,9487)	55	67	40
95 (0,9747)	75	87	52
97½ (0,9874)	80	99	60

* The % Market capitalization per cluster accounted for by the rule using WST varied for each index but averaged out at approximately 35%.

have to be held when using the new rule compared to the old rule, for the same level of explanation. However, this must be considered together with the fact, that for the second rule, there is only a slight decrease in the percentage explanation for the one sector which does not satisfy the requirement. However, there is quite a considerable difference in the case of the first rule. For example, to account for 90% of the variation, using the rule proposed in the previous section, it can be seen (Tables 2.1 and 2.2) that the α chosen should be 70, and that the one sector which does not satisfy the requirement (Sugar) has a correlation of 0,7236 compared with the desired 0,9487 - quite a considerable difference. However, using the rule proposed in this section, it can be seen (from Tables 2.7 and 2.8) that the α chosen should be 55 and the one sector which does not satisfy the requirement (Motors) has a correlation of 0,9426 instead of the required 0,9487 - almost no difference at all.

Thus, it may be concluded that the rule proposed in this section is more reliable than that of the previous section but requires the collection of more data. The choice of method to be used in practice when constructing a market capitalization index would depend on the accuracy required. However, if it is sufficient to explain approximately 80% of the variation in the ALLINDEX then it is felt that, as a general guide, the second method is more appropriate even though it requires 25% of the total data to be collected as compared with 17% if the first method were used. The reason for this statement is, it must be mentioned, completely subjective, but it is felt that the extra data required in this case would not be excessive (in a market of 1000 securities it would require data to be collected on 80 extra securities - 250 instead of 170) when contrasted with the more certain knowledge of the relation of the index to the ALLINDEX.

In order to investigate the performance of the two other kinds of indices (Price and Return) when subjected to this rule, the empirical tests were repeated for each case. Firstly, the results for the Price Indices are presented in Tables 2.9 and 2.10.

All the remarks made in the previous section concerning Tables 2.3 and 2.4 are equally relevant to Tables 2.9 and 2.10 - that is, the correlations appear smaller than those of Table 2.6, the discussion on weights, the method of constructing Table 2.10 (ignoring the worst 2 sectors of the 10 considered instead of the worst 1 when constructing Table 2.8), etcetera.

T A B L E   2.9

SECTOR / α	COAL			GOLD : WITS & OTHERS			FIN INDUSTRIAL			INSURANCE		
	r	m	N	r	m	N	r	m	N	r	m	N
80%	0,9780	87	10	0,9957	90	11	0,9293	85	15	0,9303	91	3
75%	0,9709	86	9	0,9957	90	11	0,9198	79	12	0,9303	91	3
70%	0,8924	79	8	0,9885	78	9	0,9198	79	12	0,8929	78	2
65%	0,8821	77	7	0,9885	78	9	0,9069	78	11	0,8929	78	2
60%	0,8664	69	6	0,9879	77	8	0,9069	78	11	0,8929	78	2
55%	0,8664	69	6	0,9879	77	8	0,9031	77	10	0,8929	78	2
50%	0,8664	69	6	0,9771	64	6	0,8500	57	7	0,8929	78	2
45%	0,8566	64	5	0,9709	58	5	0,8500	57	7	0,8929	78	2
WST	0,7541	60	4	0,9648	48	4	0,8500	57	7	0,8929	78	2

	BUILDING			FOOD			MOTORS			PAPER		
80%	0,9486	92	17	0,9410	97	12	0,8932	85	11	0,9620	94	8
75%	0,9415	90	16	0,8724	79	9	0,8762	79	9	0,9163	91	7
70%	0,9278	87	15	0,8724	79	9	0,8762	79	9	0,8343	84	6
65%	0,9271	76	14	0,8724	79	9	0,8762	79	9	0,8343	84	6
60%	0,9247	74	13	0,8686	77	8	0,7979	67	7	0,7789	82	5
55%	0,9247	74	13	0,8686	77	8	0,7979	67	7	0,7789	82	5
50%	0,9086	64	10	0,8686	77	8	0,7979	67	7	0,7361	76	4
45%	0,8945	55	8	0,7935	51	6	0,7466	48	5	0,7361	76	4
WST	0,8750	52	7	0,7594	49	5	0,7416	47	4	0,7361	76	4

	STORES			SUGAR		
80%	0,9954	85	8	0,9849	87	4
75%	0,9954	85	8	0,9725	84	3
70%	0,9137	79	6	0,9725	84	3
65%	0,9137	79	6	0,9725	84	3
60%	0,8963	72	5	0,9725	84	3
55%	0,8963	72	5	0,9725	84	3
50%	0,8243	66	4	0,9725	84	3
45%	0,8243	66	4	0,9327	51	2
WST	0,6366	49	3	0,9327	51	2

r ≡ Correlation between weekly returns of the selected index and ALLINDEX

m ≡ Actual % market capitalization accounted for by the selected index

N ≡ No. of shares included in the selected index

TABLE 2.10

% Variation Explained (Correlation)	Lowest Possible α	No. of shares held	% of Shares held
60 (0,7746)	45	48	29
70 (0,8367)	55	67	40
75 (0,8660)	55	67	40
80 (0,8944)	75	87	52
85 (0,9219)	80	99	60
90 (0,9487)	Cannot be obtained unless more than 90% of m.cap. per cluster is accounted for		

It should be mentioned that the results presented in Tables 2.4 and 2.10 indicate (as in the case of market capitalization indices) that using the rule proposed in this section would result in more securities having to be held than if the rule proposed in the previous section had been used. Once again, however, the rule proposed in this section provides more certain results. For example, consider in both cases that it is required to account for 80% of the variation in the ALLINDEX. Using the rule proposed in the previous section, the correlations between the two sectors excluded (Food and Sugar - see Table 2.3) and the ALLINDEX are 0,8122 and 0,4576 respectively, which are somewhat different from 0,8944. However, using the rule proposed in this sector, the correlations between the two sectors excluded (Food and Motors - see Table 2.9) and the ALLINDEX are 0,8724 and 0,8762 respectively, which are far closer to 0,8944.

Finally, the empirical tests were repeated for the Return indices and these results are presented in Tables 2.11 and 2.12, which are analagous to Tables 2.5 and 2.6 of the previous section. The analysis of these two tables is identical to that above, and the conclusions that can be drawn are the same, namely that: (i) the Return indices are not as well related to the ALLINDEX as is the case with market capitalization indices ; (ii) the rule proposed in this section results in more data having to be collected than if the rule proposed in the previous section had been adopted; and (iii) that the rule proposed in this section provides more certain results than that of the previous section. The last two mentioned conclusions, namely the extra data required and the greater certainty provided by the rule proposed in this section, are in fact the major conclusions which can be drawn from this section, and hold for all three types of stock market indices considered in this chapter.

T A B L E 2.11

SECTOR / α	COAL			GOLD : WITS & OTHERS			FIN INDUSTRIAL			INSURANCE		
	r	m	N	r	m	N	r	m	N	r	m	N
80%	0,9036	87	10	0,9797	90	11	0,9108	85	15	0,7534	91	3
75%	0,8937	86	9	0,9797	90	11	0,9687	79	12	0,7534	91	3
70%	0,8832	79	8	0,9634	78	9	0,9687	79	12	0,6389	78	2
65%	0,8689	77	7	0,9634	78	9	0,8432	78	11	0,6389	78	2
60%	0,8255	69	6	0,9636	77	8	0,8432	78	11	0,6389	78	2
55%	0,8255	69	6	0,9636	77	8	0,8309	77	10	0,6389	78	2
50%	0,8255	69	6	0,9500	64	6	0,7492	57	7	0,6389	78	2
45%	0,8214	64	5	0,9455	58	5	0,7492	57	7	0,6389	78	2
WST	0,7719	60	4	0,9207	48	4	0,7492	57	7	0,6389	78	2
	BUILDING			FOOD			MOTORS			PAPER		
80%	0,9094	92	17	0,9763	97	12	0,8943	85	11	0,8638	94	8
75%	0,8895	90	16	0,9143	79	9	0,8687	79	9	0,8316	91	7
70%	0,8844	87	15	0,9143	79	9	0,8687	79	9	0,7794	84	6
65%	0,8788	76	14	0,9143	79	9	0,8687	79	9	0,7794	84	6
60%	0,8737	74	13	0,8901	77	8	0,8268	67	7	0,7384	82	5
55%	0,8737	74	13	0,8901	77	8	0.8268	67	7	0,7384	82	5
50%	0,8330	64	10	0,8901	77	8	0,8268	67	7	0,6895	76	4
45%	0,8056	55	8	0,8206	51	6	0,7536	48	5	0,6895	76	4
WST	0,7927	52	7	0,7812	49	5	0,7738	47	4	0,6895	76	4
	STORES			SUGAR								
80%	0,9632	85	8	0,9618	87	4						
75%	0,9632	85	8	0,8827	84	3						
70%	0,9177	79	6	0,8827	84	3						
65%	0,9177	79	6	0,8827	84	3						
60%	0,8611	72	5	0,8827	84	3						
55%	0,8611	72	5	0,8827	84	3						
50%	0,8340	66	4	0,8827	84	3						
45%	0,8340	66	4	0,7946	51	2						
WST	0,8051	49	3	0,7946	51	2						

r ≡ Correlation between weekly return of the selected index and ALLINDEX

m ≡ Actual % market capitalization accounted for by the selected index

N ≡ No. of shares included in the selected index

TABLE 2.12

% Variation Explained (Correlation)	Lowest Possible α	No. of shares held	% of Shares held
50 (0,7071)	WST	42	25
60 (0,7746)	55	67	40
70 (0,8367)	65	76	46
75 (0,8660)	70	79	48
80 (0,8944)	80	99	60
85 (0,9219)	Cannot be obtained unless more than 90% of the Market Capitalization is accounted for		

2.5 CONCLUSIONS

In this chapter an attempt has been made to provide some guidelines for selecting the constituents of a stock market index. The rules proposed appear to provide indices fairly close to an "ideal" index, and at the same time provide considerable savings in the data requirements and adjustment necessary for the maintenance of a stock market index. In addition, if used, the rules provide an indication of the accuracy which can be expected, something which is not provided by other selection methods.

In the past, the selection of constituents has usually been by some form of random selection, or otherwise, by a subjective analysis of the securities involved. The first method, random selection, is undesirable since it can easily result in an "unfortunate" selection wherein most or all of the important securities are excluded. The effect of the second method is not easy to analyze. Certainly, in the hands of competent people, who know the exact reason why the index is to be constructed, the subjective analysis approach can be most satisfactory. But, the fact remains that even "experts" differ in their opinions, and it is for this reason, namely a universally acceptable index, that a more scientific exact rule has been proposed.

The rules proposed in this chapter involve the selection of those securities with the largest market capitalization. While this is probably justifiable for a market capitalization index, it can be argued that such a method is not relevant to either Price or Return indices. However, it is felt that those securities with larger market capitalizations are, in general, the more important securities, and hence are those that should be constituents of any index, be it a Price, Return or market capitalization index. In fact, it could be argued that for Price and Return indices the use of the rules proposed in this chapter for the selection of the constituents will probably provide, at worst, an index as good as any other that could be constructed using the same number of securities. As mentioned in the text, the apparent divergence of the Price and Return indices obtained using these rules from the ALLINDICES is probably due to the unsuitability of the ALLINDEX as an "ideal" index for these two methods of construction and not as a direct consequence of the inapplicability of the rules themselves.

Unfortunately, the results presented in this chapter do not indicate clearly which of the two rules is the better. On the one hand, the first rule (see page 2.9) requires far less (as much as 25% less) information to be collected and updated, while on the other hand, the second rule (see page 2.24) provides an index which is more certain to be of the required accuracy. The actual choice of the rule to be used in practice will depend on the degree of accuracy required, and the researcher will have to balance the above mentioned features of the two rules.

Finally, some comments on the practical implementation of these rules are in order. In practice, the market capitalization of each share is changing almost continuously since it is the product of price (which can vary rapidly) and the number of issued shares (which can also vary, but not nearly as often as the price). Thus, ideally all the securities quoted in a sector or cluster (depending on the rule used) should be continuously monitored, and as soon as the market capitalizations change enough either to cause a security presently excluded to have greater market capitalization than one included, or for the α requirement to be no longer met, then one should change the constituents of the index so as to satisfy the original requirements.

However, this ideal has two distinct drawbacks. Firstly, this would require the collection of data on all the securities quoted in a sector which is the situation which is being avoided and secondly, it would result in numerous changes which is not desirable. But, this re-analysis should still take place, but on some long term regular basis such as each year or every two years. This would not require a great deal of extra work and would provide an additional safety feature to prevent a repetition of the situation which was observed for the Sugar sector in Section 3.

Checking the clusterings for the second method provides a serious problem since it requires data on each of the securities in the sector to be collected over a fairly lengthy period of time - weekly data for a period of at least one year. Unfortunately, as yet there is no indication of exactly how stable the correlation matrix of stock prices is (and hence the cluster analysis) and this is an aspect of stock market research which requires further investigation. However, Blume (1971) has shown that the beta coefficients of individual securities do tend to be farily stable over time, and thus it is possible that the correlation matrix will also be so. Even if it is not, and the cluster analysis is never repeated, it is felt that the second rule proposed in this chapter will still give a better starting selection than that of the first rule, and the regular yearly or two-yearly check on the market capitalizations will probably be sufficient to maintain the desired performance.

In conclusion it must be said that as with the application of any so called mechanical rule on the stock market, the rules suggested in this chapter should only serve as firm guidelines for the construction of a stock market index. Clearly, any obvious anomalies or subsequent abnormalities in the performance of the index should be corrected by the subjective analysis of the researcher involved. Typical of this approach was the study of Haycocks and Plymen (1964) in constructing the Financial Times Actuaries index. Initially they chose those securities (650) which accounted for approximately 90 per cent of the total market capitalization. After this, certain securities were excluded for various reasons, leaving the resulting index accounting for 60 per cent of the total market capitalization. Obviously a similar "weeding out" would have to be performed when applying the methods proposed in this chapter.

CHAPTER THREE

AN EMPIRICAL COMPARISON OF THE PERFORMANCE OF DIFFERENT STOCK MARKET INDICES

3.1 INTRODUCTION

In this chapter the performances of five different types of stock market indices are empirically examined. An attempt is made to compare the relative performance of each in both bull and bear markets, and some remarks are made about the relative volatility of the five types of indices. In addition, some other interesting results are highlighted, and some of the subjective remarks made in the literature relating to the performance of particular types of stock market indices are shown to hold in practice.

In Section Two, the five types of indices examined in this chapter are defined, while in Section Three the data used in the study is discussed. Section Four describes the statistical tests performed on this data, and presents the empirical results obtained. Finally, Section Five contains a discussion of the conclusions which may be drawn from the results presented in this chapter.

3.2 TYPES OF STOCK MARKET INDICES

In this section, each of the five methods of constructing stock market indices which will be examined in this chapter, are defined. The methodology and underlying philosophy of each of the first four types of indices are given in Section 2.2 of the previous chapter, and the reader is referred to that section for further details. Only the formulae needed to compute these four indices are given below.

1) ARITHMETIC AVERAGE OF PRICE: (DJ INDEX)

$$I_t = 1/n \sum_{i=1}^{n} P_{i;t}$$

where I_t is the level of the index at time t;

$P_{i;t}$ is the price of the i^{th} security included in the index at time t; and

n is the number of securities included in the index.

In this chapter, this index will be called the DJ index after the Dow Jones Averages, which are perhaps the most famous of "arithmetic average of price" indices.

2) MARKET CAPITALIZATION INDICES: (SP INDEX)

$$I_t = \frac{\sum_{i=1}^{n} N_{i;t} * P_{i;t}}{\sum_{i=1}^{n} N_{i;o} * P_{i;o}} * \text{L.F.}$$

where $N_{i;t}$ is the number of shares issued in the i^{th} security at time t;

$P_{i;o}$ is the price of the i^{th} security at some base period, t = o; and

L.F. is a linking factor to preset the index at some desired level.

In the following analysis, this type of index will be called the SP index after the Standard and Poors Indices.

3) ARITHMETIC AVERAGE OF RETURN: (UP INDEX)

$$I_t = 1/n \sum_{i=1}^{n} P_{i;t}/P_{i;t-1} * L.F.$$

In this chapter this index will be called the UP index after the United Press International Market Indicator which is constructed in this manner.

4) GEOMETRIC AVERAGE OF RETURN: (VL INDEX)

$$I_t = (\Pi_{i=1}^{n} P_{i;t}/P_{i;t-1})^{1/n} * L.F.$$

In the following analysis this index will be called the VL index after the Value Line index.

5) THE ESE INDICES: (ESE INDEX)

The fifth type of index which will be considered in this chapter, is an index which is peculiar to the Johannesburg Stock Exchange. It is the ESE type index, which is constructed in the following manner:

$$I_t = 1/n \sum_{i=1}^{n} \left(\frac{1000}{P_{i;o}} * P_{i;t}\right) * L.F.$$

where $P_{i;o}$ is the price of the i^{th} security at the beginning of the current year,and the remaining symbols are as previously defined.

This type of index reflects the performance of an investor who allocates his funds equally (i.e. equal rand amounts) in each of the n constituent securities at the beginning of the year, and maintains that portfolio until the end of the year when he sells all his securities and reallocates equal rand amounts to each security. Thus, in effect, he sells part of his holdings in those securities which have performed best in the year, and purchases more of those securities which have performed worse. Clearly, this index is very similar to the UP type index (see Section 2.2 of Chapter 2). The difference lies in the fact that the UP index reallocates each period that the index is computed, while the ESE index reallocates annually. It can be argued that this makes the ESE a more realistic index for the average investor,since it is unlikely that an investor would generally reallocate his funds every period (especially if the index is constructed daily or weekly) whereas he might reallocate annually. In any case, except for a very active trader, the yearly reallocation is probably closer to reality than daily or weekly reallocation. While on the subject of reallocation and reality, it is worth noting that the market capitalization type indices, which appear to be the most popular at the present time, are not very realistic in terms of mirroring the performance of the average investor. As Levy (1968) points out, the average investor is much more likely to buy either an equal number of the securities he selects (i.e. the DJ type index) or an equal number of dollars (or rands) worth (as assumed by the UP and ESE indices).

The above mentioned are the five methods of constructing stock market indices, which are to be examined in this chapter. In the next section, the data, which was used to perform this analysis, is described.

3.3 THE DATA

In order to examine the performance of the different methods of constructing stock market indices, it was decided to draw a number of random samples of 50 securities from the 203 securities (see Appendix C) for which data was available over the period 22 March 1968 to 20 June 1975. In all, 8 samples were drawn, 4 of which were simple random samples and 4 of which were stratified random samples (stratified by the number of securities in each sector relative to the total number of securities in the market). Then, for each sample, five indices were constructed - one for each method of construction considered. Thus, 40 indices in all were constructed.

In order to analyze the performance of these indices, it was decided to break the data into four distinct periods:

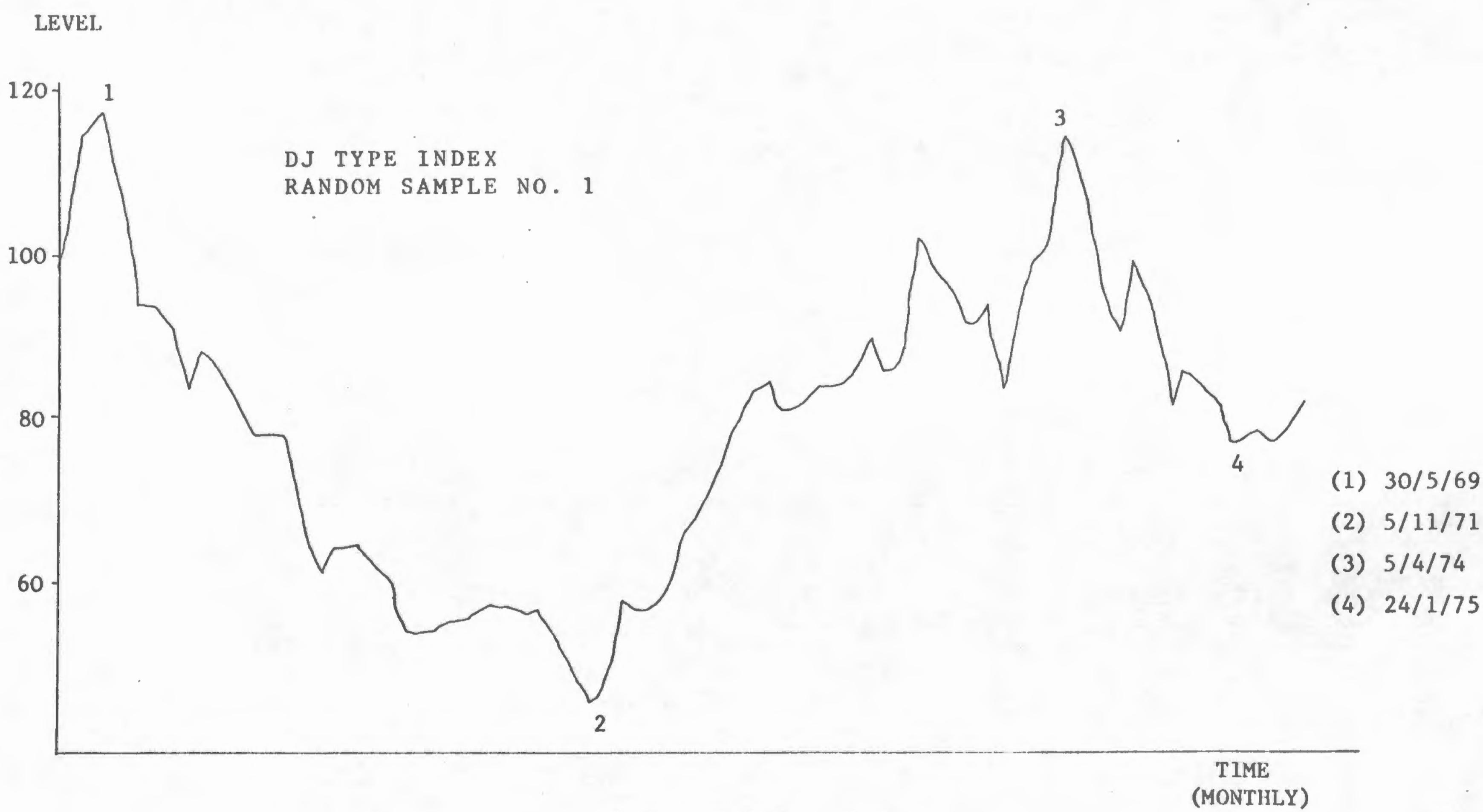
(i)	GENERAL BULL MARKET:	22 March 1968 to 30 April 1969
(ii)	GENERAL BEAR MARKET:	30 April 1969 to 5 November 1971
(iii)	GOLD BULL MARKET:	5 November 1971 to 5 April 1974
(iv)	GOLD BEAR MARKET:	5 April 1974 to 24 January 1975

These periods can be clearly seen by examining the graph (plotted monthly) of the DJ index using random sample number one, which is presented in Figure 3.1 below. The graphs for the other types of indices considered, and for the different samples drawn, all have the same basic shape, although the amount of movement varies for each. Specifically, each of the 40 indices examined reached a peak on 30 April 1969, each reached a bottom on 5 November 1971, reached another peak on 5 April 1974, and on the 21 January each again reached a bottom.

Now, in order to make the comparisons simpler, the performance in each type of market ((i) to (iv) above) was considered separately, and the analyses were performed on only a single figure per index for each sample drawn - the overall percentage return on the index in that period. That is,

$$\frac{I_t - I_o}{I_o} \times 100$$

where


LEVEL
120
100
80
60
1
2
3
4
DJ TYPE INDEX
RANDOM SAMPLE NO. 1
(1) 30/5/69
(2) 5/11/71
(3) 5/4/74
(4) 24/1/75
TIME
(MONTHLY)

F I G U R E 3.1

I_t is the level at the end of the period, and

I_o is the level at the beginning of the period.

These figures are presented for the different indices using the various samples in Tables 3.1 to 3.4 below.

T A B L E 3.1 : GENERAL BULL MARKET (NET % RETURN OVER PERIOD 22/3/68 - 30/4/69)

INDEX / SAMPLE	DJ	SP	UP	VL	ESE
1	18	13	26	23	29
2	20	26	21	18	22
3	20	26	27	24	31
4	26	25	21	19	22
5	21	24	29	25	31
6	27	29	21	20	23
7	13	16	23	20	28
8	25	22	26	22	27

T A B L E 3.2 : GENERAL BEAR MARKET (% RETURN 30/4/69 - 5/11/71)

INDEX / SAMPLE	DJ	SP	UP	VL	ESE
1	-62	-61	-45	-54	-50
2	-62	-65	-54	-60	-54
3	-48	-52	-38	-49	-41
4	-68	-67	-47	-56	-49
5	-64	-62	-47	-57	-50
6	-60	-67	-43	-52	-44
7	-48	-54	-42	-52	-45
8	-65	-61	-47	-55	-49

TABLE 3.3 : GOLD BULL MARKET (% RETURN 5/11/71 - 5/4/74)

INDEX / SAMPLE	DJ	SP	UP	VL	ESE
1	151	152	164	123	181
2	215	177	166	119	175
3	202	237	143	106	151
4	145	139	158	123	163
5	158	157	145	107	158
6	175	128	139	107	146
7	198	215	163	124	169
8	161	140	157	118	160

TABLE 3.4 : GOLD BEAR MARKET (% RETURN 5/4/74 - 24/1/75)

INDEX / SAMPLE	DJ	SP	UP	VL	ESE
1	-35	-34	-30	-34	-33
2	-26	-29	-27	-31	-29
3	-25	-25	-24	-28	-25
4	-36	-32	-29	-34	-32
5	-32	-35	-29	-34	-31
6	-29	-33	-26	-31	-29
7	-28	-29	-27	-32	-29
8	-27	-25	-25	-30	-27

3.4 TESTS AND RESULTS

The data for each type of market was initially analyzed by performing a two way analysis of variance on each of the data tables above (Tables 3.1 to 3.4). The two factors considered were the method of sampling (random or stratified) and the method of construction (DJ, SP, UP, VL, and ESE).

It must be mentioned that in order to perform these analyses of variance, the returns on the various indices must be assumed to be normally distributed. However, while this assumption is questionable since it has been shown (Fama (1965a) and Affleck-Graves (1974)) that the distributions appear to be Stable Paretian, it can nevertheless be argued that the data is not "too far removed" from normality (except for the tails of the distribution) and hence the analyses of variance should give fairly satisfactory results. Moreover, it will be shown later that a nonparametric test does not give any different results from a parametric test assuming normality, which would appear to indicate that the degree of non-normality (if indeed present) is not serious.

The analysis of variance tables for each of the four periods examined are presented in Tables 3.5 to 3.8 below.

TABLE 3.5

ANOVA TABLE FOR GENERAL BULL MARKET

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F-Ratio
Due to Sample	5,625	1	5,625	0,297
Due to Index	162,350	4	40,587	2,145
Sample ×Index	1,250	4	0,313	0,017
Error	567,750	30	18,925	
Total (Corrected)	736,975	39		

TABLE 3.6

ANOVA TABLE FOR GENERAL BEAR MARKET

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F-Ratio
Due to Sample	8,100	1	8,100	0,252
Due to Index	1563,100	4	390,775	12,161
Sample ×Index	1,900	4	0,475	0,015
Error	964,000	30	32,133	
Total (Corrected)	2537,100	39		

TABLE 3.7

ANOVA TABLE FOR GOLD BULL MARKET

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F-Ratio
Sample	680,625	1	680,625	1,217
Index	17 521,000	4	4 380,250	7,832
Sample×Index	193,000	4	48,250	0,086
Error	16 778,750	30	559,292	
Total (Corrected)	35 173,375	39		

TABLE 3.8

ANOVA TABLE FOR GOLD BEAR MARKET

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F-Ratio
Sample	2,500	1	2,500	0,227
Index	89,850	4	22,460	2,042
Sample×Index	4,750	4	1,190	0,108
Error	330,000	30	11,000	
Total (Corrected)	427,100	39		

From the analysis of variance tables presented above, it can be seen that there is no significant interaction between the method of sampling and the method of construction in each of the four markets examined. Also, the method of sampling is not significant in all four cases. That is, there is no statistical difference in the results obtained using either random sampling or stratified sampling. This is to be expected since the total population examined consisted of 200 securities, and each sample comprised of 50 securities. Thus, a significant difference in the results obtained from the two sampling methods would not be expected as the smallness of the population, relative to the sample, tends to ensure a fairly good spread among the sectors, even for the random samples.

The method of constructing the stock market index was, however, a significant factor in two of the four types of markets which were examined (the General Bear Market and the

Gold Bull Market). Moreover, in the two markets where this factor was not significant (the General Bull Market and the Gold Bear Market) the F-ratio was larger than 2. This would appear to indicate that the effect of the method of construction should be more closely examined.

In order to do this, the four types of markets under consideration were again examined individually, and pair-wise multiple comparisons performed. That is, the performance of each pair of methods of construction were compared, resulting in ten comparisons for each type of market (forty comparisons in all). These comparisons were made using the paired t-test since each of the eight samples drawn were used to construct all five types of indices.

The t-tables resulting from these tests are presented for the four different types of markets in Tables 3.9 to 3.12 below. So as to have an overall significance level of 5%, the tests were performed using the t-tables computed using the Bonferoni Inequality (hereafter called the Bonferoni t-tables) to allow for the simultaneous inferences being used. With $k = 10$ (the number of pairs) and $n = 8$ (the sample size for each test), the critical value is found to be 4,40 (cf Miller (1966)). (The asterisked values in the tables below are the significant values.)

TABLE 3.9

PAIRED t-TESTS : GENERAL BULL MARKET (22/3/68-30/4/69)

INDEX 1 \ INDEX 2	SP	UP	VL	ESE
DJ	-0,96	-1,37	-0,06	-2,07
SP		-0,58	0,55	-1,44
UP			8,21*	-4,46*
VL				-8,10*

(Note: Positive values indicate that index type 1 rose more on average than index type 2, and vice versa for negative numbers.)

TABLE 3.10

PAIRED t-TESTS : GENERAL BEAR MARKET (30/4/69-5/11/71)

INDEX 1 \ INDEX 2	SP	UP	VL	ESE
DJ	-1,05	7,40*	2,64	6,18*
SP		10,36*	4,44*	8,07*
UP			-16,84*	-4,46*
VL				14,39*

(Note: Positive values indicate that index type 1 fell more on average than index type 2, and vice versa for negative numbers.)

TABLE 3.11

PAIRED t-TESTS : GOLD BULL MARKET (5/11/71-5/4/74)

INDEX 1 \ INDEX 2	SP	UP	VL	ESE
DJ	-0,17	2,48	6,42*	1,25
SP		0,97	3,65	0,37
UP			24,71*	-5,30*
VL				-18,53*

TABLE 3.12

PAIRED t-TESTS : GOLD BEAR MARKET (5/4/74-24/1/75)

INDEX 1 \ INDEX 2	SP	UP	VL	ESE
DJ	-0,51	2,97	-2,37	0,51
SP		3,66	-1,82	1,18
UP			-25,28*	-9,00*
VL				9,03*

The analysis of the results presented in the tables above is fairly difficult and subjective since the tables do not present identical results for each type of market. However, those results, which it is felt are most important, are presented below. The conclusions which can be drawn from these results are discussed in the next section.

1. THE DJ AND SP TYPE INDICES

These indices are of special interest because they are by far the most popular methods of constructing stock market indices at present. In all four types of markets, no statistically significant difference in the performance of the indices constructed using these two methods, was detected.

2. UP, VL, AND ESE TYPE INDICES

These three indices can be considered together since all three are based on return and not price. It is extremely interesting to observe that in all four types of markets, the performance of these three indices were statistically significantly different. It is also interesting to note the direction of this difference. In the two bull markets, the ESE type indices rose more than the UP type indices, which in turn rose more than the VL type indices, while in the two bear markets, the VL indices fell more than the ESE indices, which fell more than the UP type indices.

3. COMPARISONS BETWEEN THE PRICE (DJ AND SP) AND RETURN (UP, VL, AND ESE) INDICES

In general, there does not appear to be a statistically significant difference between indices based on price and indices based on return. With the exception of the General Bear Market, there was only one significant t-statistic (between DJ and VL in the Gold Bull Market). However, in the General Bear Market, all comparisons were significant except for the DJ and VL type indices.

In addition, all the directions of difference were positive, which indicates that in this market, the price indices fell more than the return indices.

Finally, because all of the above tests (the analyses of variance and the paired t-tests) assume normality of the data, it was decided to re-examine the paired comparisons using a nonparametric test and to check whether any difference in the results was obtained. The nonparametric test chosen was the randomization test (cf Siegel (1956), page 94), which is a one hundred per cent efficient test since it uses all the information in the sample. The critical point for this test varies with each pairwise test and hence, in the tables below the results are merely presented as NS (meaning non-significant), US (meaning significant at the upper α% level), and LS (meaning significant at the lower significant level). The results of the randomization tests are presented in Tables 3.13 to 3.16 below. Since simultaneous tests are being made, a significance level of one percent, instead of five percent, was used (the Bonferoni t allows for such a correction in the test above).

TABLE 3.13

RANDOMIZATION TESTS : GENERAL BULL MARKET (22/3/68-30/4/69)

INDEX 1 \ INDEX 2	SP	UP	VL	ESE
DJ	NS	NS	NS	NS
SP		NS	NS	NS
UP			US	LS
VL				LS

(Note: US indicates that index 1 rose more than index 2 and vice versa for LS.)

TABLE 3.14

RANDOMIZATION TESTS : GENERAL BEAR MARKET (30/4/69-5/11/71)

INDEX 1 \ INDEX 2	SP	UP	VL	ESE
DJ	NS	US	NS	US
SP		US	US	US
UP			LS	LS
VL				US

(Note: US indicates that index 1 fell more than index 2 and vice versa for LS.)

TABLE 3.15

RANDOMIZATION TEST : GOLD BULL MARKET (5/11/71-5/4/74)

INDEX 2 / INDEX 1	SP	UP	VL	ESE
DJ	NS	NS	US	NS
SP		NS	US*	NS
UP			US	LS
VL				LS

TABLE 3.16

RANDOMIZATION TEST : GOLD BEAR MARKET (5/4/74-24/1/75)

INDEX 2 / INDEX 1	SP	UP	VL	ESE
DJ	NS	NS	NS	NS
SP		US*	NS	NS
UP			LS	LS
VL				US

The above results are almost identical to those results presented in Tables 3.9 to 3.12. The two differences in the sets of tables (asterisked in Tables 3.15 and 3.16 above) both had absolute t-values of over 3,5, which while not statistically significant using the Bonferoni t, is nevertheless high. Thus, it is felt that the assumption of normality used in constructing the analysis of variance tables, is reasonable, and that the comments on the different methods of constructing stock market indices presented after Table 3.12, remain valid.

3.5 CONCLUSIONS

In this section, some of the main conclusions which can be drawn from the results presented in the previous section, will be discussed. As mentioned previously, some of the conclusions are rather subjective and are not statistically significant, but wherever this is the case, the lack of statistically significance will be explicitly stated.

Firstly, it was shown that there was not a significant difference in the results obtained using either random or stratified sampling. As mentioned in Section 4, this was to be expected since the effect of the stratification would not be very marked in the case under consideration (a sample of 50 from a population of 200, which is itself well stratified over the various sectors).

Secondly, it was shown that in two of the four markets examined, the method of constructing the stock market index caused a significant difference in the performance of the index. In the remaining two markets, which did not indicate significance, the F-ratios were nevertheless large. These tests resulted in further analyses being performed to examine the differences between the methods of construction. The results obtained can be divided into three sections; (i) the relative performance of indices based on price and indices based on return; (ii) the performance of the indices based on price; and (iii) the performance of indices based on return. These three sections are discussed separately below.

(i) PRICE AND RETURN INDICES

There does not appear to be any general difference in the performance of the price and return indices. However, in a bear market it does appear as if the price indices fall more than the return indices. In a bull market there appears to be no difference in the percentage rise of the two types of indices, which (with the greater drop of the price indices in the bear markets) would appear to justify the remark that indices based on the arithmetic mean of return tend to outperform indices based on price because of an upward bias in the return indices (cf for example, Latané, Tuttle and Young (1971)). On the other hand, if one equates volatility in an index with information (i.e. the more volatile an index the more informative it is), then it would appear that the price indices are superior since they rise as much as return indices in bull markets, but fall by more in bear markets, and are thus slightly more volatile.

(ii) INDICES BASED ON PRICE (THE DJ AND SP TYPE INDICES)

The theoretical differences between arithmetic averages of price and market capitalization indices have been discussed in the literature, but most empirical studies have merely indicated that the two methods provide indices which are fairly highly correlated (for example Kantor (1972)), and have not contrasted the relative volatility of the two types of indices. In this chapter, the relative performance of the indices was examined and no significant difference was found between the performance of the DJ indices and the SP indices. However, on examination of Tables 3.9 to 3.12 it can be seen that the t-statistics for the DJ and SP comparisons are always negative, which tends to indicate that the SP type indices rise slightly more than the DJ indices in bull markets, and fall slightly more in bear markets. Thus, one might conclude (although not statistically justifiably) that the SP type indices are slightly more volatile,and hence more informative,than the DJ type indices. This observation can perhaps be justified, in the South African context at least, by noting that the larger companies are often the more popular companies with the average investor, and are thus more volatile than the smaller companies, which in turn results in the SP type indices being more volatile than the DJ type indices.

(iii) INDICES BASED ON RETURN (THE UP, VL, AND ESE TYPE INDICES)

The return indices indicated statistically significant differences in all four types of markets. However, the ordering of the indices was different for bull and bear markets. In both bull markets, the ESE type indices rose more than the UP type indices, which in turn rose more than the VL type indices, while in both bear markets the VL indices fell more than the ESE, which fell more than the UP. The following conclusions can be drawn from these results.

1) The VL indices are always at a lower level than

either the UP or ESE indices. This follows since in a rising market the VL will rise less than the other two indices, while in a falling market the VL will fall more. Thus, the empirical results presented in this chapter confirm the theoretical conclusion that the VL index should form a floor for all other return indices (cf for example Marks and Stuart (1971); Latane, Tuttle and Young (1971); Hodges and Schaefer (1974)).

2) The ESE indices appear to be better than the UP indices. This remark can be justified since the ESE indices rise more than the UP indices in the bull markets, and fall by more in the bear markets. This indicates that the ESE type indices are more volatile, and hence more informative than the UP indices. Thus, it would appear that a return index based on yearly reallocation, is preferable to a return index based on weekly reallocation,since it is more volatile. In addition, as has been mentioned previously, it can be argued that yearly reallocation is probably a truer reflection of the average investor's strategy than weekly reallocation, and hence it is suggested that the ESE type index be used in preference to a UP type index.

Finally, in concluding this chapter, it can be said that the results presented above indicate that the method used in constructing a stock market index is more important than the actual sample of securities used,if the sample is fairly well spread across the market. Unfortunately, there does not seem to be a universally most acceptable index. Rather, the choice of the method of construction should depend on the purpose for which the index is to be used. On the Joahnnesburg Stock Exchange it would appear that investors are well catered for as the two main sets of indices, the Rand Daily Mail (SP type indices) and the ESE indices,are probably the most suitable of the price and return indices respectively. It is left to the individual investor to choose which best satisfies his requirements.

CHAPTER FOUR

STOCK MARKET INDICES AND PRINCIPAL COMPONENTS

4.1 INTRODUCTION

The use of principal component analysis in the construction of economic and stock market indices is not new. Theil (1960) discussed their use in economic indices, while Feeney and Hester (1967) gave a detailed account of how this idea could be applied in the construction of stock market indices. Troskie (1970) showed how these ideas could be used on the Johannesburg Stock Exchange, and presented some examples of such indices.

The concept of an index based on principal components is intuitively most appealing. If an index is designed to measure movement in the market, then it will be most sensitive (and hence most informative) if the weights are assigned in such a way that the index has maximum variance over all linear combinations of the stocks to be included in the index. But, such a combination is simply the largest component,which can be obtained very easily. The question then arises as to why so little attention has been paid in recent years to the development of stock market indices based on principal components.

In their paper, Feeney and Hester (1967) conclude that, "we do not think that simple (linear) stock price indices are as promising a guide for investors because weights of even the largest components appear to change considerably over relatively short periods of time." While this might be a disadvantage, it is not felt that this, the instability of the weights, is as important a factor as made out by Feeney and Hester. Afterall, numerous stock market indices (for example, the ESE indices, the UPI Market Indicator, and market capitalization indices in general) do involve regular changes in the factor by which the price is weighted. What is felt to be of far more importance, is the average investors failure to accept or conceive of an index with negative weights, since such a weight implies a negative holding in a company which is completely unacceptable to most investors (even though it can be argued that this is equivalent to a short holding). Moreover, if some of the weights are negative, then it is possible for the index to go negative, as is shown in the following example.

Consider the following two security problem where the first principal component has the following form:

Security	Weight
A	0,866
B	-0,500

Now, if the index is constructed weekly and the prices of the individual securities are as given in Table 4.1 below, then the value of the index will vary as shown in the last column of Table 4.1.

TABLE 4.1

Week No.	Price of A	Price of B	Index
1	100	100	36,60
2	200	100	123,20
3	100	100	36,60
4	50	150	-31,70

This is clearly a most unsatisfactory state of affairs for a stock market index.

In this chapter it will be shown that by a simple re-statement of the principal component problem, this important disadvantage of negative weights can easily be overcome. In addition, there are further benefits to the designer of the stock market index in the form of a facility to impose additional constraints.

In Section Two, a brief general overview of principal components is presented, while in Section Three it is shown how the problem can be restated to give more satisfactory results. Section Four presents some practical examples of indices constructed in this manner, and compares them with the original principal component indices. Section Five shows some additional uses to which these indices can be put, and discusses some additional factors which can be incorporated in the new model together with some further examples. Section Six presents examples of two sector indices constructed in the above manner, which demonstrate the considerable flexibility of the method proposed in this chapter. Finally, some overall conclusions are discussed in Section Seven.

4.2 PRINCIPAL COMPONENT ANALYSIS

In this section a brief summary of the concept of principal components is given — for a detailed mathematical exposition, the reader is referred to Anderson (1958).

The problem of principal components can be briefly stated as follows: Given a set of p variables $C(= \begin{pmatrix} C_1 \\ \cdot \\ \cdot \\ \cdot \\ C_p \end{pmatrix})$ with covariance matrix Σ, find the linear combination $a = X'C$ such that the variance of a (i.e. $x'\Sigma x$) is a maximum. To do this, one extracts from Σ its characteristic roots $(\lambda_1, \ldots, \lambda_p)$ and vectors $(h_{(1)}, \ldots, h_{(p)})$; that is, one solves

$$|\Sigma-\lambda_i I| = 0$$

$$(\Sigma-\lambda_i I)h_{(i)} = 0$$

subject to $h'_{(i)}h_{(i)} = 1$

for $i = 1,2,\ldots,p$

The characteristic vector corresponding to the largest characteristic root gives the linear combination with maximum variance. Moreover, the variance of this linear combination is given by the largest characteristic root.

Now, in general, Σ is not known and must be estimated by

$$S = \sum_{i=1}^{N}(Z_i-\bar{Z})(Z_i-\bar{Z})'/N-1$$

where Z_i is the vector of the i^{th} observation on each of the p variables, $C_1,\ldots,C_p$. The characteristic roots and vectors extracted from S will be estimates of the corresponding roots and vectors of Σ. Thus, each estimated characteristic vector and characteristic root pair provides an estimate of a principal component and its associated variance.

The technique of principal components has an additional feature. Not only do we have a component which has maximum variance, but it is possible (very simply) to obtain a second component, which has maximum variance subject to the condition that it is uncorrelated (orthogonal) to the first component. This component is given by the characteristic vector of S (Σ if known) corresponding to the second largest characteristic root, which also gives the variance of this component. In fact, p such components can be constructed (if S is of full rank) - the p characteristic vectors of S - which have the useful property of being mutually orthogonal.

Finally, a further aspect of principal components which must be mentioned at this stage, concerns the choice of the data on which to perform the principal component analysis. There are generally, three schools of thought, all of which have their pros and cons.

Firstly, there are those who argue that the covariance matrix of the raw stock prices should be used since these are the variables of interest. However, it has been frequently noted that this results in the principal component being price dependent, since a security priced at R50 with a standard deviation of R5 will probably have more weight than a security priced at R2,00 with a standard deviation of R1, even though the latter is relatively far more volatile.

This type of argument led to the second group,who argue that one should work with standardized prices. This is equivalent to using the correlation matrix, and results in all securities having equivalent standard deviations. However, it can be argued that principal component analysis on the correlation matrix "has a rather artificial quality" (Morrison (1967)) since it is based on standardized scores, which are in fact dimensionless.

Thirdly, there is the proposal of Feeney and Hester (1967) that the principal component analysis be performed on the covariance matrix of returns. From theoretical considerations, this would appear to be the best alternative, but does have a disadvantage in that most investors do not seem to accept the idea of indices based on return, and tend to "think in terms of price."

In this chapter, it is not proposed to investigate the relative merits and demerits of the above three proposals. The methods proposed in Sections Three and Four are equally applicable in all three cases, and are thus independent of the choice. However, for the purposes of illustration, the opinion of the majority will be respected, and the examples presented in the next three sections all use the covariance matrix of price.

4.3 PRINCIPAL COMPONENT ANALYSIS WITH POSITIVE WEIGHTINGS

The problem of principal components can be stated in terms of a Mathematical Programming Problem as follows:

$$\begin{aligned} &\text{Max} \quad x'\Sigma x \\ &\text{subject to} \quad \textstyle\sum_{i=1}^{p} x_i^2 = 1 \qquad (4.1) \\ &\qquad x_i \ \text{unrestricted in sign} \end{aligned}$$

Clearly, if one wishes to restrict the x_i to be positive then one can simply restate the problem as

$$\begin{aligned} &\text{Max} \quad x'\Sigma x \\ &\text{S.T.} \quad \sum x_i^2 = 1 \\ &\qquad\quad x_i \geq 0 \quad \text{for} \quad i = 1,2,\ldots,p \end{aligned} \tag{4.2}$$

This then is a quadratic programming problem (since the objective function has terms of the form x_i^2 and $x_i x_j$) subject to a quadratic constraint $(\sum x_i^2 = 0)$ and the constraint that the x_i's be positive or zero. This problem can be solved using any of the methods for solving either general nonlinear programming problems or quadratic programming problems subject to quadratic constraints.

Unfortunately, the extension of this idea of positive weightings to the second, third and other components is not possible in general. Clearly, if the vector X is strictly greater than zero (that is, no element of X equals zero) then it is not possible to find another vector Y also strictly greater than zero such that X and Y are orthogonal $(\sum x_i y_i = 0)$. However, given a first component which is positive, it is possible to find an orthogonal second component if one relaxes the positive restriction on the second component. The problem can then be formulated as follows.

Find the first component (positive) by solving the Quadratic Programming Problem (4.2) above. Then, solve the following Quadratic Programming Problem:

$$\begin{aligned} &\text{Max} \quad y'\Sigma y \\ &\text{S.T.} \quad \textstyle\sum_{i=1}^{p} y_i^2 = 1 \\ &\qquad\ \textstyle\sum_{i=1}^{p} y_i x_i = 0 \\ &\qquad\ y_i \ \text{unrestricted in sign} \end{aligned} \tag{4.3}$$

The additional constraint in the above problem $(\sum y_i x_i = 0)$ is merely a linear constraint since the weights x_i are known from the solution to (4.2).

Clearly, this idea can be easily extended to obtain all the other components, provided one is prepared to forsake the $X \geq 0$ constraint on all but the first component. It can be argued that this is reasonable, since it is often hoped that the first component will provide some overall description of the market, while subsequent factors will describe other aspects of the market (e.g. contrasting various sectors etc.). In general then, the r^{th} component $(r \geq 2)$ can be found as follows:

$$\begin{aligned} &\text{Max} \quad z'\Sigma z \\ &\text{S.T.} \quad \textstyle\sum_{i=1}^{p} z_i^2 = 1 \\ &\qquad\ \textstyle\sum_{i=1}^{p} w_{ij} z_i = 0 \qquad j = 1,2,\ldots,r-1 \\ &\qquad\ z_i \ \text{unrestricted in sign} \end{aligned} \tag{4.4}$$

where w_{ij} is the weight assigned to the i^{th} variable by the j^{th} component.

In this way, the problem of principal component analysis subject to the restriction that the weights of the first component are all positive or zero, can be solved as shown above, and as many of the remaining components as desired may still be computed.

4.4 EXAMPLES

In this section some practical examples of the ideas discussed in the previous section will be presented. Before listing the results obtained, some remarks are made on the method by which the Quadratic Programming Problems were solved.

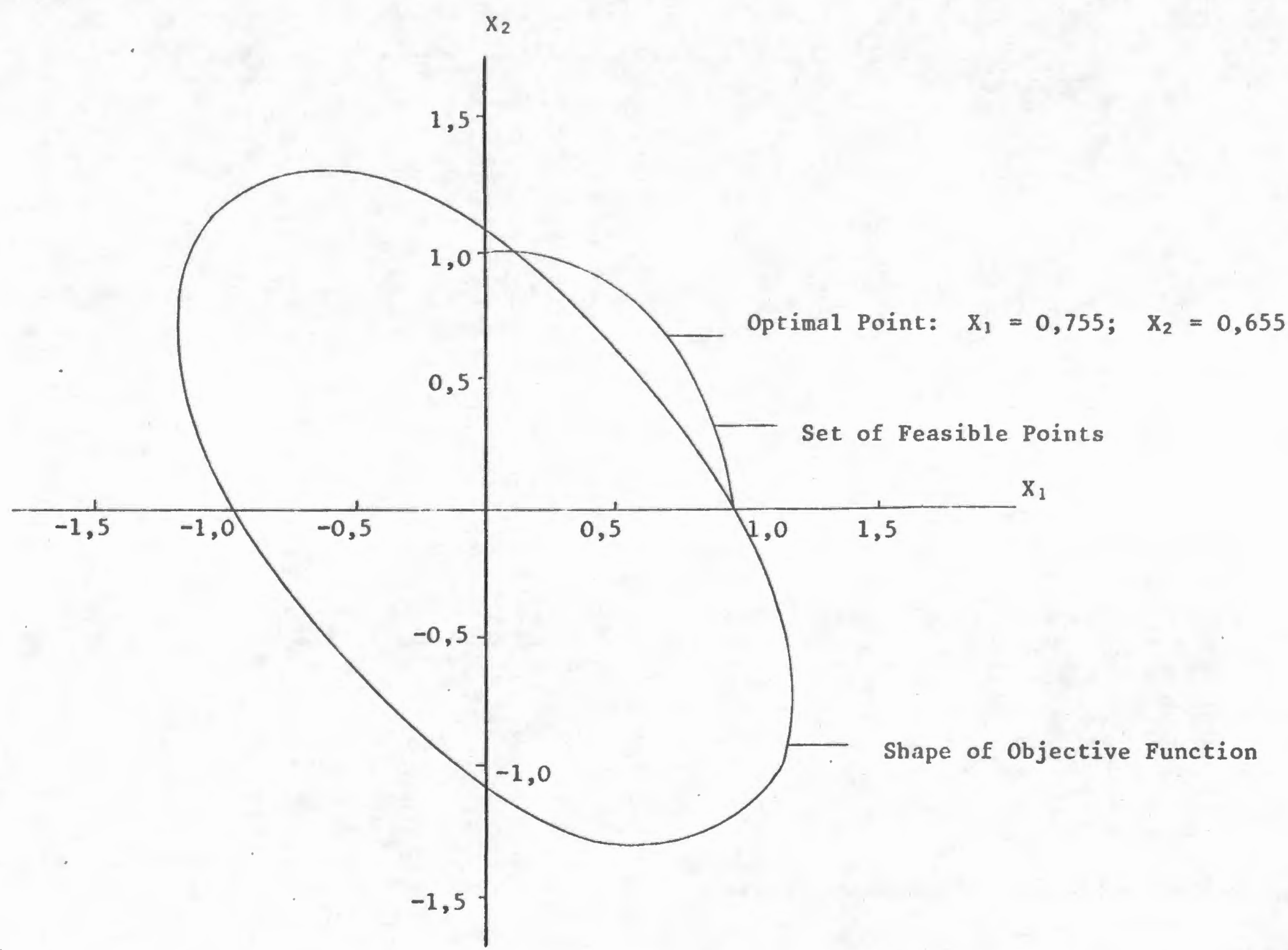
There are numerous techniques available for the optimization of a nonlinear function subject to a number of constraints, either linear or nonlinear. In addition, there are algorithms which handle Quadratic Programming problems specifically. The method used to solve the examples presented below was a general nonlinear optimization technique called the Flexible Tolerance Method (discussed in Himmelblau (1972) pages 340 to 359). The actual program used was the program listed in Himmelblau (pages 458 to 468) with a few minor alterations necessitated by the specific computer configuration employed. This program is listed in Appendix D.

The Flexible Tolerance Method is a search technique, and does not provide an exact solution. However, on numerous tests performed, it was found that the error between the solution obtained and the true optimal was seldom more than ½%. One of the reasons for this error is that the objective function is similar in shape to the constraint $\sum x_i^2 = 1$, and hence there is a fairly wide range near the true optimal solution, where there is very little change in the value of the objective function. This is perhaps best illustrated by examining the two dimensional case. The objective function and feasible region are shown in Figure 4.1 on the following page for the case where

$$\Sigma = \begin{pmatrix} 10 & 5 \\ 5 & 8 \end{pmatrix}$$

In practice, this is not thought to constitute a great drawback since the solution obtained is usually very close to the true optimal. However, no claim is made about the overall superiority of this method. Rather, it is one of several methods which could have been used, but was chosen because of convenience.

In order to illustrate the concepts presented in the previous section, several examples are presented below. All of these examples are based on basically the same problem, namely to construct an index from ten securities quoted on the Johannesburg Stock Exchange; five from the Coal Sector and five from the Gold-Witwatersrand and Others Sector (see Appendix E). The covariance matrix, over a period of approximately two years, was estimated (S) from the


X2
1,5
1,0
0,5
-0,5
-1,0
-1,5
X1
-1,5
-1,0
-0,5
0,5
1,0
1,5
Optimal Point: X1 = 0,755; X2 = 0,655
Set of Feasible Points
Shape of Objective Function

F I G U R E 4.1

weekly closing prices of these ten securities (4/1/74 to 20/2/76) and is given below.

	1	2	3	4	5	6	7	8	9	10
1.Apex	33430									
2.Clydes	2074	309								
3.T'stock	53067	3831	103298							
4.T.Natal	1901	287	2845	345						
5.W'dacht	7108	536	11356	546	1755					
6.D.Deep	-110745	-4012	-222312	-114	-20915	762268				
7.ERPM	-96188	-3815	-190293	-491	-18667	651662	574699			
8.G'vlei	-21255	-950	-40736	-351	-4539	140671	122338	29042		
9.M'vale	-33671	-1672	-63037	-830	-7017	198300	170100	39924	59085	
10.Sallies	-48431	-2293	-97172	-979	-10520	336654	290132	68627	94915	170792

The first principal component was determined using both the traditional approach (hereafter referred to as the P-Component) and the proposed approach, which restricts the weights to be nonnegative (hereafter referred to as the G-Component). The results are presented in Table 4.2 below.

From the results presented in Table 4.2, the difference between the two methods can be clearly seen. Moreover, it is interesting to note that there is not a great difference in the variance of the linear combination (the objective function) - the G-Component is approximately 5% lower than the P-Component.

As a further example, the second component of the above problem was found, and the results are presented in Table 4.3

TABLE 4.2

SECURITY	P-Component	G-Component
Apex	-0,10337	0,00000
Clydesdale	-0,00431	0,00005
Tavistock	-0,20446	0,00000
Trans Natal	-0,00099	0,00042
Welgedacht	-0,02026	0,00272
Durban Deep	0,67840	0,69181
E.R.P.M.	0,58705	0,61369
Grootvlei	0,12774	0,12787
Marievale	0,18009	0,17675
S.A. Lands	0,30576	0,31174
Variance (obj.func.)	1 641 570	1 556 792
% Variation Explained	94,61%	89,73%

below. Although the second component using the modified approach is allowed to have negative weights (as discussed in the previous section) it will still not be identical to the traditional second component, since both must be orthogonal to their respective first components.

These two examples indicate the manner in which the procedure described in Section Three may be used to perform a principal component analysis, with the additional constraint that the first component must be nonnegative.

TABLE 4.3

SECURITY	Second P-Component	Second G-Component
Apex	0,50057	0,46910
Clydesdale	0,05765	0,03753
Tavistock	0,80842	0,86202
Trans Natal	0,06286	0,10319
Welgedacht	0,12077	0,13908
Durban Deep	0,13891	0,02344
ERPM	0,16097	-0,01546
Grootvlei	0,04302	-0,01448
Marievale	-0,09059	-0,06400
S.A. Lands	0,13695	0,01929
Variance (obj.func.)	50 670	133 051
% Variation Explained	2,92%	7,67%

Finally, before concluding this section, some remarks concerning computer time involved in the solution will be made. The program obtains the first G-Component in approximately the same computer time required to obtain the first P-Component. However, a second run is required to obtain the second G-Component, and because of the additional constraint, more computer time is needed. In addition, since the method employed to solve the problem is an iterative search procedure, Himmelblau (1972) recommends that the analysis be repeated using different starting solutions. The experience gained from the examples presented in this chapter suggests that the following procedure will produce good solutions in most cases:

> First perform the traditional principal component analysis. If the first component has all non-negative elements then no further analysis is required - the solution has been found. If some of the elements are negative, then use the first component as the initial starting solution.

Using this method the computer time involved was not at all excessive - for the ten security problem the first component was found in under one minute, and the second component (using the second P-Component as an initial solution) in a similar time.

4.5 EXAMPLES WITH ADDITIONAL CONSTRAINTS

In this section it will be shown how the example presented in the previous section can be extended to allow the investigator even more control over the component.

Since the method of solution is a general nonlinear optimization method, any additional constraints which the investigator may wish to impose on the system can be easily incorporated in the problem - something which cannot be done using traditional principal component analysis. To illustrate this, some examples, which it is felt would be particularly useful in the construction of Stock Market indices, are given.

The first example deals with the limitation of the weights given to any particular security. One of the arguments against the use of principal components is that the situation frequently arises where an individual security can be assigned a very large weight - sometimes as much as 80% of the total weight (i.e. a weight of approximately 0,9 since $\sum x_i^2 = 1$). Using the procedure proposed in Section Three, this can be easily overcome by placing an upper bound on the weight given to any single security. For example, it might be decided that in the ten security problem discussed above, no individual security should have a weight of more than 0,5. This can be incorporated by formulating the model as follows:

$$\text{Max} \quad x'Sx$$

$$\text{S.T.} \quad \sum x_i^2 = 1$$

$$x_i \geq 0 \quad \text{for} \quad i = 1,2,\ldots,10$$

and

$$x_i \leq 0,5 \quad \text{for } i = 1,2,\ldots,10$$

Doing this, the following results (Table 4.4) were obtained for the first component.

TABLE 4.4

Security	First P-Component	First G-Component	Bounded First G-Component
Apex	-0,10337	0,00000	0,00002
Clydesdale	-0,00431	0,00005	0,00025
Tavistock	-0,20446	0,00000	0,00001
Trans Natal	-0,00099	0,00042	0,00443
Welgedacht	-0,02026	0,00272	0,00763
Durban Deep	0,67840	0,69181	0,50002
ERPM	0,58705	0,61369	0,49976
Grootvlei	0,12774	0,12787	0,30933
Marievale	0,18009	0,17675	0,39392
S.A. Lands	0,30576	0,31174	0,49936
Variance	1 641 570	1 556 792	1 321 490

Obviously, any upper bound which is desirable may be used, but it must be remembered that the sum of the squares of the weights must equal 1. Thus, in the ten security example, the constraint

$$X_i \leq 0{,}25 \quad \text{for} \quad i = 1,2,\ldots,10$$

would not be permissible since in that case, if all ten securities were given a weight of 0,25 (their upper bound) the sum of their squares would be 0,625, not 1 as required.

The second example illustrates the situation where one may wish to impose further constraints on the weights. A case, which could easily arise in the ten security example, would be that since the securities come from two different sectors, each sector should contribute equal weight to the index. In order to do this, it is only necessary to include the additional constraint

$$x_1+x_2+x_3+x_4+x_5 - (x_6+x_7+x_8+x_9+x_{10}) = 0.$$

However, in order to illustrate the flexibility of this method, in that any linear or nonlinear equality or inequality constraint can be included, it will be assumed that it is desired that the sum of the squares of the weights from the Coal Sector should equal the sum of the squares of the weights from the Gold Sector - that is an additional quadratic constraint will be imposed. This can be done by reformulating the model as

$$\begin{aligned} \text{Max} \quad & x' S x \\ \text{S.T.} \quad & \textstyle\sum_{i=1}^{10} x_i^2 = 1 \\ & \textstyle\sum_{i=1}^{5} x_i^2 - \sum_{j=6}^{10} x_j^2 = 0 \\ & x_i \geq 0 \quad \text{for} \quad i = 1,2,\ldots,10 \end{aligned}$$

This formulation results in the following first component being obtained (Table 4.5).

Finally, an example incorporating both of the above conditions is given. That is, the sum of the squares of the weights in each sector must be equal, and no single security is to have a weight of more than 0,5. The results for the first component (restricted to be nonnegative) are presented in Table 4.6 below.

TABLE 4.5

Security	Constrained First G-Component
Apex	0,00000
Clydesdale	0,67581
Tavistock	0,00000
Trans Natal	0,20803
Welgedacht	0,00030
Durban Deep	0,49664
ERPM	0,42932
Grootvlei	0,08115
Marievale	0,12767
S.A. Lands	0,21483
Variance	772 306

TABLE 4.6

Security	Constrained First G-Component
Apex	0,00020
Clydesdale	0,49999
Tavistock	0,00000
Trans Natal	0,50000
Welgedacht	0,00233
Durban Deep	0,48069
ERPM	0,42931
Grootvlei	0,11507
Marievale	0,15129
S.A. Lands	0,22023
Variance	772 075

It should be noted that each of the above examples could easily be repeated with the same or different constraints in order to obtain the second G-Components and, if desired, even the third, fourth, etcetera, components.

Lastly, it must be mentioned that the above technique can be used if it is required to obtain a traditional first component (that is the weights not restricted to be non-negative) subject to the imposing of additional constraints. For example, if an unrestricted in sign first component is required subject to the condition that the total weight of each sector is equal, and that no security has an absolute weight of more than 0,5, the problem can be formulated as follows:

$$\text{Max} \quad x'Sx$$

$$\text{S.T.} \quad \sum_{i=1}^{10} x_i^2 = 1$$

$$\sum_{i=1}^{5} x_i - \sum_{j=6}^{10} x_j = 0$$

$$-0{,}5 \leq x_i \leq 0{,}5 \quad \text{for} \quad i = 1,\ldots,10$$

and solved using the Flexible Tolerance Method.

4.6 CONSTRUCTING SECTOR INDICES

In this section indices are constructed for two sectors of the Johannesburg Stock Exchange using the methods proposed in this chapter. The two sectors examined are the Banking Sector and the Chemicals Sector.

For the Banking Sector, those securities for which data was available over the entire period of 13 February 1970 to 20 February 1976, were chosen as the constituents of the index. This resulted in the six securities listed in Table E.3 of Appendix E being chosen. The covariance matrix of the prices was computed using the weekly data for the period 13 February 1970 to 29 December 1972. This was done so that the performance of the index could be examined both in the period on which the principal component was based, and in the subsequent periods. The first principal component obtained from the covariance matrix of the six banking shares is presented in Table 4.7 below.

TABLE 4.7

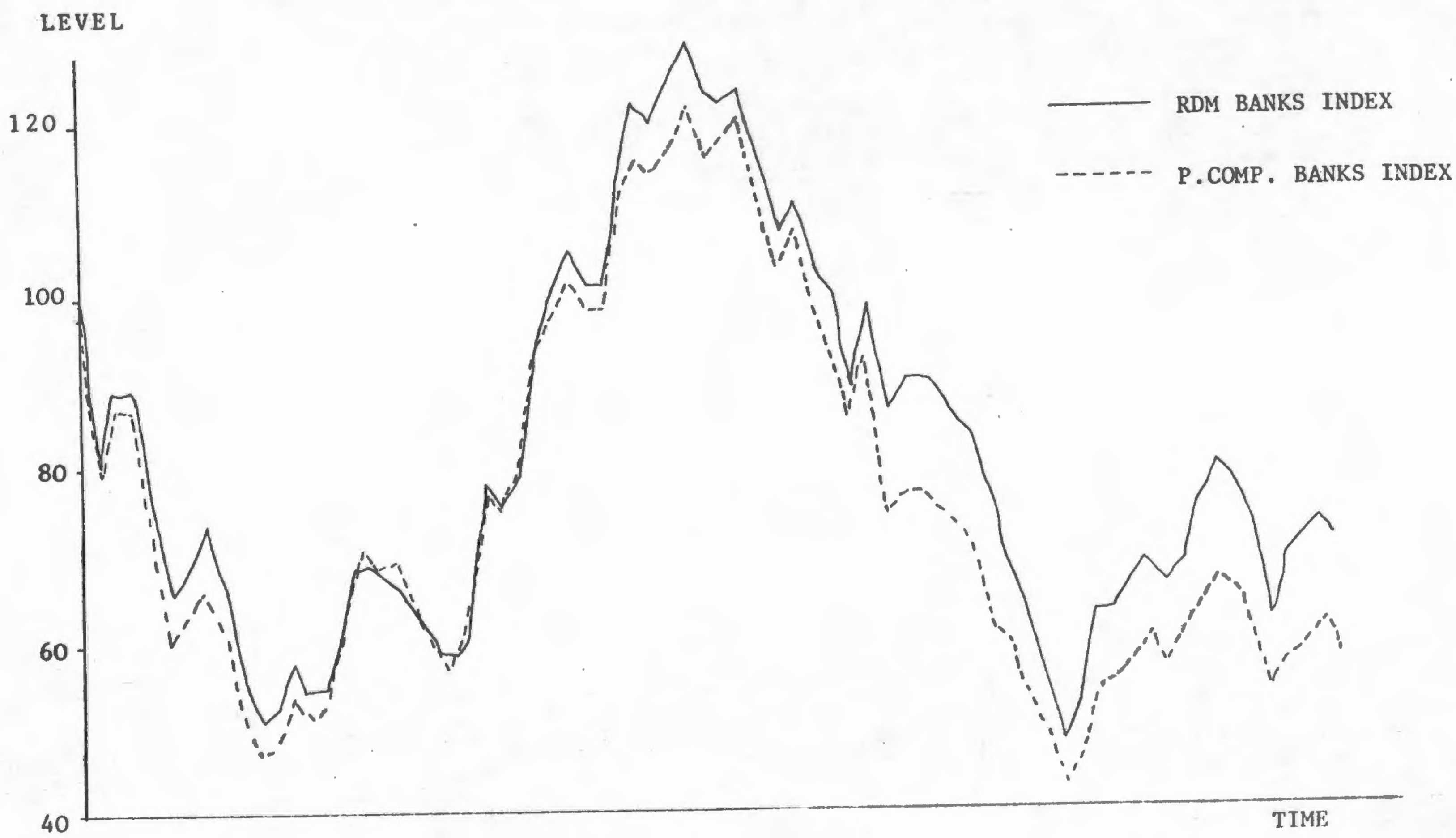
1st PRINCIPAL COMPONENT : BANKING SECTOR

Security	Weight
Bankorp	0,4016
Nedbank	0,4149
Santam	0,0969
Stanbic	0,5392
Trust Bank	0,2855
Volkskas	0,5338

As can be seen from the above table, all of the weights are positive and none are excessively large. Hence, it is unlikely that it would be desirable to impose any additional constraints on the weights, and hence the flexible tolerance method program would not be required. The graph of the index obtained using these weights is presented in Figure 4.2 below. This figure also presents the graph of the Rand Daily Mail Banking Sector Index (scaled to the same initial level as the principal component index). As can be seen, the behaviour of the two indices is very similar, and thus it can be argued that the principal component index is an acceptable index. It must be noted that it is not argued that the principal component index is a better index (although if the same securities are included, it should be a more volatile index), merely that it is an acceptable index in that it behaves as other stock market indices behave.

For the Chemicals Sector, the period for which data was available was 5 December 1969 to 20 February 1976, and the securities chosen as constituents for the index are listed in Table E.4 of Appendix E. The covariance matrix of the prices was computed using weekly data for the period 5 December 1969 to 29 December 1972. The first principal component obtained from this covariance matrix is listed in Table 4.8 below.

From Table 4.8 it can be seen that two of the weights used to compute the principal components index are negative. The graph of the index based on these weights is presented in Figure 4.3 below (plotted monthly). As can be seen the index "goes negative" from April 1971 until February 1973, after which it oscillates between positive and negative weights until August 1975. Clearly, this is a most unsatisfactory situation.



F I G U R E 4.2

TABLE 4.8

1st PRINCIPAL COMPONENT : CHEMICALS SECTOR

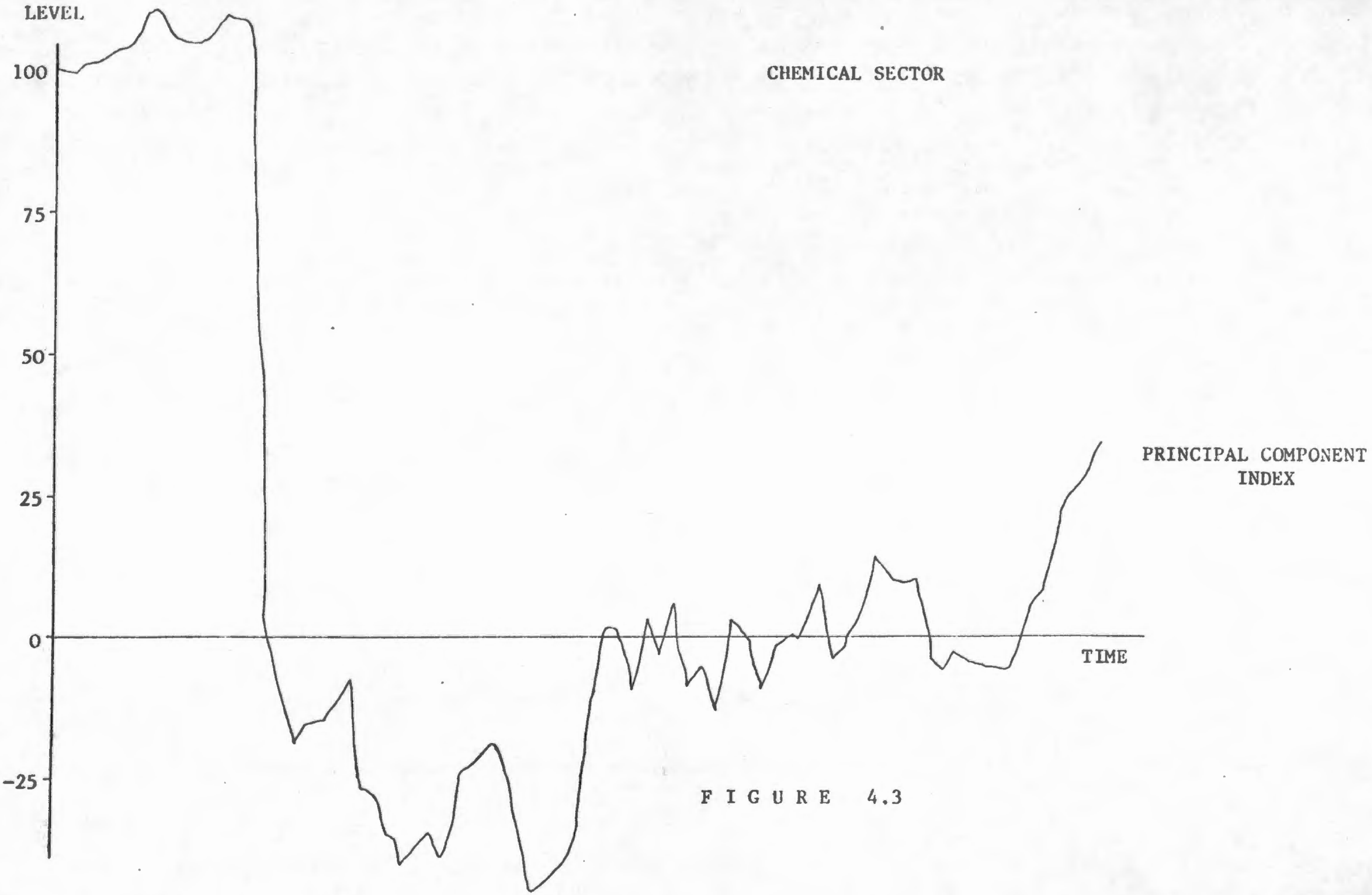
Security	Weight
AECI	-0,1459
Coal By-Products	0,9330
Fedmis	-0,3278
Sentrachem	0,0267

In this chapter, it has been recommended that if any of the weights of the first principal component are negative, then the flexible tolerance method should be used to restrict all the weights to nonnegative values. Doing this, the following weights (Table 4.9) were obtained.

TABLE 4.9

Security	Weight
AECI	0,0000
Coal By-Products	0,9994
Fedmis	0,0000
Sentrachem	0,0346

On examining Table 4.9 it is obvious that the proposed index consists, for all practical purposes, of a single security, Coal By-Products. This is certainly an unsatisfactory situation, especially as Coal By-Products is a relatively unstable security - it is liable to sudden rather violent price movements. Thus, when using the flexible


LEVEL
100
75
50
25
0
-25
CHEMICAL SECTOR
PRINCIPAL COMPONENT INDEX
TIME

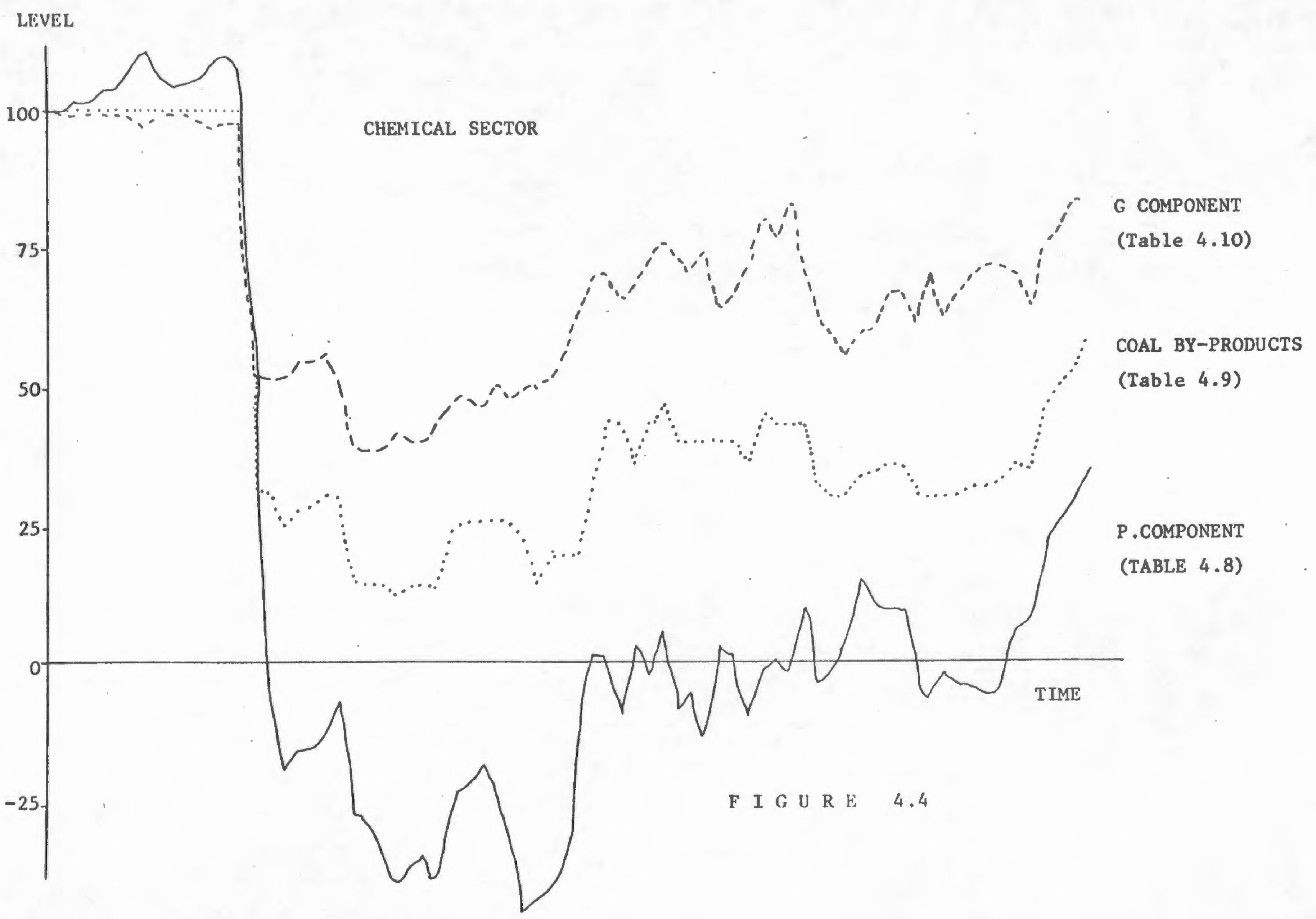
F I G U R E 4.3

tolerance method, it is suggested that an upper bound be placed on each security to be included in the index. Since only four securities are considered in this example, and since the sum of the squares of weights is constrained to be equal to one, it is proposed that the maximum weight allocated to any one security be limited to 0,7. Including these additional constraints ($X_i \leq 0,7$ for $i = 1,2,3,$ and 4) the problem was resolved, and the following weights (Table 4.10) obtained.

T A B L E 4.10

Security	Weights
AECI	0,1414
Coal By-Products	0,7000
Fedmis	0,0000
Sentrachem	0,7000

Sector indices were constructed for the period 5 December 1969 to 20 February 1976 using each of the three sets of weights obtained (Tables 4.8 - 4.10), and graphs of these indices are presented in Figure 4.4 below. As can be seen from the figure, both of the indices using the flexible tolerance method avoid the problem of the index "going negative." In addition, the restricted problem ($X_i \leq 0,7$) is less affected than the other indices by the sudden drop in Coal By-Products in April 1971.


LEVEL
100
75
50
25
0
-25
CHEMICAL SECTOR
G COMPONENT
(Table 4.10)
COAL BY-PRODUCTS
(Table 4.9)
P.COMPONENT
(TABLE 4.8)
TIME

F I G U R E 4.4

The results presented in this section indicate the method by which it is proposed that stock market indices based on principal components, should be constructed. Firstly, the first principal component should be found. If all the weights are nonnegative and any other necessary constraints are met, then these weights should be used to construct the index. If, however, the weights are not all nonnegative, or if some additional required constraints are not met, then the flexible tolerance method should be used, and will result in an index which is the most volatile subject to the imposed constraints.

4.7 CONCLUSIONS

In this chapter, the problem of principal component analysis, when it is desired to impose various additional constraints on the weights, has been examined. A method of solving such problems has been indicated, and various practical examples presented. Such a technique, it is felt, will have many applications in analyses where principal components are used, as the method gives great felxibility to the researcher. It must be emphasized that the technique is not limited to use in stock market problems. However, the latter do lend themselves very well to such applications, and this technique could prove most useful in this field of research.

From the studies presented in this chapter, it appears as if the solution method proposed (the Flexible Tolerance

Method) works well in practice, and although an iterative search method, it does provide answers very close to the optimal. But, it is possible that other methods, or in fact an exact solution method, may provide better solutions in some cases.

It should be noted that the examples presented in Section 4.4 indicated that there was only a difference of approximately 5% in the variance of the first P-Component compared with the first G-Component. From experience it appears as if this is approximately the order of the difference when only the nonnegative weighting restriction is imposed. In addition, it is obvious from mathematical or geometric considerations that the second G-Component will always have variance at least as large as the second P-Component, and thus the % variation explained by the first two P-Components will usually be close to the % variation explained by the first two G-Components.

However, when additional constraints are imposed (such as an upper bound on each weight), it is not easy to predict the exact effect on the G-Component. It is also not possible to assess the relative performance of the first two components, as this will be very dependent on the nature of the restrictions imposed, and whether in fact they are imposed on both components or only on the first component.

In conclusion, it might be said that the most important contribution of this chapter is the concept of reformulating the problem of principal components as a quadratic programming problem, and the indication that solutions to such a programming problem can be obtained, not using excessive computer time (all examples took less than five minutes).

CHAPTER FIVE

THE RANDOM WALK MODEL AND EXTENSIONS

5.1 INTRODUCTION

The major aim of most investment analysts is to determine at some point in time which securities are going to rise in the immediate future, which are going to decline, and which are going to remain at their current level. In order to do this, the analyst may employ one or several methods, ranging from a detailed fundamental analysis of the company as a whole (the quality of its management, the prospects of the industrial sector in which it operates, etcetera) to the use of technical trading rules such as those typified by the simplest of charting techniques.

There is almost universal agreement on the fact that detailed fundamental analysis is a very appropriate method of predicting the behaviour of a securities price. Unfortunately, such an analysis is extremely difficult to perform and it is debatable whether in fact,many of the variables which should be considered (such as government decisions, rises and falls in the prices of raw materials such as gold and oil, the psychological attitude of the public, etcetera) can be successfully forecast by an analyst. In short, it can be argued that a perfect fundamental analysis is impossible.

On the other hand, technical analysis (especially charting) with its fairly well defined "rules" of when to buy and when to sell can be easily performed by most investors. Moreover, books such as Markstein (1966), Cohen (1966) and Granville (1963) have tended to give the public the impression that there are vast profits to be easily made on the stock exchange. These two factors have contributed greatly to the enormous popularity of charting methods, and this popularity has led to research interest in these techniques. Initially, researchers attempted to show that certain patterns did recur and could be used to make substantive profits. When this failed, researchers turned to the opposite extreme and tried to show that charting could not be used successfully. In order to do this, they focussed their attention on the so called random walk model.

In this chapter, the random walk model is investigated and an extension which, it is felt, is more in keeping with the behaviour of the chartist, is proposed. In Section 2, the random walk model is defined and some of the more important implications discussed, while in Section 3 some important results, which have emerged from studies on the New York Stock Exchange, are considered. Analagous results for the Johannesburg Stock Exchange, which have already appeared in the literature, are discussed and the performance of South African securities are compared with those of other countries in terms of the random walk model. An extension of this model is proposed in Section 4, where results are presented for securities quoted on the Johannesburg Stock Exchange so as to ascertain the suitability of this extension. In Section 5, this extension is examined in greater detail and under different market conditions. Finally, Section 6 presents some conclusions which can be derived from the results presented in this section, and some thoughts on the position of the chartist are given.

5.2 THE RANDOM WALK MODEL

In this section the random walk model is defined, and a very brief discussion of some of the main implications of the model is given. The model has been widely discussed in the literature, and for a more comprehensive discussion the reader is referred to Fama (1965a) where the full implications of the model are discussed in great detail.

The random walk model has been defined in numerous slightly different forms in the literature. Probably the simplest and most popular form is:

$$P_t = P_{t-1} + \ell_t$$

where P_t is the price of the security at time t; and ℓ_t is the random error where it is assumed that

(i) $E(\ell_t) = 0$ and

(ii) ℓ_t and ℓ_s are independent for all $t \neq s$.

Although one of the conditions of the model as defined above is that ℓ_t and ℓ_s are independent for all $t \neq s$, most statistical tests of the model in the literature have concentrated on the correlation between ℓ_t and ℓ_s, $t \neq s$ (for example Fama (1965a)). Lack of correlation would imply independence if and only if the ℓ_t's were normally distributed, which is not necessarily true. However, if the errors are not normally distributed, they probably have stable distributions (cf. for example, Fama 1965a) and Affleck-Graves (1974)), and it is felt that the divergence from normality will not be great. Thus, in this chapter the observed correlation between ℓ_t and ℓ_s will be examined, and a lack of correlation will be assumed to show a lack of independence. In addition, it should be noted that lack of correlation, not lack of independence, is the important factor as far as the practical considerations of the random walk model are concerned.

As far as this, the practical point of view, is concerned, the importance of the random walk model stems from the fact that the model implies that future prices cannot be predicted from a knowledge of past prices and hence, that the charting of a stock market security cannot enable one to predict its future price. In short, since under the random walk hypothesis,

$$\ell_t = P_t - P_{t-1}$$

and $E(\ell_t) = 0$, the model implies that the next price change cannot be predicted from previous prices - its expected value is zero, and hence the best estimate of tomorrow's closing price is today's closing price. This implies that the return in the next period cannot be forecast (or estimated) and for this reason the variable that has usually been examined is

$$\ell_t = \log_e P_t - \log_e P_{t-1}.$$

This can be justified since the difference in the logarithms of price is the return with continuous compounding over the period considered, and the random walk model implies that this return cannot be predicted from previous prices. Thus, if the difference in logarithms is used, then acceptance of the random walk model is equivalent to acceptance of the hypothesis that $\beta = 0$ in the model

$$R_t = \beta R_{t-1} + U_t ,$$

where R_t equals the return with continuous compounding in the t^{th} period (that is, $R_t = \ell_t$);

and U_t is the random error.

The importance of this implication is stressed by Adam Smith (1967) who says:

> "If the random walk people are right, the Chartists are out of business and all the security analysts are in trouble."

Naturally, statements such as this have resulted in a vigorous debate in the literature over the applicability of the model. The two opponents can be, in very general terms, summed up as on the one hand, the academics (particularly some American researchers such as Fama, Mandelbrot, Fisher etc.) who propose that the random walk model holds, and who have presented numerous statistical results to substantiate their claims, and on the other hand, some practicing investment analysts who argue that charting works for them, but who keep their profitable patterns and their net figures of profit and loss to themselves for obvious reasons. At present, a stalemate appears to have developed with neither side prepared to concede defeat or to provide new results.

5.3 EMPIRICAL RESULTS IN THE LITERATURE

As mentioned previously, numerous authors have presented statistical results in support of the random walk model in the literature of the last twenty years. Among the most important are the studies of Mandelbrot (1963), Osborne (1959), Cootner (1962), Granger and Morgenstern (1963), Young (1971) and Schwartz and Whitcomb (1977), but perhaps the most comprehensive of these was that of Fama (1965a) who examined the model in great detail for the 30 Dow Jones Industrial Average stocks. He computed the empirical autocorrelation between ℓ_t $(\log_e P_t - \log_e P_{t-1})$ and ℓ_{t-s} for a number of values of s as well as for different periods (that is, using daily data, data every 4 days, etcetera) and used these to test for independence.

In addition he tried an alternative approach. Runs tests were performed for each of the three different types of runs which occur in stock prices (positive, zero, and negative). The results of both of the above mentioned tests showed that the random walk model held (over the time period examined) for almost all securities, and it was argued that, even for those securities for which the autocorrelation was statistically significantly different from zero, this autocorrelation was so small (usually less than 0,15) as to be almost useless for prediction purposes.

Similar studies based on Fama's approach have been performed on most of the major European exchanges, and were presented by Solnik (1973). Solnik's findings were similar to Fama's (although he used an observation period of one week instead of four days), and he concluded that securities on the major European exchanges had a similar behaviour to those quoted on the New York Stock Exchange, and that the random walk model seemed to be an appropriate model.

With the considerable interest in the applicability of the random walk model on the New York Stock Exchange, and the major European exchanges, it was only natural that the behaviour of South African securities should be examined. A study based on exactly the same criteria as that of Fama (1965a) and Solnik (1973) was performed by Affleck-Graves (1974) for a sample of fifty securities quoted on the Johannesburg Stock Exchange using a differencing interval (or observation period) of one week. The results obtained were summarized in Affleck-Graves and Money (1975), and showed that there appeared to be very little evidence of nonzero autocorrelation. Of the seven (out of 50) securities which had statistically significant autocorrelation, only three had an empirical autocorrelation of more than 0,2. Thus, it was concluded that the results obtained for the Johannesburg Stock Exchange were similar to those of the New York Stock Exchange and the major European exchanges, and that they demonstrated considerable support for the random walk model.

In order to compare the performance of the different exchanges, the following table is reproduced from Affleck-Graves and Money (1975).

TABLE 5.1

Author	Exchange	Differencing Interval	Average Autocorrelation (Lag of 1 Period)	S.D.
Fama	New York	4 days	-0,038	0,058
Solnik	London	1 week	-0,055	0,060
Solnik	Paris	1 week	-0,049	0,060
Solnik	Frankfurt	1 week	0,056	0,060
Graves & Money	Johannesburg	1 week	-0,018	0,061
Fama	New York	9 days	-0,057	0,086
Solnik	London	2 weeks	0,005	0,090
Graves & Money	Johannesburg	2 weeks	0,000	0,087
Fama	New York	16 days	-0,009	0,116
Solnik	London	4 weeks	0,020	0,120
Graves & Money	Johannesburg	3 weeks	0,037	0,107

5.4 AN EXTENSION OF THE RANDOM WALK MODEL

From the discussion and results presented in the previous section it is clear that considerable statistical proof exists in support of the random walk model. As most statistical tests presented in the literature have been adequately performed, one cannot argue against the results obtained. Thus, as the random walk model has been proved to be true for almost all securities on most of the world's major exchanges, it is argued that this should be accepted as fact by all investment analysts.

The question which now arises is: What of the Chartists? How can their claims of success be reconciled with the random walk model? The answers to these questions possibly lie in the fact that the random walk model examines only the past price of the security itself. In practice, the successful chartist usually contrasts the behaviour of numerous factors and does not view the stock price in isolation. Typically, an overall market index, an index of the relevant sector of the market, and the volume of securities traded would be plotted in addition to the price of the individual security. In deciding to buy or sell the security, the chartist would almost certainly consider all of these factors together with others, such as the overall economic climate, the G.N.P., the Gold Price, etcetera. Thus, it can be argued that while the random walk model is valid, it does not preclude the success of a chartist.

Therefore, in order to examine the chartists position more realistically, the following model is examined in this section.

$$R_t = \alpha + \beta_1 R_{t-1} + \beta_2 S_{t-1} + \beta_3 M_{t-1} + \beta_4 V_{t-1} + e_t$$

where

$R_t = \log_e P_t - \log_e P_{t-1}$ is the return with continuous compounding of the security in period t;

$S_{t-1} = \log_e I_{t-1} - \log_e I_{t-2}$ is the return with continuous compounding on the sector index (I) in period t-1;

$M_{t-1} = \log_e J_{t-1} - \log_e J_{t-2}$ is the return with continuous compounding on the market index (J) in period t-1;

V_{t-1} is the number of shares (in millions) of the security traded in the $(t-1)^{th}$ period;

α, β_1, β_2, β_3, β_4, are the regression parameters which are to be estimated; and

e_t is the random error where it is assumed that

(i) $E(e_t) = 0$, and

(ii) e_t and e_s are uncorrelated for all $t \neq s$.

That is, it is hypothesized that the return on the security in the i^{th} period is a function of the return on the security, the return on the sector index, the return on the market index, and the volume traded, in the previous period.

In order to contrast the behaviour of the above model with the results obtained in Affleck-Graves (1974), the same sample of 50 securities was used. Unfortunately, three of these securities have been delisted, and data on a fourth was unavailable. Hence, the results presented hereunder are for the remaining 46 securities only. A list of these securities is given in Appendix F together with the number of weeks of data available for each security. The data available consisted of the weekly closing price of each security and the weekly closing level of the relevant indices for the period 22/3/68 to 20/2/76*. The sector indices used in the study were the Rand Daily Mail (RDM) indices, and the market index used was the RDM '100' industrial average. Although the latter is an "industrials only" index and does not include any of the Mining or Financial securities, it was used for the market index since no general index of the Johannesburg Stock Exchange was considered satisfactory, and since the RDM '100' is probably the most widely quoted of all indices on this exchange.

For each of the 36 securities under consideration, a multiple regression analysis was performed and estimates of α, β_1, β_2, β_3, and β_4 in the above model were obtained. In addition, an estimate of the correlation between R_t and

---

* Data was not available for all securities and indices from 22/3/68 and the number of weeks of data available for each case is given in Appendix F. The data for all securities ended on the same date, namely 20/2/76.

R_{t-1} (which is in fact the correlation between e_t and e_{t-1} in the original random walk model, and which is the statistic traditionally used to examine the suitability of the random walk model) was obtained. Also, the multiple correlation coefficient (the correlation between the observed R_t and the estimated R_t obtained using the above model) was estimated and gives an indication of the suitability of the estimated regression equation. These results are presented in Table 5.2 below together with the results of stepwise regression analyses. The second last column of Table 5.2 (marked "STEPWISE") lists those variables which were chosen to be included in the regression equation using the Backward Elimination Procedure, while the final column gives the multiple correlation coefficient of the stepwise regression equation.

It can be seen from Table 5.2 (the first column - marked r) that 39 of the 46 securities can be classified as obeying the random walk model. That is, only 7 out of the 46 securities examined showed a significant disagreement with the random walk model, which is the same result as obtained in Affleck-Graves and Money (1975). This is slightly more rejections than would be expected by chance, but most of those securities which indicated disagreement, still had correlations of less than 0,2, and thus are not felt to contradict the random walk model too severely. For a more detailed discussion of this result the reader is referred to either Affleck-Graves (1974) or Affleck-Graves and Money (1975).

TABLE 5.2

SHARE	r	R	$\hat{\alpha}$	$\hat{\beta}_1$ (R_{t-1})	$\hat{\beta}_2$ (S_{t-1})	$\hat{\beta}_3$ (M_{t-1})	$\hat{\beta}_4$ (U_{t-1})	STEPWISE	R_s
Vaal Reefs	0,011	0,1204	-0,0033	-0,0560	0,0489	0,0714	0,4158	β_4	0,1125
Zandpan	0,069	0,1638*	-0,0050	-0,0533	0,1752	-0,0258	0,1632	β_4	0,1291
Harmony	-0,062	0,2202*	-0,0062	-0,2060	0,1250	0,1181	0,2903	$\beta_1;\beta_4$	0,2017
Pres.Brand	-0,054	0,1196	-0,0017	-0,2066	0,1837	0,0635	0,3339	all excl.	-
Kloof	-0,092	0,1175	0,0015	-0,1949	0,1657	0,0004	-0,0239	all excl.	-
Wes.Drie	0,007	0,0429	0,0017	0,0366	-0,0436	0,0704	0,0545	all excl.	-
Messina	-0,034	0,1880*	-0,0027	-0,1436	0,0424	0,4040	0,0853	$\beta_1;\beta_3$	0,1836
Union Tin	-0,238*	0,2799*	0,0019	-0,2567	0,2326	0,3860	-0,3787	$\beta_1;\beta_3$	0,2649
Cons.Murch	-0,041	0,2315*	0,0001	-0,5128	0,5140	0,4594	0,1101	$\beta_1;\beta_2;\beta_3$	0,2305
Anglovaal	0,096	0,3914*	-0,0031	-0,1041	0,3695	0,5125	2,0170	$\beta_2;\beta_3$	0,3736
Charter	-0,017	0,1608*	-0,0007	-0,0588	-0,0402	0,4089	-0,0149	β_3	0,1436
Johnnies	0,078	0,2952*	-0,0011	-0,1814	0,2730	0,5844	0,1292	$\beta_1;\beta_2;\beta_3$	0,2951
Rand Sel.	0,036	0,2578*	-0,0034	-0,2219	0,2045	0,4148	0,1582	$\beta_1;\beta_2;\beta_3$	0,2561
UCI	0,022	0,2147*	-0,0010	-0,1715	0,3171	0,2137	0,1088	$\beta_1;\beta_2$	0,1952
Amic	0,092	0,2341*	-0,0024	-0,0672	0,0068	0,4147	0,2183	β_3	0,2238
Barlows	-0,005	0,1721*	-0,0037	-0,1786	0,0615	0,4161	0,0345	$\beta_1;\beta_3$	0,1627
Lonrho	0,061	0,1221	0,0019	0,0448	-0,0216	0,2736	-0,2872	β_3	0,1050
Fugit	0,021	0,2615*	-0,0024	-0,3655	0,4983	0,3203	0,0409	$\beta_1;\beta_2;\beta_3$	0,2583
Guardian	-0,132*	0,2487*	-0,0042	-0,2619	0,0330	0,5419	0,0024	$\beta_1;\beta_3$	0,2481
Glen Anil	0,121*	0,2360*	-0,0087	-0,0435	0,0610	0,5565	0,2807	$\beta_3;\beta_4$	0,2341
Rand Mines	0,059	0,1951*	-0,0066	-0,0483	-0,0164	0,5741	0,2217	β_3	0,1737
Nedbank	0,044	0,2249*	-0,0010	-0,1362	0,1480	0,4488	0,0420	β_3	0,1911
Trust Bank	-0,051	0,2840*	-0,0058	-0,3123	0,4615	0,2243	0,0162	$\beta_1;\beta_2$	0,2747
Volkskas	0,031	0,1625	-0,0006	-0,1215	0,1949	0,1581	-0,0424	β_2	0,1242
Oudemeester	-0,184*	0,3047*	-0,0054	-0,2982	-0,0824	0,6625	0,0945	$\beta_1;\beta_3;\beta_4$	0,3029
Anglo Alpha	-0,085	0,1739*	-0,0038	-0,1643	0,1247	0,1020	0,1738	$\beta_1;\beta_2$	0,1429
Blue Circle	-0,033	0,2016*	-0,0025	-0,1230	0,2265	0,2647	0,1521	β_3	0,1659
AECI	-0,060	0,2312*	-0,0010	-0,2247	0,0248	0,3935	0,0088	$\beta_1;\beta_3$	0,2310
Kaap Kunene	-0,089	0,1541*	-0,0064	-0,1867	0,1353	0,0847	0,2272	$\beta_1;\beta_4$	0,1331
Tiger Oats	0,045	0,2303*	0,0019	-0,1712	0,1317	0,3604	-0,1085	$\beta_1;\beta_3$	0,2125

SHARE	r	R	$\hat{\alpha}$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	STEPWISE	R_s
Searles	-0,018	0,2000*	0,0009	-0,0840	-0,0076	0,6177	0,6395	β_3	0,1836
Afcol	-0,024	0,2060*	-0,0071	-0,2057	0,1176	0,4642	0,1862	$\beta_1;\beta_3$	0,1771
Russel	0,049	0,1934*	0,0003	-0,1227	0,0904	0,5498	-0,0014	β_3	0,1738
Hub.Davies	-0,110*	0,2525*	-0,0053	-0,2159	0,0651	0,5491	0,2115	$\beta_1;\beta_3$	0,2470
Stew.Lloyds	0,007	0,2234*	0,0026	-0,0851	0,0503	0,4867	-0,1276	β_3	0,2070
Toyota	0,087	0,2700*	-0,0066	-0,0809	0,0247	0,6427	0,8359	$\beta_3;\beta_4$	0,2624
Sappi	-0,116*	0,1321	-0,0005	-0,1198	-0,0547	0,1523	0,0008	β_1	0,1157
Skye	-0,082	0,1681*	-0,0104	-0,1389	-0,0785	0,6228	0,3893	$\beta_1;\beta_3$	0,1533
Twins	-0,062	0,2217*	-0,0071	-0,1603	0,0429	0,6102	0,2311	$\beta_1;\beta_3$	0,2007
Saan	0,006	0,1907*	-0,0004	-0,0053	-0,0967	0,4676	-0,2401	β_3	0,1822
Vaderland	-0,183*	0,2527*	-0,0041	-0,2194	0,0682	0,3918	0,1815	$\beta_1;\beta_3$	0,2446
Bonmore	0,027	0,2633*	-0,0014	-0,0909	0,0383	0,5821	0,0328	β_3	0,2499
Pick 'n Pay	0,020	0,2802*	0,0011	-0,1560	-0,1303	0,8281	0,2397	$\beta_1;\beta_3$	0,2716
Tongaat	0,001	0,2027*	0,0089	-0,0702	0,0607	0,2722	-0,3195	$\beta_3;\beta_4$	0,1955
Romatex	-0,034	0,1526*	-0,0022	-0,0743	-0,0457	0,4776	-0,0025	β_3	0,1269
Utico	0,048	0,2510*	-0,0018	-0,0451	0,0611	0,3195	-0,1430	β_3	0,2421

* in the r column indicates rejection of the null hypothesis

$$H_0 : \rho = 0$$

* in the R column indicates rejection of the null hypothesis

$$H_0 : P = 0$$

Now, the values presented in the first column (marked r) of Table 5.2, are, in fact, the multiple correlation coefficients of the model

$$R_t = \alpha + \beta R_{t-1} + e_t ,$$

and hence can be compared with R in the second column, which gives the multiple correlation coefficient when fitting the extended model proposed at the beginning of this section.

On examining the second column of Table 5.2 it can be seen that in 39 out of the 46 securities examined, the null hypothesis of no relationship between the estimated and observed return was rejected, compared with 7 when the random walk model was considered. This would appear to indicate that considerably more information is available to the investment analyst if he considers a market index, a sector index and the volume traded, in addition to the price. This is bourne out by the average of the absolute values of r and R, which are 0,062 and 0,2109 respectively. Whether this is sufficient information to confirm that charting techniques can work remains debatable, and will be discussed in more detail at the end of this section.

While discussing Table 5.2, it is of interest to examine various other results which emerge from an examination of the table. Firstly, it is interesting to examine the signs of the regression estimates when using the multiple regression procedure. These results are summarized in Table 5.3 below:

TABLE 5.3

Coefficient	No. of Estimates with -ve signs (out of 46)
$\hat{\alpha}$	35
$\hat{\beta}_1$	44
$\hat{\beta}_2$	11
$\hat{\beta}_3$	1
$\hat{\beta}_4$	12

From the above table it can be seen that the coefficient of the return on the security $(\hat{\beta}_1)$ was negative for almost all securities, while the coefficients of the sector index $(\hat{\beta}_2)$, the market index $(\hat{\beta}_3)$, and the volume traded $(\hat{\beta}_4)$ were usually positive. The generally negative estimate of β_1, is to be expected since the empirical results of the tests on the random walk model indicate that any autocorrelation is likely to be negative, and it has been argued, in the literature, that this occurs because the stock returns are generated by a random walk process with a reflecting barrier (Niederhoffer and Osborne (1966)), and also because of investors preference (on the New York Stock Exchange) for "round eights" (Osborne (1962).

A second interesting result which emerges from Table 5.2 concerns the stepwise procedures and, in particular, the variables chosen to be in the stepwise regression equations. These results are summarized in Table 5.4 below.

TABLE 5.4

Variable	No. of times chosen by Stepwise Procedure
R_{t-1}	23
S_{t-1}	9
M_{t-1}	34
V_{t-1}	8

Table 5.4 indicates that R_{t-1} and M_{t-1} are probably the more important variables under consideration. That is, the return on the security and the return on the market in the previous period appear to be important, while the return on the sector index and the volume traded are less important. The relative unimportance of the sector index is interesting and is probably due to the fact that the return on the security and the return on the sector index are fairly highly correlated (as much as 0,75 in some cases), and hence the stepwise procedure omits one of them (since it contributes little additional information), which is usually the sector index. It should be noted that since the correlations between R_{t-1} and S_{t-1} are sometimes fairly high, an alternative form of regression such as Ridge Regression should be considered. However, in the case at hand, the main point of interest is whether a fit exists or not. This is measured by R^2, which is maximized when using ordinary least squares, and which will, in fact, be lowered when using Ridge Regression. Thus, in this case, Ridge estimates were not obtained, but it is stressed that if the particular β

estimates are the point of interest then such an analysis should be performed.

Finally, it is of interest to examine the relative behaviour of different sections of the market. The securities analyzed in this chapter, can be divided into four main groups; the Gold Shares (1-6); the Mineral Shares (7-9); the Financial Shares (10-24); and the Industrial Shares (25-46). The relative behaviour of these four groups is summarized in Table 5.5 below:

TABLE 5.5

SECTION	No. of SECURITIES	$\overline{\lvert r \rvert}$	$\bar{R}$	No. of SIGN. R's	PROPORTION OF TIMES CHOSEN BY STEPWISE PROCEDURE			
					β_1	β_2	β_3	β_4
GOLD	6	0,049	0,1307	2	0,17	0,00	0,00	0,50
MINERALS	3	0,104	0,2331	3	1,00	0,33	1,00	0,00
FINANCE	15	0,058	0,2307	13	0,47	0,47	0,80	0,07
INDUSTRIALS	22	0,062	0,2162	21	0,55	0,05	0,86	0,18

From the above table it can be seen that the Gold shares do not seem to be suitably represented by the model proposed at the beginning of this section. However, the remaining shares appear to have significant multiple correlation coefficients in general (that is, non zero multiple correlation coefficients), which appear to be of approximately the same order for each of the three groups. Also, each of the groups (including the Gold group) appear to have approximately the same average r (correlation between R_t and R_{t-1}) which indicates agreement as far as the random walk model is concerned. (It is argued that this is true even for the minerals group since, although the average is somewhat higher than the averages for the other groups, it is due to one security only, namely Union Tin, while the remaining two securities in the group have r's consistent with the other groups.)

As regards the stepwise procedures, Table 5.5 tends to confirm the results obtained from Table 5.3 and Table 5.4, namely that R_{t-1} and M_{t-1} appear to be the more important variables, especially if the Gold shares are ignored. It should be noted that the lack of conformity of the Gold Section is probably due to the fact that the RDM '100' is not a suitable market index to use when analyzing the Gold shares.

There are, it is felt, three main conclusions which can be drawn from the discussion and results presented in this section.

Firstly, it would appear that of the four variables considered the most important factor, as far as prediction purposes are concerned, appears to be the general market factor. The next most important factor is the individual security factor, followed by the sector factor. Finally, the volume of shares traded in the particular security does not appear to be a very relevant factor.

Secondly, the random walk model of stock market prices does appear to hold in practice. Certainly the evidence presented in the literature is confirmed by the results presented in this section, and is overwhelmingly in favour of the model. Thus, it is argued that the validity of the random walk model should be accepted by all investors. However, it can be argued from the chartists' point of view that, while the random walk model is in fact valid, it does not affect charting for two main reasons:

(i) the chartist does not merely examine the price history of a security in isolation; and

(ii) the patterns which the chartist uses as "buy" or "sell" signals are mathematically complex functions of the previous prices and, while the random walk model does state that no linear relationship exists between tomorrow's price and previous prices, it does not preclude the existence of a nonlinear relationship.

As far as (ii) is concerned, it would be extremely difficult for the proponents of the random walk model to prove that no relationship, linear or nonlinear, existed between successive returns and hence, it must be concluded that this rather negative argument can never be disproved.

Argument (i), which can be used by the chartist to explain why the random walk model does not affect him, is the one which was examined in this section, namely that a security is not viewed in isolation. In order to investigate this assertion, an extension of the random walk model was proposed,which hypothesized that the return on a security is a linear function of the return in the previous period on (i) the security itself, (ii) an index of the sector in which the security is quoted, (iii) an overall market index, and (iv) the volume of shares traded in the particular security in the previous period.

The results of this investigation provide the third main conclusion of this section, namely that the results obtained indicate that for almost all securities, a significant regression equation was found (that is, the hypothesis: $H_0 : P = O$, was rejected). Thus, it can be argued that a statistically significant linear relationship exists between the return on a security in the present period, and the return on the security, the sector index, the market index, and the volume traded in the previous period.

However, it must be stressed that in the previous sentence, the emphasis must be on the phrase "statistically significant." The fact remains that, in general, the observed multiple correlation coefficients were very small, with the average being only 0,21. This implies that on average, under 5% of the variation in R_t could be explained by means of a linear relationship of R_{t-1}, S_{t-1}, M_{t-1}, and V_{t-1}. Moreover, there was not a single case in which the observed multiple correlation coefficient was greater than 0,4 (that is, approximately 16% of the variation in R_t explained). Clearly then, it is extremely doubtful whether such an equation could be satisfactorily used for prediction purposes, and it is not the contention of this chapter that it should be used for such a purpose.

Nevertheless, it does indicate that some relationship, albeit slight, does exist and that it is too large and persistent to be due to chance alone. When considered with the chartists' claim that the relationships they use are more complicated than simple linear relationships, it can be argued that the results presented in this section indicate that there might well be foundation in the claim that superior profits can be made using charting techniques.

5.5 THE BEHAVIOUR OF THE RANDOM WALK MODEL AND ITS EXTENSION IN DIFFERENT TYPES OF MARKETS.

The final investigation of the random walk model to be discussed in this chapter concerns the behaviour of this model, and the extension proposed in the previous section in different types of markets (that is, bull and bear markets). Since it was shown in the previous section that the six Gold securities behaved somewhat differently to the remaining forty securities, these two groups were considered separately.

First, in order to establish the different types of markets which existed for the Gold shares between 22/3/68 and 20/2/76, the "RDM Klerksdorp Gold Sector Index" was graphed (Figure 5.1). From this graph it was decided to divide the period into three main subperiods as follows:

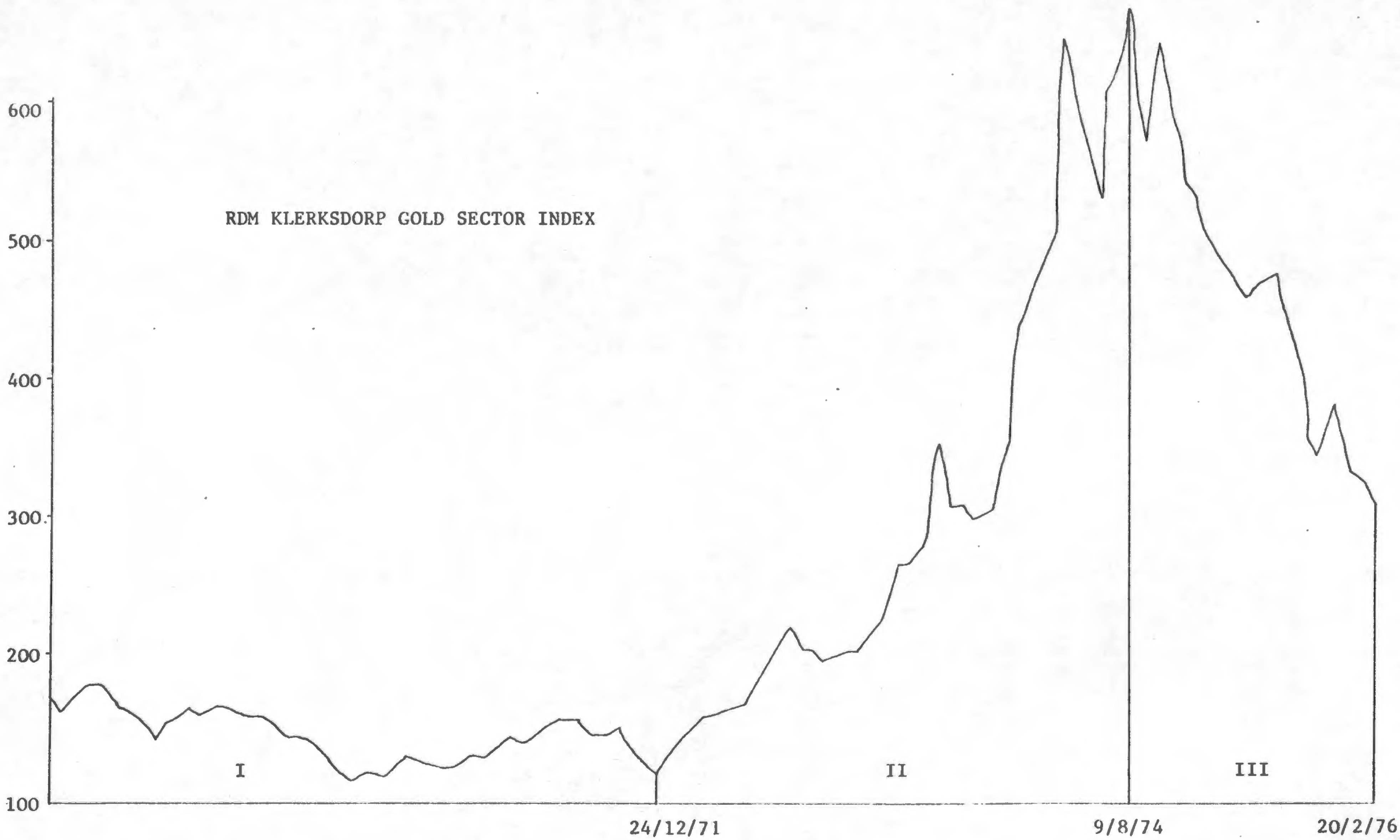
22/3/68 to 24/12/71	Dormant Period
24/12/71 to 9/8/74	Bull Market
9/8/74 to 20/2/76	Bear Market

An analysis similar to that discussed in the previous section was performed for each of the six securities, and for each subperiod, and the results are presented in Table 5.6 below.

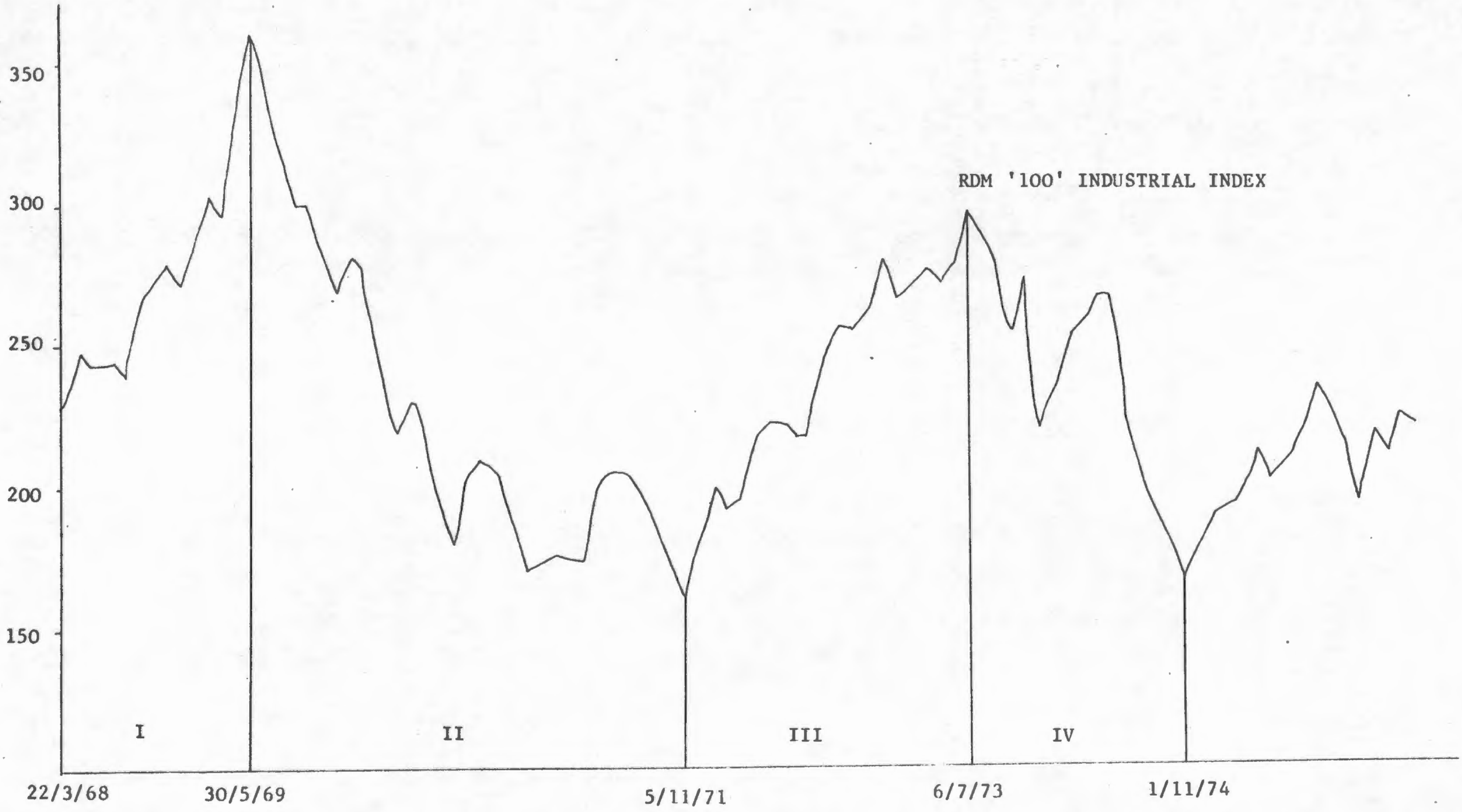
Also, to examine the remaining securities, the "RDM 100 Industrials Index" was graphed (Figure 5.2). From this graph it was decided to divide the period into four main subperiods:

22/3/68 to 30/5/69	First Bull Market
30/5/69 to 5/11/71	First Bear Market
5/11/71 to 6/7/73	Second Bull Market
6/7/73 to 1/11/74	Second Bear Market

The data from 8/11/74 to 20/6/76 was ignored for this section of the study. The results for these securities are also presented in Table 5.6 below:


RDM KLERKSDORP GOLD SECTOR INDEX
600
500
400
300
200
100
I
II
III
24/12/71
9/8/74
20/2/76

F I G U R E 5.1



RDM '100' INDUSTRIAL INDEX
350
300
250
200
150
I
II
III
IV
22/3/68
30/5/69
5/11/71
6/7/73
1/11/74

F I G U R E 5.2

TABLE 5.6

SHARE	PERIOD 1		PERIOD 2		PERIOD 3		PERIOD 4	
	r	R	r	R	r	R	r	R
Vaal Reef	-0,057	0,1443	-0,024	0,1305	0,007	0,2202		
Zandpan	-0,022	0,2163	0,088	0,1474	-0,039	0,1939		
Harmony	-0,146*	0,2216*	-0,085	0,1657	-0,140	0,2096		
Pres.Brand	-0,197*	0,3029*	-0,037	0,1002	-0,036	0,1859		
Kloof	-0,151*	0,1590	-0,124	0,2282	-0,081	0,0853		
Wes.Drie	0,064	0,1621	-0,112	0,1352	-0,011	0,1654		
Messina	-0,007	0,2324	-0,021	0,1080	0,058	0,2072	-0,161	0,3820*
Union Tin	-0,235	0,3332	-0,081	0,1816	-0,319*	0,3979*	-0,155	0,3127
Cons.Murch	-0,145	0,2222	-0,007	0,2818*	-0,039	0,3008	-0,235	0,3664
Anglovaal	-0,012	0,3628	0,135	0,3750*	0,106	0,3678*	0,035	0,5718*
Charter	-0,090	0,3422	-0,070	0,2109	-0,034	0,2991	0,002	0,2295
Johnnies	0,071	0,4050*	0,146	0,2658	-0,232*	0,3183	0,226	0,4762*
Rand Sel.	-0,048	0,1964	-0,032	0,3816*	0,028	0,2661	0,175	0,3189
UC Inv.	-0,150	0,2114	-0,017	0,3214*	0,072	0,1447	0,153	0,4167*
Amic	-0,119	0,3238	0,176	0,2161	0,114	0,3625*	0,132	0,2951
Barlows	-0,058	0,0610	-0,102	0,2262	-0,028	0,2008	0,177	0,2923
Lonrho	-	-	-0,058	0,1588	0,057	0,2039	0,104	0,2028
Fugit	0,017	0,3439	-0,109	0,2312	0,168	0,2888	0,007	0,4603*
Guardian	-	-	-0,235*	0,3225*	-0,103	0,1666	0,011	0,3915*
Glen Anil	-	-	0,056	0,2368	-0,007	0,4847*	-0,043	0,2835
R.M. Props.	-0,045	0,3039	0,047	0,2485	-0,008	0,2344	0,323	0,3841*
Nedbank	-	-	0,111	0,3477*	-0,140	0,2865	-0,065	0,2830
Trust Bank	-	-	-0,063	0,3226*	0,079	0,3042	-0,192	0,2686
Volkskas	-	-	-0,002	0,2818	0,155	0,1912	-0,129	0,3207
Oudemester	-0,236	0,4179*	-0,041	0,1787	0,073	0,2796	-0,341*	0,4886*
Anglo Alpha	-0,173	0,3953*	-0,032	0,1385	-0,145	0,3814*	-0,087	0,1879
Blue Circle	0,041	0,0731	-0,053	0,4653*	-0,039	0,1736	0,081	0,3828*
AECI	-0,284*	0,3884*	0,014	0,2524	-0,104	0,3074	0,000	0,3787*
Kaap Kunene	-0,280*	0,3104	-0,077	0,2204	-0,208	0,2637	-0,059	0,3261
Tiger Oats	-0,193	0,3552	0,026	0,2303	-0,134	0,1623	0,203	0,2612
Searles	-0,189	0,2680	0,168	0,3329*	-0,052	0,1109	0,080	0,4160*
Afcol	-0,279	0,4294*	0,060	0,2844*	-0,137	0,2202	-0,080	0,3241
Russel	-0,276	0,3833	0,153	0,2150	0,060	0,2344	-0,052	0,2855
Hub.Davies	-0,419*	0,5089*	-0,089	0,1642	-0,312*	0,3313*	0,109	0,2947
Stew.Lloyds	-0,247	0,2831	0,055	0,2709	0,228*	0,3523*	0,060	0,3592
Toyota	-0,155	0,1648	0,135	0,3032*	0,111	0,1825	0,102	0,4159*
Sappi	-0,273*	0,3093	-0,057	0,0847	-0,076	0,0968	-0,170	0,2322
Skye	-	-	-0,101	0,2292	-0,186	0,2550	0,140	0,2881
Twins	-	-	-0,116	0,2615	-0,134	0,3045	-0,076	0,3088
Saan	0,071	0,1782	-0,080	0,1816	-0,056	0,1925	0,120	0,2868
Vaderland	-0,052	0,1667	-0,109	0,2387	-0,139	0,2468	-0,377*	0,4998*
Bonmore	-0,115	0,1728	-0,021	0,2495	0,098	0,1950	0,118	0,4353*
Pick 'n Pay	-	-	-0,085	0,3265*	0,006	0,4886*	0,195	0,3402
Tongaat	-0,081	0,3044	0,069	0,2557	0,056	0,3139	-0,093	0,1697
Romatex	-0,084	0,2961	-0,029	0,1995	-0,114	0,1857	0,050	0,3806*
Utico	-0,039	0,1767	0,034	0,3665*	0,188	0,3062	0,097	0,5702*

(Note: * indicates rejection of the null hypothesis - see Table 5.2)

In order to examine the behaviour of the Gold securities more closely, the data presented in Table 5.6 is summarized by:

(i) the number of securities in the group for which data was available for the required period (the column marked N);

(ii) the average for all N securities of the absolute value of the correlation (r) between R_t and R_{t-1} for each period (the column marked $\overline{|r|}$);

(iii) the average of the multiple correlation coefficients (R) obtained when fitting the extended model proposed in the previous section (the column marked ($\overline{R}$);

(iv) the number of positive correlations (r) observed (the column marked +ve r);

(v) the number of significant correlations - i.e. the number of correlations which did not fall within 2 standard deviations of zero (the column marked SIG r); and

(vi) the number of significant multiple correlation coefficients observed - using the F ratio (the column marked SIG R).

These results are presented in Table 5.7 below.

TABLE 5.7

PERIOD	N	$\overline{\lvert r \rvert}$	$\bar{R}$	+ve r	SIG r	SIG R
1	6	0,106	0,2010	1	3	2
2	6	0,078	0,1512	1	0	0
3	6	0,052	0,1767	1	0	0
TOTAL	18	0,079	0,1763	3	3	2

The results presented in Table 5.7 above confirm the results of the previous section, namely that for the Gold securities, neither the correlation between successive returns (r) nor the multiple correlation coefficient from the extended model (R) are very significant, that the random walk model appears to be a satisfactory model for these securities, and that the extended model does not appear to be more suitable than the random walk model (in fact there are fewer significant multiple correlation coefficients than significant correlations).

One point which merits further attention is that the only period in which disagreement with the random walk model was observed was the first period, that is the "Dormant Period." This is a somewhat strange result since it indicates that successive returns are not independent (i.e. are dependent) in the dormant period, and yet are independent in both the bull and the bear markets. This is probably due to the "steplike" progress of the market mentioned in the previous section, and could be considered as a fairly strong argument in favour of those who contend that the random walk model does not really summarize what is seen by the chartist.

In order to examine the behaviour of the remaining 40 securities, the following tables (Tables 5.8 - 5.11) were also produced from Table 5.6. The notation is the same as for Table 5.7.

TABLE 5.8 : ALL 40 SECURITIES

PERIOD	N	$\overline{\lvert r \rvert}$	$\overline{R}$	+ve r	SIG r	SIG R
1	31	0,145	0,2877	4	4	6
2	40	0,077	0,2542	15	1	13
3	40	0,110	0,2653	17	4	8
4	40	0,125	0,3472	24	2	16
TOTAL	151	0,112	0,2886	60	11	43

TABLE 5.9 : MINING GROUP

PERIOD	N	$\overline{\lvert r \rvert}$	$\overline{R}$	+ve r	SIG r	SIG R
1	3	0,129	0,2626	0	0	0
2	3	0,036	0,1905	0	0	1
3	3	0,139	0,3020	1	1	1
4	3	0,184	0,3537	0	0	1
TOTAL	12	0,122	0,2772	1	1	3

TABLE 5.10 : FINANCIAL GROUP

PERIOD	N	$\overline{|r|}$	$\overline{R}$	+ve r	SIG r	SIG R
1	9	0,068	0,2834	2	0	1
2	15	0,091	0,2765	6	1	6
3	15	0,089	0,2746	8	1	3
4	15	0,118	0,3463	11	0	6
TOTAL	54	0,094	0,2952	27	2	16

TABLE 5.11 : INDUSTRIAL GROUP

PERIOD	N	$\overline{|r|}$	$\overline{R}$	+ve r	SIG r	SIG R
1	19	0,184	0,2938	2	4	5
2	22	0,073	0,2477	9	0	6
3	22	0,121	0,2538	8	2	4
4	22	0,122	0,3469	13	2	9
TOTAL	85	0,123	0,2856	32	8	24

In analyzing the results presented in the tables above, the discussion will be broken into two sections. The first deals with the overall group of 40 securities, while the second deals with the three subgroups - Mining, Financial and Industrial.

COMBINED MINING, FINANCIAL AND INDUSTRIAL GROUPS

Of the 151 regression equations obtained, approximately 28% indicated a significant relationship for the extended model at the 5% level (that is rejection of the null hypothesis that the multiple correlation coefficient (P) equals zero). On the other hand, only 7% of the 151 situations examined indicated disagreement with the random walk model at approximately the same level of significance (that is, had a correlation coefficient more than 2 standard deviations from zero). This would tend to indicate that the extended model indicates greater support for the chartist than that suggested by the random walk model.

Compared with the results presented in Table 5.2, it can be seen that when considering the shorter periods and the different types of markets, the estimates of r and R were on average higher (0,112 and 0,2886 respectively) than when considering the entire period (0,062 and 0,2109 respectively). However, fewer significant r's and R's were found in the shorter periods (7% and 28% respectively) compared with the entire period (18% and 93% respectively). This indicates that there is not much (if in fact any) improvement in the "fit" of the respective models when considering the shorter periods and the specific types of markets. Once again, this is a somewhat strange result since the type of market was chosen a posteriori and a better fit might have been expected. Once again, this strangeness can probably be attributed to the "step like" feature of the market. This fact is further bourne out by a consideration of the number of positive values of the correlation (r) between R_t and R_{t-1}. It might have been thought a priori, that within a particular type of market (bull or bear) the correlation between successive returns would be positive. However, this is not the case, and there is, in fact, a preponderance of negative correlations (91 out of 151). Moreover, of the eleven significant correlations, only one was positive.

INDIVIDUAL GROUPS

The results from the individual groups, as expected, present very similar results to those discussed above, with no really startling differences. Perhaps the most interesting point is that for all three groups, the fourth period considered had a higher R value on average than the remaining three periods. Combining the groups, this R (for the fourth period) was highest of all 4 periods considered for 22 of the 40 securities examined. In addition, the second and fourth periods each had a greater number of significant R's than the first and third periods. As the first and third periods represent bull markets while the second and fourth periods indicate bear markets, it would appear that the extended model is perhaps slightly more suitable in a bear market than in a bull market.

There are, it is felt, two main conclusions which can be drawn from the results presented in this section. Firstly, it does not appear that there is a marked increase in the suitability of the extended model when considering specific types of markets only. Moreover, consideration of specific types of markets does not lead to a greater rejection of the random walk model. Secondly, there appears to be considerable support for the "step-like progress of the market" theory, and it would appear that even in a strong bull or bear market, if the random walk model does not hold, the correlation between successive returns is likely to be negative. This point illustrates perhaps one of the greatest differences between the chartist and the random walk believer, namely that the random walk believer is interested in predicting next week's return for every week, while the chartist is merely interested in predicting whether the security is worth buying or selling, and will often be in a position where he has no idea whether the security is going to rise, fall or remain static in a given week, and as such will make no decision.

5.6 CONCLUSIONS

In concluding this chapter, it might be said that the results presented indicate that the random walk model almost certainly holds and therefore should be accepted by all investors as a fait accompli. Afterall, the argument used by the proponents of the random walk model is that it is worthless to study a chart of the price history of a security in isolation, ignoring all other factors, and this statement is likely to find agreement among almost all stock market investors.

What can be argued is that the random walk model does not mirror the behaviour of a chartist, and as such should not be used as proof against the applicability of charting techniques. In this chapter it is shown that by a very simple extension of the random walk model, a statistically significant relationship can be found, and it is argued that this extension is more in keeping with the behaviour of the chartist than is the random walk model.

It is important that the interpretation of the statistical significance of the relationship is clearly understood. It is not claimed that this relationship should be used for prediction purposes. In fact, sound statistical reasoning would advocate that this relationship should certainly not be used for prediction. But, it is important to note that there is evidence of some relationship (however slight). When one considers the complexity of the chartist's signals in comparison to the simplicity of the model examined in this section, it is indeed possible that a far better relationship exists using the chartist's model than is evidenced by the degree of relationship observed for the extended model. An example of this is the volume of securities traded in the previous week, which was found to be a relatively unimportant variable in the regression model discussed in this chapter. However, the regression analysis uses each week's volume as a specific number and, as there can be wide fluctuations in the volume traded from week to week, this probably accounts for the variables apparent lack of importance. But the chartist might use this variable more subjectively, considering it to be merely light, medium, or heavy, or as much higher or lower than the previous week.

Thus, it is felt that the results presented in this chapter, while not proving that charting techniques do work, does provide some indication that they may perform better than has been suggested, and therefore provides far more hope for the chartist than has been forthcoming in the recent literature.

CHAPTER SIX

THE MEASUREMENT OF RISK FOR SECURITIES QUOTED ON THE JOHANNESBURG STOCK EXCHANGE

6.1 INTRODUCTION

In recent years the use of the so called 'Beta Coefficient' as a measure of the risk inherent in a particular security has become increasingly popular in the United States, Europe, and even, to a more limited degree, in South Africa. The reason for this is, it is felt, that investors are becoming more aware of the importance of the inter-relationships between securities. Unfortunately, as William Sharpe argued (Sharpe (1970)), the notion of covariance with the market and with other securities (a suitable statistical measure) lacks intuitive appeal. For this reason, Sharpe proposed an alternative, namely the volatility of the security's rate of return relative to changes in the market as a whole. This measure of the volatility has come to be called the Beta Coefficient, and has received considerable attention in the literature.

In this chapter, the beta coefficients for some of the securities quoted on the Johannesburg Stock Exchange are examined. The results presented indicate that the securities appear to behave differently to those quoted on other exchanges (such as New York), and appear to lead to some rather puzzling conclusions. On re-examining the simple statistical model used to compute these beta coefficients, it can be seen why this situation arises, and this leads to the proposal of an alternative measurement of risk which, it is argued, has most of the advantages of the beta coefficient , and also provides a more realistic measure of the relative risk of the security.

In the following section, a brief description of what beta coefficients are, and how they may be used in practice, is given. Then, in Section 3, some empirical beta coefficients are examined and the results of some statistical tests presented. Section 4 contains an examination of the statistical model used to obtain the beta coefficients, an explanation of why this model does not always provide realistic beta coefficients, and a proposal for a new measure of volatility. Empirical results are presented for this new proposal, and it is argued that these results are intuitively more appealing than those obtained using the "traditional" beta coefficients. Finally, in Section 5 the results presented in this section are summarized and some overall conclusions given.

6.2 BETA COEFFICIENTS AND VOLATILITY

The idea of using beta coefficients as a measure of risk had its origin in the work of William Sharpe (1964) concerning market equilibrium under conditions of risk, and has been extended and developed by numerous researchers into the so called "market model." This model can be briefly summarized as follows.

The return on a security from period t-1 to period t is assumed to be a linear function of a market factor (henceforth referred to as the market return) common to all securities, and an independent factor which is unique to each security. Symbolically, this relationship can be written as

$$Y_t = \alpha + \beta X_t + e_t$$

where Y_t is the return on the security in the t^{th} period;

α is a parameter (in fact, α is the expected return on the security when there is no movement in the market);

β is a parameter (the slope of the line);

X_t is the return on the market in period t; and

e_t is the disturbance or error term and is assumed to have zero expectation and to be independent of all e_s, $s \neq t$.

The parameters α and β can be easily estimated using the statistical technique of regression analysis, provided that one has some past data on the return of the security under consideration, and the return on the market. For this reason, the market factor is usually assumed to be well described by some general overall market index.

The parameter β can then (it is argued) be used as a measure of the volatility (Treynor (1965)) of that security relative to the market since, if β is greater than one, then when the market rises, it is obvious that the return on the security will rise more rapidly than the return on the market. Similarly, if the market falls, the return on the security will fall more rapidly than the return on the market and thus, the security can be regarded as being more volatile, and hence more risky, than the market. On the other hand, if β is less than one, then in a rising market the security will rise slower than the market, and in a falling market will fall less than the market and thus it is not as risky as the market. This interpretation has been used extensively in the literature (e.g. Levy and Sarnat (1972)). In addition, the beta coefficients for a number of securities may be computed and used to compare the <u>relative</u> riskiness of those securities.

This model (often called the market model) has received much discussion in the fairly recent literature on the behaviour of stock market prices. Most of the discussion has tended to favour the use of these coefficients, but not all of the arguments presented have been completely convincing. Perhaps the most important results which have appeared are that:

(i) the linearity assumption appears to be fairly well satisfied (Fama, Fisher, Jensen and Roll (1969));

(ii) the beta coefficients appear to be fairly stable over time (Blume (1971) and Levy (1971)), especially as the length of the period under consideration increases (Baesel (1974));

(iii) even if the beta coefficients do change over time, the ranking of the securities in order of risk does not (Gross (1974));

(iv) the beta coefficients do give a fairly good measure of the risk inherent in a security (Modigliani and Pogue (1974)); and

(v) the value of beta in any period can be related to some fundamental characteristics of the firm in that period (Beaver, Kettler and Scholes (1970) and Rosenberg and McKibben (1973)).

However, it should be noted that some of these results do require further substantiation.

6.3 STATISTICAL TESTS OF THE MARKET MODEL

In this section some of the statistical considerations of the market model, which have been presented in the literature, are examined and analagous results for securities quoted on the Johannesburg Stock Exchange are presented.

Using ordinary least squares, the consistent estimation of the parameters α and β of the market model depends on the following assumptions concerning the model:

(i) $E(e_t) = 0$;

(ii) $E(e_t e_s) = 0$ for all $t \neq s$;

(iii) $E(e_t^2) = \sigma^2$ for all t ; and

(iv) e_t is independent of X_t for all t.

The first two assumptions are, it is felt, not likely to be seriously violated in practice. They have been shown to be satisfactorily obeyed for securities quoted on the New York Stock Exchange (Fama et al (1969)), and since results for the random walk model (which while not exactly analagous to the market model is based on very similar principles) have been shown to be very similar for the New York Stock Exchange and the Johannesburg Stock Exchange (Affleck-Graves and Money (1975)), it is felt that this similarity will hold for the first two assumptions of the market model as well.

Should the third assumption be violated, known as heteroskedasticity, then ordinary least squares is no longer applicable, and this factor has evoked much discussion in the literature,where it is generally argued and accepted that, if heteroskedasticity is present, it is likely to be evidenced by the violation of assumption (iv); that is, that e_t is in fact dependent on X_t. This follows since, in the case of stock price data, it is likely that the variance of e_t will be dependent on X_t and hence will not be σ^2 for all t. However, Fama, Fisher, Jensen and Roll (1969) concluded that the regression assumption of homoscedasticity was very well supported by the evidence they presented, and this conclusion was further substantiated by Martin and Klemkosky (1975).

But, Praetz (1969) found that 35 out of the 37 price series studied from the Sydney Stock Exchange exhibited significant heteroskedasticity at the one precent level.

As there appears to be a difference in the results presented for the New York Stock Exchange and the Sydney Stock Exchange, it was decided to investigate the situation on the Johannesburg Stock Exchange (JSE). In order to do this only the Industrial section of the JSE was examined, and it was assumed that the RDM (Rand Daily Mail) Industrial Index was a suitable measure of the market factor. Only the Industrial Section of the market was examined since no overall market index was considered suitable. It is not felt that any different behaviour would have been observed if all sections of the market had been considered.

The 94 securities used in the Index Selection Analysis of Chapter 2 for the six industrial sectors examined (Building, Food, Motors, Paper, Stores and Sugar - see Appendix A, Tables A.5 to A.10) were used. The weekly returns* on each of these securities, and on the RDM index, were computed for the period 22nd March 1968 to 20th February 1976, as follows:

---

* Dividends were not considered - c.f. Sharpe and Cooper (1972); Martin and Klemkosky (1975).

$$Y_{i;t} = \log_e P_{i;t} - \log_e P_{i;t-1}$$

where $Y_{i;t}$ is the return on the i^{th} security in the t^{th} period;

$P_{i;t}$ is the price of the i^{th} security at the end of period t; and

$P_{i;t-1}$ is the price of the i^{th} security at the end of the previous period.

The difference in the logarithms of price was used as this is a measure of the return under continuous compounding (Fama (1965a)).

Using the return on the RDM index (computed in the same manner as the return on the individual securities) for the return on the market (X_t), the parameters α and β were estimated for each of the 94 securities examined, using ordinary least squares regression. These results, together with three additional statistics, are presented in Table 6.1 below.

The estimated α and β coefficients presented in Table 6.1 might at first glance appear to be very similar to those obtained in studies of the New York Stock Exchange (for example Blume (1971)). The alpha coefficients are all fairly close to zero, as was found for the NYSE, which is to be expected since these coefficients represent the expected return on the security when there is no movement in the market. In fact, the largest alpha coefficient, that

TABLE 6.1

SHARE	$\hat{\alpha}$	$\hat{\beta}$	R	RS	S.D. RATIO
Alex Blaikie	0,001	0,741	0,287	0,060	2,58
Anglo Alpha	0,000	0,689	0,395	0,000	1,74
Bellandia	-0,002	-0,073	0,025	0,506	-2,92
Blue Circle	-0,001	0,701	0,291	-0,103	2,41
Boumat	0,002	1,052	0,403	-0,023	2,61
Brick and Pott.	0,002	1,008	0,265	0,025	3,80
Bruynzeel	-0,001	0,964	0,351	0,050	2,75
Buffalo	0,001	0,056	0,030	0,440	1,87
Coronation Brick	0,003	0,207	0,100	0,215	2,07
Everite	0,002	0,205	0,161	0,129	1,27
Golden Brown	0,002	0,465	0,169	0,047	2,75
Grinaker	0,001	0,914	0,411	0,004	2,22
Gypsum	-0,001	0,094	0,061	-0,211	1,54
Johnstone	0,002	0,603	0,327	0,047	1,84
Katzenellenbogen	0,002	0,248	0,099	0,198	2,51
LTA	-0,002	1,107	0,397	0,073	2,79
Masonite	-0,002	0,365	0,156	0,020	2,34
Murray & Roberts	-0,001	1,207	0,522	-0,030	2,31
Plascon Evans	0,001	0,517	0,202	0,099	2,56
Plate Glass	0,000	0,966	0,523	0,048	1,85
Premier Cement	0,001	-0,027	0,024	-0,820	-1,13
Pretoria Cement	0,000	0,705	0,426	0,009	1,65
Rhodesian Cement	0,000	0,175	0,111	0,067	1,58
Rhodesian Brick	-0,002	-0,019	0,168	0,752	-0,11
Sinclair	0,002	0,237	0,058	0,117	4,09
Af.Prods.	0,003	0,539	0,294	0,123	1,83
Bakers	-0,001	0,275	0,170	-0,026	1,62
Becketts	0,000	0,584	0,239	-0,044	2,44
Crown Mills	0,001	0,347	0,171	0,100	2,03
I & J	-0,000	1,122	0,526	-0,039	2,13
Imperial C.S.	0,001	0,621	0,252	-0,058	2,46
Jabula	0,003	0,217	0,140	0,159	1,55
Lewis	0,002	0,398	0,208	0,074	1,91
Monis & Fatti	-0,001	0,295	0,158	-0,071	1,87
Picardi	0,005	0,392	0,080	0,232	4,90
Premier Mill.	0,002	0,887	0,455	-0,145	1,95
Simba	-0,004	0,984	0,318	0,080	3,09
Stein Bros.	0,000	0,578	0,279	-0,024	2,07
Tiger Oats	0,001	1,048	0,643	-0,053	1,63
Union C.S.	0,000	0,057	0,049	0,039	1,16
Assoc. Eng.	-0,002	0,346	0,183	-0,073	1,89
Aurochs	0,001	0,220	0,087	0,073	2,53
Autolec	0,001	0,215	0,112	0,108	1,92
Bus Bodies	0,001	0,433	0,205	-0,016	2,11
Cap. Gold	0,000	0,251	0,097	0,066	2,59
Curries	-0,000	0,543	0,282	0,009	1,93
Dunlop	-0,001	0,800	0,357	0,114	2,24
Eriksen	0,000	0,181	0,082	0,044	2,21
Gen. Tire	0,000	0,636	0,322	0,027	1,98
Gen Tire A	0,000	0,646	0,328	0,038	1,97

SHARE	$\hat{\alpha}$	$\hat{\beta}$	R	RS	S.D. RATIO
Lawson	-0,002	1,456	0,418	0,094	3,48
Lucys	-0,003	0,570	0,222	-0,008	2,57
Mac. Rodway	0,003	1,083	0,486	-0,023	2,23
NFS Motors	0,000	0,219	0,139	0,067	1,58
Putco	0,002	0,445	0,213	0,018	2,09
Robbs	-0,000	0,471	0,263	-0,011	1,79
Tollgate	-0,001	0,575	0,251	0,023	2,29
Toyota	-0,000	1,234	0,502	0,074	2,46
Trans. & Eng.	-0,000	0,140	0,080	-0,046	1,75
Welfit	0,000	0,220	0,111	0,092	1,98
Wm. Hunt	0,000	0,730	0,316	0,020	2,31
Copi	0,003	0,457	0,217	0,004	2,11
Coates	-0,001	0,210	0,110	-0,131	1,91
Cons. Glass	0,001	0,358	0,237	0,043	1,51
Ev. Haddon	-0,000	0,284	0,161	-0,052	1,76
Kohler	0,001	0,181	0,101	0,124	1,79
Metal Box	0,001	0,599	0,337	-0,018	1,78
Metal Closures	0,000	0,418	0,212	-0,006	1,97
Prem. Paper	0,000	0,191	0,128	0,081	1,49
Press Supp.	0,002	-0,025	0,023	-0,728	-1,09
Rhodesian Pulp	-0,000	0,044	0,074	-0,174	0,59
Sappi	-0,000	0,811	0,416	0,045	1,95
Trio Rand	0,003	0,349	0,134	0,031	2,60
Edgars	-0,000	0,740	0,431	-0,024	1,72
Foschini	-0,001	0,580	0,329	0,049	1,76
Garlicks	-0,001	0,350	0,268	0,001	1,31
Greatermans	-0,002	1,089	0,573	-0,022	1,90
Greatermans 'A'	-0,002	1,148	0,575	-0,044	2,00
Harrowes	-0,002	0,574	0,193	-0,032	2,97
Hepworths	-0,000	0,035	0,030	-0,281	1,17
John Orr	-0,002	0,327	0,172	-0,131	1,90
Lewis Fosch.	-0,001	0,911	0,435	-0,046	2,09
O.K.	-0,001	1,168	0,645	0,025	1,81
Spitz	0,002	0,866	0,255	0,133	3,40
Stuttaford	-0,001	0,396	0,269	-0,053	1,47
Truworths	-0,000	0,240	0,164	-0,029	1,46
Woolworths	0,001	1,306	0,615	0,020	2,12
Woolworths 'A'	0,001	1,311	0,685	0,004	1,91
Crookes	0,003	0,323	0,159	0,154	2,03
Huletts	0,001	0,951	0,471	0,036	2,02
Illovo	0,002	0,710	0,246	0,087	2,89
C.G. Smith	0,004	0,632	0,312	0,134	2,03
Swazi Sugar	0,006	-0,031	0,012	-0,732	-2,58
Tongaat	0,001	0,738	0,389	0,026	1,90
AVERAGE	0,000	0,544	0,259	0,013	2,09

of Swaziland Sugar, was 0,006, which indicates an estimated return of 0,6% on Swaziland Sugar when there is no market movement.

On a closer examination of the beta coefficients, it becomes apparent that the majority of securities have coefficients less than one - 80 out of the 94 examined (approximately 85%). This means that according to the results presented in the second column of Table 6.1, only 15% of the securities are more risky or more volatile than the market itself, while the remainder are less volatile. This is indeed a surprising result, especially when it is remembered that the market movement has been measured by a large index, which is generally supposed to "flatten out" the random movements, and which has often been quoted in the literature as being a "sluggish" type of market indicator. This result is similar to that obtained by Altman, Jacquillat and Levasseur (1974) for the Paris Bourse. They found that, for the period 1964-1971, on average, approximately 70% of the beta coefficients were less than one.

In order to examine this situation more closely, the multiple correlation coefficient, R, was computed for each regression equation, and these results are presented in the third column of Table 6.1. This statistic (R) can be used to determine the suitability of the model since the square of R (called the coefficient of determination) gives the percentage of the variation in the dependent variable (the return on the security in this case) explained by the independent variable (the return on the market). It should be noted that, since the market model has only one independent variable (or regressor), the multiple correlation coefficient is in fact the ordinary correlation between X and Y, that is the returns on the market and the security.

In order to test the suitability of the market model, the null hypothesis that the population multiple correlation coefficient (P) equals zero (that is that there is no fit) can be tested for each security examined. Since the sample size for each security is large (413 weeks), it follows that even a very low value of the sample multiple correlation coefficient (R) will result in rejection of the null hypothesis. In fact, the critical value of R is 0,0961. From the results presented in Table 6.1 it can be seen that 71 of the securities examined had significant fit. However, the percentage variation explained by the regression equation, when the multiple correlation coefficient is 0,0961, is less than 1%. Thus, while statistically significant, such an equation is clearly not of much practical use to an investor. Probably, the minimum explanation which would be satisfactory from a practical point of view (bearing in mind the findings of King (1966), namely that the market factor accounts for approximately 31% of the movement of securities on the New York Stock Exchange) is of the order of 10%, which would correspond to an R of 0,316. Using this value of R as a guide it is found that only 32 of the 94 securities examined had suitable fits; that is, for only 34% of the securities examined did the market model explain more than 10% of the variation in the return of the security.

The literature on beta coefficients has suggested that a possible reason for the poor estimation is that heteroskedasticity is present and thus, that classical least squares is not applicable. In order to examine the question of heteroskedasticity, a simple test proposed by Gorringe (Johnston (1963)) was employed. This test is based on the assumption that heteroskedasticity is evidenced by an increase in the variance of the error term as the independent variable (the return on the RDM index in this case) increases and vice versa. Thus, the test consists of computing the Spearman Rank Correlation between the absolute values of the residual error and the corresponding market return (Martin and Klemkosky (1975)). This test was performed on each of the 94 securities under examination, and the results are presented in the fourth column of Table 6.1. If the rank correlation is significantly different from zero, then it can be concluded that heteroskedasticity is present. The critical value of the rank correlation coefficient can be obtained using t tables and the following formula

$$t = r_s \sqrt{\frac{N-2}{1-r_s^2}}$$

where t is distributed as the Student's t with $N-2$ degrees of freedom, and r_s is the Spearman's Rank Correlation coefficient. Using this formula and a 5% level of significance, it is found that an absolute value of the rank correlation of more than 0,096 is statistically significant (that is different from zero).

On examination of Table 6.1 it can be seen that 28 of the 94 securities examined displayed significant heteroskedasticity. In addition, it appears as if there is some relationship between the multiple correlation coefficient R and the Spearman Rank Correlation. Specifically, it appears as if the better the fit, that is the higher R, the lower the rank correlation, that is the less the degree of heteroskedasticity, and vice versa. In order to examine this aspect more closely, the following table, Table 6.2, was constructed, once again using the argument that the multiple correlation coefficient must be greater than 0,316 to be of practical use.

TABLE 6.2

	$R \geq 0{,}316$	$R < 0{,}316$	
Significant RS: $\lvert RS \rvert \geq 0{,}096$	2	26	28
Non Significant RS: $\lvert RS \rvert < 0{,}096$	30	36	66
	32	62	94

Regarding Table 6.2 as a 2×2 contingency table, the chi-square value can be calculated (using the formula $[\lvert ad-bc \rvert - \frac{1}{2}n]^2 . n/n_1 n_2 n_3 n_4$) and is found to be 11,20, which indicates significant dependence (χ^2_1 at the 95% level is 3,84).

Thus, it can be concluded that the multiple correlation coefficient (which gives an indication of the "fit") and the Spearman's Rank Correlation coefficient (which gives an indication of heteroskedasticity) are not independent. In fact, from Table 6.1, and Figure 6.1 on the following page, one can conclude that as the fit gets poorer (that is, R decreases) one is more likely to find heteroskedasticity present. In fact, in only 2 out of the 28 cases of significant heteroskedasticity was the multiple correlation coefficient sufficiently high to be of practical use. What is more, it can be shown (see Appendix G) that if heteroskedasticity is present, and if ordinary least squares is used, then the multiple correlation coefficient R is in fact overestimated. Also, of the 28 securities which appeared to have significant heteroskedasticity, 10 had a negative Spearman's rank correlation coefficient, which would indicate that for those securities, as the RDM index rises, the variance of the error term decreases, which would be a most surprising result.

Thus, from these findings it is felt that the apparent heteroskedasticity observed in approximately 30% of the securities examined, does not really exist, but is actually a result of the fact that the market model does not appear to be a very suitable model for a large number of the securities quoted on the Johannesburg Stock Exchange.

It should be noted that as Brown (1977) has indicated,

F I G U R E   6.1

the test used above does not prove conclusively that there is no heteroskedasticity. Rather, it examines a particular form of heteroskedasticity. However, it is likely that if heteroskedasticity was present in the market model, it would be of the form examined and hence, it is argued, the results presented above (and those presented in Martin and Klemkosky (1975)) are valid. The most important point is the poorness of the fit of the market model, which makes the whole question of heteroskedasticity rather academic.

6.4 THE S.D. RATIO

The results presented in the previous section indicate that the market model does not appear to be a very suitable model for use on the Johannesburg Stock Exchange (certainly for the case when the RDM '100' Industrial Average is used as a market indicator), since only 32 of the 94 securities examined displayed a satisfactory degree of fit. It is interesting to note that the average value of R (for all 94 securities) is 0,259, which is much lower than the average of 0,41 found by Altman et al (1974) for the Paris Bourse over the period 1964-1971, and the average of 0,529 found by Blume (1971) for the New York Stock Exchange (although the latter was based on monthly data which, in general, has the effect of increasing R - Baesel (1974)). Thus, it would appear as if the market model is not as applicable in the South African context as it is on other world exchanges, although in all cases the average values of R are rather low.

In order to examine how this lack of fit can affect the beta coefficients it was decided to examine the results presented in Table 6.1 more closely.

If the estimated beta coefficients for the various securities are plotted against the corresponding multiple correlation coefficients (the R's) as in Figure 6.2, then it becomes apparent that as the fit improves, that is, as the multiple correlation coefficient increases, the beta coefficient increases and vice versa. In fact, from Figure 6.2 it can be clearly seen that R and $\hat{\beta}$ are not independent. This is not an altogether surprising result since, for the single regressor case, the estimate of the beta coefficient is

$$\hat{\beta} = S_{XY}/S_X^2 ,$$

where S_{XY} is the estimated covariance between X and Y, and S_X^2 is the estimated variance of X, while the estimated multiple correlation coefficient is, in this case, the ordinary sample correlation coefficient (r), so that

$$R = r = S_{XY}/S_X S_Y$$

where S_Y is the estimated standard deviation of Y. Therefore,

$$\hat{\beta} = R(S_Y/S_X) .$$

Now, the above result has important implications for the users of beta coefficients since it implies the following rules:

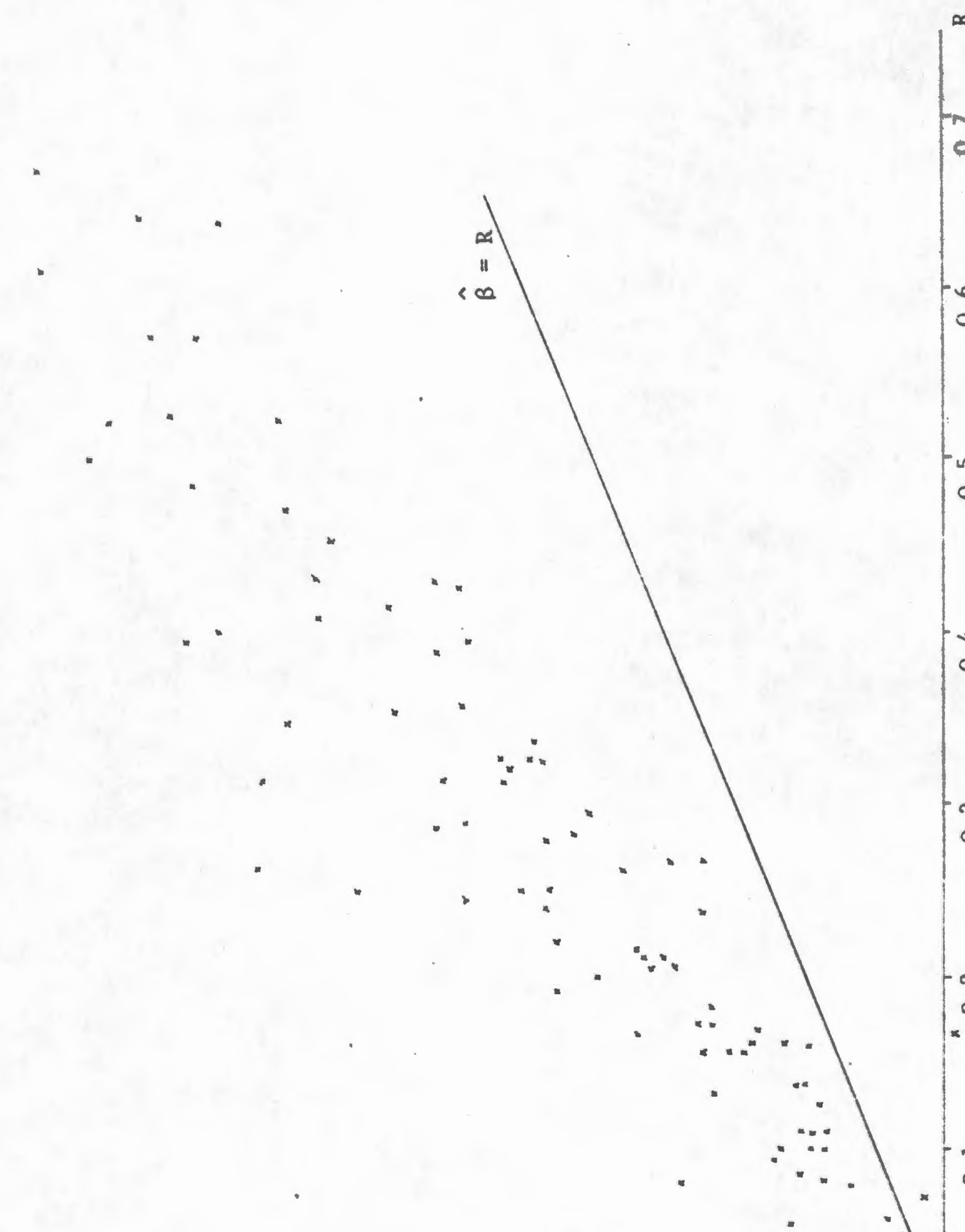


FIGURE 6.2

(i) if $S_Y > S_X$, then $\hat{\beta} > R$;

(ii) if $S_Y = S_X$, then $\hat{\beta} = R$; and

(iii) if $S_Y < S_X$, then $\hat{\beta} < R$;

where S_Y is the estimated standard deviation of the return on the security and S_X is the estimated standard deviation of the market return. It has long been accepted that the standard deviation can be used as a measure of risk, and so the above rules imply that, if $\hat{\beta} > R$, then the security has a greater standard deviation than the market, and hence has greater risk than the market and vice versa. The question that must now be answered is how this affects the interpretation of the beta coefficients?

A beta coefficient greater than one is said to indicate that the security is a high risk security (that is, is more risky than the market). This is certainly true since R is bounded above by one and hence, if the beta coefficient is greater than one, then beta must certainly be greater than R, which implies that $S_Y > S_X$ and hence that the security is more risky than the market. On the other hand, if $\hat{\beta}$ is less than one, then one cannot say that the security is less risky than the market since $\hat{\beta}$ might still be greater than R, thus implying that $S_Y > S_X$ and that the security is more risky than the market even though $\hat{\beta} < 1$. The low $\hat{\beta}$ might arise, not from the fact that the security rises or falls less than the market, but due to the fact that the market model does not "fit" the data that well. This follows since, if the fit is not very good, then the regression analysis (which minimizes the sum of the squares of the vertical distances) would tend to fit a more horizontal line to the data regardless of the relative riskiness of the security, that is to set $\hat{\beta}$ close to zero. Thus, as the fit gets poorer, the beta coefficient will get smaller and will almost certainly be less than one if the fit is bad (that is the multiple correlation coefficient is less than, say, 0,2) even though the security may be considerably more risky than the market.

Thus, it is argued, both the beta coefficient and the multiple correlation coefficient should be considered in classifying a security as a high risk or a low risk share. This can be done by plotting the line $\hat{\beta} = R$ as has been done in Figure 6.2. Then, all securities above this line can be regarded as high risk, while those below the line can be regarded as low risk securities.

One important point which should be noted is that the multiple correlation coefficients listed in Table 6.1 are in fact the absolute values of R. If the beta coefficient is negative, then the sample correlation coefficient between X and Y is also negative, and is the multiple correlation coefficient in this, the single regressor, case. Thus, in determining the relative riskiness of the security, the absolute value of the beta coefficient should be plotted against the absolute value of the multiple correlation coefficient, since a negative beta coefficient does not indicate more or less risk than the market, but that the security and the market tend to move in opposite directions.

What is now required is some measure of the relative riskiness of the security; that is, how far the beta coefficient is above or below the line $\hat{\beta} = R$. This can be done by considering the ratio $\hat{\beta}/R$. But,

$$\hat{\beta}/R = S_Y/S_X .$$

If the multiple correlation coefficient (R) equals one, that is if the market model provides a perfect fit for the data, then

$$\hat{\beta} = S_Y/S_X .$$

Thus, it is argued, when R is less than one the ratio S_Y/S_X (hereafter referred to as the S.D. ratio) should be used as a measure of the security's relative risk and not $\hat{\beta}$, which will be influenced (decreased) by the lack of fit (that is, the low R). The only information provided by the beta coefficient, which is not given by the S.D. ratio, is whether the security moves with the market (that is $\hat{\beta} > 0$) or counter to the market ($\hat{\beta} < 0$). This information can be incorporated in the S.D. ratio by prefixing it with the same sign as the covariance between the return on the security and the return on the market.

The S.D. ratio (S_Y/S_X or $\hat{\beta}/R$) has been calculated for each of the 94 securities examined in this chapter, and the results are presented in the fifth column of Table 6.1. Examination of the results indicates that 92 of the 94 securities examined have S.D. ratios greater than one, which indicates that they are more risky than the market.

Clearly, it will be of interest to test whether the observed S.D. ratio is greater than one, equal to one, or less than one; that is, whether the security is more risky, as risky, or less risky than the market. This can be done by testing the null hypothesis

$$H_0: \sigma_x^2 = \sigma_y^2$$

against the alternative

$$H_1: \sigma_x^2 \neq \sigma_y^2 \ .$$

Since the two series under consideration, the return on the security and the return on the market, are not independent, the usual F-test is not strictly applicable. Nevertheless, the above null hypothesis may still be tested as described below (Kenney and Keeping (1951)):

Reject H_0 if

$$S_Y^2/S_X^2 > [A - (A^2-1)^{\frac{1}{2}}]^{-1}$$

or $$S_Y^2/S_X^2 < [A + (A^2-1)^{\frac{1}{2}}]^{-1}$$

where $$A = \frac{N-2+2(1-r^2)t^2_{\alpha;N-2}}{N-2} \ ;$$

N = the number of observations ;

r = the correlation between X and Y (that is, in this case R); and

$t_{\alpha;N-2}$ = the student's t-distribution with N-2 degrees of freedom.

It should be noted that the above test assumes that X and Y have a bivariate normal distribution. As has been discussed previously, it is possible that security returns have stable distributions, but it is felt that they are sufficiently normal for the above test to be used.

It should also be noted that the simpler F-test can be used as an approximate test if the sample size is large, and if the correlation between the series is small. That is, it can be assumed that

$$S_Y^2/S_X^2 \sim F_{N-1,N-1}$$

For the data analyzed in this chapter, the sample size is large (413), and the correlation between the series is generally fairly small (only 10 of the 94 correlations are greater than 0,5, and none are greater than 0,7). This will often be the case in stock market data such as this, and hence the F-test can often be used as an approximation. In fact, for the securities examined in this chapter, the results obtained using the approximate F-test, and those of the exact test, are identical. The null hypothesis was rejected for all 94 securities thus indicating that none of the securities can be considered to have the same risk as the market, and only two securities (Rhodesian Brick and Rhodesian Pulp) can be classified as having less risk than the market. These results are, it is felt, intuitively more appealing than those obtained from the analysis of the beta coefficients, where it was found that only 14 of the 94 securities could be regarded as having more risk than the market.

Thus, it is argued, the S.D. ratio gives a far better indication of the riskiness of a security than its associated beta coefficient. Moreover, it retains the main advantage of the beta coefficients, namely that it enables a comparison of the relative riskiness of two securities - a security with a high S.D. ratio is relatively more risky than a security with a lower S.D. ratio. It should be noted that in many cases the ordering obtained using the beta coefficients is different from that obtained using the S.D. ratio. For example, the beta coefficient of Plate Glass is 0,966, while that of Plascon Evans is 0,517 which would indicate that Plate Glass is relatively more risky than Plascon Evans. However, on examining the S.D. ratios it is found that the S.D. ratio of Plate Glass is 1,85 while that of Plascon Evans is 2,56 indicating the opposite of the above, namely that Plate Glass is less risky than Plascon Evans. Since S_x^2 (the estimated variance of the market return) is the same for both securities, the lower S.D. ratio of Plate Glass compared to that of Plascon Evans indicates that the standard deviation of Plate Glass is less than the standard deviation of Plascon Evans, which (it is argued) indicates that the S.D. ratio is a more suitable measure of the riskiness of a security than the beta coefficient.

The S.D. ratio also has two other important advantages over the beta coefficient. Firstly, it is independent of the "degree of fit" (or of the multiple correlation coefficient) and, in fact, does not use the market model at all. Secondly, as the ratio of any two variances from normal populations can be tested using the above tests, the S.D. ratio can be calculated for any two securities, and it can then be decided whether the first security is more risky, less risky or has approximately the same risk as the second security.

Finally, mention must be made of the fact that, for the results presented in Table 6.1, each variance was estimated using 413 weeks of data and that such a large sample results in a narrow confidence interval. In practice, an investor will often be interested in the risk profiles of the securities over a much shorter period, for example a year (52 data points), which will result in a wider confidence interval, and hence the possible classification of more securities as having approximately the same degree of risk as the market as a whole.

6.5 CONCLUSIONS

In this chapter an attempt has been made to examine the behaviour of the so called beta coefficients for securities quoted on the Johannesburg Stock Exchange. The analysis of the market model has been approached from a statistical point of view, resulting in the emergence of some interesting findings which cause some doubt as to the validity of the use of beta coefficients in general. An alternative measure, the S.D. ratio, is proposed which, it is argued, is a more suitable measure of the risk inherent in a particular security relative to the market as a whole.

The first conclusion which can be drawn from the results presented in this chapter, is that the market model does not appear to be a very satisfactory model for securities quoted on the Johannesburg Stock Exchange. In general, the multiple correlation coefficients (indicating the degree of fit) tend to be low, and it would appear that only about one third of the securities have a multiple correlation coefficient sufficiently large to be of any possible practical use. Even for these securities, the multiple correlation coefficients are generally very low compared with what is recommended by most texts on regression analysis and hence, all results based on the analysis of the market model should be used with the utmost caution.

It could possibly be argued that the lack of fit observed was due to the unsuitability of the RDM index as a measure of the overall industrial market, and not due to the failure of the market model. However, the RDM Industrial index is a market capitalization type index containing approximately one hundred securities from the industrial section of the market and does incorporate many of the large companies (that is companies with large market capitalizations) and, apropos the remarks made in Chapter Two, it is extremely doubtful whether the performance of the index would be very different from that of any other market capitalization type index incorporating most of the larger companies. Moreover, studies of other major exchanges have also indicated a fairly low average multiple correlation coefficient.

The second interesting result to emerge from the analysis presented in this chapter,concerns the investigation into the presence of heteroskedasticity in the market model. The results obtained using a fairly simple statistical test indicate that approximately thirty percent of the securities examined displayed significant heteroskedasticity. However, closer examination revealed that the degree of heteroskedasticity appeared to be dependent on the multiple correlation coefficient, in that, in general, the degree of heteroskedasticity (absolute value of Spearman's Rank Correlation Coefficient) was high only when the multiple correlation coefficient (that is, the degree of fit) was low, and when the multiple correlation coefficient was high, the degree of heteroskedasticity was low. Thus, it was found that the multiple correlation coefficient had considerable bearing on the results, and it can be concluded that, provided the market model is appropriate, there appears to be very little evidence of heteroskedasticity. If, however, the model is inappropriate then clearly the question of heteroskedasticity does not even arise.

It should be mentioned that there are other more powerful tests for heteroskedasticity (Johnston (1963)) that could have been used, but it is felt that the results presented above do not make such an investigation worthwhile. In fact, the degree of heteroskedasticity, even when significant, was fairly low, and of the 26 cases with an absolute rank correlation of more than 0,1, ten were negative and sixteen positive, which again appears to indicate that the poorness of fit is a more important factor than the presence of heteroskedasticity.

The third and final conclusion, which can be drawn from the results presented in this chapter, concerns the proposed S.D. ratio. The results presented in this chapter indicate that the S.D. ratio provides a better and intuitively more appealing estimate of the risk inherent in a particular security than does the traditional beta coefficient. In addition, the S.D. ratio has further advantages in that it does not depend on the suitability of the market model, and that it allows for a direct comparison between the relative riskiness of any two securities.

C H A P T E R   S E V E N

THE MARKOWITZ AND SHARPE PORTFOLIO SELECTION MODELS: A COMPARISON ON THE JSE

7.1 INTRODUCTION

The problem of which securities to include in a portfolio is a question of fundamental interest to all stock market investors. It is surprising therefore, that most early research into stock market behaviour (for example, Bachelier (1900), Cowles (1933), etcetera) concentrated exclusively on the behaviour of individual securities in isolation. This apparent oversight was corrected by Markowitz (1952), who formulated a mathematical model for portfolio selection which took into account both the return the investor wished to derive from the portfolio, as well as the risk he was prepared to accept. Thus a model was proposed which could be easily adapted to a particular investor's requirements,and this made it intuitively very appealing.

These ideas of portfolio selection models were entrenched by the work of Tobin (1958) who independently derived many of Markowitz' results, and the publishing of a book by Markowitz (1959), which extended and elaborated on the methods proposed in his earlier paper.

Since then, there has been a great deal of interest and research into various alternative portfolio selection models. Among the more notable early contributions were those of Farrar (1962), Sharpe (1963), Baumol (1963), Fama (1965b), Sharpe (1967), and Hastie (1967). More recently, research interest has been in portfolio revision models (as opposed to the Markowitz single period models) with works such as Smith (1967), Mossin (1968), Hakansson (1971), and Chen, Jen and Zionts (1971) making valuable contributions. Unfortunately, the multiperiod models require vast amounts of input and for this reason are still regarded as generally rather impractical, and are often impossible to solve numerically. Moreover, Fama (1968) has shown that most investors can be regarded as single period utility maximizes, and as such it can be argued that the single period models still have considerable advantages over the multi-period models. For this reason, only single period models will be discussed in this chapter.

The two models which will be examined are the original Markowitz model, and a simplification of this model which will be called Sharpe's Index Model. The reason for this is that although the Index Model was originally proposed by Markowitz (1959), the considerable simplifications and advantages of this model were first fully discussed by Sharpe (1963).

The considerable importance of the original Markowitz Model is bourne out by the fact that it is still in practical use today and, in fact, the concepts underlying the model have been extended to many other areas of the theory of finance, besides portfolio selection. On the other hand, Sharpe's Index Model has resulted in the establishment of a new theory, called Capital Market Theory (c.f. for example Sharpe (1964), Lintner (1965) and Fama (1968)), which has developed into an extremely important field of stock market research, incorporating such concepts as systematic and unsystematic risk, beta coefficients, market equilibrium, etcetera. Since this model is merely a simplification of the original Markowitz model, it is surprising that so little discussion of the differences in the results obtained using these two models has been forthcoming in the literature. In this chapter, it is this aspect of the portfolio selection problem which will be investigated.

In Section 2, the mathematical formulation of these two models is presented as well as some basic definitions required in the subsequent analysis. Section 3 discusses the data which was used in this study, while Section 4 presents the first empirical results, those for both the original Markowitz Model and Sharpe's Index Model, assuming no upper limit is imposed on the funds which can be invested in any one security. In Section 5, an alternative index approach using principal components is discussed, and empirical results analagous to those of Section 4 are given. Section 6 considers the effect of different upper bounds on the amount which can be invested in any one security, while finally, Section 7 presents some overall conclusions which can be drawn from the results presented in this chapter.

7.2 THE MODELS

In this section the two portfolio selection models which will be empirically examined in this chapter are formulated, and a brief summary of some previous comparisons between these two models are given. The formulation below is brief, with only the most basic terms and notation being defined. For a more detailed formulation the reader is referred to one of the numerous text books which now cover the field of portfolio selection and, in particular, to Sharpe (1970), which formed the basis of most of the work presented below.

Both the original Markowitz Model and Sharpe's Index Model assume that there are basically two factors to be considered in choosing a portfolio. They are:

(i) the expected return on the portfolio; and
(ii) the risk associated with this return (which can be measured by the standard deviation of the return).

From consideration of these two factors, Markowitz (1959) proposed the following definitions.

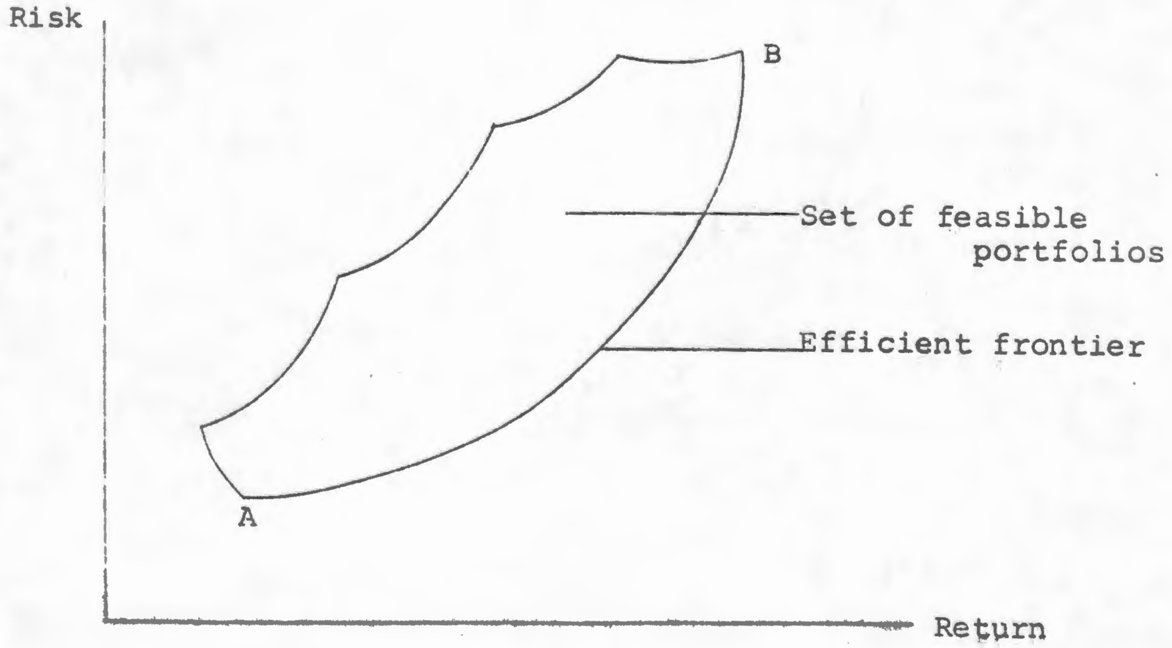
Definition 7.1: Efficient Portfolio

A portfolio is said to be "efficient" if it is impossible to obtain a greater expected return without incurring greater risk, and it is impossible to obtain smaller risk without decreasing expected return.

(The definition of an "inefficient portfolio" follows directly from the above definition, as being any portfolio which is not efficient.)

Definition 7.2: Efficient Frontier

The set of all efficient portfolios forms a boundary which is called the efficient frontier (see Figure 7.1 below).



F I G U R E 7.1

It is assumed that no rational investor will choose an inefficient portfolio, and the portfolio selection problem is therefore to derive the set of all efficient portfolios (that is, the efficient frontier) so that the individual can choose (from this efficient set) the single portfolio which best meets his return/risk requirements.

The set of efficient portfolios is obtained as follows:

$$\text{Minimize} \quad -\lambda E_p + \sigma_p^2$$

$$\text{for all} \quad \lambda \geq 0;$$

$$\text{Subject to} \quad \sum_{i=1}^{n} X_i = 1;$$

$$X_i \geq 0, \quad i = 1,2,\ldots,n \;;$$

plus any other linear equality constraints imposed by the individual investor; plus

$$L_i \leq X_i \leq U_i \quad \text{for all} \quad i = 1,2,\ldots,n \;;$$

where

$$E_p = \sum_{i=1}^{n} X_i E_i \;, \quad \text{and}$$

$$\sigma_p^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} X_i X_j \sigma_{ij} \;, \quad \text{and where}$$

n = the number of securities considered for the portfolio;

X_i = the proportion of funds invested in the i^{th} security $(i = 1,2,\ldots,n)$;

E_i = the expected return on the i^{th} security $(i = 1,2,\ldots,n)$;

σ_{ij} = covariance between the i^{th} and j^{th} securities $(i = 1,2,\ldots,n \;;\; j = 1,2,\ldots,n)$;

E_p = the expected return on the portfolio;

σ_p = the standard deviation of the portfolio;

U_i = the upper bound on the proportion of funds to be invested in security i, and

L_i = the lower bound.

It should be noted that each value of $\lambda \geq 0$ yields a different efficient portfolio, and there are thus an infinite number of efficient portfolios. Fortunately, however, many of these portfolios constitute holding the same securities but in differing proportions. Thus, the problem is to find only the corner portfolios (that is the portfolios where a security either enters or leaves the portfolio) as the entire efficient frontier can be found once this (finite) set of corner portfolios is known. Since σ_p^2 contains terms of the form X_i and $X_i X_j$, the above problem is a quadratic programming problem. Algorithms have been proposed for the solution of such problems, with those of Markowitz (1959), Wolfe (1959), and Sharpe (1970) being among the more popular methods. The solution of these algorithms yield the set of all corner portfolios.

Unfortunately, use of the above model is, in general, limited by the large number of estimates required - the n expected returns and the n(n-1)/2 distinct covariances must be estimated. To overcome this difficulty, Markowitz (1959) proposed that the returns of the different securities are related only through their common relationship with some basic underlying factor, which can be measured by some overall market index. Formally then, the model assumes that the return on security j at time t (R_{jt}) is linearly related to some index I_t at the same time t, as follows:

$$R_{jt} = \alpha_j + \beta_j I_t + u_{jt}$$

where α_j and β_j are parameters (which must be estimated), and u_{jt} is a stochastic term with zero mean and variance $\sigma^2_{u_j}$ for all t.

If the model further assumes that,

(i) $Cov(u_{j;t}; I_t) = 0$ for all $j = 1,2,\ldots,n$ and for all t;

(ii) $Cov(u_{j;t}; u_{j;t+s}) = 0$ for all $s \neq 0$; and

(iii) $Cov(u_{j;t}; u_{i;t}) = 0$ for all $i \neq j$ (the j^{th} security's u is uncorrelated with any other security's u - that is, the two securities are only related through their mutual relationship to the index);

then it can be shown that

$$E_j = \alpha_j + \beta_j E_I$$

$$\sigma^2_j = \beta^2_j \sigma^2_I + \sigma^2_{u_j}$$

$$\sigma_{ij} = \beta_i \beta_j \sigma^2_I, \quad i \neq j$$

where E_I = the expected level of the index, and

σ^2_I = the variance of the index.

Hence, the analyst need only estimate

(a) the parameters α_j, β_j and $\sigma^2_{u_j}$ for each security, and

(b) E_I and σ^2_I.

Thus, the model requires only 3n+2 estimates which is considerably less than n(n-1)/2 + n, for large n.

In addition, Sharpe (1963) showed that if an index is used for this purpose, it becomes unnecessary to multiply out all the entries of the covariance matrix. He showed that, by setting

$$\beta_p = \sum_{j=1}^{n} X_j \beta_j \ ,$$

the model could be rewritten as

$$\text{Minimize} \quad -\lambda\left(\sum_{j=1}^{n} \alpha_j X_j + \beta_p E_I\right) + \left(\beta_p \sigma_I^2 + \sum_{i=1}^{n} X_i \sigma_{u_i}^2\right)$$

$$\text{for all} \quad \lambda \geq 0 \ ;$$

$$\text{Subject to} \quad \sum_{i=1}^{n} X_i \beta_i = \beta_p \ ;$$

$$\sum_{i=1}^{n} X_i = 1 \ ;$$

$$X_i \geq 0 \ ;$$

plus any other linear constraints or bounds.

Since the only quadratic terms which appear in the above formulation are the squared ones, the covariance matrix has been reduced to a diagonal form, which simplifies the solution of the problem. This theory can easily be extended to the case of more than one index (c.f. Sharpe (1970)).

Before concluding this section, a brief description of three comparative studies of the Markowitz and Index Models which have appeared in the literature, will be given. The three studies are those of Wallingford (1967), Cohen and Pogue (1967) and Affleck-Graves and Money (1976). All three studies found that the Markowitz full covariance model was the most appropriate model. In addition, Cohen and Pogue found that the single index model was superior to the two index model, while Wallingford found that the two index model generated more efficient portfolios than the single index model. Affleck-Graves and Money, on the other hand, found that there was very little difference between the one and two index models, but that a five index model provided a substantial improvement although it was still not as satisfactory as the original Markowitz Model. Thus, at present, there appears to be no clear cut evidence as to exactly what empirical differences exist between the models. This is even more obvious when one considers that Cohen and Pogue imposed an upper limit of 0,05 on the proportion of funds which could be invested in any one security, while the study of Affleck-Graves and Money imposed no upper limit. As Francis and Archer (1971) conclude,

> "considering the contradictory evidence .... it is not possible to generalize about the relative efficiency of various simplified models."

7.3 THE DATA

The models examined in this chapter are more suitable for long term portfolio selections (say, longer than 6 months) than for short term "in and out" trading. Thus, the data used for the empirical tests presented in this chapter differs from the data used in the remainder of this thesis, in that annual data was used as opposed to the weekly data used elsewhere.

Data was collected on 157 securities quoted on the Johannesburg Stock Exchange over the period January 1962 to December 1976. These 157 securities are listed in Appendix H and were taken as the universe of all possible securities. The yearly return on each security was computed for the period 1962 to 1976, using the formula

$$R_i(t) = \frac{P_i(t) + D_i(t) - P_i(t-1)}{P_i(t-1)}$$

where $R_i(t)$ is the return on the i^{th} security in the t^{th} period ;

$P_i(t)$ is the price of the i^{th} security at the end of the t^{th} period; and

$D_i(t)$ is the total of all dividends paid to the i^{th} security in the t^{th} period.

As was mentioned in the previous section, use of the Markowitz Model necessitates a large number of estimates.

Therefore, the tests in this chapter are based on samples of size 50 from the universe of 157 securities. Five random samples, each of size 50, were drawn and those securities chosen for each sample are listed in Tables I.1 to I.5 of Appendix I.

In performing empirical tests on portfolio selection models, the researcher is faced with an additional problem of estimation, namely how to obtain estimates of the return on individual securities for periods in the past, since the return on these periods is now known. To overcome this problem, the results presented in this chapter refer to the portfolio which would have been selected for the year 1976 (that is which would have been purchased in January 1976 and sold in December 1976). The best estimate of the risk associated with each security's return, was assumed to be the standard deviation of the annual return on that security over the previous fourteen years (that is, 1962-1975 inclusive). The expected return on the individual securities for the year 1976 were taken as the actual return that was achieved in that year. In other words, the results presented indicate the portfolio that would have been selected if each of the estimated returns had in fact been perfect. It must be stressed that this does not mean that, since perfect information was used, the risk was zero and thus that the risk aspect of the problem was ignored. Rather, each return was still assumed to have a risk associated with it (the standard deviation over the past 14 years). This policy was adopted for all five samples and for all of the empirical results presented.

The method by which the indices used for the Index Models were constructed must also be mentioned. As suggested by Cohen and Pogue (1967), an aggregate performance index which is more pertinent to the particular universe of 157 securities used, was constructed rather than using one of the standard published indices. This index is the unweighted average of the return on all securities in the universe, and was constructed for each of the years 1962 to 1976. The actual value of this index in 1976 was taken as an estimate of the expected level of the index in that year, as was discussed above for the individual securities. An estimate of the standard deviation of the index for 1976 was calculated using the level of the index in the previous 14 years. Also, since the return on each security, as well as the level of the index are known for the fourteen year period 1962 to 1975 (and would have been known at the beginning of 1976), estimates of β_i and $\sigma^2_{u_i}$ for the model

$$R_i = \alpha_i + \beta_i I + u_i ,$$

were obtained by regressing R_i on I, the level of the index. These estimates were obtained for each of the securities in each of the five samples and were then used as input for the model.

Now, in addition to considering the single index model, two multi-index models were examined; a two index model and a four index model. For the two index case, the universe of 157 securities was broken into what were felt to be two distinct classes, one containing all 57 Mining shares (numbers 1 to 57 in Appendix H), and the other all 100 Industrial and Financial shares (numbers 58 to 157 in Appendix H). Indices of these two sections were then constructed in exactly the same manner as the aggregate index discussed above. In addition, the covariance between the two indices was estimated. Once again, estimates of β_{i1}, β_{i2}, and $\sigma^2_{u_i}$ in the model

$$R_i = \alpha_i + \beta_{i1}I_1 + \beta_{i2}I_2 + u_i$$

were computed for each security in each of the five samples, using classical regression techniques.

Finally, for the four index problem, the universe of 157 securities was divided into four distinct groups as follows:

I_1 = Gold Index (34 securities - numbers 1 to 34 in Appendix H)

I_2 = Minerals Index (23 securities - numbers 35 to 57 in Appendix H)

I_3 = Financials Index (26 securities - numbers 58 to 83 in Appendix H)

I_4 = Industrials Index (74 securities - numbers 84 to 157 in Appendix H).

As before, indices of these five sections were constructed by computing the arithmetic average of the returns on all securities included in the respective subdivisions. In addition, the covariance between each pair of indices was estimated, and estimates of the parameters $\beta_{i1}, \beta_{i2}, \beta_{i3}, \beta_{i4}$, and $\sigma^2_{u_i}$ in the model

$$R_i = \alpha_i + \beta_{i1}I_1 + \beta_{i2}I_2 + \beta_{i3}\ I_3 + \beta_{i4}I_4 + u_i,$$

were again computed using classical regression techniques.

Before concluding this section, mention must be made of the algorithms used to solve both the original Markowitz Model and Sharpe's Index Model problems. The algorithms used in obtaining all results presented in this chapter were those of Sharpe (1970). These algorithms were programmed and listings, together with a brief description of how they may be used, can be found in Appendix J.

7.4 EMPIRICAL RESULTS FOR THE CASE OF NO UPPER BOUNDS

In this section, the results of the first and most obvious comparison between the two models are presented, namely for the case where no upper limit (or restriction) is imposed on the proportion of funds to be invested in any one security. As mentioned in the previous section, the analysis is based on five different random samples of 50 securities each.

Using the computer programs listed in Appendix J, efficient frontiers were obtained when using the Markowitz

Model, the one index model, the two index model, and the four index model, for each of the five samples under consideration. Sketches of these efficient frontiers are presented in Figures 7.2 to 7.6 below.

On studying these sketches, it is apparent that the general conclusion of the three studies discussed in Section 2, namely that the Markowitz Model generates the dominant efficient frontier, is not bourne out by the results obtained for these five samples. In fact, only two samples (numbers 1 and 3) indicated agreement with this statement, while samples 2, 4, and 5 indicated that the Markowitz Model did not generate the dominant frontier. On considering this result, it is apparent that there is no theoretical reason why the Markowitz Model should generate a dominant frontier and that, as demonstrated by these five samples, it will not, in general, do so.

What is of far more importance, is the fact that Sharpe's Index Model is a simplification of the Markowitz Model. Hence, the sketches should be used to examine the effect of this simplification; that is, to determine how close the efficient frontiers generated by different index models are to the efficient frontier generated by the Markowitz Model. Visually, such a comparison is rather difficult, but it appears as if the index models do not (with the exception of Sample 5) provide an efficient frontier very similar to that obtained when using the original Markowitz Model.

FIGURE 7.2

SAMPLE 1

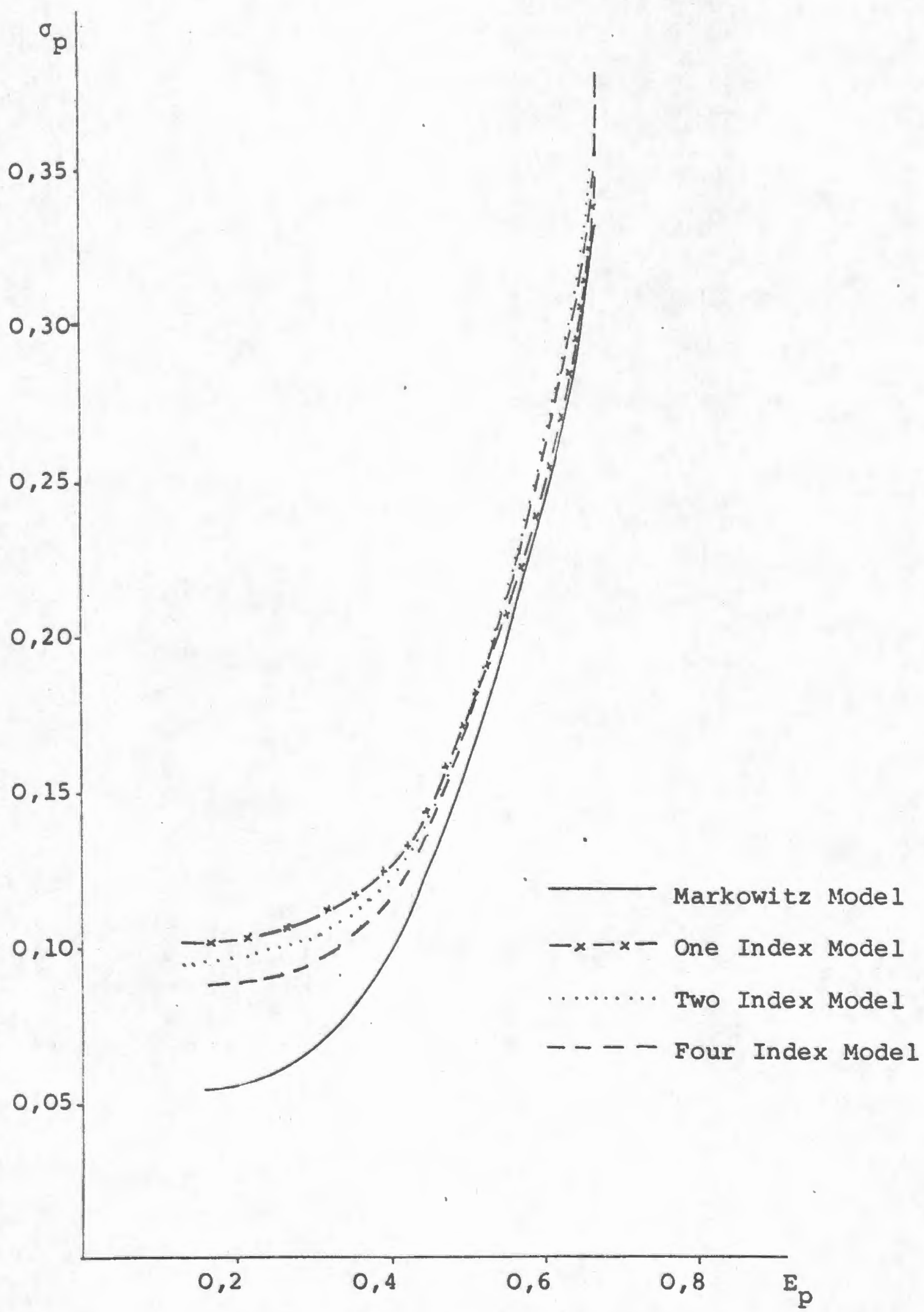


FIGURE 7.3

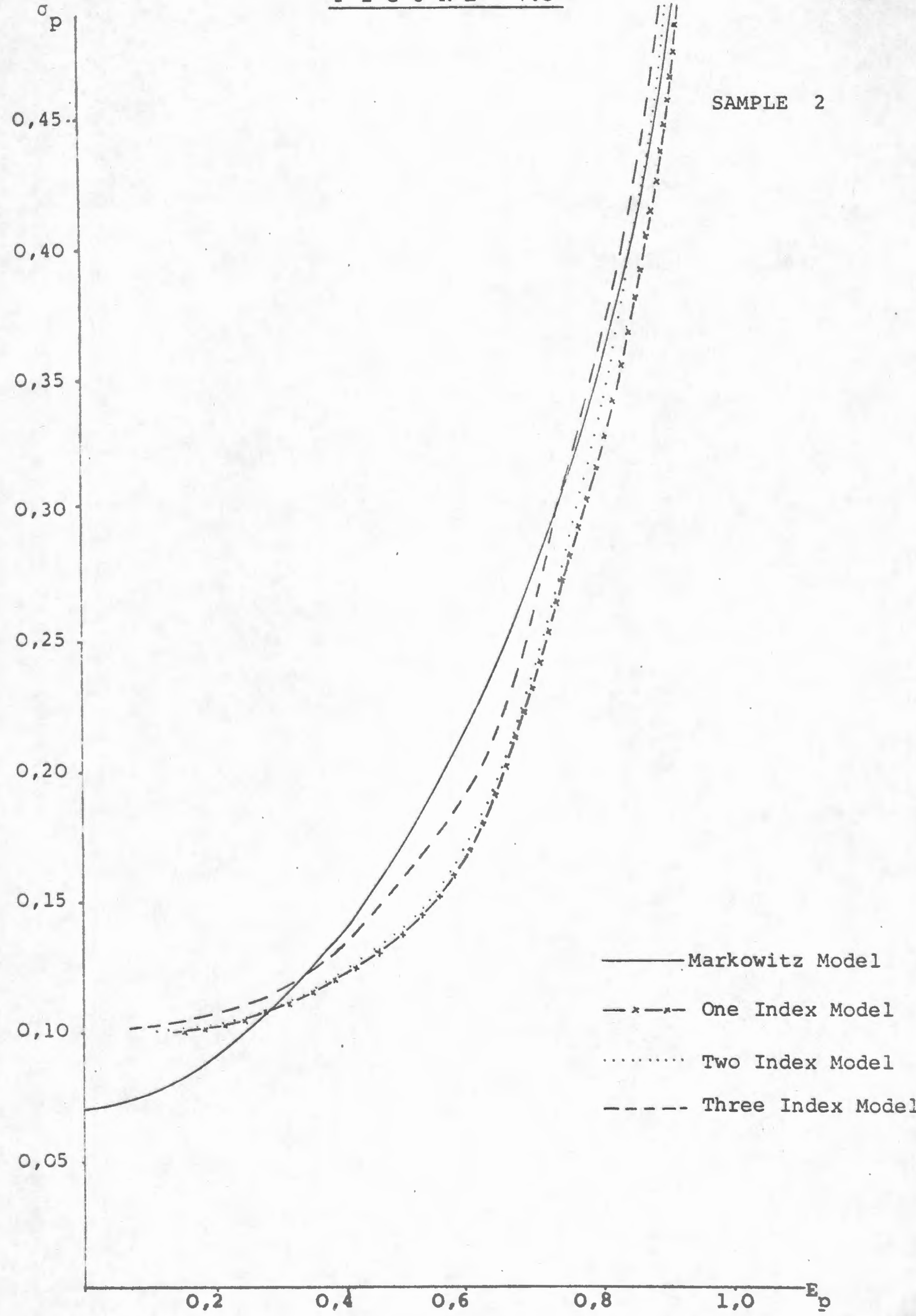

σp
0,45
0,40
0,35
0,30
0,25
0,20
0,15
0,10
0,05
0,2
0,4
0,6
0,8
1,0
Ep
SAMPLE 2
Markowitz Model
One Index Model
Two Index Model
Three Index Model

FIGURE 7.4

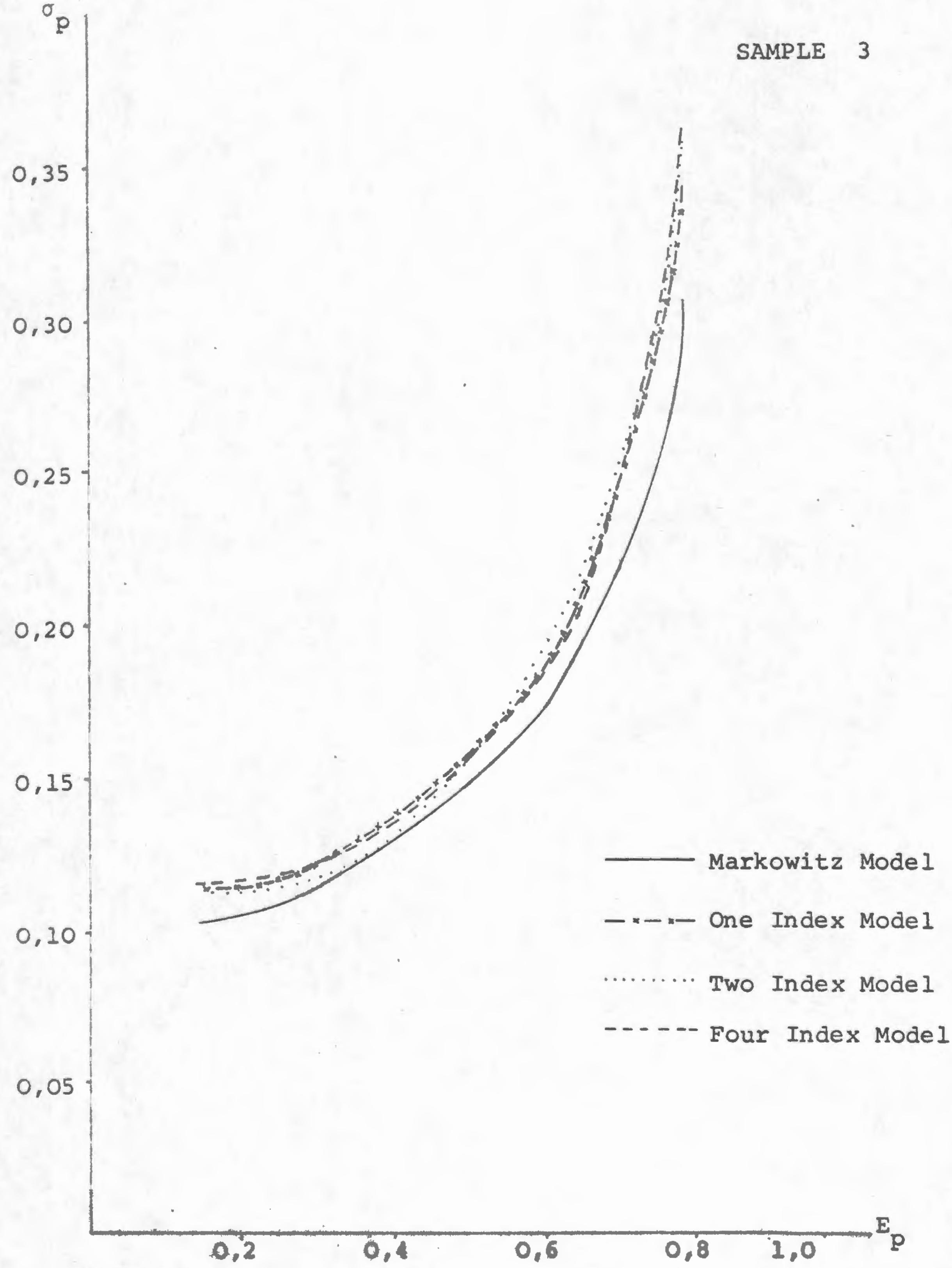
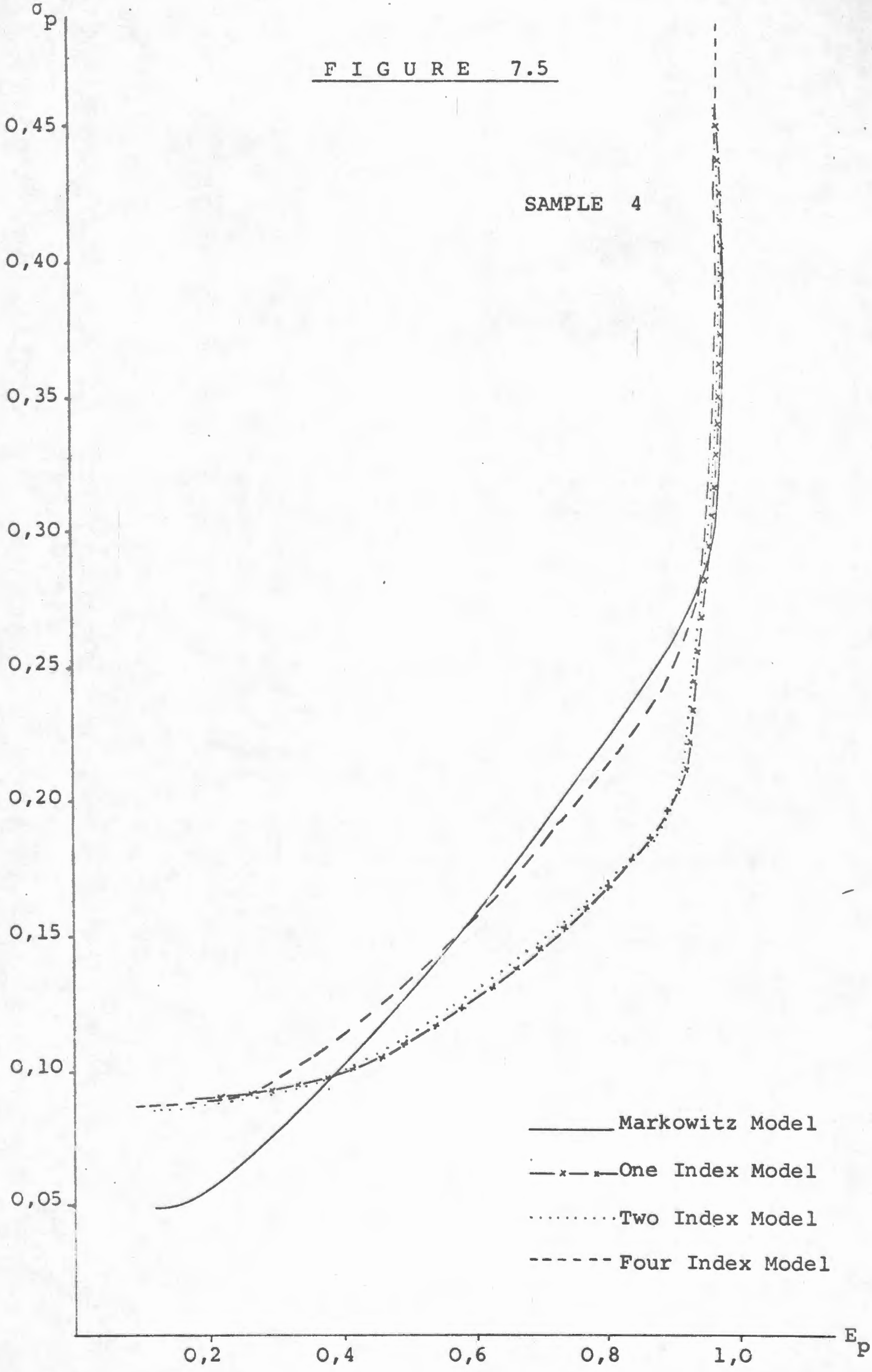
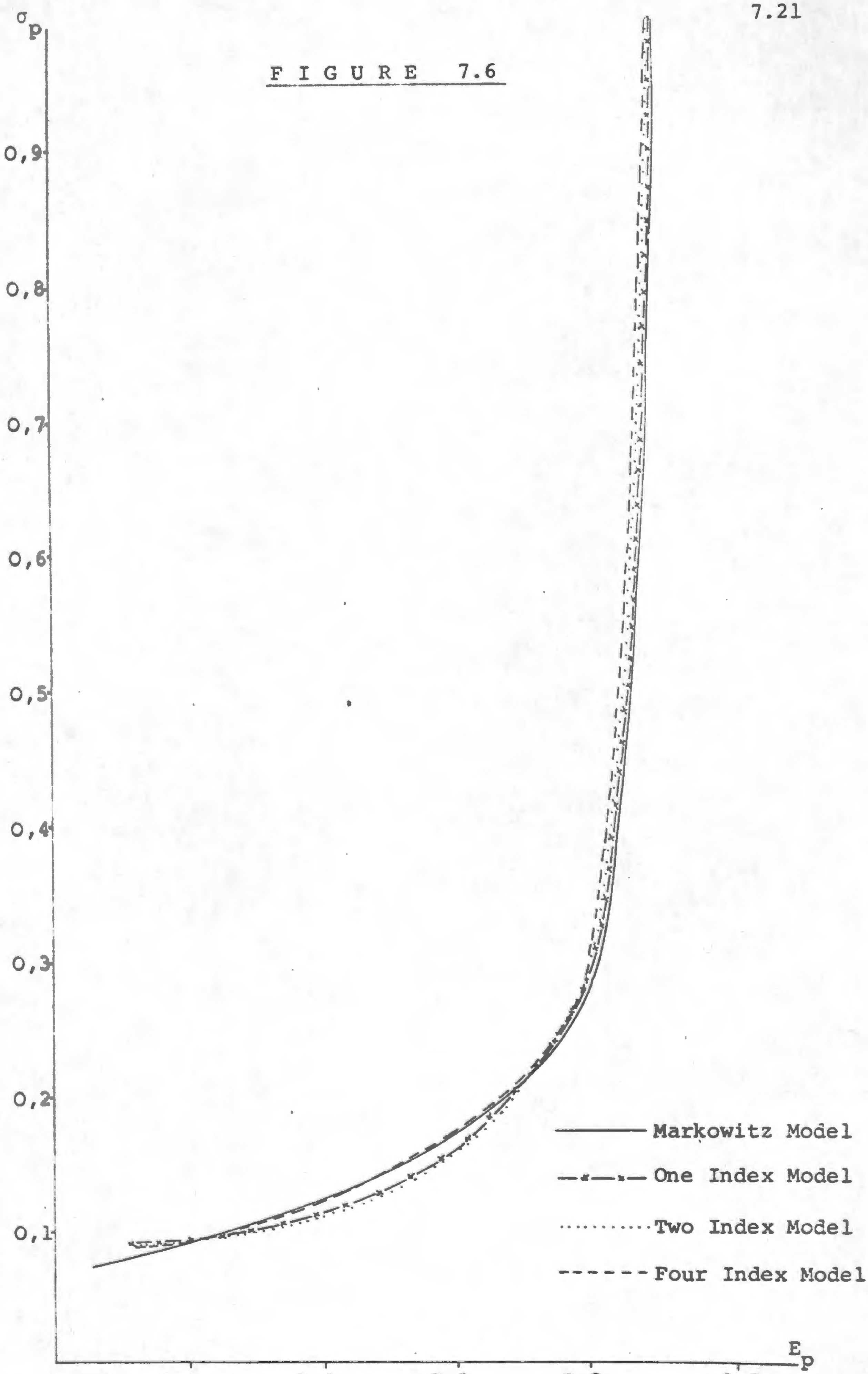

σp
SAMPLE 3
0,35
0,30
0,25
0,20
0,15
0,10
0,05
Markowitz Model
One Index Model
Two Index Model
Four Index Model
0,2
0,4
0,6
0,8
1,0
Ep

FIGURE 7.5



σ_p
0,45
0,40
0,35
0,30
0,25
0,20
0,15
0,10
0,05
SAMPLE 4
Markowitz Model
One Index Model
Two Index Model
Four Index Model
0,2
0,4
0,6
0,8
1,0
E_p

FIGURE 7.6


σp
0,9
0,8
0,7
0,6
0,5
0,4
0,3
0,2
0,1
Ep
0,2
0,4
0,6
0,8
1,0
Markowitz Model
One Index Model
Two Index Model
Four Index Model

In addition, it would appear that the sketches tend to confirm the findings of Affleck-Graves and Money (1976), that the one index model is not out-performed by the two index model. In fact, it appears as if the two models produce very similar efficient frontiers (for all five samples) and this could possibly account for the conflicting findings of Wallingford (1967) and Cohen and Pogue (1967). This is a somewhat strange result since, intuitively, it can be argued that increasing the number of indices will increase the accuracy of the Index Model. However, in the case of the one and two index models, the additional information obtained from using two indices is not sufficient to result in a significant improvement of the results. On examining the four index model, it is apparent that this model does produce an efficient frontier closer to the Markowitz efficient frontier than either the one or two index models, although it is not possible to say exactly how much closer it is.

In order to examine the questions raised above more closely, it was decided to select a typical portfolio from each of the models under examination, and to compare the performance of these portfolios. In order to choose a "typical portfolio", a value of 0,08969 was chosen for λ. This value corresponds to Farrar's (1962) average risk aversion factor for 23 American mutual funds. Thus, while obviously not the exact value of λ that every investor would use, the choice of 0,08969 does enable the choice of a single portfolio which will probably be in the region of those considered by most investors, and can therefore be used for comparison purposes. The portfolios that would have been chosen (using this value of λ) by each of the four models considered (the Markowitz, one index, two index, and four index models) for each of the five samples, are presented in Tables 7.1 to 7.5 below.

TABLE 7.1 SAMPLE 1

SECURITY	% of Funds to be Invested in Each Security			
	Markowitz Model	One Index	Two Index	Four Index
Vryheid	13,14	9,47	7,72	11,75
Wankie	14,62	6,24	5,95	7,63
MTD	7,35	17,01	17,29	15,53
Ass Mang	31,36	18,93	16,53	15,89
S A Mang		8,39	8,55	8,43
Rooiberg	2,21	3,68	4,33	3,60
Zaaiplaats	18,96	20,82	17,27	15,70
Tweefontein	3,18	10,68	13,19	13,15
Rand Carbide	9,18	3,45	8,09	6,04
Becketts		1,33	1,08	2,28
E_p	36,86	39,61	38,21	38,08
σ_p	9,46	13,10	12,30	11,72

TABLE 7.2 SAMPLE 2

SECURITY	Markowitz Model	One Index	Two Index	Four Index
W R Cons	11,58	1,89	7,60	10,50
Amal Coll	29,20	23,99	22,83	18,21
Clydesdale		17,23	17,66	9,78
Nat Ants	16,23	23,24	22,56	18,26
Welgedacht	2,53	11,12	11,25	9,40
Rooiberg	6,84	5,50	5,11	3,86
Cons Murch	3,33	1,38	0,93	
Msauli		2,01	1,91	
Anglos	16,90			8,83
T C Lands		2,13		2,07
Trade & Ind	13,39	10,46	10,15	17,86
Anglo Alpha		1,05		
Woolworths				1,23
E_p	41,20	53,90	54,80	48,85
σ_p	14,04	14,94	15,21	15,70

TABLE 7.3 SAMPLE 3

SECURITY	Markowitz Model	One Index	Two Index	Four Index
W R Cons		2,02		1,33
Clydesdale		22,49	18,49	11,60
Tavistock	45,91	35,33	31,39	33,04
Mid Wits	3,60			
T C Lands		2,32	1,18	6,44
Trade & Ind	19,85	11,79	12,18	18,66
Eveready		11,93	16,97	7,58
Massey Ferg	7,03			0,84
Greatermans	2,06			
Tongaat		0,60	1,71	
Remgro			1,63	3,61
Tweefontein	21,55	13,52	16,45	16,90
E_p	58,70	53,05	49,99	53,20
σ_p	16,96	16,26	15,32	16,33

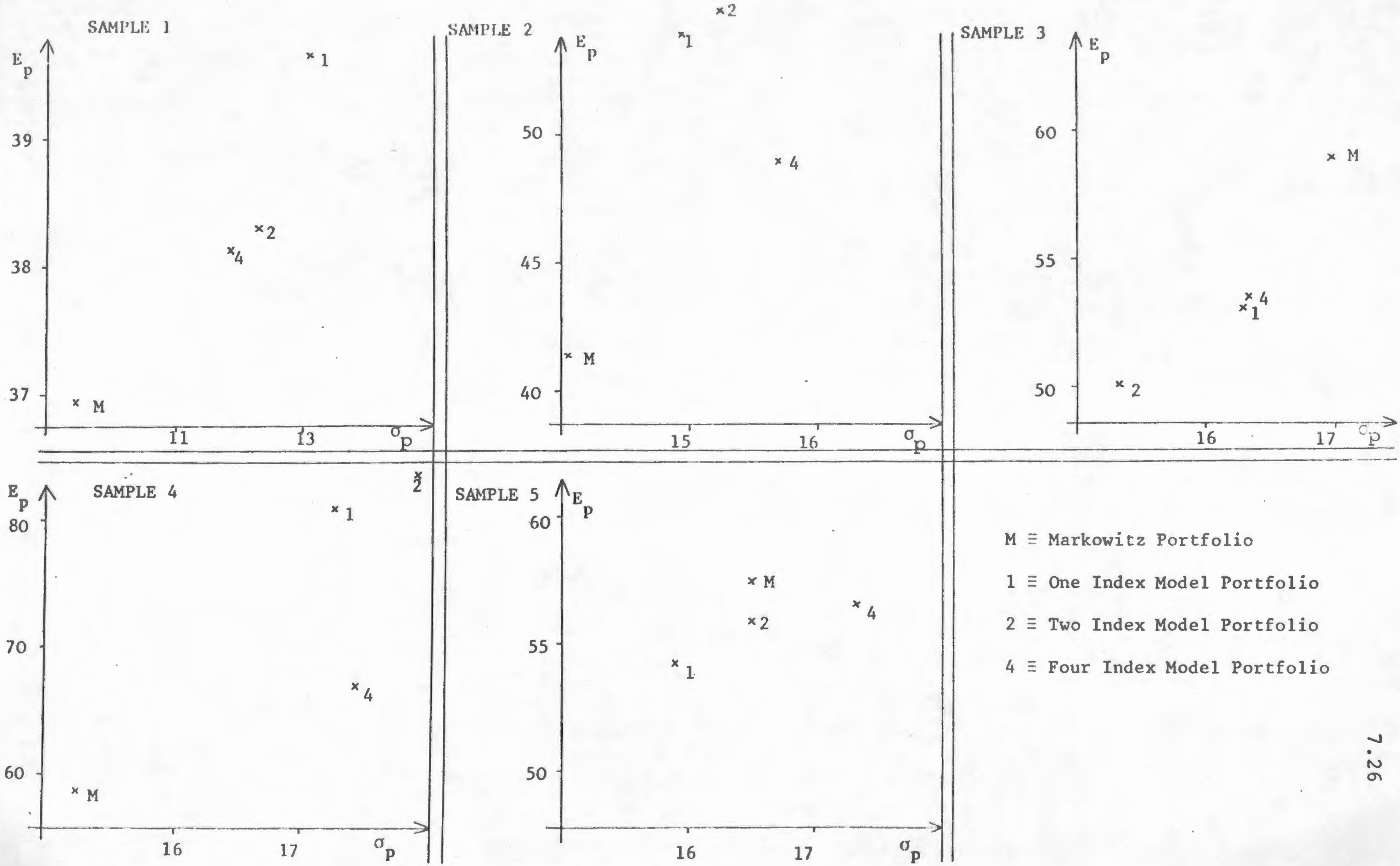
TABLE 7.4 SAMPLE 4

SECURITY	Markowitz Model	One Index	Two Index	Four Index
Randfontein	6,67	2,90	8,02	9,69
Wes Drie	22,79			0,67
A T Coll	1,36	18,57	17,12	9,59
Apex		18,15	17,98	12,45
Nat Ammo	43,96	19,96	22,22	20,31
Vierfontein		2,23	5,40	1,41
Witbank	2,76	20,06	19,23	12,23
Cons Gold	3,61			8,05
PP Cement		5,73	0,61	
I L Back	16,28	4,92	5,14	5,46
Cullinan	2,57			3,59
Lion		7,48	4,28	15,83
Remb Beh				0,52
E_p	57,97	80,27	83,21	66,20
σ_p	15,25	17,26	17,91	17,42

TABLE 7.5 SAMPLE 5

SECURITY	Markowitz Model	One Index	Two Index	Four Index
Randfontein	2,25	2,55	5,69	7,11
Pres Brand	5,03			
Amal Coll		28,11	28,03	25,93
Tavistock	47,94	33,56	32,52	31,41
S A Mang	6,92	9,32	10,22	7,31
Rooiberg	16,00	6,63	6,72	6,21
Cons Murch		2,30	2,45	
T C Lands		4,19	1,29	2,11
Trade & Ind	19,71	11,78	11,92	17,26
Natal Chem	2,15			
Woolworths				2,66
Safmarine		1,56	1,16	
E_p	57,28	54,00	55,60	56,48
σ_p	16,47	15,91	16,50	17,32

From these tables it would appear that the "visual" conclusions discussed above are confirmed, namely that the four index models provide the best approximation to the Markowitz Model (expected value of the four index portfolio is closer than the expected value of either the one index or two index portfolio, to the expected value of the Markowitz portfolio, for all five samples). Also, plotting the expected return and standard deviation of return for all four models in risk return space (as presented in Figure 7.7 below), it can clearly be seen that the four index model provides more satisfactory results than either the one index or two index models. It can also be seen that it is often difficult to distinguish between one and two index models, or to determine which provides the better approximation in general.


SAMPLE 1
SAMPLE 2
SAMPLE 3
SAMPLE 4
SAMPLE 5
E_p
σ_p
M ≡ Markowitz Portfolio
1 ≡ One Index Model Portfolio
2 ≡ Two Index Model Portfolio
4 ≡ Four Index Model Portfolio

F I G U R E 7.7

Finally, Table 7.6 below presents some interesting averages derived from the five samples.

TABLE 7.6

AVERAGE	Markowitz	One Index	Two Index	Four Index
E_p	50,40	56,17	56,36	52,56
σ_p	14,44	15,49	15,45	15,70
$\% E_p$		12,28	12,49	5,06
$\% \sigma_p$		10,11	9,26	10,28
$\|\% E_p\|$		18,42	19,60	9,37
$\|\% \sigma_p\|$		13,12	13,13	11,76

In this table, the row $\% E_p$ refers to the average (for all five samples) percentage difference between the expected return on the portfolios obtained using the Index Model and the Markowitz Model. That is,

$$\% E_p = 1/5 \sum_{j=1}^{5}((E_{I;j} - E_{M;j})/E_{M;j})\times 100,$$

where $E_{I;j}$ is the expected return on the portfolio obtained using the Index Model for the j^{th} sample, and $E_{M;j}$ is the expected return on the Markowitz Model portfolio for the j^{th} sample.

The $\% \sigma_p$ is similarly defined. The row $|\% E_p|$ is similarly defined except that the absolute percentage difference is used; that is

$$|\% E_p| = 1/5 \sum_{j=1}^{5}(|E_{I;j} - E_{M;j}|/E_{M;j})\times 100,$$

with the $|\% \sigma_p|$ being similarly defined.

It is felt that the averages of most interest in Table 7.6 are the final two rows, that is the averages of the absolute percentage differences in the returns and the standard deviations of the portfolios. From the figures presented in the table, it can be deduced that, on average, the expected return using the four index model would differ from the expected return on the Markowitz Model by approximately 9% (compared to approximately 18% and 20% for the one and two index models respectively), and that the standard deviation would differ by approximately 12% (compared to 13% for both the one and two index models). These results confirm the belief that the four index model is more suitable than either the one or two index models.

Unfortunately, with only 14 years of data available, it was not considered feasible to examine index models with more than four indices (five estimates are already required) unless further data was available. However, the results presented in this section do give some indication of the magnitude of the error which can be expected when using the simplified approach with the different index models. It is reasonable to assume that as more and more indices are added, the errors will get smaller and smaller. But, as the addition of extra indices necessitates additional estimates, it would appear that for the case of no upper bounds, the original Markowitz approach should be used unless a fairly large multi-index model is considered feasible. Also, it has been established that there is probably very little extra benefit to be derived from the use of a two index model, and therefore, if no upper bounds are to be imposed on the proportion of funds invested in any one security, it is better to use the simpler one index model.

7.5 PRINCIPAL COMPONENT INDICES AND PORTFOLIO SELECTION

The method used in the construction of the indices used in Sharpe's Index Models will obviously have some impact on the performance of these models. For this reason, it was decided to investigate the effects of using principal component indices for the various index models. These principal component indices have an additional advantage in that the problem is numerically easier to solve if the indices are uncorrelated, as was argued by Sharpe (1963).

In order to use this approach, a principal component analysis was performed on each of the 5 samples (of 50 securities each). Then, four indices were constructed using the weights of the first four components, for each of the five samples under consideration. For the one index model, only the first principal component index was used. For the two index model, the first and second principal component indices were used, while for the four index model the first four principal component indices were used. It should be noted that the principal component indices are based on the particular sample under consideration, and not on the universe of 157 securities as was the case in the previous section. Results analogous to those of the previous section, are presented for the principal component index situation below.

Firstly, Figures 7.8 to 7.12 provide sketches of the efficient frontiers obtained using the Markowitz Model (which is obviously the same efficient frontier as that presented in Section 4), and the one index, two index, and four index models using the principal component indices. A visual comparison between Figures 7.2 to 7.6 and Figures 7.8 to 7.12 is rather difficult, but it does not appear as if the use of principal component indices results in a significant difference. In fact, with the possible exception of Sample 1, it appears as if the principal component index models result in efficient frontiers that are even further from the Markowitz efficient frontier than those obtained using the indices computed in the previous section.

To examine the situation more closely, individual portfolios were once again chosen for each of the models and for all five samples, using a value of 0,08969 for λ. The relevant statistics for these portfolios are presented in Tables 7.7 to 7.11 below.

FIGURE 7.8

SAMPLE 1

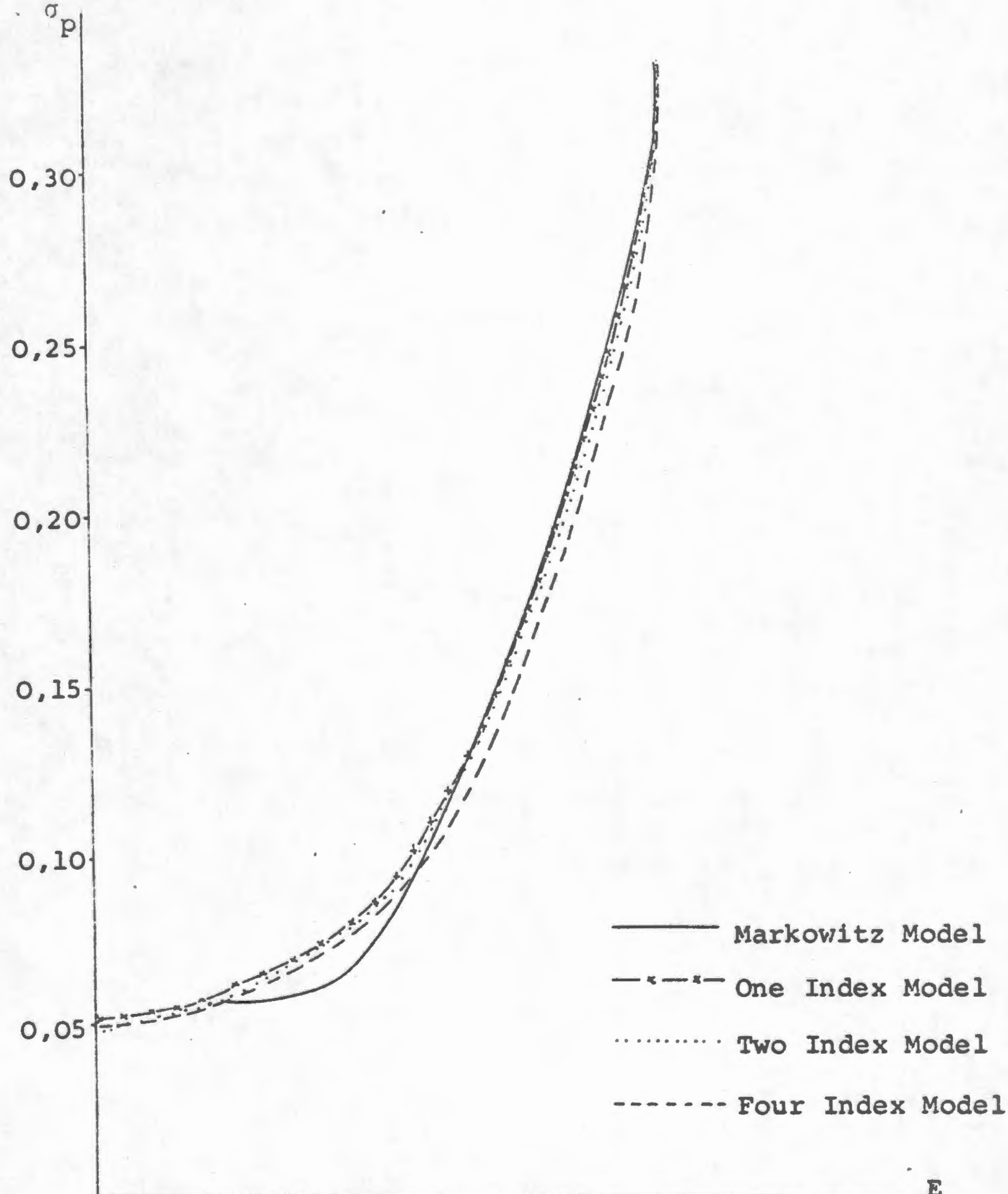
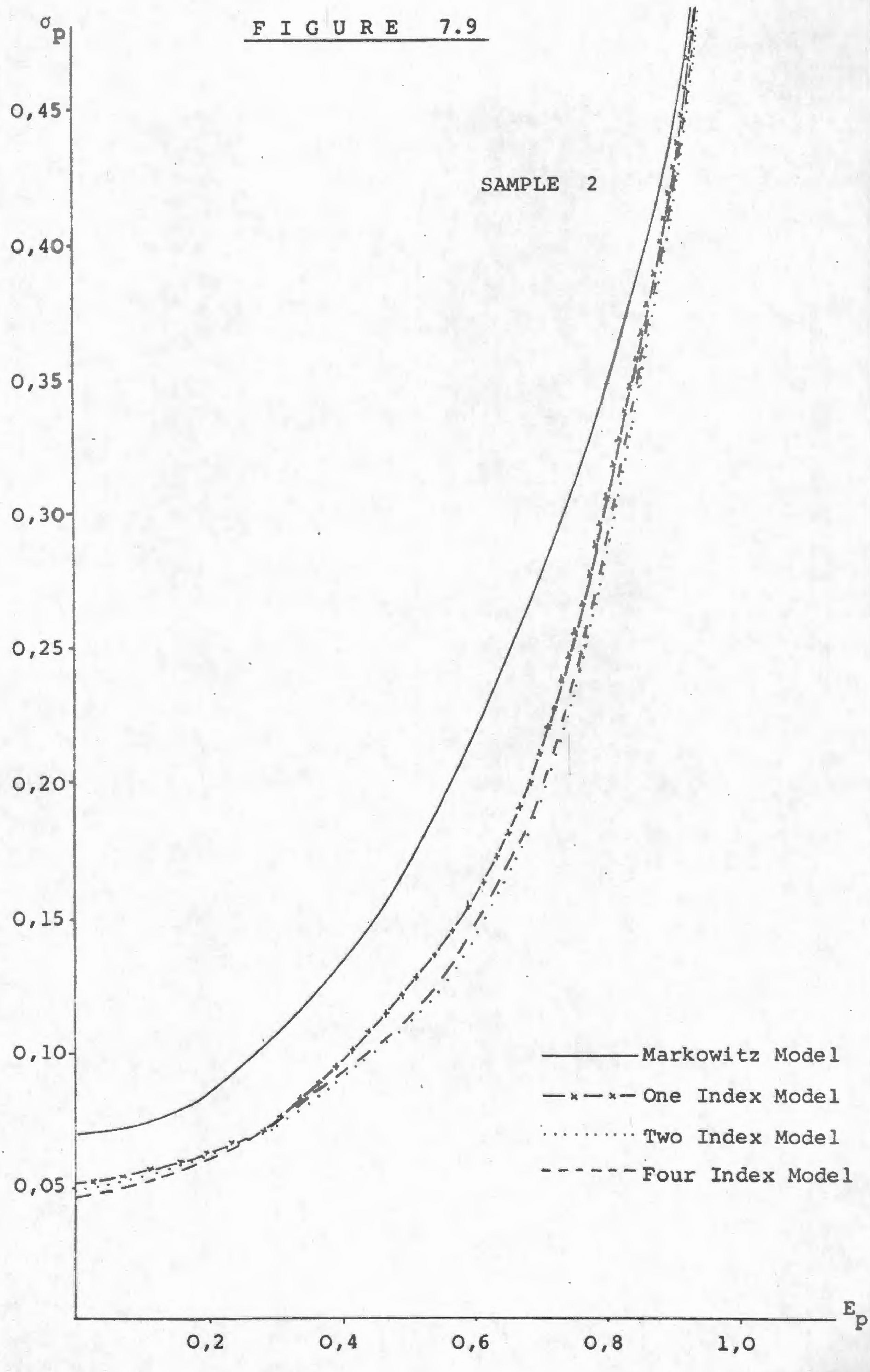


FIGURE 7.9



σp
0,45
0,40
0,35
0,30
0,25
0,20
0,15
0,10
0,05
SAMPLE 2
Markowitz Model
One Index Model
Two Index Model
Four Index Model
0,2
0,4
0,6
0,8
1,0
Ep

FIGURE 7.10

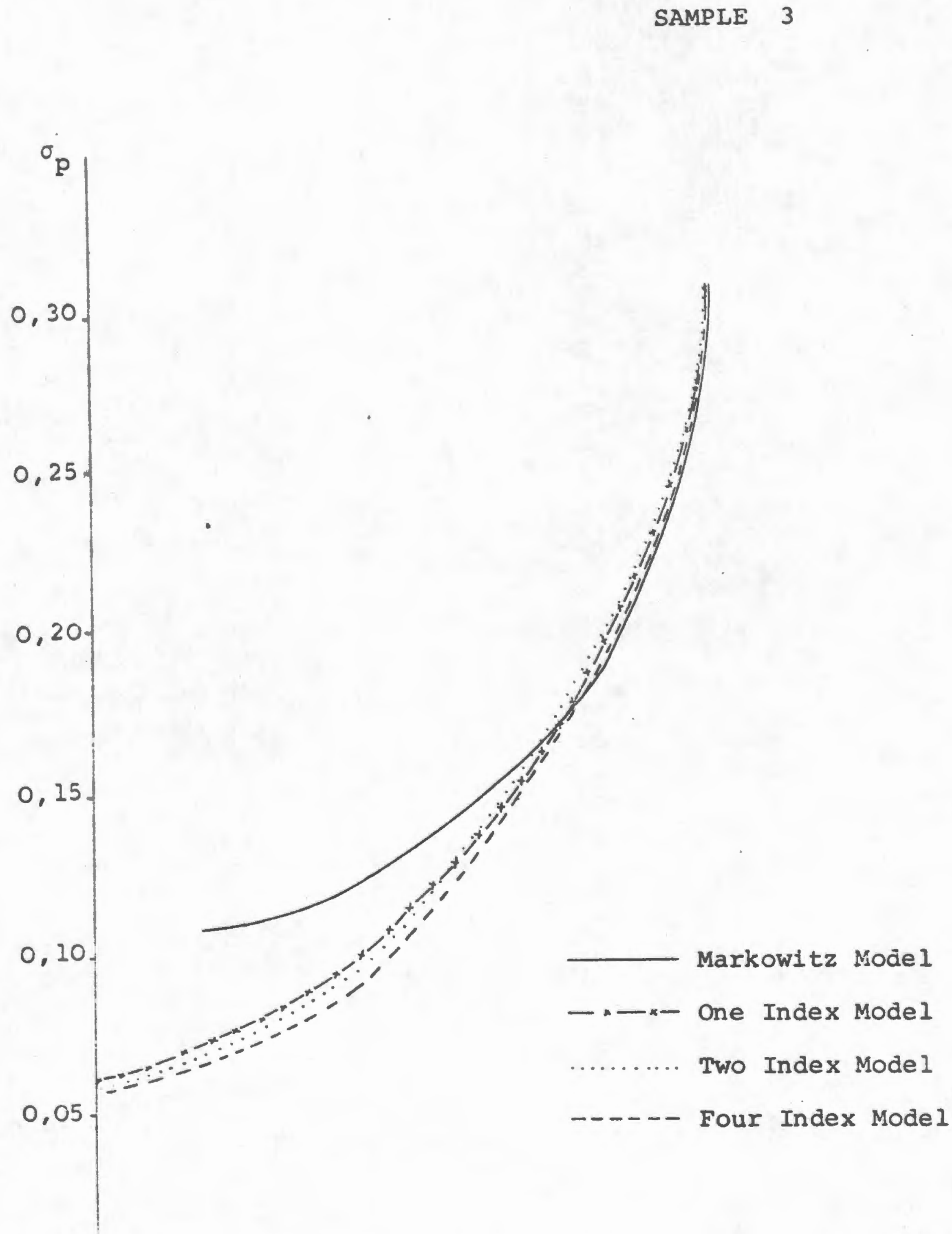

SAMPLE 3
σp
0,30
0,25
0,20
0,15
0,10
0,05
0,2
0,4
0,6
0,8
1,0
Ep
Markowitz Model
One Index Model
Two Index Model
Four Index Model

FIGURE 7.11

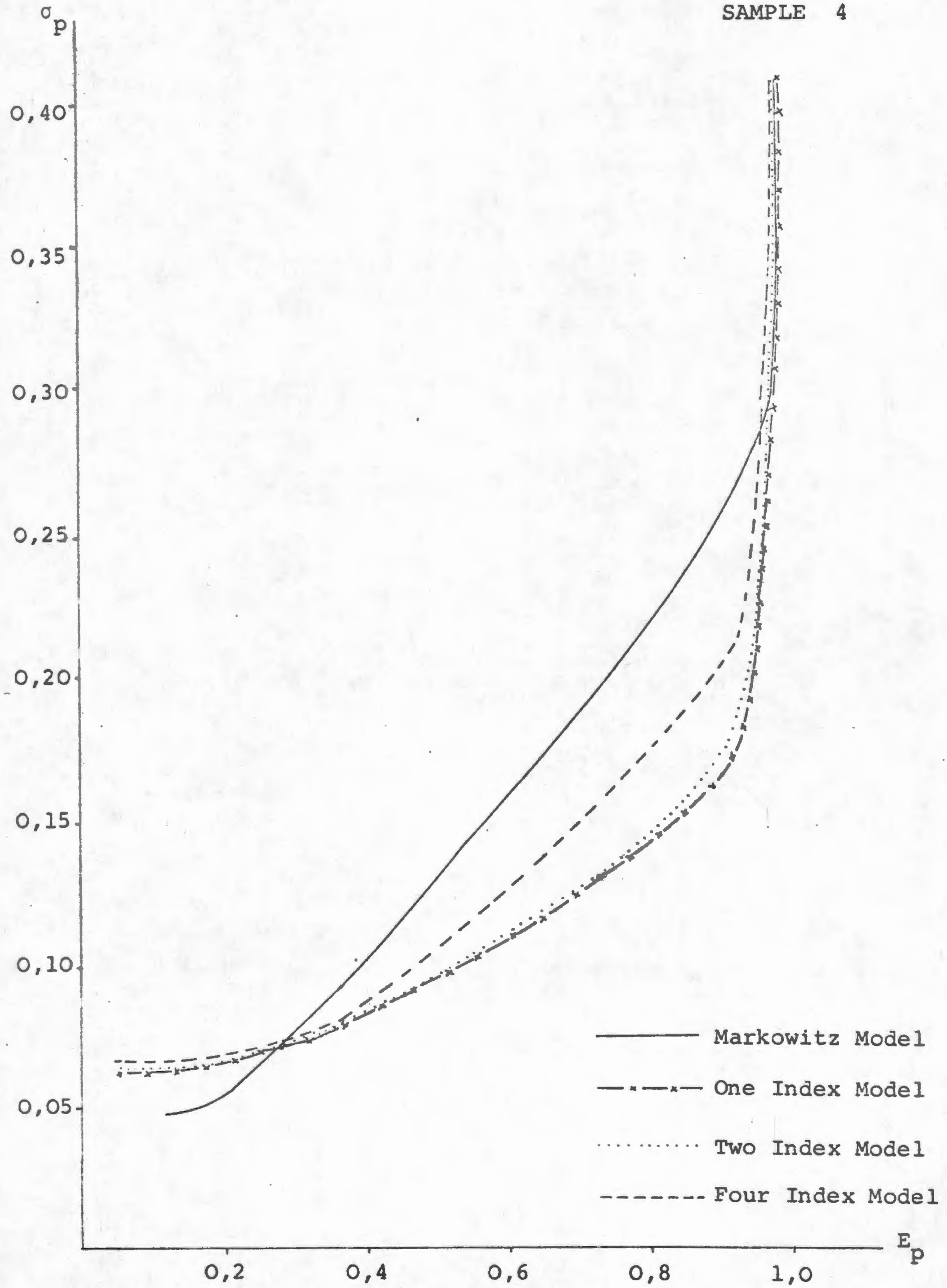

SAMPLE 4
σ_p
0,40
0,35
0,30
0,25
0,20
0,15
0,10
0,05
E_p
0,2
0,4
0,6
0,8
1,0
Markowitz Model
One Index Model
Two Index Model
Four Index Model

FIGURE 7.12

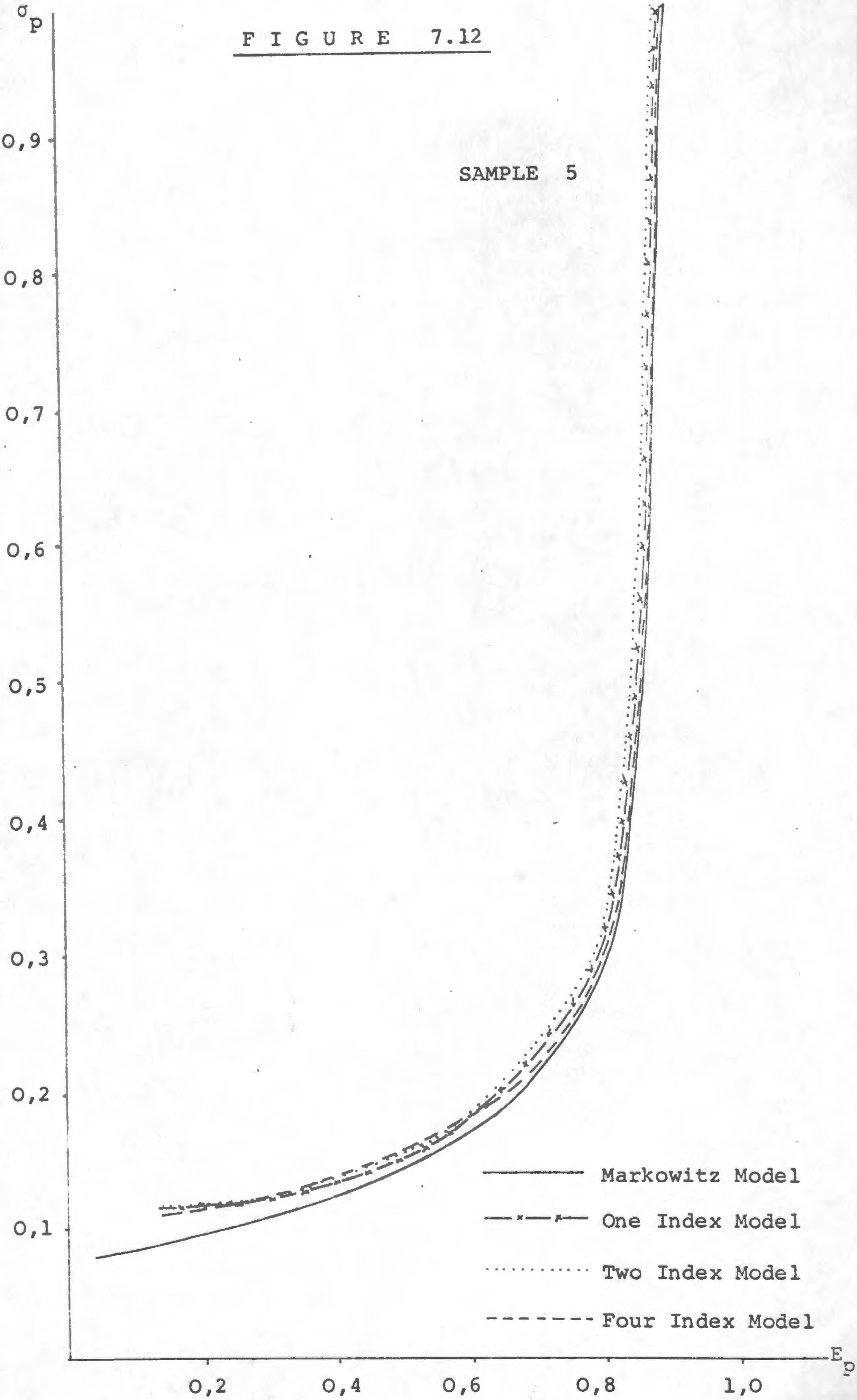

SAMPLE 5
σ_p
0,9
0,8
0,7
0,6
0,5
0,4
0,3
0,2
0,1
E_p
0,2
0,4
0,6
0,8
1,0
Markowitz Model
One Index Model
Two Index Model
Four Index Model

TABLE 7.7 SAMPLE 1

SECURITY	% of Funds to be Invested in Each Security			
	Markowitz Model	One Index	Two Index	Four Index
Vryheid	13,14	7,60	7,89	11,46
Wankie	14,62	4,91	6,66	6,15
MTD	7,35	7,18	9,83	9,36
Ass Máng	31,36	15,07	13,83	12,38
S A Mang		5,65	4,82	2,77
Rooiberg	2,21	5,31	5,30	4,90
Zaaiplaats	18,96	22,40	21,95	21,60
Tweefontein	3,18	11,58	10,90	13,01
Rand Carb	9,18	4,68	5,99	8,23
Becketts		7,55	6,53	5,78
Ed L Bateman		5,03	4,49	4,36
Busaf		1,09		
Safmarine		1,95	1,81	
E_p	36,86	40,89	39,98	40,75
σ_p	9,46	11,75	11,21	11,20

TABLE 7.8 SAMPLE 2

SECURITY	Markowitz Model	One Index	Two Index	Four Index
Grootvlei				1,74
W R Cons	11,58	13,62	13,49	14,20
Stilfontein		2,77	2,70	0,11
Amal Coll	29,20	15,98	15,57	18,72
Clydesdale		11,03	11,83	10,22
Nat Ants	16,23	24,35	24,02	22,28
Welgedacht	2,53	16,14	15,89	13,59
Rooiberg	6,84	3,90	4,11	9,45
Cons Murch	3,33			2,94
Msauli		1,59	1,63	1,74
Anglos	16,90			
T C Lands		2,39	2,25	
Trade & Ind	13,39	8,23	8,51	3,75
Prem Mill				1,26
E_p	41,20	57,48	57,28	54,29
σ_p	14,04	13,40	13,21	13,48

TABLE 7.9 SAMPLE 3

SECURITY	Markowitz Model	One Index	Two Index	Four Index
Clydesdale		15,89	10,03	8,72
Tavistock	45,92	37,97	36,87	37,32
Mid Wits	3,60		4,44	2,77
T C Lands		4,03	2,65	1,61
Tweefontein	21,54	14,70	13,23	13,52
Trade & Ind	19,85	10,92	10,78	13,48
Eveready		10,24	11,81	10,18
Massey Ferg	7,03	0,50	2,86	3,09
Greatermans	2,06			
Tongaat		2,51	5,01	7,97
Remgro		3,24	2,32	1,34
E_p	58,70	52,70	48,83	49,80
σ_p	16,96	14,59	13,47	13,25

TABLE 7.10 SAMPLE 4

SECURITY	Markowitz Model	One Index	Two Index	Four Index
Randfontein	6,67	8,27	7,14	8,13
Wes Drie	22,79			
A T Coll	1,36	19,82	19,12	14,67
Apex		20,14	18,46	15,49
Nat Ammo	43,96	27,31	24,96	26,93
Witbank	2,76	20,99	21,11	12,83
Cons Gold	3,61			5,22
PP Cement			0,52	
I L Back	16,28	3,47	5,97	2,75
Nat Chem			0,24	
Cullinan	2,57			4,16
Lion			2,00	4,91
Remb Beh			0,48	4,91
E_p	57,97	91,69	88,23	76,30
σ_p	15,25	17,03	17,00	16,61

TABLE 7.11 SAMPLE 5

SECURITY	Markowitz	One Index	Two Index	Four Index
Randfontein	2,25	3,88	2,99	2,14
P Brand	5,03			
Amal Coll		26,53	23,91	10,00
Tavistock	47,94	38,47	37,90	43,33
S A Mang	6,92	8,90	10,86	5,75
Rooiberg	16,00	5,74	3,56	13,36
Cons Murch		2,28	0,52	
Fed Mynbou				10,10
T C Lands		2,52	4,37	
Trade & Ind	19,71	10,69	13,51	8,79
Nat Chem	2,15			
Dorman Long				0,64
Woolworths			0,40	3,89
Huletts				1,47
Remgro		0,50	1,56	0,53
Safmarine		0,49	0,42	
E_p	57,28	56,91	56,06	51,24
σ_p	16,47	15,26	14,73	14,75

On comparing the above tables with Tables 7.1 to 7.5, it is immediately obvious that the use of the principal component indices does not improve the performance of the index models and, in fact, produces worse results than the index models of the previous section (in the region of $\lambda = 0{,}08969$ at least). These results are confirmed by the averages presented in Table 7.12 below, which is analogous to Table 7.6 of the previous section.

T A B L E 7.12

AVERAGE	Markowitz Model	One Index	Two Index	Four Index
E_p	50,40	59,93	58,08	54,48
σ_p	14,44	14,41	13,92	13,86
% E_p		19,55	16,15	9,65
% σ_p		2,00	-1,41	-1,80
$\lvert\% E_p\rvert$		23,90	23,73	19,93
$\lvert\% \sigma_p\rvert$		12,35	13,41	12,72

Thus it must be concluded, that principal component indices are not very suitable for use in the Sharpe Index Models, and that use of an index based on the arithmetic average of return is likely to result in a better approximation to the Markowitz efficient frontier.

7.6 THE EFFECT OF UPPER BOUNDS

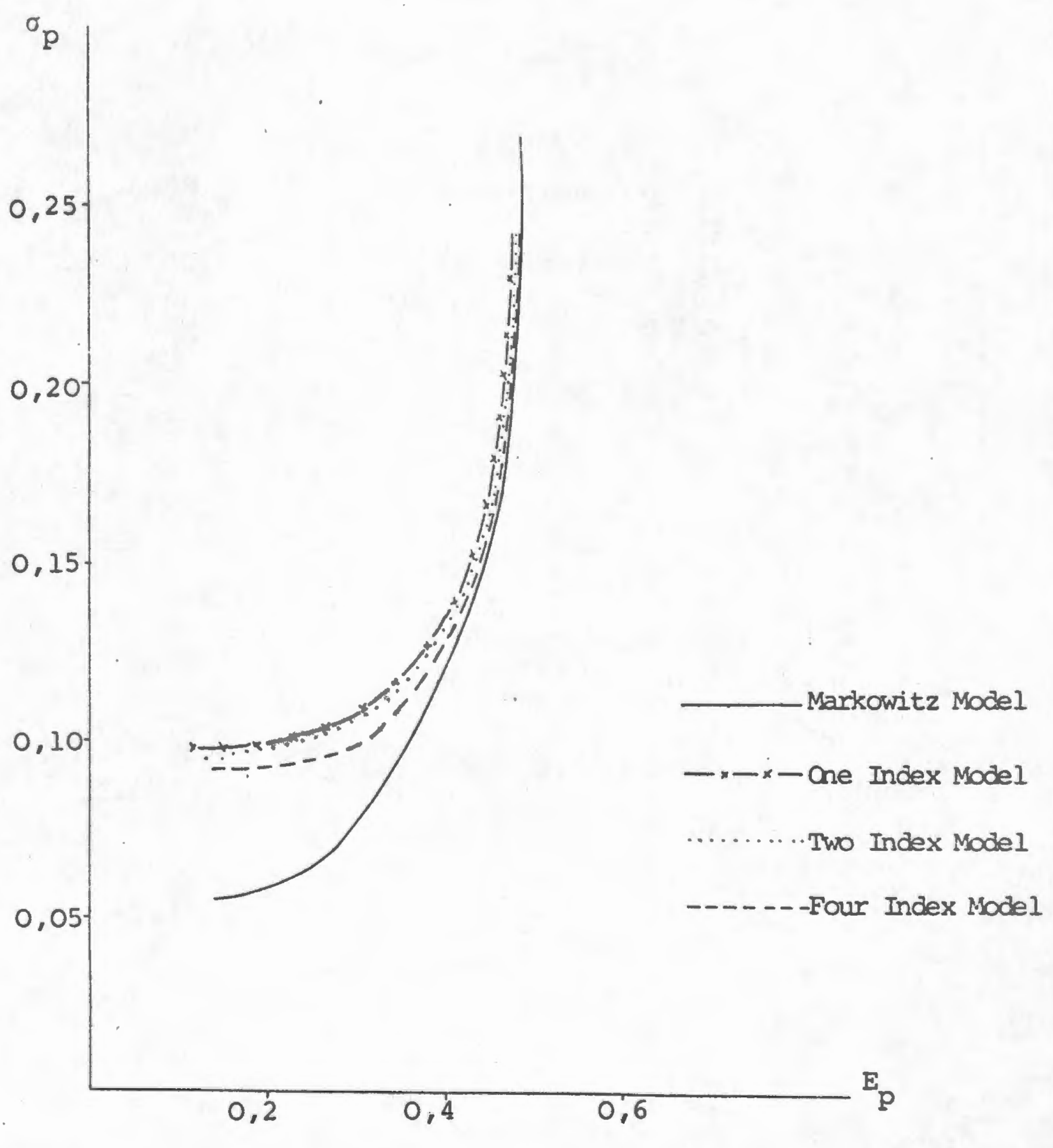
In this section, the performance of the models when different upper bounds are imposed on the proportion of funds that may be invested in any one security, is investigated. Two distinct cases are examined, namely an upper bound of 0,25 on each security, and an upper bound of 0,10 on each security. The reason for not considering a higher upper bound (such as 0,50), is that in almost all cases, except for the first few corner portfolios, the efficient frontier generated is identical to that obtained in the unbounded case. This occurs because (in the unbounded case) it seldom occurs that more than 50% of the total funds are allocated to any one security. A smaller upper bound than 0,10 was also not investigated, since it was felt that a detailed examination of the two cases considered would be sufficient to ascertain differences in the behaviour of the models under the condition of no upper bound and the condition of an imposed upper bound.

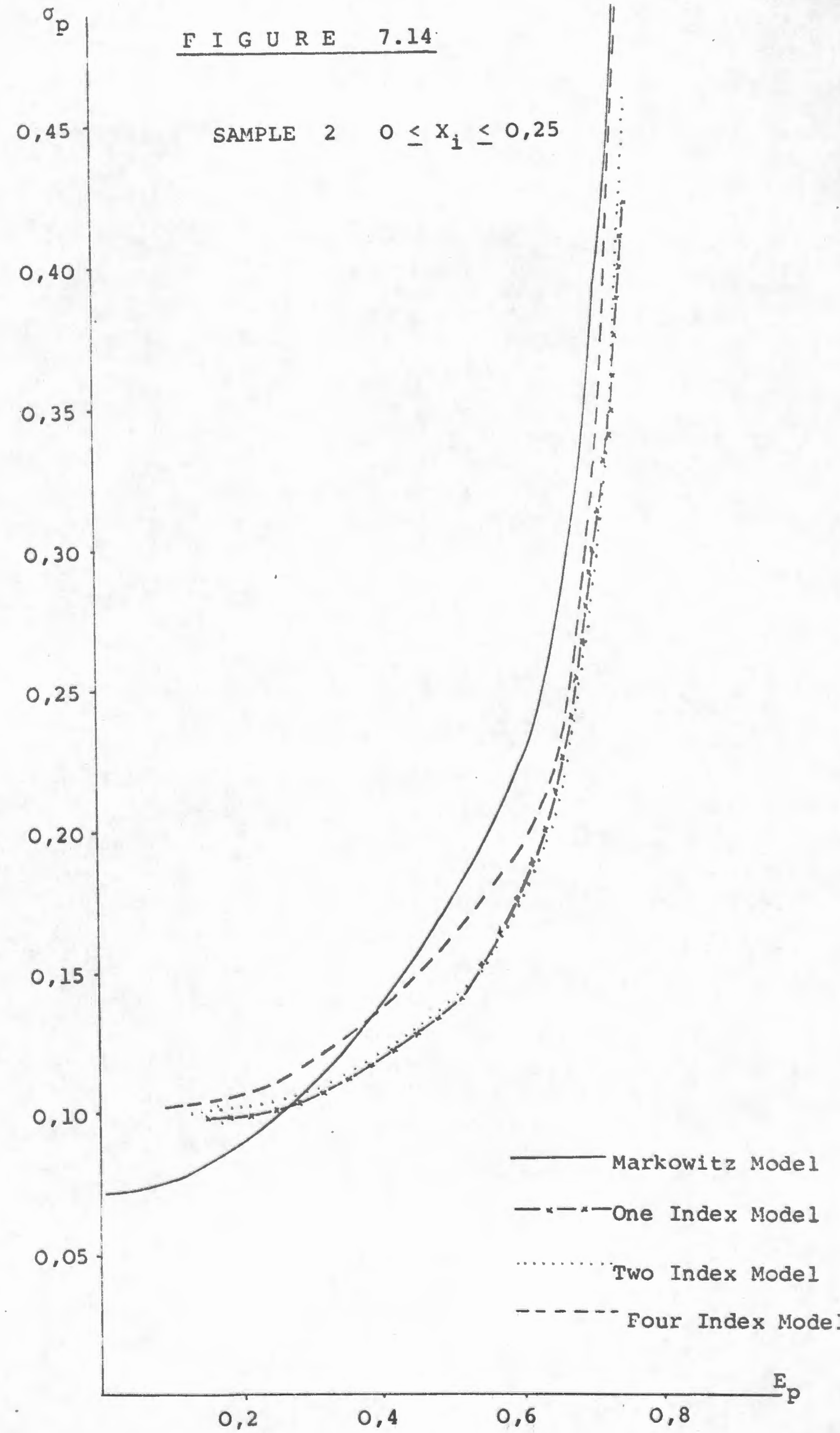
It should be noted that for any of the models under consideration, the unbounded efficient frontier will obviously dominate the bounded efficient frontier and hence, if possible, the investor should choose his portfolio from the unbounded model. However, if the total funds available are very large, or if some external restrictions are imposed on the model, then it becomes necessary to use an upper bounded model. Since this is usually the case for large institutional investors (such as mutual funds), the upper bounded models are of considerable practical importance.

Figures 7.13 to 7.17 below present sketches of the efficient frontier for the Markowitz, one index, two index, and four index models respectively, when an upper bound of 0,25 is imposed. A visual comparison with Figures 7.2 to 7.6 indicates that the index models do not appear to be any closer to the Markowitz efficient frontier in the upper bounded case than they are in the unbounded case (with the possible exception of Sample 3). In fact, in some cases the upper bounded case appears to be much worse than the unbounded case (namely Samples 4 and 5).

FIGURE 7.13

SAMPLE 1 $0 \leq X_i \leq 0,25$





σ_p
FIGURE 7.14
SAMPLE 2 $0 \leq X_i \leq 0,25$
0,45
0,40
0,35
0,30
0,25
0,20
0,15
0,10
0,05
E_p
0,2
0,4
0,6
0,8
Markowitz Model
One Index Model
Two Index Model
Four Index Model

FIGURE 7.15

SAMPLE 3 $0 \leq X_i \leq 0,25$

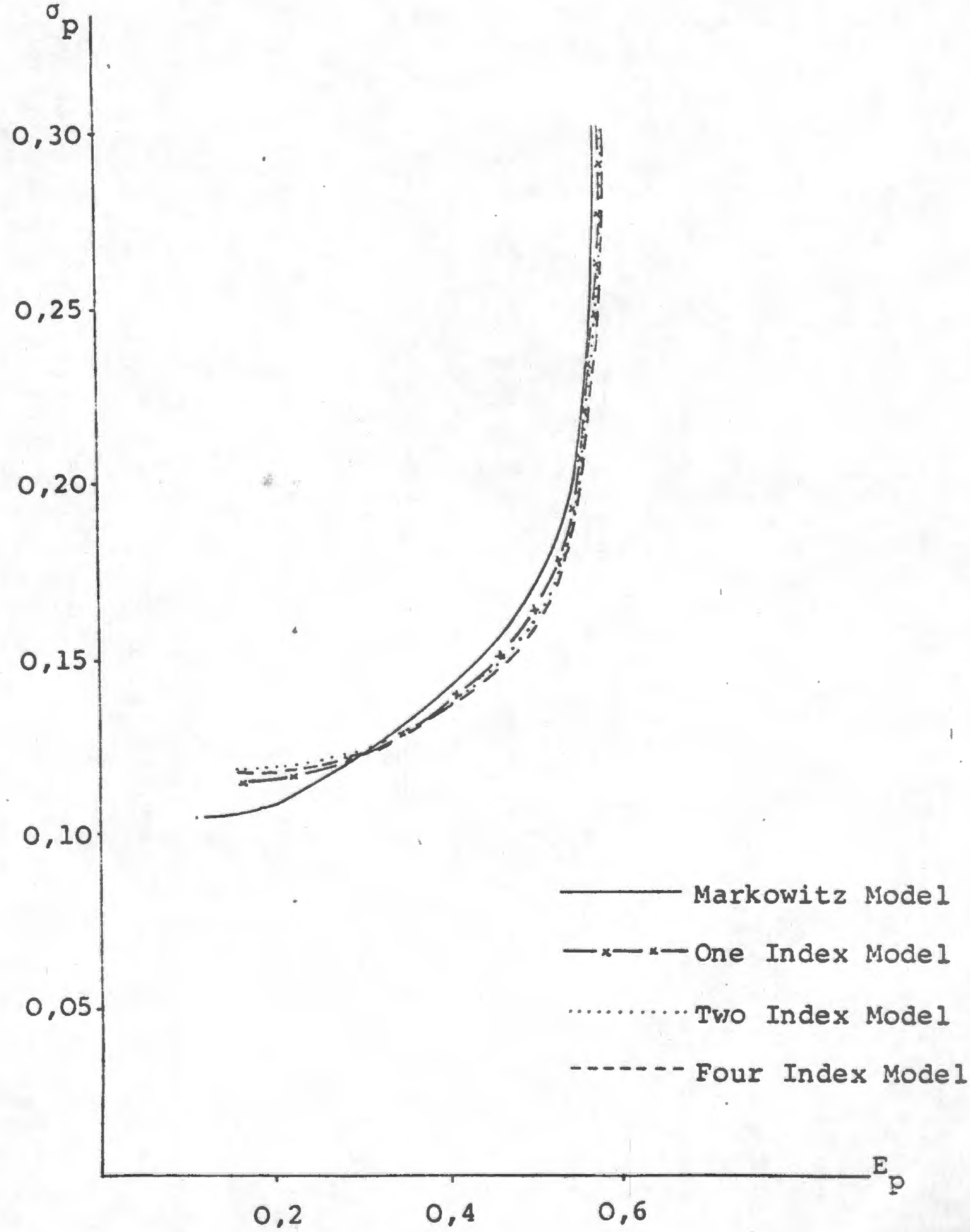


FIGURE 7.16

SAMPLE 4 $0 \leq X_i \leq 0{,}25$

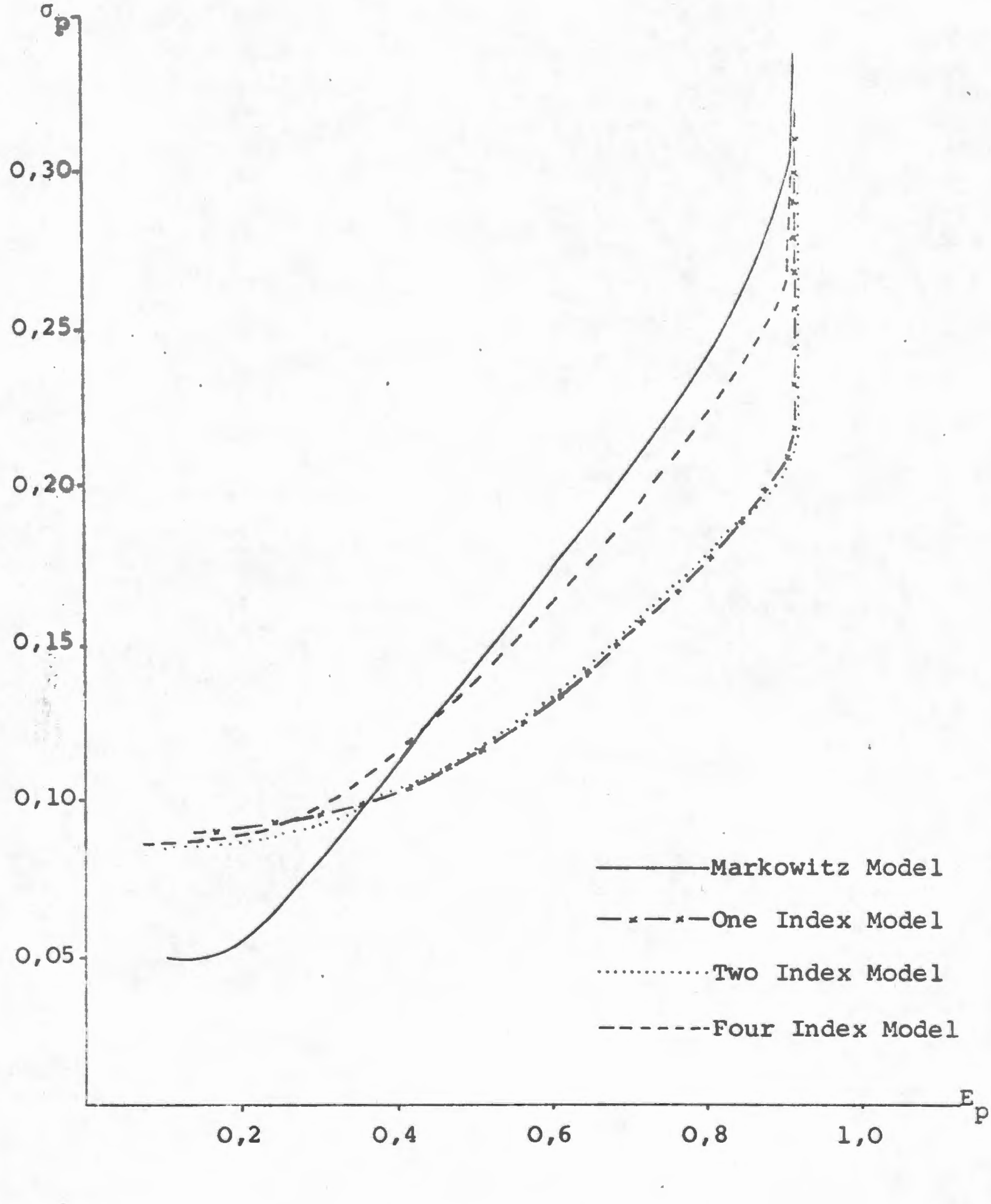
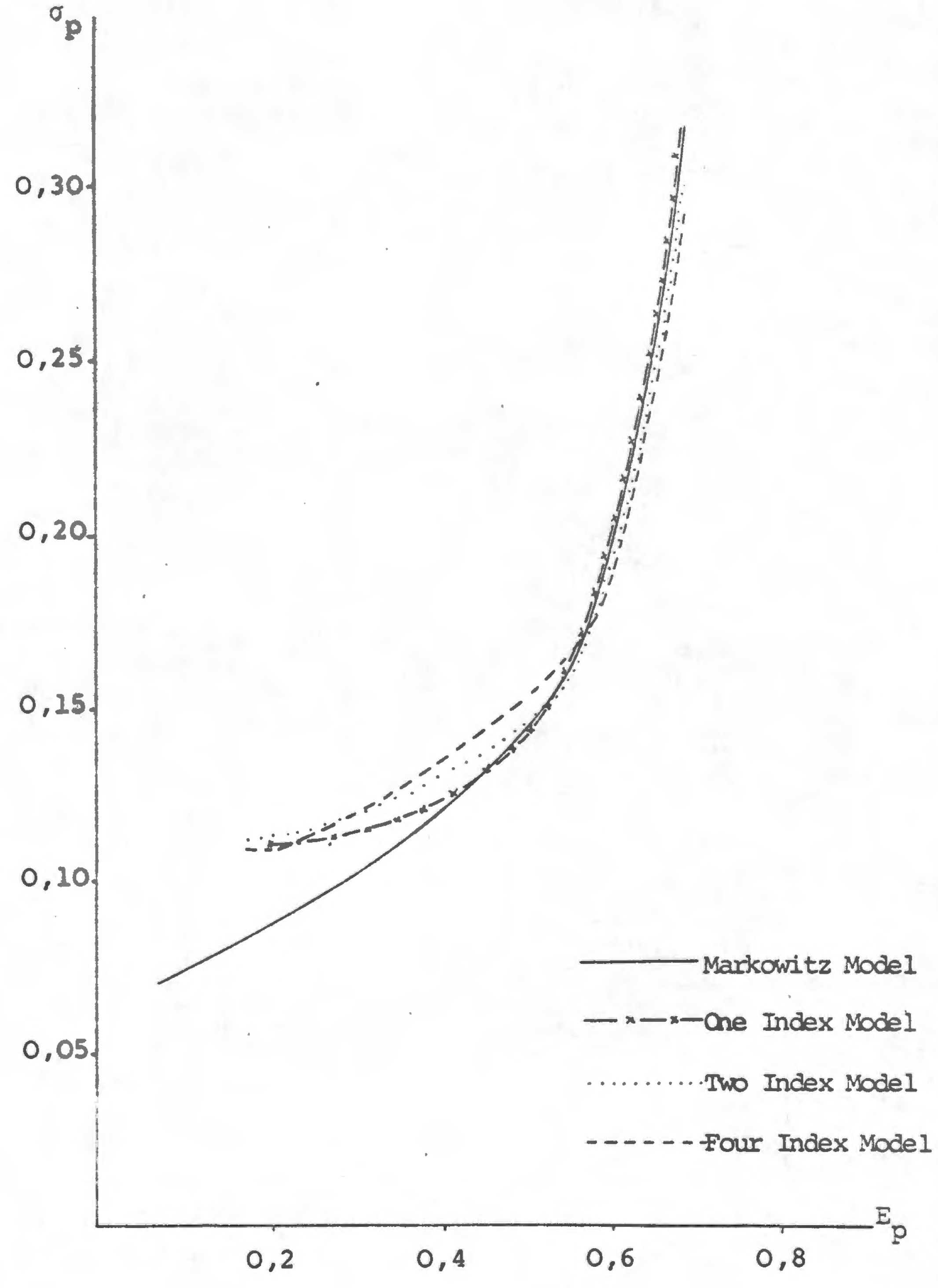


FIGURE 7.17

SAMPLE 5 $0 \leq X_i \leq 0{,}25$



In order to examine the differences between the models in the region of interest to the "average" investor in more detail, individual portfolios were once again chosen using a value of 0,08969 for λ. These portfolios are presented in Tables 7.13 to 7.17 below.

TABLE 7.13 SAMPLE 1

SECURITY	% of Funds to be invested in Each Security			
	Markowitz Model	One Index	Two Index	Four Index
Vryheid	14,51	9,47	7,72	11,75
Wankie	15,90	6,23	5,94	7,61
MTD	11,46	17,01	17,29	15,53
Ass Mang	25,00	18,92	16,53	15,88
S A Mang	0,16	8,39	8,55	8,43
Rooiberg	1,68	3,68	4,33	3,61
Zaaiplaat	20,86	20,84	17,28	15,71
Tweefontein	3,54	10,69	13,19	13,16
Rand Carb	6,89	3,44	8,09	6,04
Becketts		1,33	1,08	2,28
E_p	37,01	39,62	38,22	38,09
σ_p	9,87	13,10	12,30	11,72

TABLE 7.14 SAMPLE 2

SECURITY	Markowitz Model	One Index	Two Index	Four Index
W R Cons	11,53	1,89	7,60	10,51
Amal Coll	25,00	23,99	22,81	18,21
Clydesdale		17,22	17,66	9,77
Nat Ants	19,76	23,25	22,57	18,27
Welgedacht	2,80	11,13	11,26	9,41
Rooiberg	6,32	5,50	5,11	3,86
Cons Murch	2,33	1,38	0,93	
Msauli		2,01	1,91	
Anglos	20,45			8,79
T C Lands		2,13		2,08
Trade & Ind	11,81	10,46	10,15	17,87
Anglo Alpha		1,04		
Woolworths				1,23
E_p	40,81	53,91	54,81	48,87
σ_p	13,92	14,95	15,21	15,71

TABLE 7.15 SAMPLE 3

SECURITY	Markowitz Model	One Index	Two Index	Four Index
W R Cons		2,04		0,82
Clydesdale	4,49	25,00	20,36	17,54
Tavistock	25,00	25,00	25,00	25,00
T C Lands		1,64	0,32	6,20
Tweefontein	25,00	14,54	17,64	17,15
Trade & Ind	19,68	12,52	12,45	19,03
Eveready	13,10	16,74	19,92	9,19
Massey Ferg	12,73			1,22
Tongaat		2,46	2,65	
Remgro		0,06	1,66	3,85
E_p	47,69	47,28	46,62	49,21
σ_p	15,53	15,12	14,52	15,47

TABLE 7.16 SAMPLE 4

SECURITY	Markowitz Model	One Index	Two Index	Four Index
Randfontein	5,49	2,90	8,03	9,71
Wes Drie	13,03			0,64
A T Coll		18,58	17,13	9,60
Apex	1,06	18,17	17,99	12,45
Nat Ammo	25,00	19,98	22,23	20,33
Vierfontein	8,41	2,20	5,38	1,39
Witbank	16,93	20,07	19,24	12,24
Cons Gold	11,68			8,04
PP Cement		5,72	0,60	
I L Back	13,99	4,92	5,14	5,46
Cullinan	4,41			3,79
Lion		7,46	4,26	15,83
Remb Beh				0,52
E_p	51,23	80,30	83,25	66,25
σ_p	13,96	17,27	17,92	17,43

TABLE 7.17 SAMPLE 5

SECURITY	Markowitz Model	One Index	Two Index	Four Index
Randfontein	2,73	1,89	6,31	6,88
P Brand	6,39			
Amal Coll	16,19	25,00	25,00	25,00
Tavistock	25,00	25,00	25,00	25,00
S A Mang	13,15	12,59	13,64	9,93
Rooiberg	15,68	8,08	7,95	7,19
Cons Murch		3,61	3,55	
Fed Mynbou				0,78
Gen Mining				3,48
T C Lands		5,12	1,31	2,25
Trade & Ind	20,86	13,50	13,50	17,41
Nat Chem		0,14		
Woolworths				1,84
Remgro		0,97		0,05
Safmarine		4,10	3,74	0,19
E_p	48,28	48,64	51,76	51,63
σ_p	14,96	14,87	15,92	16,30

Examination of the above tables indicates that when an upper bound of 0,25 is imposed, the index models do not perform better (that is, do not produce a portfolio closer to the Markowitz portfolio) than in the unbounded case, as was surmised by Affleck-Graves and Money (1976). This fact is confirmed by comparing the "averages" presented in Table 7.18 below, with those of Table 7.6. In fact, a comparison of these tables indicates that in the bounded case, the performance is somewhat worse than in the unbounded case.

TABLE 7.18

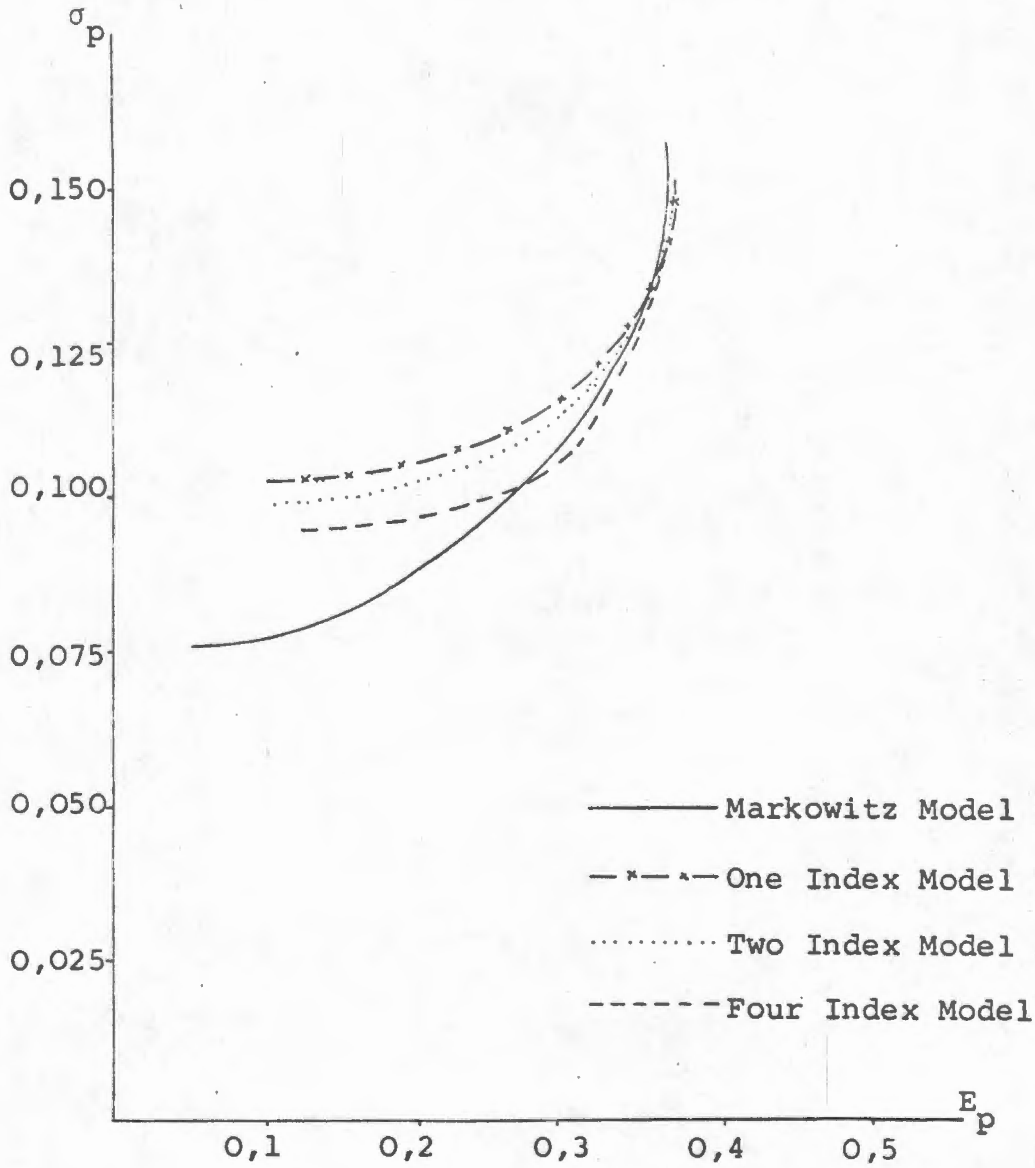
AVERAGE	Markowitz Model	One Index	Two Index	Four Index
E_p	36,86	53,95	54,93	50,81
σ_p	13,65	15,06	15,17	15,33
$\% E_p$		19.16	21,01	12,42
$\% \sigma_p$		12,12	12,44	13,01
$\|\% E_p\|$		19,50	21,91	12,42
$\|\% \sigma_p\|$		13,42	15,04	13,16

Analogous results are presented below for the case where the upper bound is 0,10 (that is, a maximum of 10% of the total funds may be invested in any one security). Firstly, Figures 7.18 to 7.22 below present sketches of the efficient frontiers for each of the five samples. As can be seen from the sketches presented in this section and in Section 4, the efficient frontier tends to get more and more compressed as the upper bound decreases. Also, the index models do not appear to produce efficient frontiers which are visually much closer to the Markowitz efficient frontier, for the case where the upper bound is 0,10.

As in the previous analyses, individual portfolios were selected using a value of 0,08969 for λ. These results are presented in Tables 7.19 to 7.23 below and the relevant averages are presented in Table 7.24.

F I G U R E 7.18

SAMPLE 1 $0 \leq X_i \leq 0{,}10$



F I G U R E 7.19

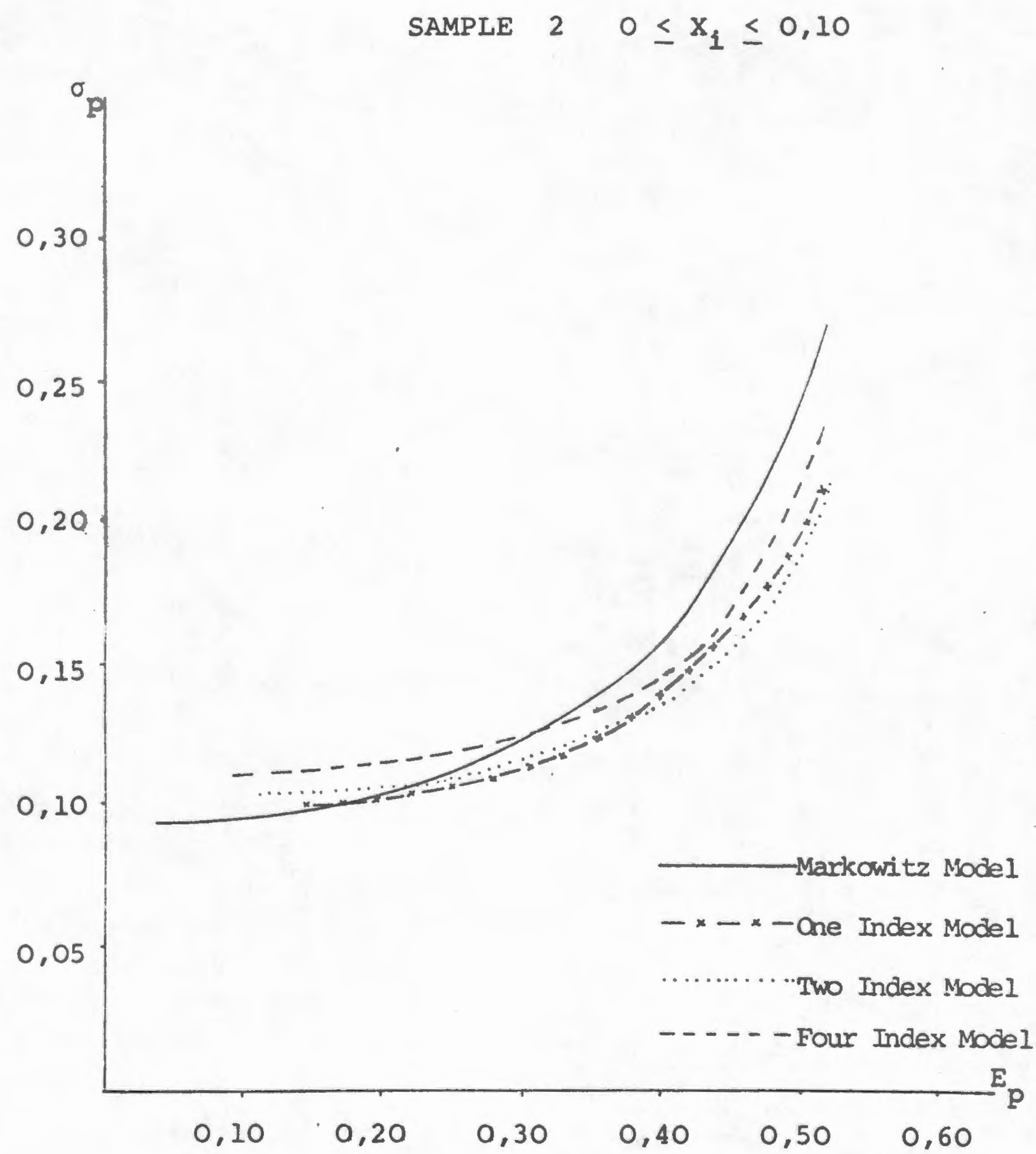

SAMPLE 2 0 ≤ X_i ≤ 0,10
σ_p
0,30
0,25
0,20
0,15
0,10
0,05
E_p
0,10
0,20
0,30
0,40
0,50
0,60
Markowitz Model
One Index Model
Two Index Model
Four Index Model

FIGURE 7.20

SAMPLE 3 $0 \leq X_i \leq 0{,}10$

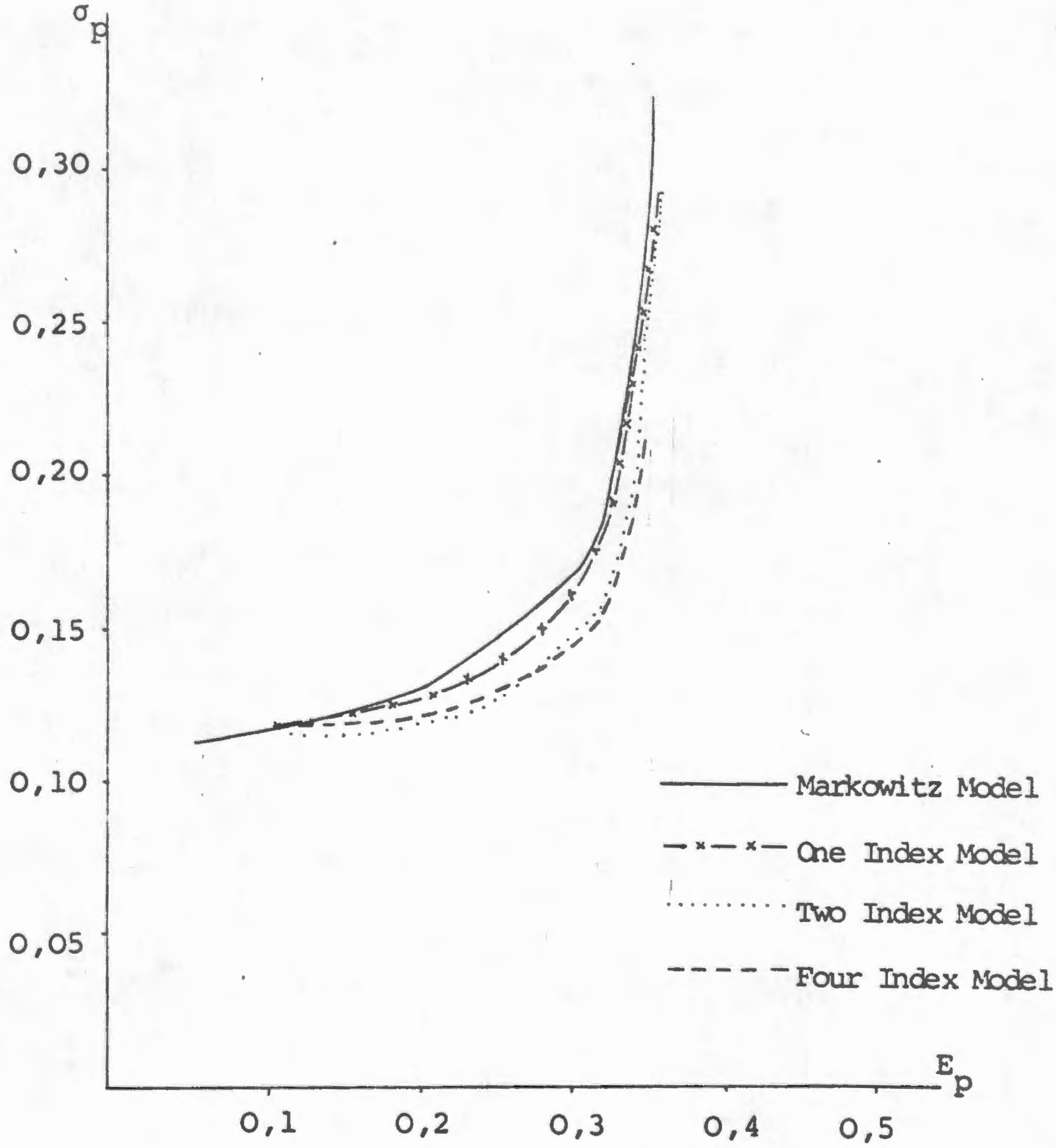


FIGURE 7.21

SAMPLE 4 $0 \leq X_i \leq 0{,}10$

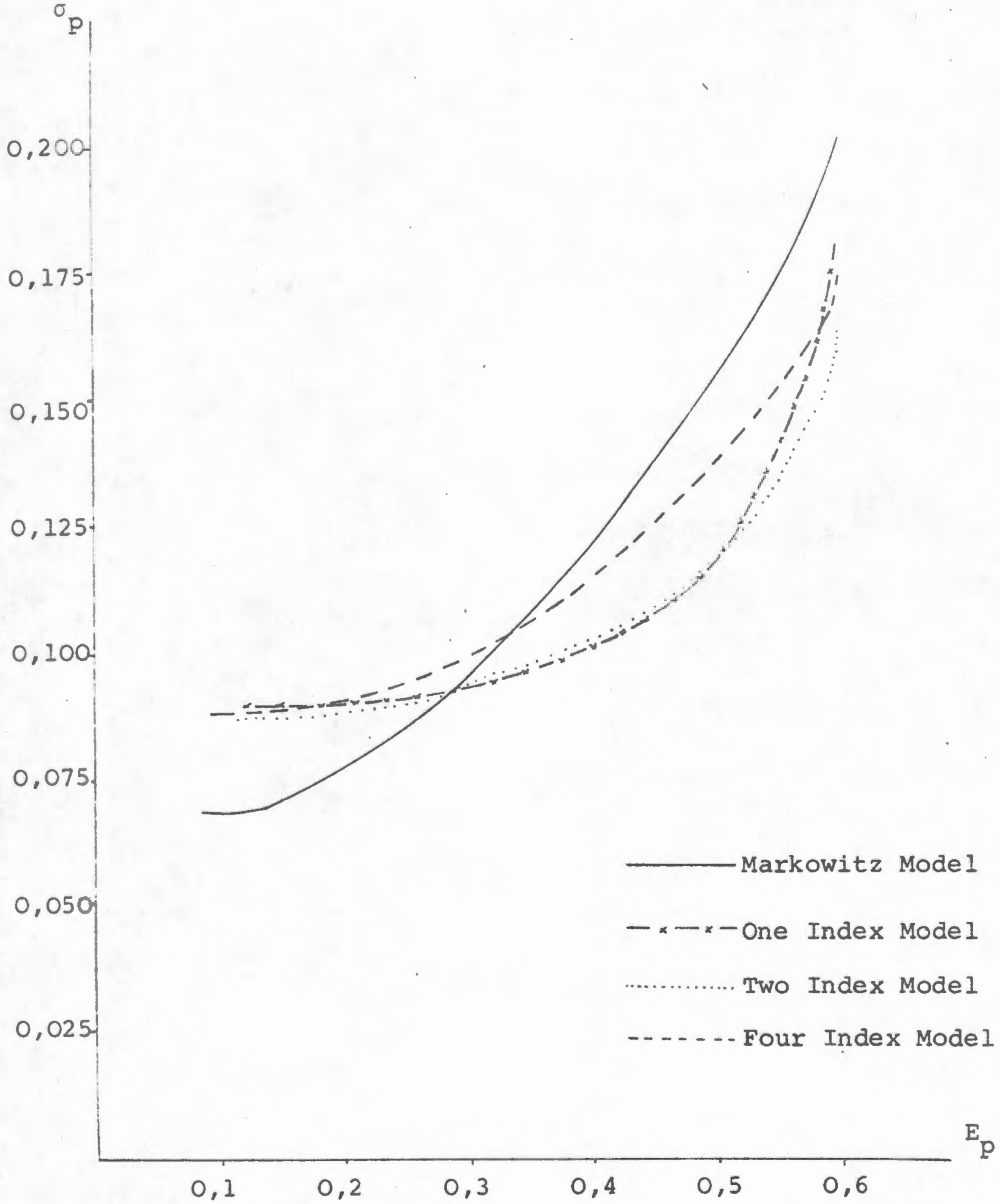


FIGURE 7.22

SAMPLE 5 $0 \leq X_i \leq 0,10$

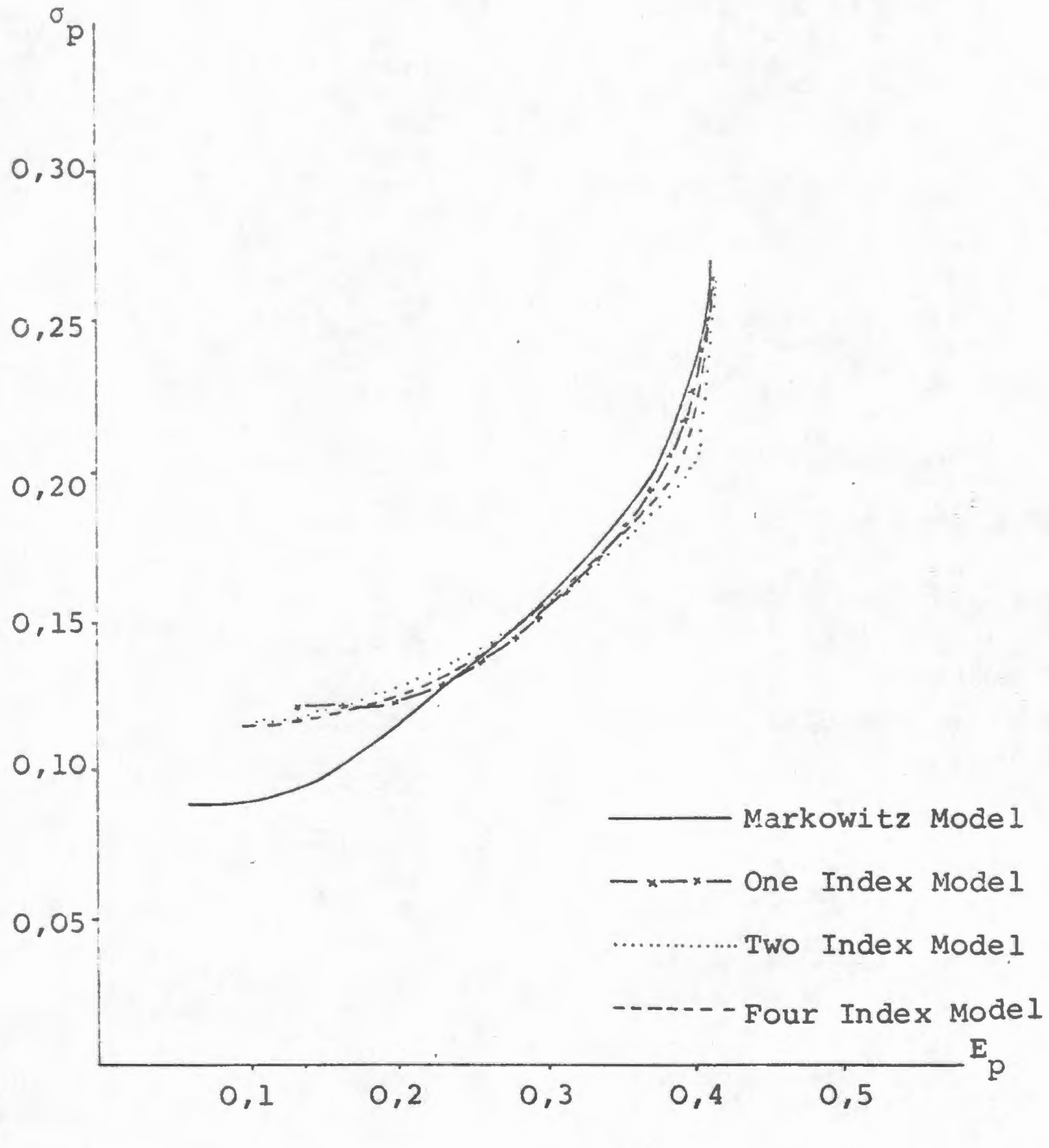


TABLE 7.19 SAMPLE 1

SECURITY	% of Funds to be Invested in Each Security			
	Markowitz Model	One Index	Two Index	Four Index
Winkels	1,15			
Vryheid	10,00	10,00	10,00	10,00
Wankie	10,00	10,00	10,00	10,00
MTD	10,00	10,00	10,00	10,00
Ass Mang	10,00	10,00	10,00	10,00
S A Mang	10,00	10,00	10,00	10,00
Rooiberg	8,76	6,81	6,86	5,62
Zaaiplaat	10,00	10,00	10,00	10,00
Tweefontein	10,00	10,00	10,00	10,00
Rand Carb	10,00	10,00	10,00	10,00
Becketts	10,00	7,19	7,81	8,00
Busaf		0,88	0,21	
Tongaat		2,78	2,52	0,97
Safmarine	0,09	2,34	2,60	5,41
E_p	34,21	33,47	33,60	33,45
σ_p	12,05	12,42	12,12	11,51

TABLE 7.20 SAMPLE 2

SECURITY	Markowitz Model	One Index	Two Index	Four Index
W R Cons	10,00	2,25	10,00	10,00
Winkels				0,40
Amal Coll	10,00	10,00	10,00	10,00
Clydesdale	10,00	10,00	10,00	10,00
Nat Ants	10,00	10,00	10,00	10,00
Welgedacht	8,51	10,00	10,00	10,00
Messina	9,51			
Rooiberg	10,00	8,42	7,72	6,67
Cons Murch	3,42	4,11	3,28	1,58
Msauli		2,81	2,54	0,21
Anglos	10,00	7,14	6,64	10,00
T C Lands		3,04		8,35
De Beers		4,18	3,25	1,14
Trade & Ind	10,00	10,00	10,00	10,00
Anglo Alpha		6,30	5,29	1,30
Un Steel	8,92	9,99	10,00	8,25
Sappi		1,76	1,28	
Woolworths				2,10
E_p	34,74	38,50	40,58	39,24
σ_p	13,78	13,38	13,95	14,58

TABLE 7.21 SAMPLE 3

SECURITY	Markowitz Model	One Index	Two Index	Four Index
W R Cons		2,36	5,70	5,69
Leslie	4,80			
Winkels	1,75			
Clydesdale	10,00	10,00	10,00	10,00
Tavistock	10,00	10,00	10,00	10,00
Gen Mining	10,00	10,00	7,81	7,83
Mid Wits	8,43	6,90	5,41	4,62
T C Lands		2,03	0,27	5,18
Tweefontein	10,00	10,00	10,00	10,00
Trade & Ind	10,00	10,00	10,00	10,00
Bakers	10,00	8,91	8,28	4,00
Eveready	10,00	10,00	10,00	10,00
Massey Ferg	10,00	7,07	8,65	9,26
Greatermans		0,47	0,38	0,54
Tongaat	2,89	10,00	10,00	7,28
Remgro	2,13	2,26	3,49	5,60
E_p	22,15	24,90	26,50	28,06
σ_p	13,55	13,48	13,28	13,84

TABLE 7.22 SAMPLE 4

SECURITY	Markowitz Model	One Index	Two Index	Four Index
Randfontein	5,20	1,74	6,76	6,87
Wes Drie	10,00		5,33	5,67
A T Coll	8,69	10,00	10,00	10,00
Apex	10,00	10,00	10,00	10,00
Nat Ammo	10,00	10,00	10,00	10,00
Vierfontein	10,00	10,00	10,00	10,00
Witbank	10,00	10,00	10,00	10,00
Cons Gold	10,00			6,46
PP Cement		10,00	10,00	8,20
I L Back	10,00	8,12	7,66	6,26
Nat Chem		0,31		
Cullinan	8,89	6,39	3,60	5,01
Un Steel	6,33	6,63	6,65	
Cons Glass		4,87		
Lion	0,89	10,00	10,00	10,00
Remb Beh		1,94		1,53
E_p	45,70	50,69	53,25	52,17
σ_p	14,03	12,15	13,12	14,53

TABLE 7.23 SAMPLE 5

SECURITY	Markowitz Model	One Index	Two Index	Four Index
Randfontein			3,11	3,12
Winkels	3,21		0,57	
P Brand	10,00		1,07	2,09
Westn Hldgs			9,28	10,00
Blyvoor	1,67			
Amal Coll	10,00	10,00	10,00	10,00
Tavistock	10,00	10,00	10,00	10,00
De Beers		0,82		
Messina	0,80	0,74		
S A Mang	10,00	10,00	10,00	10,00
Rooiberg	10,00	10,00	10,00	9,34
Cons Murch	2,92	6,36	5,84	3,67
Fed Mynbou	4,52	0,17		2,06
Gen Min	10,00	10,00	10,00	10,00
T C Lands		4,25		
Trade & Ind	10,00	10,00	10,00	10,00
Nat Chem	4,47	3,43	2,94	2,16
Dorbyl		3,92	0,53	
Stew & Lloyds	4,19	8,03	7,47	9,67
Woolworths		0,08		
Remgro		2,99	0,81	
Safmarine	8,22	9,21	8,38	7,89
E_p	20,58	27,03	27,29	26,11
σ_p	11,39	14,01	14,63	14,28

TABLE 7.24

AVERAGE	Markowitz Model	One Index	Two Index	Four Index
E_p	31,48	34,92	36,24	35,96
σ_p	12,96	13,09	13,42	13,75
$\% E_p$		12,67	16,76	15,69
$\% \sigma_p$		1,85	4,36	6,48
$\|\% E_p\|$		13,53	17,47	16,58
$\|\% \sigma_p\|$		8,58	7,75	8,27

The main conclusion which can be drawn from the results presented in this section is that the index models do not provide efficient frontiers closer to the Markowitz efficient frontier when an upper bound is imposed on the proportion of funds which can be invested in any one security. Examination of Tables 7.6, 7.18, and 7.24 indicates that, in fact, it appears as if the index models perform best in the unbounded case (the four index model has an average absolute error of about 9%) and better for the 0,25 bounded model ($|\% E_p| = 12,42$) than for the 0,10 bounded model ($|\% E_p| = 16,58$). Also, it appears that as the bound decreases, the three different index models examined tend to perform more similarly. But, although a visual examination does confirm that, in general, the four index model generates an efficient frontier closer to Markowitz frontier than either the one or two index models, it is not nearly as marked as in the unbounded case.

7.7 CONCLUSIONS

In this chapter, the difference in the performance of the Markowitz Model and the Sharpe's Index Model was examined under a number of different conditions. The main conclusion which can be drawn is that the index models do not approach the Markowitz efficient frontier as closely as might be hoped, either when a low upper bound is imposed on the proportion of funds to be invested in any one security or when principal component indices are used. Thus, it would appear that an investor wishing to use one of the portfolio selection models, should strive to use the Markowitz Model, as this remains probably the most realistic of all the single period models. If, for estimation purposes, it is considered infeasible to use this model (either past data is not available or subjective estimates of the risk are preferred) then it is almost certain that the investor will have to resort to the use of an index model. The results presented in this chapter indicate that in such circumstances the error in the expected value of the portfolio chosen (that is, the difference between the expected value of the portfolio chosen using the index model and the portfolio which would have been chosen if the Markowitz Model had been used), is likely to be of the order of 9 percent in the unbounded case when using a four index model, and is likely to increase as the upper bound is decreased (12% in the 0,25 upper bound case and 16% in the 0,10 case). This error could be in either direction but the index model's expected return is more likely to be higher than the Markowitz portfolio than lower, and more so as the upper bound decreases.

The second conclusion which can be drawn from the results presented in this chapter, is that the four index model is superior to both the one and two index models in almost all the cases considered (this can be seen visually from the sketches). However, as the upper bound decreases, this superiority decreases and, in fact, in the 0,1 upper bound case it appears that in the region of interest to the "average" investor (λ = 0,08969) the one index model produces superior results to the four index model ($|\% E_p|$ of 13,53 compared to 16,58). In addition, it is apparent that in almost all instances the one index model is not outperformed by the two index model. In fact, the opposite is often true, namely that the one index model outperforms the two index model (certainly from an examination of the averages presented in Tables 7.6, 7.18, and 7.24). As the two index model requires more input, it can be concluded that the one index model is superior.

Finally, it is interesting to note the "flattening out" effect of the index models for the low return efficient portfolios. In most cases, the index models straighten out (that is, become nearly horizontal) at a much higher risk (standard deviation) than the corresponding Markowitz Model. This suggests that the index models will probably perform worse for very risk averse investors than the results presented in this chapter (for the "average investor") indicate. This is not nearly as marked in the principal component indices examples, and indicates that the use of such indices may be more appropriate for very risk averse investors.

CHAPTER EIGHT

REGRESSION ANALYSIS WITH TRAPEZOIDAL DATA

8.1 INTRODUCTION

A problem which frequently arises in the analysis of economic time series, especially stock market data, is that of trapezoidal data. That is, the data is available for the different variables from differing starting dates. Pictorially, the observations may occur as follows (t_j indicates period j):

SERIES	PERIOD t_1	t_2	t_3	t_4	t_5 ...	t_m ...	t_M
1	$R_{1;1}$	$R_{1;2}$	$R_{1;3}$	$R_{1;4}$	$R_{1;5}$ ...	$R_{1;m}$ ...	$R_{1;M}$
2		$R_{2;2}$	$R_{2;3}$	$R_{2;4}$	$R_{2;5}$ ...	$R_{2;m}$ ...	$R_{2;M}$
3					$R_{3;5}$ ...	$R_{3;m}$ ...	$R_{3;M}$
.						.	.
.					.	. ...	.
.					.	.	.
.						.	
.							
N						$R_{N;m}$	$R_{N;M}$

It should be noted that each $R_{i;j}$ indicates the set of observations on both the dependent and independent variables from series i in period j (that is, each $R_{i;j}$ is a $k\times1$ vector where k is the total number of variables in the system). This is, perhaps, best illustrated by means of an example.

Suppose that it is desired to find the general relationship between the price of a security and various factors which it is felt might influence the price, such as the issue price of the security and the Gross National Product. Then, in this case, the series would consist of the various securities in the sample, and each $R_{i;j}$ would contain three values; (i) the price of the i^{th} security in the j^{th} period; (ii) the issue price of the i^{th} security; and (iii) the G.N.P. in period j.

There are obviously many real world examples where such a situation can arise. Moreover, the missing data may be due to a variety of reasons some of which can be overcome by further data collection. However, there remain situations in which further data collection is impossible. The problem mentioned above is a good example of such a situation; it is clearly impossible to obtain information about a security prior to its issue and, since securities are issued and listed at different dates, the trapezoidal problem arises where further data collection is not possible. Moreover, even in cases where the data is obtainable, it may be difficult, expensive, and time consuming to collect the additional data.

Thus, the following three questions arise.

(i) How much of the available data should be used?

(ii) How much information is lost by not having the entire set of data?

(iii) Which of the available data should be used in the analysis?

8.2 SOME POSSIBLE APPROACHES

Before discussing the actual results obtained, some of the alternatives open to the investigator will be considered.

A. USE OF ALL THE OBSERVATIONS: Since there are progressively fewer observations on the successive series, more weight is given to the first series than to all the remaining series. In addition, more weight is given to the second series than to the third and remaining series, etcetera and this unequal weighting of the series can result in a distortion of the results. This probably accounts for the reason why this method has seldom been used in practice, since in many cases, a method which gives equal weight to each series may be more desirable. The easiest way of achieving this is to choose an equal number of observations per series, and such methods are discussed below.

B. THE LAST CROSS-SECTION: (i.e., the last few observations in each series.) Only analyze those periods for which data is available on all the series. In the pictorial example above, this is equivalent to using periods t_m to t_M only. This appears to be the most popular method in practice but has the disadvantage of ignoring all history prior to period m.

C. SIMPLE RANDOM SAMPLING: Draw simple random samples of equal size (usually $M-m+1$) from each series. This results in an equal number of observations per series and still allows for the use of some of the historical data.

D. SYSTEMATIC RANDOM SAMPLING: Draw a systematic sample (that is, every k^{th} observation) from each series, choosing different k's so that an equal number of observations are chosen from each series.

E. CLUSTER SAMPLING: Divide each series into an equal number of clusters on the basis of time and then draw one observation at random from each cluster. That is, the first cluster for series one will consist of $R_1, R_2, \ldots, R_q$, the second cluster will consist of $R_{q+1}, \ldots, R_p$, etcetera. This method ensures an equal number of observations per series and a good historical spread.

F. MOVING AVERAGES: For each series, the observations may be transformed to a new data set comprising of centered moving averages of length α (the same for all series). Then, a simple random sample or cluster sample may be chosen from each transformed series such that an equal number of observations are chosen for all series.

G. AVERAGING METHODS: These involve either averaging over each entire series, thus obtaining one set of figures

($\bar{R}_i$) for each series or else taking the average of each cluster for each series. However, since the series have differing numbers of observations and since the variance of the average is given by σ^2/n_i, the averaging methods result in the introduction of heteroskedasticity. While there are methods of overcoming this, they result in giving more weight to the series with more observations (that is, less variance) than to those with fewer observations. This is similar to using the entire trapezoidal data set (A above) and is often the situation which one is trying to avoid. For this reason, these averaging methods were not investigated in the simulation study below.

8.3 THE SIMULATION STUDY

In order to examine the problem of trapezoidal data in regression analysis, a set of data obeying a given function with a specified distribution of the error term was generated using a random number generator. The simulated regression equation was based on a hypothetical stock market problem. The function used was

$$P_{ij} = 700 + Q_i - 5S_i + 10F_i - 3G_j + 0{,}1T_{ij} + e_{ij}$$

where

- P_{ij} is the price of the i^{th} security at the end of the j^{th} period;
- Q_i is the total number of shares issued in the i^{th} security (that is, the issued volume);

S_i is the number of companies engaged in similar business interests and quoted on the exchange;

F_i is the issue price of the i^{th} security;

G_j is the Gross National Product at the end of period j;

T_{ij} is the time (at the end of period j) since the issue of the i^{th} security; and

e_{ij} is the random error which was assumed to be normally distributed with zero mean and standard deviation 40.

For the purpose of the simulation, it was assumed that a new security was issued every two weeks, and that the system was observed for 120 weeks with the last security being issued in the 109^{th} week. Thus, there were 55 securities in total, each with a given fixed value of Q, S, and F. The G.N.P. of South Africa was used and the desired error term generated using a random number generator. It should be noted that the number of observations per series (that is, per security) varied from 120 for the first security, 118 for the second security, down to 12 for the 55^{th} and last security.

Using the simulated data, each of the alternatives discussed in Section 2 above (except for G) were examined and the results are presented in Table 8.1 below.

TABLE 8.1

METHOD	RUN	CONST.	Q	S	F	G	T	R^2
ALL OBS.	-	695,6 (4,79)	0,86 (1,29)	-4,96 (0,03)	9,92 (0,38)	-2,83 (0,32)	0,12 (0,03)	0,922 (39,95)
LAST 12 OBS. IN EACH SERIES	-	663,1 (58,30)	0,91 (3,17)	-4,97 (0,06)	9,95 (0,84)	-1,51 (2,42)	0,13 (0,05)	0,915 (40,59)
SIMPLE RANDOM	1	691,9 (11,02)	3,78 (3,10)	-5,01 (0,06)	9,82 (0,82)	-2,80 (0,46)	0,11 (0,06)	0,919 (39,67)
SAMPLE OF SIZE 12	2	717,0 (11,34)	2,65 (3,10)	-4,97 (0,06)	10,76 (0,82)	-4,01 (0,46)	0,11 (0,06)	0,918 (39,64)
FROM EACH SERIES	3	697,3 (12,43)	-1,88 (3,35)	-5,02 (0,07)	11,26 (0,89)	-2,75 (0,52)	0,10 (0,07)	0,908 (42,96)
SYSTEMATIC SAMPLE	1	704,1 (11,90)	2,45 (3,14)	-4,99 (0,06)	9,50 (0,87)	-3,33 (0,55)	0,17 (0,07)	0,915 (40,39)
OF SIZE 12	2	715,4 (12,32)	1,11 (3,10)	-4,96 (0,06)	9,58 (0,88)	-3,98 (0,56)	0,17 (0,07)	0,915 (40,35)
FROM EACH SERIES	3	696,9 (11,39)	-0,19 (3,17)	-4,87 (0,04)	9,92 (0,85)	-3,00 (0,55)	0,16 (0,07)	0,914 (40,26)
CLUSTER SAMPLE	1	692,2 (11,53)	4,73 (3,08)	-4,97 (0,06)	10,36 (0,82)	-2,86 (0,47)	0,09 (0,06)	0,919 (39,49)
OF SIZE 12	2	694,9 (11,70)	0,57 (3,12)	-4,97 (0,06)	10,17 (0,83)	-2,70 (0,48)	0,04 (0,06)	0,918 (39,94)
FROM EACH SERIES	3	707,4 (11,71)	-0,89 (3,13)	-4,96 (0,06)	10,01 (0,83)	-3,40 (0,48)	0,10 (0,06)	0,917 (40,06)
MOVING AVG. OF	1	698,6 (7,27)	2,04 (1,94)	-5,03 (0,04)	9,93 (0,51)	-2,86 (0,31)	0,09 (0,04)	0,975 (21,53)
LENGTH 3; SIMPLE	2	692,2 (7,73)	1,77 (2,07)	-4,96 (0,04)	9,55 (0,54)	-2,57 (0,31)	0,05 (0,04)	0,971 (22,79)
RANDOM SAMPLE OF SIZE 9	3	708,2 (7,63)	0,07 (2,06)	-4,94 (0,04)	9,70 (0,55)	-3,30 (0,32)	0,09 (0,04)	0,971 (22,84)

METHOD	RUN	CONST	Q	S	F	G	T	R^2
MOVING AVG. OF	1	690,5 (7,97)	-0,06 (2,05)	-4,92 (0,04)	10,35 (0,57)	-2,63 (0,33)	0,07 (0,04)	0,969 (23,87)
LENGTH 3; CLUSTER	2	692,7 (7,53)	0,90 (2,03)	-4,90 (0,04)	10,14 (0,54)	-2,81 (0,31)	0,12 (0,04)	0,971 (22,57)
SAMPLE OF SIZE 9	3	695,3 (8,05)	-0,21 (2,18)	-4,92 (0,04)	10,16 (0,58)	-2,84 (0,33)	0,10 (0,04)	0,967 (24,23)
MOVING AVG. OF	1	695,6 (7,81)	5,86 (2,13)	-4,96 (0,04)	9,84 (0,57)	-2,93 (0,33)	0,07 (0,04)	0,986 (17,63)
LENGTH 5; RANDOM	2	694,8 (7,60)	-0,62 (2,10)	-4,91 (0,04)	9,93 (0,56)	-2,74 (0,32)	0,09 (0,04)	0,983 (17,36)
SAMPLE OF SIZE 5	3	690,4 (8,13)	2,26 (2,14)	-5,03 (0,04)	10,66 (0,57)	-2,62 (0,34)	0,08 (0,04)	0,983 (17,85)
MOVING AVG. OF	1	706,5 (8,23)	-1,56 (2,23)	-4,88 (0,04)	9,23 (0,59)	-3,20 (0,34)	0,08 (0,04)	0,981 (17,45)
LENGTH 5; CLUSTER	2	697,3 (8,63)	1,27 (2,35)	-4,96 (0,05)	9,98 (0,62)	-2,88 (0,36)	0,09 (0,05)	0,979 (18,01)
SAMPLE OF SIZE 5	3	693,4 (7,97)	0,64 (2,15)	-4,94 (0,04)	9,78 (0,57)	-2,73 (0,33)	0,12 (0,04)	0,982 (17,21)

NOTE: The figures in the table above represent the estimated regression coefficients and the figures in brackets, the standard error of the estimates. The final column presents the coefficient of determination and the estimate of the variance of the error term (that is, the standard error of the estimate - S.E.E.) is given in parenthesis.

Some interesting observations are immediately apparent on examination of Table 8.1. Firstly, almost all of the point estimates of the coefficients are within two standard deviations of their true values, which indicates that none of the proposed methods have biased the results unduly. Secondly, as far as comparisons are concerned, the standard deviations of the estimates of the regression coefficients are of most interest. Clearly, the smaller these standard deviations, the better. On this basis, it is apparent that the first method (that is, using all the available observations) is the most appropriate. However, this is to be expected because of the nature of the simulated data (well behaved) and hence, the effect of giving more weight to those series with a greater number of observations does not have a marked effect on the regression estimates. It should be stressed that this might not be the case with less "well behaved" data.

On examining Table 8.1 it is rather difficult to evaluate the relative merits and demerits of the different sampling methods. The standard deviations are similar for most schemes except for the moving average methods, which would be expected to have lower standard deviations since they are based on averages, not data points.

Hence, in order to examine the different methods more closely, a table (Table 8.2 below) of the mean absolute difference of the point estimates from their true values was constructed for each of the methods examined.

TABLE 8.2

METHOD	NO.OF RUNS	CONST	Q	S	F	G	T
All Obs.	1	4,40	0,14	0,04	0,08	0,17	0,02
Last 12	1	36,90	0,09	0,03	0,05	1,49	0,03
Random Sample	3	9,27	2,44	0,02	0,73	0,49	0,01
Systematic Sample	3	7,53	0,92	0,06	0,33	0,44	0,07
Cluster Sample	3	6,77	2,02	0,03	0,18	0,28	0,02
M.A. α=3 Random Sample	3	5,80	0,91	0,04	0,27	0,29	0,02
M.A. α=3 Cluster Sample	3	7,17	0,79	0,09	0,22	0,24	0,02
M.A. α=5 Random Sample	3	6,40	2,58	0,05	0,30	0,24	0,02
M.A. α=5 Cluster Sample	3	5,27	1,06	0,07	0,34	0,20	0,02

From Table 8.2, it is fairly obvious that what appears to be the most popular method in practice, namely using the last cross-section of the observations (12 periods in the simulated example case) does not provide the best results. In particular, the estimates of the constant term and the coefficient of G are very poor - further, by far, from their true values than when using any of the other methods (see also Table 8.1). Moreover, from the estimates of the standard deviations of the regression coefficients presented in Table 8.1, it is clear that this method also provides estimates with larger standard deviations. It should be noted that Q, S, and F are fixed values for each series and hence the sampling procedure does not greatly affect the standard deviations of their regression estimates.

The results presented in Table 8.2 indicate that the systematic sampling method and the cluster sampling method appear to provide more accurate estimates than the random sampling method. In addition, it appears as if the systematic method is worse than the cluster method. Also, from theoretical considerations, one would be very apprehensive about advocating the use of systematic sampling in general, unless the researcher was certain that there were no cyclical effects. For these reasons, it is argued that the cluster sampling method is more appropriate than either the random or systematic sampling methods.

The moving average schemes also present very encouraging results. The standard deviations are smaller (based on averages of length α), but since the moving averages of each series are of the same length (α for all series) the problem of heteroskedasticity is not introduced. The main disadvantage that could be raised against this method is that it uses averages and not individual data points. Once again it is difficult to tell which sampling scheme is the best but, intuitively, it is likely that the cluster method will be more reliable in practice.

In conclusion it can be said that the results obtained from the simulation study indicate that either the entire data set (if the data is well behaved), the cluster sampling method (if it is desired to use individual data points), or the moving average method with cluster sampling should be used. Unless the researcher has very good reasons for doing so, the common practice of taking only the last cross-section of observations available for all series should be avoided.

8.4 EXAMPLE

As an illustration of how the ideas discussed above might be used in a "real world" example, the following problem was examined. Suppose that it is desired to find a general equation defining the behaviour of South African Gold Shares. Note that a general equation for all shares is required and not a prediction equation for one share in isolation. Obviously, there are many variables which might affect the prices of gold shares, but for the purposes of this example only five variables were considered, and a linear relationship of the form below was assumed:

$$P_{i;j} = \alpha_0 + \alpha_1 Vol_i + \alpha_2 DJ_j + \alpha_3 GP_j + \alpha_4 RES_i + \alpha_5 GPTM_i + e_{i;j},$$

where

Vol_i represents the number of shares (in millions) issued in the i^{th} gold share;

DJ_j is the closing price of the Dow Jones Industrial Average at the end of the j^{th} week;

GP_j is the Gold Price (in dollars) on the London Metal Exchange at the end of the j^{th} week;

RES_i is the estimated gold reserves (in millions of tons) at the beginning of 1973 of the i^{th} gold share;

$GPTM_i$ is the average number of grams of gold per ton mined in 1972 for the i^{th} gold share;

$P_{i;j}$ is the closing price of the i^{th} gold share at the end of the j^{th} week;

$\alpha_0;\alpha_1,...,\alpha_5$ are the regression coefficients to be estimated; and

$e_{i;j}$ is the random error.

The data used for P, VOL, RES, and GPTM was from 16 gold mining companies quoted on the Johannesburg Stock Exchange (see Table K.1 in Appendix K). The period considered for P, DJ, and GP was 5th January 1973 to 20th February 1976, a period of 163 weeks. Thus, in total there were 2 608 (16×163) observations for the proposed regression analysis.

Initially, the regression analysis was performed on the full set (i.e. 2 608 observations) hereafter referred to as the ENTIRE DATA SET. Then, in order to investigate the situation if some of the data had not been available, it was decided to assume that the history on successive shares was less by 8 weeks than its predecessor; that is, 163 weeks were assumed available for the first share; the first eight weeks were NOT available for the second share -

only the remaining 155 weeks were available; 147 weeks were available for the third share; etcetera; until only 43 weeks of data were available for the last (16^{th}) share. This data was used as the trapezoidal data set and the various procedures discussed above (in Sections 2 and 3) were examined. The results are presented in Table 3 below.

TABLE 8.3

METHOD	CONST α_0	VOL α_1	DJ α_2	GP α_3	RES α_4	GPTM α_5	R^2	S.E.E.
Entire Data Set	-1094	-33,74	-1,62	10,12	141,2	194,9	0,82	566
All Trapezoidal Data	-563	-37,67	-1,81	7,06	160,2	196,8	0,82	568
Last 40 wks on all Shares	-5345	-29,31	1,34	18,52	148,3	207,6	0,91	368
Simple random sample of size 40	-65	-36,93	-2,48	5,09	180,7	210,7	0,84	578
Cluster sample of size 40	-374	-32,64	-2,56	7,81	164,7	200,1	0,83	568
Moving Average of Length 3 weeks: Random Sample of size 40	-192	-40,71	-2,27	6,22	156,9	210,6	0,84	547
Moving Average of Length 3 weeks: Cluster Sample of size 40	-384	-33,33	-2,58	8,07	162,2	207,0	0,84	548
Moving Average of Length 5 weeks: Random Sample of size 35	-84	-38,39	-2,53	6,25	168,3	208,6	0,85	520
Moving Average of Length 5 weeks: Cluster Sample of size 35	-933	-37,85	-2,03	9,95	154,0	202,3	0,85	517

The results presented in Table 8.3 confirm the conclusions arrived at from the simulation study. What are required, are regression estimates based on all or some of the trapezoidal data which are as close as possible to the estimates obtained using the entire data set. Clearly, the best results (excluding the moving average methods) were obtained when using all the trapezoidal data. However, it must be mentioned that, in this example, the observations on the independent variables (VOL, DJ, GP, RES, and GPTM) were of very much the same order for each of the 16 shares. Thus, the effect of giving more weight to some shares and less to others does not have a very detrimental effect on the estimation of the regression coefficients. But, if the independent variables were of very different order for the different shares, the results might not have been as satisfactory.

Of the remaining methods, it would appear as if the Cluster Sampling method is the most appropriate by far, as it is clearly superior to using either a Simple Random sample or the last 40 observations on each share. The results obtained using the moving average methods with cluster sampling are very good (perhaps better than the results obtained using all the trapezoidal data), but they do have the disadvantage of using averages in the regression analysis and subsequently having to correct for the standard error of the estimate, and the standard deviations of the estimates of the regression coefficients.

8.5 CONCLUSIONS

When faced with a situation in which one has a trapezoidal set of data and it is either not possible or else is too expensive to obtain further information, the researcher has numerous options open to him. In this paper, some of these alternatives have been examined and some important conclusions may be drawn.

Firstly, both from the simulation study and the practical example (a well behaved data set at that), it is obvious that the hitherto most popular method of using only the last cross section for which data is available on all variables, does not provide satisfactory estimates. Thus, this practice should not be adopted unless the researcher has very strong reasons for doing so, and it must be stressed that the results obtained will probably only describe the behaviour over the recent cross section and NOT the entire period as desired. If the behaviour over the entire period is of interest, better estimates can certainly be obtained.

Secondly, as far as the accuracy of the estimates is concerned, the most appropriate method (of those considered) is to transform the data of each series by means of a centered moving average (of order 3 or 5 as considered above, or some other order if preferable). Then, each transformed series can be divided into an equal number of clusters and one value sampled at random from each cluster, thus obtaining an equal number of observations per series.

Finally, if the researcher is not prepared to perform his analysis on averages rather than the actual data, and if the data for each variable is of approximately the same order for each series, then it would be preferable to use the entire trapezoidal data set. If however, the data is not of the same order, or if the researcher has strong reasons for requiring equal weight to be given to each series, then the most appropriate method would be to divide each series into a number of clusters and draw one value at random from each cluster.

A P P E N D I X   A

T A B L E   A.1 - COAL SECTOR

1. Anglo-Transvaal Collieries Ltd.
2. Apex Mines Ltd.
3. The Clydesdale (Transvaal) Collieries Ltd.
4. Natal Ammonium Collieries (1946) Ltd.
5. Natal Anthracite Colliery Ltd.
6. Natal Coal Exploration Co. Ltd.
7. Newcastle-Platberg Colliery Ltd.
8. The Nigel Gold Mining Co. Ltd.
9. Tavistock and South Witbank Collieries Ltd.
10. Trans-Natal Coal Corporation Ltd.
11. Vierfontein Colliery Ltd.
12. Vryheid Coronation Ltd.
13. Wankie Colliery Co. Ltd.
14. Welgedacht Exploration Co. Ltd.
15. Witbank Colliery Ltd.
16. Witbank Consolidated Coal Mines Ltd.
17. Zuinguin Natal Collieries Ltd.

TABLE A.2 - GOLD - WITWATERSRAND AND OTHERS

1. Durban Roodepoort Deep Ltd.
2. East Daggafontein Mines Ltd.
3. East Rand Proprietary Mines Ltd.
4. Eastern Transvaal Consolidated Mines Ltd.
5. Falcon Mines Ltd.
6. Government Gold Mining Areas (Modderfontein) Cons. Ltd.
7. The Grootvlei Proprietary Mines Ltd.
8. Marievale Consolidated Mines Ltd.
9. The Randfontein Estates Gold Mining Co. Witwatersrand Ltd.
10. Simmer and Jack Mines Ltd.
11. The South African Land and Exploration Co. Ltd.
12. South Roodepoort Main Reef Areas Ltd.
13. Village Main Reef Gold Mining Company (1934) Ltd.
14. Vlakfontein Gold Mining Co. Ltd.
15. West Rand Consolidated Mines Ltd.
16. Witwatersrand Nigel Ltd.

TABLE A.3 - FINANCIAL INDUSTRIAL SECTOR

1. Abercom Investments Ltd.
2. Anglo American Industrial Corporation Ltd
3. Anglo-Transvaal Industries Ltd.
4. Barlow Rand Ltd.
5. Bonuskor Beperk.
6. Brick and Clay Holdings Ltd.
7. Bromain Holdings Ltd.
8. Calan Ltd.
9. Comair Holdings Ltd.
10. De Beers Industrial Corporation Ltd
11. Diamond Royalties and Holdings Ltd.
12. Federale Volksbeleggings Bpk.
13. The Finance Company for Industry Holdings Ltd
14. Golden Arrow Investments Ltd.
15. Hippo Holdings Co. Ltd.
16. Industrial and Commercial Holdings Group Ltd.
17. Industrial Investment Co. Ltd.
18. Industrial Selections Ltd.
19. McDonald Forman and Co. Ltd.
20. Malcomess-Bakke Ltd.
21. Metje and Ziegler Ltd.
22. Mitchell Cotts Ltd.
23. Premier Industries Ltd.
24. Protea Holdings Ltd.
25. Rentmeesterbeleggings Bpk.
26. Ryan Nigel Holdings Ltd.
27. Sakers Finance and Investment Corporation Ltd.
28. Sand Consolidated Investment Ltd
29. South Atlantic Corporation Ltd
30. Suiderland Development Corporation Ltd.
31. Teal Holdings Ltd.
32. Tollman Hotels and Tourist Industries Ltd.
33. Trade and Industry Acceptance Corporation Ltd.
34. Turf Holdings Ltd.
35. W & A Investment Corporation Ltd.

TABLE A.4 - INSURANCE SECTOR

1. Guardian Assurance Holdings (South Africa) Ltd.
2. LibertyLife Association of Africa Ltd.
3. The Marine & Trade Insurance Co. Ltd.
4. Protea Assurance Co. Ltd.
5. S.A. Eagle Insurance Co. Ltd.
6. Union National South British Insurance Ltd.

TABLE A.5 - BUILDING SECTOR

1. Alex Blaikie Holdings Ltd.
2. Anglo-Alpha Cement Ltd.
3. Bellandia Homes Investment Ltd.
4. Blue Circle Cement Ltd.
5. Boumat Ltd.
6. The Brick and Potteries Company Ltd.
7. Bruynzeel Plywoods Ltd.
8. Buffalo Timber and Hardware Co. Ltd.
9. Coronation Brick Free State Ltd.
10. Everite Ltd.
11. Golden Brown Brick and Tile Co. Ltd.
12. Grinaker Holdings Ltd.
13. Gypsum Industries Ltd.
14. W.F. Johnstone and Co. Ltd.
15. Katzenellenbogen Ltd.
16. L.T.A. Ltd.
17. Masonite (Africa) Ltd.
18. Murray & Roberts Holdings Ltd.
19. Plascon-Evans Paints Ltd.
20. Plate Glass and Shatterprufe Industries Ltd.
21. The Premier Portland Cement (Rhodesia) Ltd.
22. Pretoria Portland Cement Co. Ltd.
23. Rhodesian Cement Ltd.
24. Rhodesian Brick & Potteries Ltd.
25. Sinclair Holdings Ltd.

TABLE A.6 - FOOD SECTOR

1. African Products Manufacturing Co. Ltd.
2. Bakers South Africa Ltd.
3. T.W. Beckett and Company Ltd.
4. Crown Mills Holdings Ltd.
5. Irvin & Johnson Ltd.
6. Imperial Cold Storage and Supply Co. Ltd.
7. Jabula Foods Ltd.
8. H. Lewis and Co. Ltd.
9. Monis and Fattis Industries Ltd.
10. Picardi Canners Ltd.
11. The Premier Milling Co. Ltd.
12. Simba-Quix Ltd.
13. Stein Brothers (Holdings) Ltd.
14. Tiger Oats and National Milling Co. Ltd.
15. The Union Cold Storage of South Africa Ltd.

TABLE A.7 - MOTORS SECTOR

1. Associated Engineering S.A. Ltd.
2. Aurochs Investment Co. (S.A.) Ltd.
3. Autolec Ltd.
4. Bus Bodies (S.A.) Ltd.
5. Capital Gold & Exploration Co. Ltd.
6. Currie Motors (1946) Ltd.
7. Dunlop (South Africa) Ltd.
8. Eriksen Consolidated Holdings Ltd.
9. The General Tire & Rubber Co. (S.A.) Ltd.
10. Lawson Motors Group Ltd.
11. Lucy's Holdings Ltd.
12. McCarthy Rodway Ltd.
13. Northern Free State Motors Ltd.
14. Public Utility Transport Corporation Ltd.
15. Robbs Holdings Ltd.
16. Tollgate Holdings Ltd.
17. Toyota (South Africa) Ltd.
18. Transport & Engineering Investment Corporation Ltd.
19. Welfit Oddy Holdings Ltd.
20. Williams Hunt South Africa Ltd.

TABLE A.8 - PAPER AND PULP SECTOR

1. Canadian Overseas Packaging Industries Ltd.
2. Coates Brothers (South Africa) Ltd.
3. Consolidated Glass Works Ltd.
4. Evelyn Haddon & Co. Ltd.
5. Kohler Brothers Ltd.
6. The Metal Box Company of South Africa Ltd.
7. Metal Closures Group South Africa Ltd.
8. Premier Paper Mills Ltd.
9. Press Supplies Holdings Ltd.
10. Rhodesian Pulp & Paper Industries Ltd.
11. Sappi Ltd.
12. Trio-Rand (S.A.) Bpk.

TABLE A.9 - STORES SECTOR

1. Edgars Stores Ltd.
2. Foschini Ltd.
3. Garlick Ltd.
4. Greatermans Stores Ltd.
5. Harrowe's Ltd.
6. Hepworths Ltd.
7. John Orr Holdings Ltd.
8. Lewis Foschini Investment Co. Ltd.
9. O.K. Bazaars (1929) Ltd.
10. M. & S. Spitz Footwear Holdings Ltd.
11. Stuttaford & Co. Ltd.
12. Truworths Ltd.
13. Woolworths Holdings Ltd.

TABLE A.10 - SUGAR SECTOR

1. Crookes Brothers Ltd.
2. Huletts Corporation Ltd.
3. Illovo Sugar Estates Ltd.
4. C.G. Smith Sugar Ltd.
5. Swaziland Sugar Milling Co. Ltd.
6. Tongaat Group Ltd.

APPENDIX B

THE CLUSTER ANALYSIS RESULTS USED IN CHAPTER TWO

Cluster analysis is defined by Tryon and Bailey (1970) as "the general logic, formulated as a procedure, by which we objectively group together entities on the basis of their similarities and differences." Obviously such a technique is of considerable interest to the stock market researcher who can use it in numerous ways. For example, it could be applied to the market as a whole to determine which securities should be grouped together (that is, to form the different sectors), it could be applied to some measure of risk which changes over time (for example the S.D. ratio discussed in Chapter 6) to group the securities according to risk profile, etcetera.

In Chapter 3, the technique of cluster analysis was used to determine, within a sector, which securities had similar price movements. This was done using the program P1M of the BMDP statistical package (see Dixon (1975)). The algorithm used was the average linkage algorithm (discussed, for example, in Anderberg (1973)) with the sample correlation between each pair of securities being used as the distance measurement.

Unfortunately, there is no clearcut evidence on which clustering algorithm is in fact the best or the most suitable. The BMDP program P1M provides two additional alternative algorithms, the single linkage algorithm and the complete linkage algorithm. Of the three algorithms available, the single linkage algorithm and the average linkage algorithm are probably the most suitable for use on the sample correlation coefficients of raw stock market prices.

Both of these methods have their advantages and disadvantages. The single linkage algorithm has the definite disadvantage of tending to form clusters which are "strung out in long sausage like shapes" (Hartigan (1975)) which, it is argued, is highly undesirable for use in general stock market applications. On the other hand, the average linkage algorithm, when applied to the correlation matrix, has the disadvantage that the correlations are not strictly additive. However, it is felt that for the application discussed in Chapter 2, the average linkage method is the more suitable because it is preferable to have some average measure of the association within a cluster (albeit a somewhat nebulous measurement) rather than to have a strung out "sausage like" cluster in which securities grouped in the same cluster might really be "far" apart.

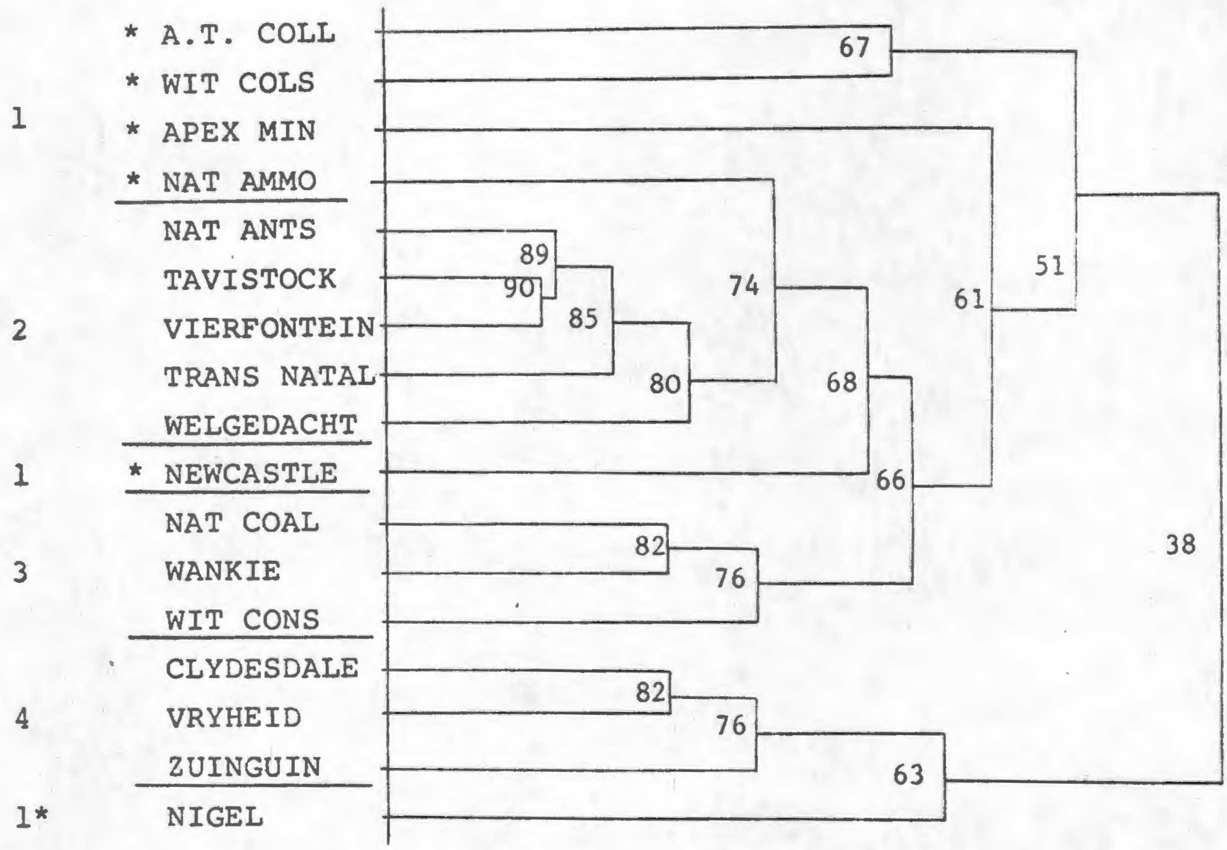
Finally, it should be mentioned that there are other clustering techniques besides the three alternatives given in the BMDP program. It might be that in some particular case one or more of these alternatives may be more appropriate than the average linkage method used in Chapter 2. In such a case, these alternatives should obviously be used. However, as Hwei-Ju Chen et al (1974) point out, the subject is relatively undeveloped and filled with unsolved problems and thus, in the absence of both a clearcut rule as to which is the most suitable algorithm and alternative computer programs, the average linkage algorithm was used in the analysis of Chapter 2.

The dendrograms (see Sokal and Sneath (1963)) for each sector examined in Chapter 2 are presented below. These dendrograms were constructed from the output of the program BMDP.P1M where the average correlations were scaled as follows:

Correlation	Value
-1,0	0
-0,8	10
-0,6	20
-0,4	30
-0,2	40
0,0	50
0,2	60
0,4	70
0,6	80
0,8	90
1,0	100

The individual clusters are indicated by means of a solid line under the last security falling in the given cluster.

FIGURE B.1 : COAL SECTOR



Note: The * securities were combined into one group - those securities that did not form a cluster or fall into any of the other clusters.

FIGURE B.2 : GOLD: WITWATERSRAND & OTHERS SECTOR

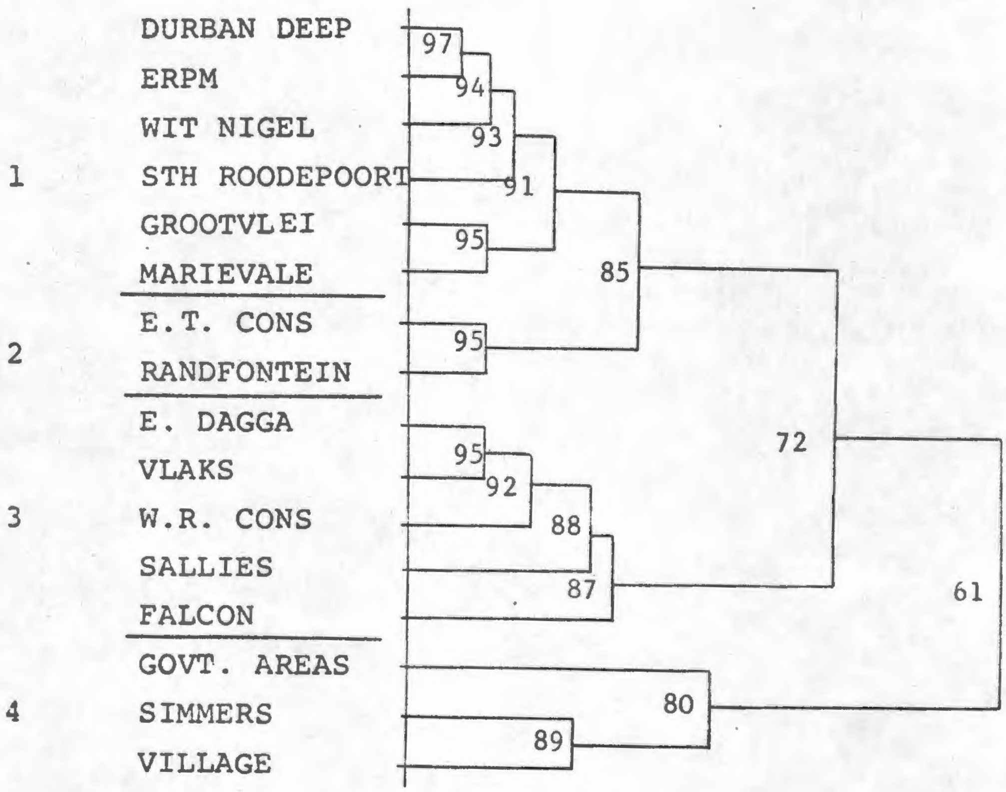


FIGURE B.3 : FINANCIAL INDUSTRIAL SECTOR

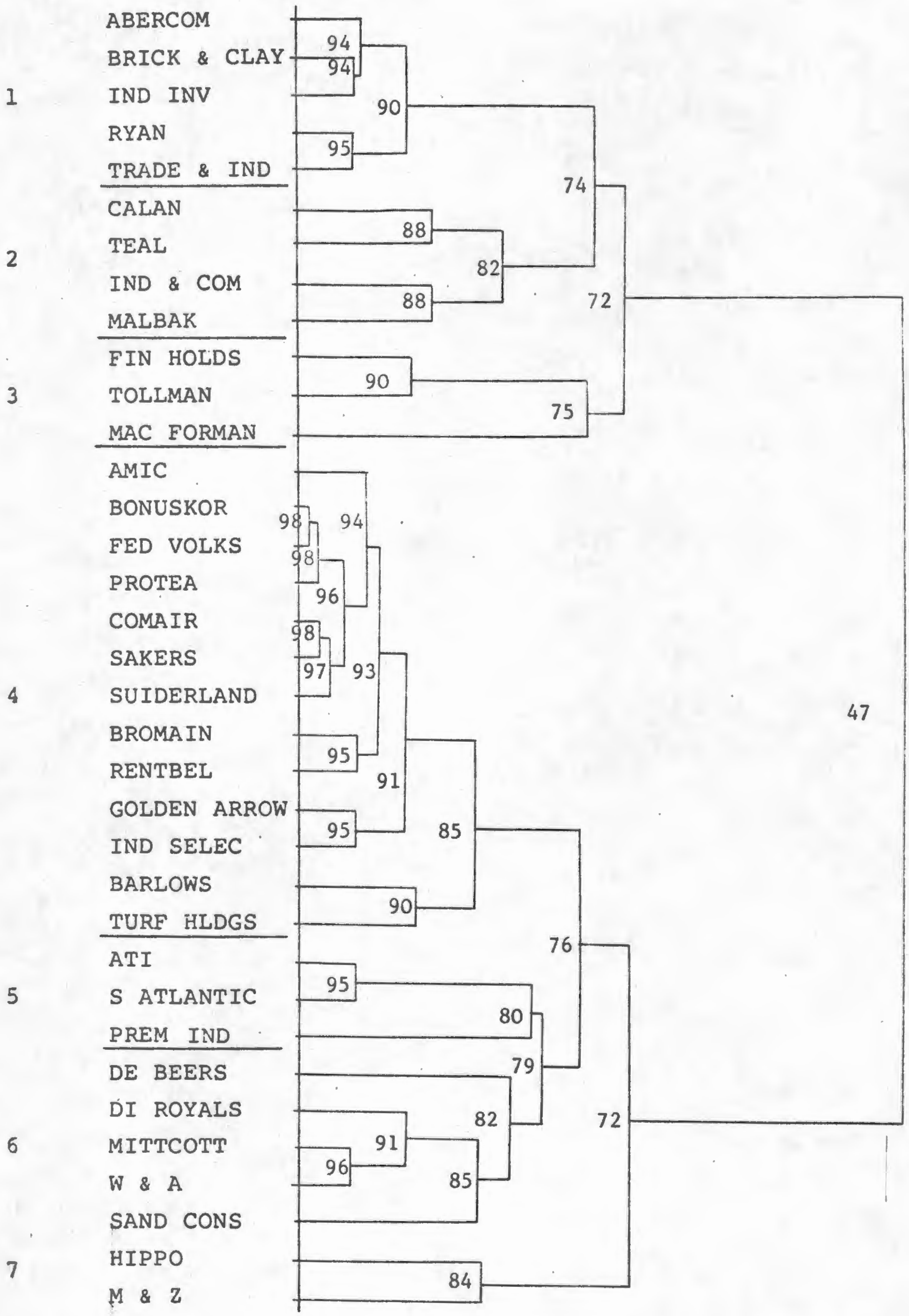


FIGURE B.4 : FINANCIAL INSURANCE SECTOR

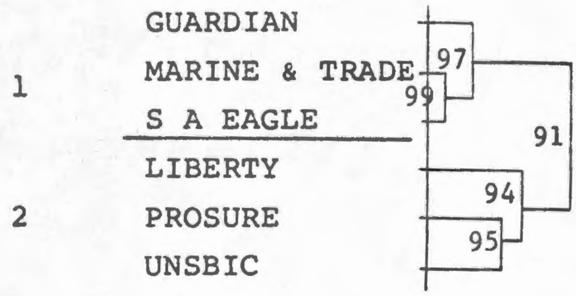
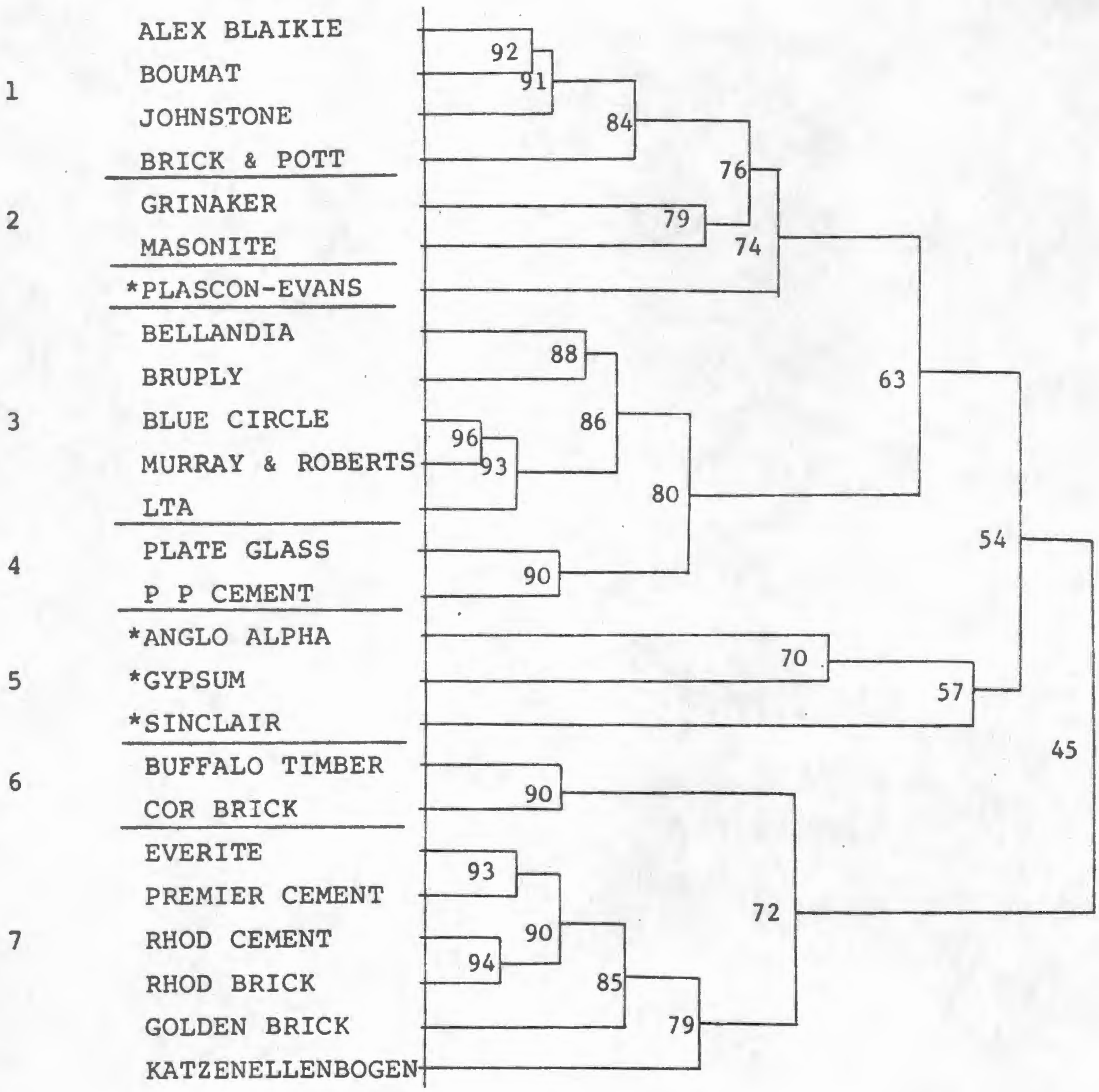


FIGURE B.5 : BUILDING SECTOR



* These securities were combined into one group - those that did not fall into any cluster.

FIGURE B.6 : FOOD SECTOR

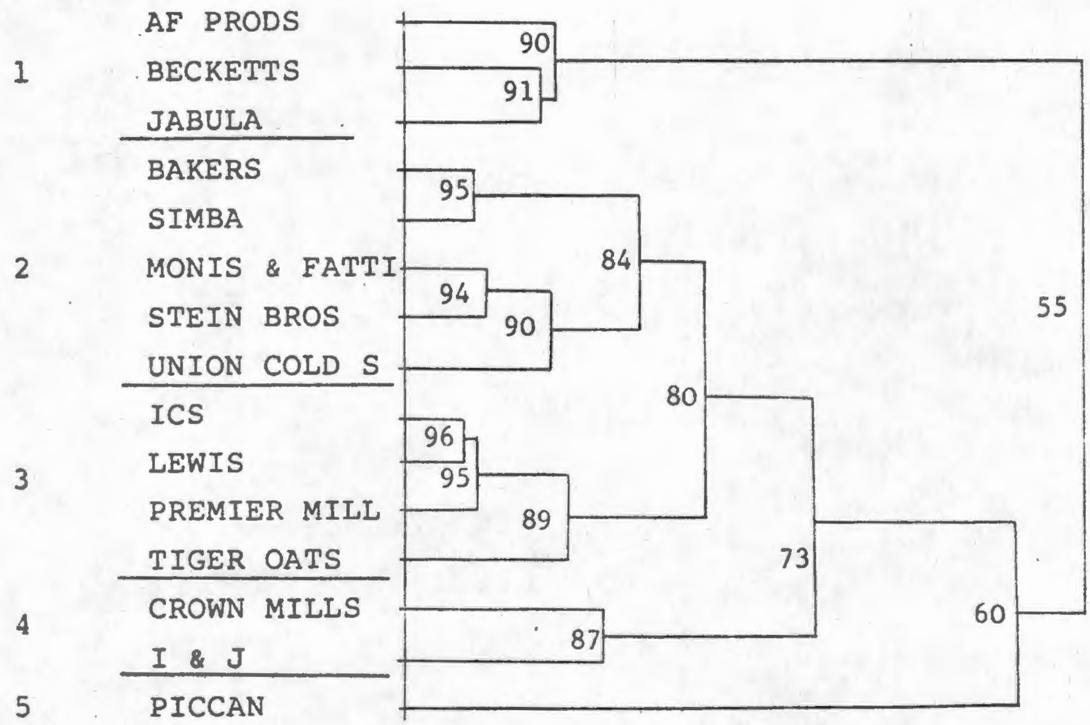


FIGURE B.7 : MOTORS SECTOR

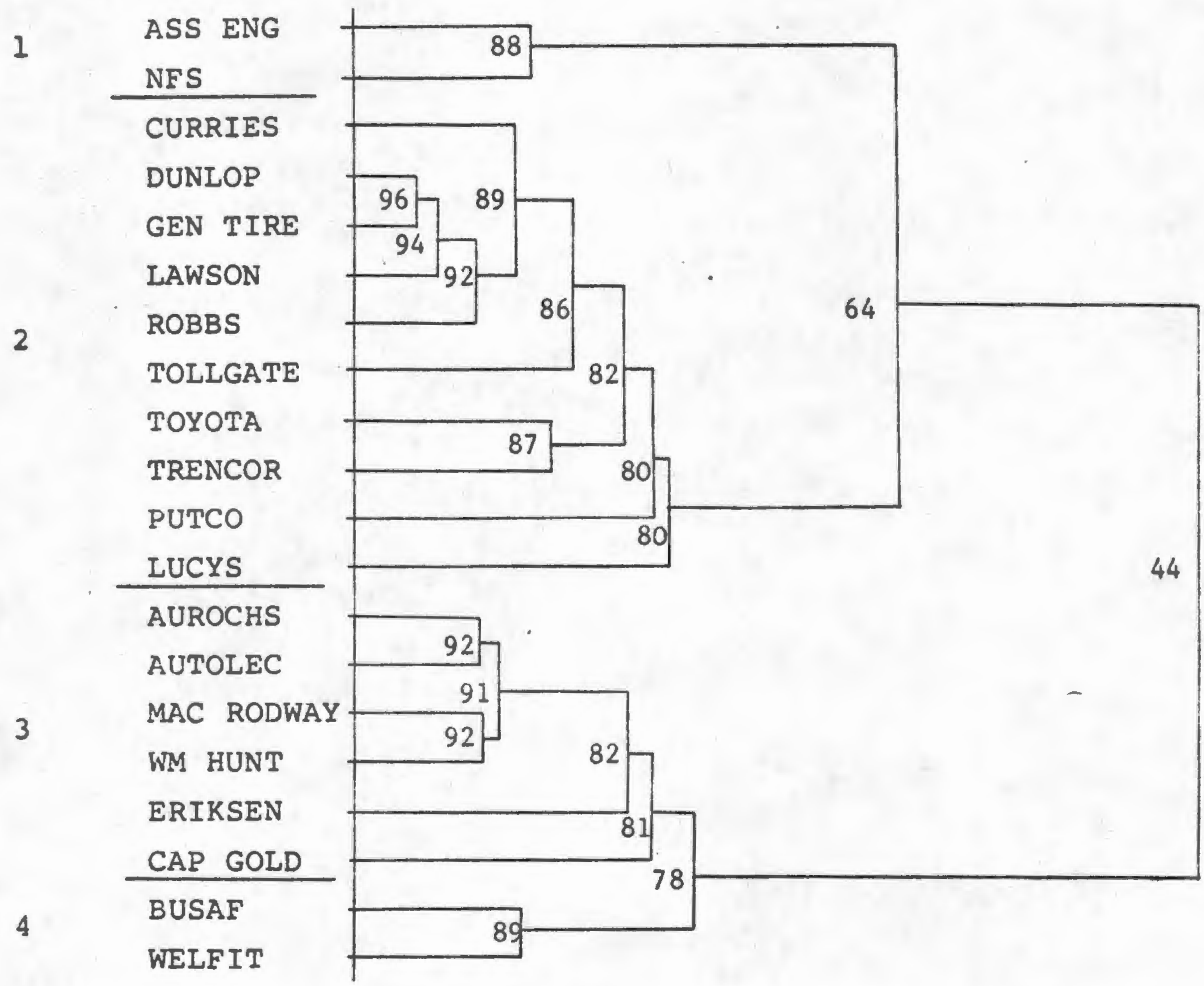


FIGURE B.8 : PAPER & PULP SECTOR

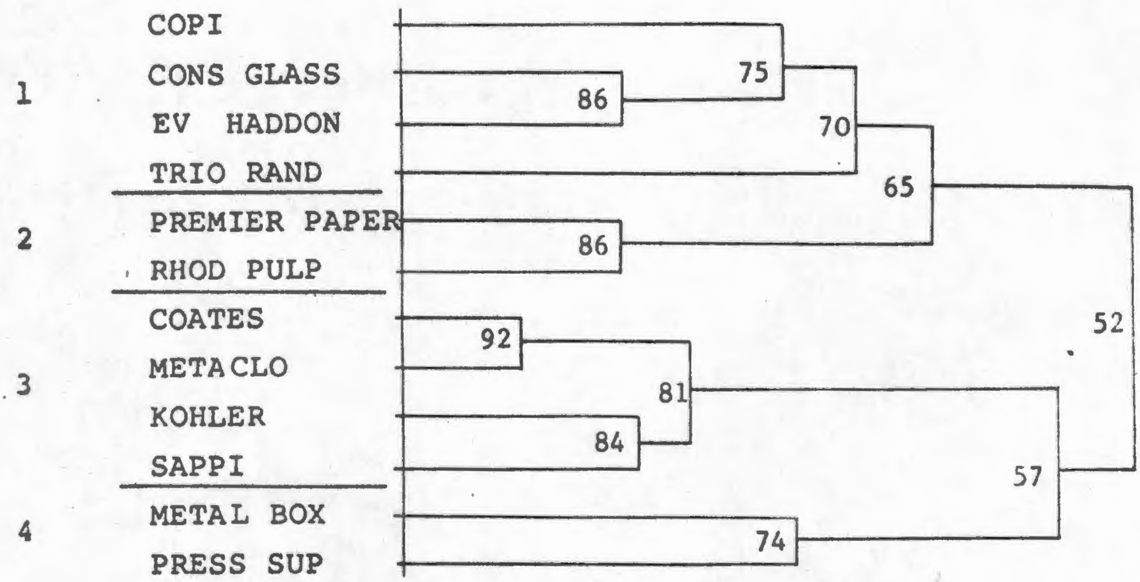
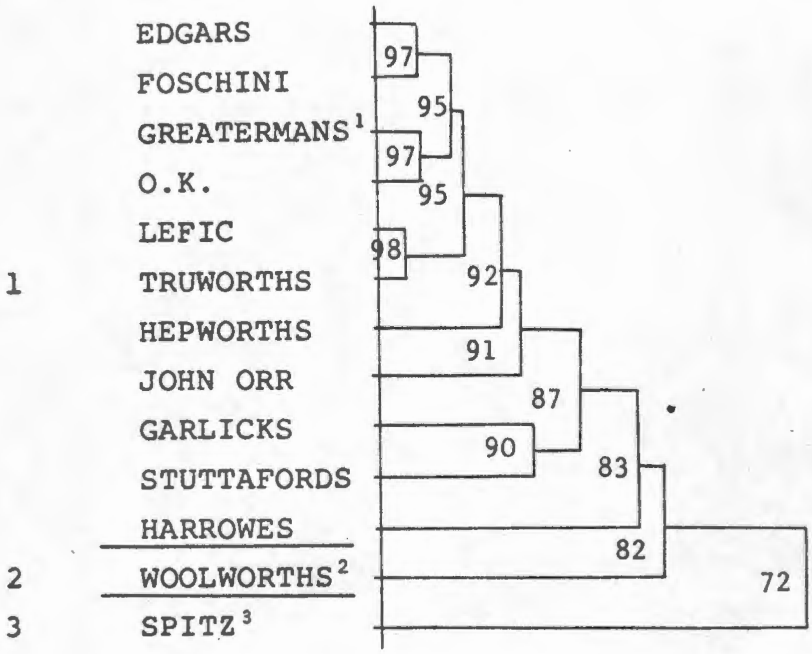


FIGURE B.9 : STORES SECTOR

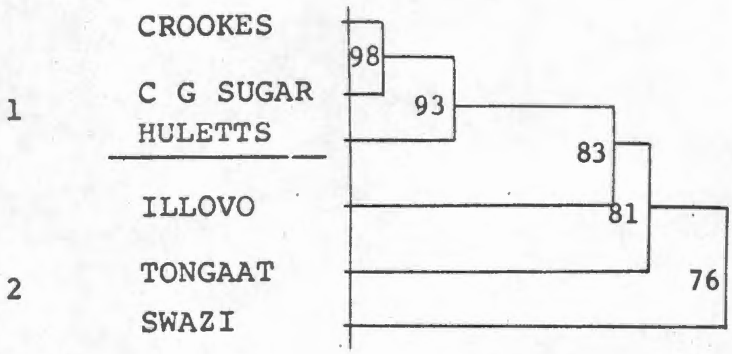


Notes: (1) Greatermans comprised of both the ordinary and the 'A' shares.

(2) Woolworths comprised of both the ordinary and the 'A' shares and as these had large market capitalization it was decided to include them as a cluster on their own.

(3) Spitz was left as a cluster on its own as it does not fit into any of the other clusters satisfactorily.

FIGURE B.10 : SUGAR SECTOR



A P P E N D I X C

SECURITIES USED IN CONSTRUCTING INDICES OF CHAPTER THREE

1. Amalgamated Collieries of South Africa Ltd.
2. The Clydesdale (Transvaal) Collieries Ltd.
3. Trans-Natal Coal Corporation Ltd.
4. Wankie Colliery Co. Ltd.
5. Witbank Colliery Ltd.
6. Anglo American Investment Trust Ltd.
7. The Consolidated Diamond Mines of South West Africa Ltd.
8. De Beers Consolidated Mines Ltd.
9. Industrial Diamonds of South Africa (1945) Ltd.
10. East Rand Proprietary Mines Ltd.
11. The Grootvlei Proprietary Mines Ltd.
12. Marievale Consolidated Mines Ltd.
13. The Randfontein Estates Gold Mining Co. Witwatersrand Ltd.
14. Bracken Mines Ltd.
15. Kinross Mines Ltd.
16. Winkelhaak Mines Ltd.
17. Buffelsfontein Gold Mining Co. Ltd.
18. Hartebeesfontein Gold Mining Co. Ltd.
19. Southvaal Holdings Ltd.
20. Vaal Reefs Exploration and Mining Co. Ltd.
21. Zandpan Gold Mining Co. Ltd.
22. Free State Geduld Mines Ltd.
23. Harmony Gold Mining Co. Ltd.
24. President Brand Gold Mining Co. Ltd.
25. President Steyn Gold Mining Co. Ltd.
26. St. Helena Gold Mines Ltd.
27. Welkom Gold Mining Co. Ltd.
28. Western Holdings Ltd.
29. East Driefontein Gold Mining Co. Ltd.
30. Elsburg Gold Mining Co. Ltd.
31. Kloof Gold Mining Co. Ltd.
32. West Driefontein Gold Mining Co. Ltd.
33. Western Areas Gold Mining Co. Ltd.
34. Western Deep Levels Ltd.
35. M.T.D. (Mangula) Ltd.
36. The Messina (Transvaal) Development Co. Ltd.
37. Minerals and Resources Corporation Ltd.
38. Palabora Mining Co. Ltd.
39. Zambia Copper Investments Ltd.
40. Associated Manganese Mines of South Africa Ltd.
41. S.A. Manganese Amcor Ltd.
42. Lydenburg Platinum Ltd.
43. Potgietersrust Platinums Ltd.
44. Union Platinum Mining Co. Ltd.
45. Waterval (Rustenburg) Platinum Mining Co. Ltd.
46. The Rooiberg Minerals Development Co. Ltd.
47. Union Tin Mines Ltd.
48. Zaaiplaats Tin Mining Co. Ltd.

49. Consolidated African Mines Ltd.
50. Consolidated Murchison Co. Ltd.
51. The Griqualand Exploration and Finance Co. Ltd.
52. The Northern Lime Co. Ltd.
53. Anglo American Gold Investment Co. Ltd.
54. Anglo Transvaal Consolidated Investment Co. Ltd.
55. Charter Consolidated Ltd.
56. Consolidated Gold Fields Ltd.
57. Geduld Investments Ltd.
58. General Mining and Finance Corporation Ltd.
59. Johannesburg Consolidated Investment Co. Ltd.
60. Middle Witwatersrand (Western Areas) Ltd.
61. Rand Selection Corporation Ltd.
62. Sentrust Beperk.
63. Transvaal Consolidated Land and Exploration Co. Ltd.
64. U.C. Investments Ltd.
65. Union Corporation Ltd.
66. The Vereeniging Estates Ltd.
67. Abercom Investments Ltd.
68. Anglo American Industrial Corporation Ltd.
69. Anglo-Transvaal Industries Ltd.
70. Barlow Rand Ltd.
71. Bonuskor Beperk.
72. Calan Ltd.
73. De Beers Industrial Corporation Ltd.
74. Federale Volksbeleggings Beperk.
75. Industrial Selections Ltd.
76. Lonrho Ltd.
77. Protea Holdings Ltd.
78. Rennies Consolidated Holdings Ltd.
79. South Atlantic Corporation Ltd.
80. Unisec Group Ltd.
81. AFC Investments Ltd.
82. First Union General Investment Trust Ltd.
83. Hesperus Holdings Ltd.
84. Priority Investments Trust Ltd.
85. Bank Holdings Corporation of S.A. Ltd.
86. Barclays National Bank Ltd.
87. Nedbank and Syfrets UAL Holdings Ltd.
88. Santam Bank Ltd.
89. Standard Bank Investment Corporation Ltd.
90. The Trust Bank of Africa Ltd.
91. Volkskas Bpk.
92. Guardian Assurance Holdings (South Africa) Ltd.
93. Liberty Life Association of Africa Ltd.
94. S.A. Metropolitan Life Assurance Co. Ltd.
95. Mutual and Federal Insurance Co. Ltd.
96. S.A. Eagle Insurance Co. Ltd.
97. Anglo American Properties Ltd.
98. Glen Anil Development Corporation Ltd.
99. Gold Fields Property Co. Ltd.
100. Mondorp Bpk.
101. Retco Ltd.
102. Sorec Ltd.
103. Unisec Developers and Contractors Ltd.

104. The Oudemeester Group Ltd.
105. Picardi Beleggings Bpk.
106. The South African Breweries Ltd.
107. South African Distilleries and Wines Ltd.
108. Stellenbosch Wine Trust Ltd.
109. Anglo-Alpha Cement Ltd.
110. Blue Circle Cement Ltd.
111. Boumat Limited.
112. Bruynzeel Plywoods Ltd.
113. Murray and Roberts Holdings Ltd.
114. Placor Holdings Ltd.
115. Pretoria Portland Cement Co. Ltd.
116. Schachat Holdings Ltd.
117. A.E.C.I. Ltd.
118. Coal By-products and Investments Ltd.
119. Federale Kunsmis Beperk.
120. Sentrachem Ltd.
121. Charmfit Holdings Ltd.
122. Delswa Ltd.
123. Dugson Holdings Ltd.
124. Veka Ltd.
125. Kaap Kunene Beleggings Bpk.
126. Marine Products Ltd.
127. Ovenstone Investments Ltd.
128. South West Africa Fishing Industries Ltd.
129. African Products Manufacturing Co. Ltd.
130. Asokor Ltd.
131. The Premier Milling Co. Ltd.
132. Stein Brothers (Holdings) Ltd.
133. Tiger Oats and National Milling Co. Ltd.
134. Eddels Holdings Ltd.
135. Searles Holdings Ltd.
136. Silverton Tannery Ltd.
137. M. & S. Spitz Footwear Holdings Ltd.
138. Associated Furniture Companies Ltd.
139. Ellerine Holdings Ltd.
140. Russell Holdings Ltd.
141. Steel and Barnett Ltd.
142. World Furnishers Group Ltd.
143. Aberdare Cables Africa Ltd.
144. African Oxygen Ltd.
145. Asea Electric South Africa Ltd.
146. Dorman Long Vanderbijl Corporation Ltd.
147. C.J. Fuchs Ltd.
148. Highveld Steel and Vanadium Corporation Ltd.
149. Hubert Davies and Co. Ltd.
150. Metkor Investments Ltd.
151. Oldham and Son (Africa) Ltd.
152. Stewarts and Lloyds of South Africa Ltd.
153. Datsun Nissan Beleggings Maatskappy Bpk.
154. Dunlop (South Africa) Ltd.
155. The General Tire and Rubber Co. (S.A.) Ltd.
156. McCarthy Rodway Ltd.
157. Toyota (South Africa) Ltd.

158. Wesco Investments Ltd.
159. Williams Hunt South Africa Ltd.
160. Carlton Paper Corporation Ltd.
161. Consolidated Glass Works Ltd.
162. The Metal Box Company of South Africa Ltd.
163. National Amalgamated Packaging Ltd.
164. Sappi Ltd.
165. Adcock Ingram (Chemists) Ltd.
166. Dermacult Ltd.
167. Skye Ltd.
168. South African Druggists Ltd.
169. Twins Pharmaceutical Holdings Ltd.
170. Argus Printing and Publishing Co. Ltd.
171. C.N.A. Investments Ltd.
172. Hortors Ltd.
173. South African Associated Newspapers Ltd.
174. Vaderland Beleggings Bpk.
175. Bonmore Investments Ltd.
176. Edgars Consolidated Investments Ltd.
177. Edgars Stores Ltd.
178. O.K. Bazaars (1929) Ltd.
179. Pick 'n Pay Stores Ltd.
180. Woolworths Holdings Ltd.
181. Gledhow Sugar Co. Ltd.
182. Hulett's Corporation Ltd.
183. Reynolds Brothers Ltd.
184. Swaziland Sugar Milling Co. Ltd.
185. Tongaat Group Ltd.
186. The Carpet Manufacturing Company (South Africa) Ltd.
187. Consolidated Textile Mills Investment Corporation Ltd.
188. Mooi River Textiles Ltd.
189. Romatex Ltd.
190. The Lion Match Co. Ltd.
191. Rembrandt Beherende Beleggings Bpk.
192. Rembrandt Group Ltd.
193. Tegniese Beleggings Korporasie Bpk.
194. Utico Holdings Ltd.
195. Empisal (South Africa) Ltd.
196. Gallo (Africa) Ltd.
197. Marshall Industrials Ltd.
198. Sterns Diamonds Ltd.
199. Woolfson's Holdings Ltd.
200. Advance Holdings Ltd.
201. Mathieson and Ashley Ltd.
202. Spectro Beherend Bpk.
203. Trek Beleggings Bpk.

A P P E N D I X D

THE COMPUTER PROGRAMS FOR THE FLEXIBLE TOLERANCE METHOD

The computer programs used to solve the Flexible Tolerance Method are those listed in Himmelblau (1972) with a few minor alterations necessitated by the particular computer configuration used and the particular problem under discussion. For further details on the workings of these programs, the reader is referred to Himmelblau (1972).

The programs consist of a main program and five subroutines, one of which must be written by the user. This subroutine, called PROB, must have three entry points, called PROB1, PROB2, and PROB3 respectively. This subroutine must be written so that PROB1 contains specification of all the equality constraints, written in " = 0" form. PROB2 must contain all inequality constraints which must be in the " $\geq$ 0" form. Finally, PROB3 must specify the objective function which must be of the minimization type. This is perhaps best illustrated by an example, namely the last example presented in Chapter 4. That is, the problem is:

$$\text{Max} \quad x' \Sigma x$$

$$\text{S.T.} \quad \sum x_i^2 = 1$$

$$0 \leq x_i \leq 0{,}7 \ .$$

Thus, PROB1 must specify $1 - \sum x_i^2 = 0$;

PROB2 must specify $0,7 - x_i \geq 0$,

$x_i \geq 0$; and

PROB3 must specify $-x'Sx$.

The subroutine for this example is listed below.

```
      SUBROUTINE PROB(X,R,COVAR,NX)
      DIMENSION X(50),R(100),COVAR(50,50)
      ENTRY PROB1(X,R,COVAR,NX)
      R(1) = 1.0
      DO 1 I=1,NX
      R(1) = R(1) - X(I)**2
1     CONTINUE
      GO TO 5
      ENTRY PROB2(X,R,COVAR,NX)
      DO 2 I=1,NX
      R(I+1) = X(I)
      R(I+NX+1) = 0.7 - X(I)
2     CONTINUE
      GO TO 5
      ENTRY PROB3(X,R,COVAR,NX)
      KI = NX + NX + 2
      R(KI) = 0.0
      DO 4 I=1,NX
      DO 3 J=1,NX
      R(KI) = X(I)*X(J)*COVAR(I,J) + R(KI)
3     CONTINUE
4     CONTINUE
      R(KI) = -1.0 * R(KI)
5     RETURN
      END
```

The main program and four remaining subroutines are listed below. The card deck for these programs was provided by Associate Professor F. Jackson of the Department of Applied Mathematics of the University of Cape Town.

```
C
C     PROGRAM FLEXIPLEX.
C
C      * * * * * * * P R O G R A M   F L E X I P L E X * * * * * * *
C
C
C
C      NX        TOTAL NUMBER OF INDEPENDENT VARIABLES
C      NC        TOTAL NUMBER OF EQUALITY CONSTRAINTS
C      NIC       TOTAL NUMBER OF INEQUALITY CONSTRAINTS
C      SIZE      EDGE LENGTH OF THE INITIAL POLYHEDRON
C      CONVER    CONVERGENCE CRITERION FOR TERMINATION OF THE SEARCH
C      ALFA      THE REFLECTION COEFFICIENT
C      BETA      THE CONTRACTION COEFFICIENT
C      GAMA      THE EXPANSION COEFFICIENT
C      X(I)      THE ASSUMED VECTOR TO INITIATE THE SEARCH
C      FDIFER    THE TOLERANCE CRITERION FOR CONSTRAINT VIOLATION
C      ICONT     A COUNTER TO RECORD STAGE COMPUTATIONS
C      NCONT     A COUNTER TO PRINT INFORMATION EVERY (NX+1) STAGE
C      LOW       AN INDEX TO IDENTIFY INFORMATION RELATED TO THE LOWEST
C                VALUE OF OBJ. FUNCTION IN MOST RECENT POLYHEDRON
C      LHIGH     AN INDEX TO IDENTIFY INFORMATION RELATED TO LARGEST VALUE
C                OF OBJ. FUNCTION IN MOST RECENT POLYHEDRON
C      LSEC      AN INDEX TO IDENTIFY INFORMATION RELATED TO THE SECOND
C      LARGEST VALUE OF OBJ. FUNCTION IN MOST RECENT POLYHEDRON
C
C                                ********************
C
      DIMENSION X(50),X1(50,50),X2(50,50),R(100),SUM(50),F(50),SR(50),
     1ROLD(100),COVAR(50,50)
      COMMON/ONE/NX,NC,NIC,STEP,ALFA,BETA,GAMA,IN,INF,FDIFER,SEQL,K1,K2,
     1K3,K4,K5,K6,K7,K8,K9,X,X1,X2,R,SUM,F,SR,ROLD,SCALE,FOLD
     2,COVAR
      COMMON/TWO/LFEAS,L5,L6,L7,L8,L9,R1A,R2A,R3A
C     PROBLEM IDENTIFICATION HEADER IS READ IN AFTER THIS CARD
C     PARAMETERS FOR THE PROBLEM ARE READ IN AFTER THIS CARD
      READ 1, NX,NC,NIC,SIZE,CONVER
      READ(5,100) NPR
100   FORMAT(I5)
      WRITE(5,99)
99    FORMAT(///,' COVARIANCE MATRIX',///)
      DO 102 I=1,NX
      READ(NPR,101) (COVAR(I,J),J=1,NX)
101   FORMAT(F12.3)
      WRITE(5,101) (COVAR(I,J),J=1,NX)
102   CONTINUE
      ALFA=1.
      BETA=0.5
```

```
      GAMA=2.
      ND = NX
      ND2 = 1
   10 CONTINUE
      DO 605 I=1,100
      R(I) = 0.0
605   CONTINUE
C     PERMANENT DATA FOR THE PROBLEM SHOULD BE READ IN AFTER THIS CARD
C     TEMPORARY DATA FOR THE PROBLEM, SUCH AS VARIABLE COEFFICIENTS OR
C     NEW PARAMETERS SHOULD BE READ IN AFTER THIS CARD
      STEP=SIZE
C     THE ASSUMED INITIAL VECTOR IS READ IN AFTER THIS CARD
      IZ = IZ + 1
      IF ( IZ .GT. ND2 )  GO TO 9999
      READ(20,650) (X(J),J=1,ND)
650   FORMAT(F12.5)
11    PRINT 100
      PRINT 759
      PRINT 756,NX,NC,NIC,SIZE,CONVER
      K1=NX+1
      K2=NX+2
      K3=NX+3
      K4=NX+4
      K5=NX+5
      K6=NC+NIC
      K7=NC+1
      K8=NC+NIC
      K9=K8+1
      N=NX-NC
      N1=N+1
      IF(N1.GE.3) GO TO 50
      N1=3
      N=2
50    N2=N+2
      N3=N+3
      N4=N+4
      N5=N+5
      N6=N+6
      N7=N+7
      N8=N+8
      XN=N
      XNX=NX
      XN1=N1
      R1A=0.5*(SQRT(5.)-1.)
      R2A=R1A*R1A
      R3A=R2A*R1A
      L5=NX+5
      L6=NX+6
      L7=NX+7
      L8=NX+8
      L9=NX+9
      ICONT=1
      NCONT=1
      PRINT 115
      PRINT 116, (X(J),J=1,NX)
      FDIFER=2.*(NC+1)*STEP
      FOLD=FDIFER
      IN=N1
      CALL SUMR
      SR(N1)=SQRT(SEQL)
```

```
      PRINT 763,FDIFER,SR(N1)
      IF(SR(N1).LT.FDIFER) GO TO 341
      CALL WRITEX
      PRINT 757
      INF=N1
      STEP=0.05*FDIFER
      CALL FEASBL
      PRINT 764
      PRINT 116, (X2(INF,J),J=1,NX)
      PRINT 765,SR(INF)
      IF(FOLD.LT.1.0E-09) GO TO 80
341   PRINT 35
      PRINT 758,ICONT,FDIFER
      CALL WRITEX
      FTER=R(K9)
C     COMPUTE CENTROID OF ALL VERTICES OF INITIAL POLYHEDRON
237   STEP1=STEP*(SQRT(XNX+1.)+XNX-1.)/(XNX*SQRT(2.))
      STEP2=STEP*(SQRT(XNX+1.)-1.)/(XNX*SQRT(2.))
      ETA=(STEP1+(XNX-1.)*STEP2)/(XNX+1.)
      DO 4 J=1,NX
      X(J)=X(J)-ETA
4     CONTINUE
      CALL START
      DO 9 I=1,N1
      DO 9 J=1,NX
      X2(I,J)=X1(I,J)
9     CONTINUE
      DO 5 I=1,N1
      IN=I
      DO 6 J=1,NX
6     X(J)=X2(I,J)
      CALL SUMR
      SR(I)=SQRT(SEQL)
      IF(SR(I).LT.FDIFER) GO TO 8
      CALL FEASBL
      IF(FOLD.LT.1.0E-08) GO TO 80
    8 CALL PROB3(X,R,COVAR,NX)
      F(I) = R(K9)
    5 CONTINUE
1000  STEP=0.05*FDIFER
      ICONT=ICONT+1
C     SELECT LARGEST VALUE OF OBJECTIVE FUNCTION FROM POLYHEDRON VERTICES
      FH=F(1)
      LHIGH=1
      DO 16 I=2,N1
      IF(F(I).LT.FH) GO TO 16
      FH=F(I)
      LHIGH=I
16    CONTINUE
C     SELECT MINIMUM VALUE OF OBJECTIVE FUNCTION FROM POLYHEDRON VERTICES
41    FL=F(1)
      LOW=1
      DO 17 I=2,N1
      IF(FL.LT.F(I)) GO TO 17
      FL=F(I)
      LOW=I
17    CONTINUE
      DO 36 J=1,NX
36    X(J)=X2(LOW,J)
      IN=LOW
```

```
      CALL SUMR
      SR(LOW) = SQRT(SEQL)
      IF(SR(LOW).LT.FDIFER) GO TO 87
      INF=LOW
      CALL FEASBL
      IF(FOLD.LT.1.0E-08) GO TO 80
      CALL PROB3(X,R,COVAR,NX)
      F(LOW)=R(K9)
      GO TO 41
87    CONTINUE
C     FIND CENTROID OF POINTS WITH I DIFFERENT THAN LHIGH
      DO 19 J=1,NX
      SUM2=0.
      DO 20 I=1,N1
20    SUM2=SUM2+X2(I,J)
19    X2(N2,J)=1./XN*(SUM2-X2(LHIGH,J))
      SUM2=0.
      DO 36 I=1,N1
      DO 36 J=1,NX
      SUM2=SUM2+(X2(I,J)-X2(N2,J))**2
36    CONTINUE
      FDIFER=(NC+1)/XN1*SQRT(SUM2)
      IF(FDIFER.LT.FOLD) GO TO 98
      FDIFER=FOLD
      GO TO 198
98    FOLD=FDIFER
198   CONTINUE
      FTER=F(LOW)
137   NCONT=NCONT+1
      IF(NCONT.LT.4*N1) GO TO 37
      IF(ICONT.LT.1500) GO TO 337
      FOLD=0.5*FOLD
337   NCONT=0
      PRINT 35
      PRINT 755,ICONT,FDIFER
      CALL WRITEX
37    IF(FDIFER.LT.CONVER) GO TO 81
C     SELECT SECOND LARGEST VALUE OF OBJECTIVE FUNCTION
      IF(LHIGH.EQ.1) GO TO 43
      FS=F(1)
      LSEC=1
      GO TO 44
43    FS=F(2)
      LSEC=2
44    DO 18 I=1,N1
      IF(LHIGH.EQ.I) GO TO 18
      IF(F(I).LT.FS) GO TO 18
      FS=F(I)
      LSEC=I
18    CONTINUE
C     REFLECT HIGH POINT THROUGH CENTROID
      DO 61 J=1,NX
      X2(N3,J)=X2(N2,J)+ALFA*(X2(N2,J)-X2(LHIGH,J))
61    X(J)=X2(N3,J)
      IN=N3
      CALL SUMR
      SR(N3)=SQRT(SEQL)
89    IF(SR(N3).LT.FDIFER) GO TO 82
      INF=N3
      CALL FEASBL
```

```
      IF(FOLD.LT.1.0E-08) GO TO 80
   82 CALL PROB3(X,R,COVAR,NX)
      F(N3)=R(K9)
      IF(F(N3).LT.F(LOW)) GO TO 84
      IF(F(N3).LT.F(LSEC)) GO TO 92
      GO TO 60
92    DO 93 J=1,NX
93    X2(LHIGH,J)=X2(N3,J)
      SR(LHIGH)=SR(N3)
      F(LHIGH)=F(N3)
      GO TO 1000
C     EXPAND VECTOR OF SEARCH ALONG DIRECTION THROUGH CENTROID AND
C     REFLECTED VECTOR
84    DO 23 J=1,NX
      X2(N4,J)=X2(N3,J)+GAMA*(X2(N3,J)-X2(N2,J))
23    X(J)=X2(N4,J)
      IN=N4
      CALL SUMR
      SR(N4)=SQRT(SEQL)
      IF(SR(N4).LT.FDIFER) GO TO 25
      INF=N4
      CALL FEASBL
      IF(FOLD.LT.1.0E-08) GO TO 80
   25 CALL PROB3(X,R,COVAR,NX)
      F(N4)=R(K9)
      IF(F(LOW).LT.F(N4)) GO TO 92
      DO 26 J=1,NX
26    X2(LHIGH,J)=X2(N4,J)
      F(LHIGH)=F(N4)
      SR(LHIGH)=SR(N4)
      GO TO 1000
60    IF(F(N3).GT.F(LHIGH)) GO TO 64
      DO 65 J=1,NX
65    X2(LHIGH,J)=X2(N3,J)
64    DO 66 J=1,NX
      X2(N4,J)=BETA*X2(LHIGH,J)+(1.-BETA)*X2(N2,J)
66    X(J)=X2(N4,J)
      IN=N4
      CALL SUMR
      SR(N4)=SQRT(SEQL)
      IF(SR(N4).LT.FDIFER) GO TO 67
      INF=N4
      CALL FEASBL
      IF(FOLD.LT.1.0E-08) GO TO 80
   67 CALL PROB3(X,R,COVAR,NX)
      F(N4)=R(K9)
      IF(F(LHIGH).GT.F(N4)) GO TO 68
      DO 69 J=1,NX
      DO 69 I=1,N1
69    X2(I,J)=0.5*(X2(I,J)+X2(LOW,J))
      DO 70 I=1,N1
      DO 71 J=1,NX
71    X(J)=X2(I,J)
      IN=I
      CALL SUMR
      SR(I)=SQRT(SEQL)
      IF(SR(I).LT.FDIFER) GO TO 72
      INF=I
      CALL FEASBL
      IF(FOLD.LT.1.0E-08) GO TO 80
```

```
 72 CALL PROB3(X,R,COVAR,NX)
70    F(I)=R(I,9)
      GO TO 1000
60    DO 75 J=1,NX
75    X2(LHIGH,J)=X2(N4,J)
      SR(LHIGH)=SR(N4)
      F(LHIGH)=F(N4)
      GO TO 1000
81    PRINT 760,ICONT,FDIFER
      CALL WRITEX
      PRINT 761
      GO TO 10
80    PRINT 760,ICONT,FDIFER
      CALL WRITEX
      PRINT 762
      GO TO 10
1     FORMAT(3I5,F10.5,E10.3)
2     FORMAT(8F10.5)
35    FORMAT(/,40X,48H * * * * * * * * * * * * * * * * * * * * * * * * )
100   FORMAT(1H1,//)
115   FORMAT(//, 41H THE STARTING VECTOR SELECTED BY USER IS )
116   FORMAT(8E16.6)
755   FORMAT(//, 35H THE COMPUTATION TIME IN SECONDS = E12.5)
756   FORMAT(//,10X,40H NUMBER OF INDEPENDENT VARIABLES          I5,/,10X
     1,40H NUMBER OF EQUALITY CONSTRAINTS          I5,/,10X,40H NUMBER O
     2F INEQUALITY CONSTRAINTS          I5,/,10X,40H SIZE OF INITIAL POLY
     3HEDRON                  E12.5,/,10X,40H THE DESIRED CONVERGENCE IS
     4          E12.5,/,10X,40H THE COMPUTATION TIME IN SECONDS
     5E12.5)
757   FORMAT(//,71H THE INITIAL X VECTOR DOES NOT SATISFY THE INITIAL TO
     1LERANCE CRITERION )
758   FORMAT( /,10X,27H STAGE CALCULATION NUMBER = I5, 20X,27H THE TOLE
     2RANCE CRITERION = E14.6)
 759 FORMAT(30H  TEST OF FLEXIBLE TOLERANCE METHOD
     1                                 )
760   FORMAT(//, 39H TOTAL NUMBER OF STAGES CALCULATIONS = I5, 10X, 25H
     1THE CONVERGENCE LIMIT = E14.6)
761   FORMAT(//,50X,25H THESE ARE FINAL ANSWERS )
762   FORMAT(//,50X,29H THESE ARE NOT FINAL ANSWERS )
763   FORMAT(//,10X,40H THE INITIAL TOLERANCE CRITERION IS      E12.5,/,
     110X,40H THE SUM OF VIOLATED CONSTRAINTS IS       E12.5)
764   FORMAT(//,70H THE VECTOR FOUND BY PROGRAM WHICH SATISFIES THE INIT
     1IAL TOLERANCE IS )
765   FORMAT(/, 31H SUM OF VIOLATED CONSTRAINTS = E17.7)
9999  STOP
      END
```

```
      SUBROUTINE FEASBL
C
C*****SUBROUTINE FEASBL MINIMIZES THE SUM OF THE SQUARE VALUES OF THE
C     VIOLATED CONSTRAINTS.  IT IS CALLED EVERY TIME THE COMBINED VALUE
C     OF THE VIOLATED CONSTRAINTS EXCEEDS THE VALUE OF THE TOLERANCE
C     CRITERION FOR THE CURRENT STAGE
C
      DIMENSION X(50),X1(50,50),X2(50,50),R(100),SUM(50),F(50),SR(50),
     1ROLD(100), R1(100), R2(100),R3(100), FLG(10), H(50)
     2,COVAR(50,50)
      COMMON/ONE/NX,NC,NIC,STEP,DUM1,DUM2,DUM3,IN,INF,FDIFER,SEQL,K1,K2
     1K3,K4,K5,K6,K7,K8,K9,X,X1,X2,R,SUM,F,SR,ROLD,SCALE,FOLD
     2,COVAR
      COMMON/TWO/LFEAS,L5,L6,L7,L8,L9,R1A,R2A,R3A
100   FORMAT(8E16.8)
C
C      ***********************************************************
C      TOL = PRESET TOLERANCE CRITERION ON CONSTRAINTS.
C
      TOL = 0.000000010000
C
C      ***********************************************************
C
      ALFA=1.
      BETA=0.5
      GAMA=2.
      XNX=NX
      ICONT=0
      LCHEK=0
      ICHEK=0
25    CALL START
      DO 3 I=1,K1
      DO 4 J=1,NX
4     X(J)=X1(I,J)
      IN=I
      CALL SUMR
3     CONTINUE
C     SELECT LARGEST VALUE OF SUM(I) IN SIMPLEX
28    SUMH=SUM(1)
      INDEX=1
      DO 7 I=2,K1
      IF(SUM(I).LE.SUMH) GO TO 7
      SUMH=SUM(I)
      INDEX=I
7     CONTINUE
C     SELECT MINIMUM VALUE OF SUM(I) IN SIMPLEX
      SUML=SUM(1)
      KOUNT=1
      DO 8 I=2,K1
      IF(SUML.LE.SUM(I)) GO TO 8
      SUML=SUM(I)
      KOUNT=I
8     CONTINUE
C     FIND CENTROID OF POINTS WITH I DIFFERENT THAN INDEX
      DO 9 J=1,NX
      SUM2=0.
      DO 10 I=1,K1
10    SUM2=SUM2+X1(I,J)
      X1(K2,J)=1./XNX*(SUM2-X1(INDEX,J))
C     FIND REFLECTION OF HIGH POINT THROUGH CENTROID
      X1(K3,J)=2.*X1(K2,J)-X1(INDEX,J)
9     X(J)=X1(K3,J)
      IN=K3
      CALL SUMR
      IF(SUM(K3).LT.SUML) GO TO 11
C     SELECT SECOND LARGEST VALUE IN SIMPLEX
```

```
      IF(INDEX.EQ.1) GO TO 38
      SUMS=SUM(1)
      GO TO 39
38    SUMS=SUM(2)
39    DO 12 I=1,K1
      IF((INDEX-I).EQ.0) GO TO 12
      IF(SUM(I).LE.SUMS) GO TO 12
      SUMS=SUM(I)
12    CONTINUE
      IF(SUM(K3).GT.SUMS) GO TO 13
      GO TO 14
C  FORM EXPANSION OF NEW MINIMUM IF REFLECTION HAS PRODUCED ONE MINIMUM
11    DO 15 J=1,NX
      X1(K4,J)=X1(K2,J)+2.*(X1(K3,J)-X1(K2,J))
15    X(J)=X1(K4,J)
      IN=K4
      CALL SUMR
      IF(SUM(K4).LT.SUML) GO TO 16
      GO TO 14
13    IF(SUM(K3).GT.SUMH) GO TO 17
      DO 18 J=1,NX
18    X1(INDEX,J)=X1(K3,J)
17    DO 19 J=1,NX
      X1(K4,J)=0.5*X1(INDEX,J)+0.5*X1(K2,J)
19    X(J)=X1(K4,J)
      IN=K4
      CALL SUMR
      IF(SUMH.GT.SUM(K4)) GO TO 6
C  REDUCE SIMPLEX BY HALF IF REFLECTION HAPPENS TO PRODUCE A LARGER VAL
C  LUE THAN THE MAXIMUM
      DO 20 J=1,NX
      DO 20 I=1,K1
20    X1(I,J)=0.5*(X1(I,J)+X1(KOUNT,J))
      DO 29 I=1,K1
      DO 30 J=1,NX
30    X(J)=X1(I,J)
      IN=I
      CALL SUMR
29    CONTINUE
5     SUML=SUM(1)
      KOUNT=1
      DO 23 I=2,K1
      IF(SUML.LT.SUM(I)) GO TO 23
      SUML=SUM(I)
      KOUNT=I
23    CONTINUE
      SR(INF)=SQRT(SUM(KOUNT))
      DO 27 J=1,NX
27    X(J)=X1(KOUNT,J)
      GO TO 26
6     DO 31 J=1,NX
31    X1(INDEX,J)=X1(K4,J)
      SUM(INDEX)=SUM(K4)
      GO TO 5
16    DO 21 J=1,NX
      X1(INDEX,J)=X1(K4,J)
21    X(J)=X1(INDEX,J)
      SUM(INDEX)=SUM(K4)
      SR(INF)=SQRT(SUM(K4))
      GO TO 26
```

```
14      DO 22 J=1,NX
        X1(INDEX,J)=X1(K3,J)
22      X(J)=X1(INDEX,J)
        SUM(INDEX)=SUM(K3)
        SR(INF)=SQRT(SUM(K3))
20      ICONT=ICONT+1
        DO 30 J=1,NX
30      X2(INF,J)=X(J)
        IF(ICONT.LT.2*K1) GO TO 50
        ICONT=0
        DO 24 J=1,NX
24      X(J)=X1(K2,J)
        IN=K2
        CALL SUMR
        DIFER=0.
        DO 57 I=1,K1
57      DIFER=DIFER+(SUM(I)-SUM(K2))**2
        DIFER=1./(K7*XNX)*SQRT(DIFER)
        IF(DIFER.GT.TOL) GO TO 50
C       IF FLEXIBLE SIMPLEX METHOD FAILED TO SATISFY THE CONSTRAINTS WITHIN
C       THE TOLERANCE CRITERION FOR THE CURRENT STAGE, THE SEARCH IS
C       PERTURBED FROM THE POSITION WHERE THE X VECTOR IS STUCK AND THEN
C       FEASBL IS REPEATED ONCE MORE FROM THE BEGINNING
51      IN=K1
        STEP=20.*FDIFER
        CALL SUMR
        SR(INF)=SQRT(SEQL)
        DO 52 J=1,NX
52      X1(K1,J)=X(J)
        DO 53 J=1,NX
        FACTOR=1.
        X(J)=X1(K1,J)+FACTOR*STEP
        X1(L9,J)=X(J)
        IN=L9
        CALL SUMR
        X(J)=X1(K1,J)-FACTOR*STEP
        X1(L5,J)=X(J)
        IN=L5
        CALL SUMR
56      IF(SUM(L9).LT.SUM(K1)) GO TO 54
        IF(SUM(L5).LT.SUM(K1)) GO TO 55
        GO TO 97
54      X1(L5,J)=X1(K1,J)
        SUM(L5)=SUM(K1)
        X1(K1,J)=X1(L9,J)
        SUM(K1)=SUM(L9)
        FACTOR=FACTOR+1.
        X(J)=X1(K1,J)+FACTOR*STEP
        IN=L9
        CALL SUMR
        GO TO 56
55      X1(L9,J)=X1(K1,J)
        SUM(L9)=SUM(K1)
        X1(K1,J)=X1(L5,J)
        SUM(K1)=SUM(L5)
        FACTOR=FACTOR+1.
        X(J)=X1(K1,J)-FACTOR*STEP
        IN=L5
        CALL SUMR
        GO TO 56
```

```
C     ONE DIMENSIONAL SEARCH BY GOLDEN SECTION ALONG EACH COORDINATE
97    H(J)=X1(L9,J)-X1(L5,J)
      X1(L6,J)=X1(L5,J)+H(J)*R1A
      X(J)=X1(L6,J)
      IN=L6
      CALL SUMR
      X1(L7,J)=X1(L5,J)+H(J)*R2A
      X(J)=X1(L7,J)
      IN=L7
      CALL SUMR
      IF(SUM(L6).GT.SUM(L7)) GO TO 68
      X1(L8,J)=X1(L5,J)+(1.-R3A)*H(J)
      X1(L5,J)=X1(L7,J)
      X(J)=X1(L8,J)
      IN=L8
      CALL SUMR
      IF(SUM(L8).GT.SUM(L6)) GO TO 76
      X1(L5,J)=X1(L6,J)
      SUM(L5)=SUM(L6)
      GO TO 75
76    X1(L9,J)=X1(L8,J)
      SUM(L9)=SUM(L8)
      GO TO 75
68    X1(L9,J)=X1(L6,J)
      X1(L8,J)=X1(L5,J)+R3A*H(J)
      X(J)=X1(L8,J)
      IN=L8
      CALL SUMR
      STEP=SIZE
      SUM(L9)=SUM(L6)
      IF(SUM(L7).GT.SUM(L8)) GO TO 71
      X1(L5,J)=X1(L8,J)
      SUM(L5)=SUM(L8)
      GO TO 75
71    X1(L9,J)=X1(L7,J)
      SUM(L9)=SUM(L7)
75    IF(ABS(X1(L9,J)-X1(L5,J)).GT.0.01*FDIFER) GO TO 97
      X1(K1,J) = X1(L7,J)
      X(J) = X1(L7,J)
      SUM(K1)=SUM(L5)
      SR(INF)=SQRT(SUM(K1))
      IF(SR(INF).LT.FDIFER) GO TO 760
53    CONTINUE
      ICHEK=ICHEK+1
      STEP=FDIFER
      IF(ICHEK.LE.2) GO TO 25
      FOLD=1.0E-09
      PRINT 853
      PRINT 850
      PRINT 851, (X(J),J=1,NX)
      PRINT 852, FDIFER, SR(INF)
      GO TO 46
760   DO 761 J=1,NX
      X2(INF,J)=X1(K1,J)
761   X(J)=X1(K1,J)
50    IF ( SR(INF) .GT. FDIFER )  GO TO 28
C     MODIFIED LAGRANGE INTERPOLATION FOR TIGHT INEQUALITIES
      IF(SR(INF).GT.0.) GO TO 35
      CALL PROB3(X,R,COVAR,NX)
      FINT=R(K9)
```

```
      DO 159 J=1,NX
159   X(J)=X2(INF,J)
      CALL PROB2(X,R,COVAR,NX)
      DO 40 J=K7,K8
40    R1(J)=R(J)
      DO 41 J=1,NX
41    X(J)=X1(KOUNT,J)
      CALL PROB2(X,R,COVAR,NX)
      DO 42 J=K7,K8
42    R3(J)=R(J)
      DO 43 J=1,NX
      H(J)=X1(KOUNT,J)-X2(INF,J)
43    X(J)=X2(INF,J)+0.5*H(J)
      CALL PROB2(X,R,COVAR,NX)
      FLG(1)=0.
      FLG(2) = 0.
      FLG(3) = 0.0
      DO 44 J=K7,K8
      IF(R3(J).GE.0.) GO TO 44
      FLG(1)=FLG(1)+R1(J)*R1(J)
      FLG(2)=FLG(2)+R(J)*R(J)
      FLG(3)=FLG(3)+R3(J)*R3(J)
44    CONTINUE
      SR(INF)=SQRT(FLG(1))
      IF(SR(INF).LT.FDIFER) GO TO 35
      ALFA1=FLG(1)-2.*FLG(2)+FLG(3)
      BETA1=3.*FLG(1)-4.*FLG(2)+FLG(3)
      RATIO=BETA1/(4.*ALFA1)
      DO 45 J=1,NX
45    X(J)=X2(INF,J)+H(J)*RATIO
       IN=INF
       CALL SUMR
       SR(INF)=SQRT(SEQL)
       IF(SR(INF).LT.FDIFER) GO TO 465
       DO 49 I=1,20
       DO 48 J=1,NX
48     X(J)=X(J)-0.05*H(J)
       CALL SUMR
       SR(INF)=SQRT(SEQL)
       IF(SR(INF).LT.FDIFER) GO TO 465
49     CONTINUE
   465 CALL PROB3(X,R,COVAR,NX)
       IF(FINT.GT.R(K9)) GO TO 46
       SR(INF)=0.
       GO TO 35
46     DO 47 J=1,NX
47     X2(INF,J)=X(J)
35     CONTINUE
       DO 335 J=1,NX
335    X(J)=X2(INF,J)
850    FORMAT(//108H IT IS NOT POSSIBLE TO SATISFY THE VIOLATED CONSTRAIN
     1T SET FROM THIS VECTOR. THE SEARCH WILL BE TERMINATED. /68H PLEASE
     2 CHOOSE A NEW STARTING VECTOR AND REPEAT SOLUTION AGAIN         )
851    FORMAT(//,63H THE VECTOR FOR WHICH THE CONSTRAINTS COULD NOT BE SA
     1TISFIED IS /,(8E16.6))
852    FORMAT(//,27H THE TOLERANCE CRITERION = E14.6,20X, 49H THE SQUARE
     1ROOT OF THE CONSTRAINTS SQUARED IS = E16.6)
853    FORMAT(//,81H * * * * * * SUBROUTINE  FEASBL  FAILS TO FIND A FEAS
     1IBLE POINT * * * * * * * *   )
       RETURN
       END
```

```
      SUBROUTINE START
      DIMENSION A(50,50)
      DIMENSION X(50),X1(50,50),X2(50,50),R(100),SUM(50),F(50),SR(50),
     1ROLD(100),COVAR(50,50)
      COMMON/ONE/NX,NC,NIC,STEP,ALFA,BETA,GAMA,IN,INF,FDIFER,SEQL,K1,K2,
     1K3,K4,K5,K6,K7,K8,K9,X,X1,X2,R,SUM,F,SR,ROLD,SCALE,FOLD
     2,COVAR
      COMMON/TWO/LFEAS,L5,L6,L7,L8,L9,R1A,R2A,R3A
      VN=NX
      STEP1=STEP/(VN*SQRT(2.))*(SQRT(VN+1.)+VN-1.)
      STEP2 = STEP/(VN*SQRT(2.))*(SQRT(VN+1.)-1.)
      DO 1 J=1,NX
 1    A(1,J)=0.
      DO 2 I=2,K1
      DO 4 J=1,NX
 4    A(I,J)=STEP2
      L=I-1
      A(I,L)=STEP1
 2    CONTINUE
      DO 3 I=1,K1
      DO 3 J=1,NX
 3    X1(I,J)=X(J)+A(I,J)
      RETURN
      END


      SUBROUTINE SUMR
C*****THIS SUBROUTINE COMPUTES THE SUM OF THE SQUARE VALUES OF THE
C     VIOLATED CONSTRAINTS IN ORDER TO BE COMPARED WITH THE TOLERANCE
C     CRITERION
C
      DIMENSION X(50),X1(50,50),X2(50,50),R(100),SUM(50),F(50),SR(50),
     1ROLD(100),COVAR(50,50)
      COMMON/ONE/NX,NC,NIC,STEP,ALFA,BETA,GAMA,IN,INF,FDIFER,SEQL,K1,K2,
     1K3,K4,K5,K5,K7,K8,K9,X,X1,X2,R,SUM,F,SR,ROLD,SCALE,FOLD
     2,COVAR
      COMMON/TWO/LFEAS,L5,L6,L7,L8,L9,R1A,R2A,R3A
      SUM(IN)=0.
      CALL PROB2(X,R,COVAR,NX)
      SEQL=0.
      IF(NIC.EQ.0) GO TO 4
      DO 1 J=K7,K8
      IF(R(J).GE.0.) GO TO 1
      SEQL=SEQL+R(J)*R(J)
 1    CONTINUE
 4    IF(NC.EQ.0) GO TO 3
      CALL PROB1(X,R,COVAR,NX)
      DO 2 J=1,NC
 2    SEQL=SEQL+R(J)*R(J)
 3    SUM(IN)=SEQL
 5    RETURN
      END
```

```
      SUBROUTINE WRITEX
      DIMENSION X(50),X1(50,50),X2(50,50),R(100),SUM(50),F(50),SR(50),
     1ROLD(100),COVAR(50,50)
      COMMON/ONE/NX,NC,NIC,STEP,ALFA,BETA,GAMA,IN,INF,FDIFER,SEQL,K1,K2,
     1K3,K4,K5,K6,K7,K8,K9,X,X1,X2,R,SUM,F,SR,ROLD,SCALE,FOLD
     2,COVAR
      COMMON/TWO/LFEAS,L5,L6,L7,L8,L9,R1A,R2A,R3A
      CALL PROB3(X,R,COVAR,NX)
      PRINT 1,R(K9)
1     FORMAT(/, 28H OBJECTIVE FUNCTION VALUE = E17.7)
      PRINT 2,(X(J),J=1,NX)
2     FORMAT(/, 29H THE INDEPENDENT VECTORS ARE /(6E17.7) )
      IF(NC.EQ.0) GO TO 6
      CALL PROB1(X,R,COVAR,NX)
      PRINT 3, (R(J),J=1,NC)
    3 FORMAT(/,'THE EQUALITY CONSTAINTS ARE',/(6E16.6))
6     IF(NIC.EQ.0) GO TO 5
      CALL PROB2(X,R,COVAR,NX)
      PRINT 4,(R(J),J=K7,K6)
    4 FORMAT(/,'THE INEQUALITY CONSTRAINT VALUES ARE',/(6E16.6))
5     RETURN
      END
```

APPENDIX E

THE SECURITIES USED IN THE EXAMPLES PRESENTED IN CHAPTER FOUR

E.1 Coal Sector

Apex	-	Apex Mines Ltd
Clydesdale	-	The Clydesdale (Transvaal) Collieries Ltd.
Tavistock	-	Tavistock and South Witbank Collieries Ltd.
Trans Natal	-	Trans-Natal Coal Corporation Ltd.
Welgedacht	-	Welgedacht Exploration Co. Ltd.

E.2 Gold - Witwatersrand and Others Sector

Durban Deep	-	Durban Roodepoort Deep Ltd.
ERPM	-	East Rand Proprietary Mines Ltd.
Grootvlei	-	The Grootvlei Proprietary Mines Ltd.
Marievale	-	Marievale Consolidated Mines Ltd.
S.A. Lands	-	The South African Land and Exploration Co. Ltd.

E.3 Banks Sector

1. Bank Holdings Corporation of S.A. Ltd.
2. Nedbank and Syfrets UAL Holdings Ltd.
3. Santam Bank Ltd.
4. Standard Bank Investment Corporation Ltd.
5. The Trust Bank of Africa Ltd.
6. Volkskas Bpk.

E.4 Chemical Sector

1. A.E.C.I. Ltd.
2. Coal By-Products and Investments Ltd.
3. Federale Kunsmis Beperk
4. Sentrachem Ltd.

APPENDIX F

NO	SHARE	NO OF WEEKS
1	Vaal Reefs Exploration & Mining Co. Ltd	414
2	Zandpan Gold Mining Co. Ltd.	414
3	Harmony Gold Mining Co. Ltd.	414
4	President Brand Gold Mining Co. Ltd.	414
5	Kloof Gold Mining Co. Ltd.	414
6	West Driefontein Gold Mining Co. Ltd.	414
7	The Messina (Transvaal) Development Co. Ltd.	414
8	Union Tin Mines Ltd.	414
9	Consolidated Murchison Co. Ltd.	414
10	Anglo Transvaal Consolidated Investment Co. Ltd.	414
11	Charter Consolidated Ltd.	414
12	Johannesburg Consolidated Investment Co. Ltd.	414
13	Rand Selection Corporation Ltd.	414
14	U.C. Investments Ltd.	414
15	Anglo American Industrial Corporation Ltd.	414
16	Barlow Rand Ltd.	414
17	Lonrho Ltd.	352
18	First Union General Investment Trust Ltd.	409
19	Guardian Assurance Holdings (S.A.) Ltd.	375
20	Glen Anil Development Corporation Ltd.	378
21	Rand Mines Properties Ltd.	405
22	The Oudemeester Group Ltd.	414
23	Anglo-Alpha Cement Ltd.	414

NO	SHARE	NO OF WEEKS
24	Blue Circle Cement Ltd.	414
25	AECI Ltd.	414
26	Kaap Kunene Beleggings Bpk.	414
27	Tiger Oats & National Milling Co. Ltd.	414
28	Searles Holdings Ltd.	411
29	Associated Furniture Companies Ltd.	409
30	Russell Holdings Ltd.	409
31	Hubert Davies & Co. Ltd.	414
32	Stewarts & Lloyds of South Africa Ltd.	414
33	Toyota (South Africa) Ltd.	414
34	Sappi Ltd.	414
35	Skye Ltd.	364
36	Twins Pharmaceutical Holdings Ltd.	372
37	South African Associated Newspapers Ltd.	414
38	Vaderland Beleggings Bpk.	414
39	Bonmore Investments Ltd.	411
40	Pick 'n Pay Stores Ltd.	389
41	Tongaat Group Ltd.	414
42	Romatex Ltd.	414
43	Utico Holdings Ltd.	414
44	Nedbank and Syfrets UAL Holdings Ltd.	314
45	The Trust Bank of Africa Ltd.	314
46	Volkskas Bpk.	314

APPENDIX G

THE EFFECT OF HETEROSKEDASTICITY ON THE MULTIPLE CORRELATION COEFFICIENT

In the simple regression model

$$y = \alpha + \beta x + e$$

the square of the multiple correlation coefficient can be computed as follows:

$$R^2 = 1 - \frac{S^2_{y.x}}{S^2_y} .$$

where $S^2_{y.x}$ is an estimate of the variance due to the regression, and S^2_y is the variance of the dependent variable and is independent of the form of the line or the method of estimation of the parameters.

Goldberger (1964) discusses the case of pure heteroskedasticity with a single regressor and T observations on X. That is

$$E(e_t^2) = \sigma^2 k_t$$

and hence Ω has the form

$$\Omega = \begin{bmatrix} k_1 & & & & \\ & k_2 & & 0 & \\ & & \cdot & & \\ & & & \cdot & \\ & 0 & & & k_T \end{bmatrix}$$

where $\Sigma k_t = T$. Goldberger also assumes without loss of generality that $\Sigma_t x_t = 0$. Now, it can be shown (Goldberger) that if ordinary least squares is used to estimate α and β instead of generalized least squares then,

$$E(S^2_{y.x}) = \sum_{t=1}^{T} \sigma^2 k_t \left(\frac{1 - x_t^2/\Sigma_i x_i^2 - 1/T}{T-2} \right)$$

$$= \sigma^2 \sum_{t=1}^{T} k_t \left(\frac{1 - x_t^2/\Sigma_i x_i^2 - 1/T}{T-2} \right) \qquad (G.1)$$

The question as to how R^2 will be biased depends on whether the summation in (G.1) above is greater than or less than one. Now, the summation term can be simplified as follows:

$$\sum_{t=1}^{T} k_t \left(\frac{1 - x_t^2/\Sigma_i x_i^2 - 1/T}{T-2} \right) = \frac{\Sigma k_t}{T-2} - \frac{\Sigma k_t x_t^2}{(T-2)\Sigma_i x_i^2} - \frac{\Sigma k_t}{T(T-2)}$$

$$= \frac{T}{T-2} - \frac{\Sigma k_t x_t^2}{(T-2)\Sigma_i x_i^2} - \frac{T}{T(T-2)}, \text{ since } \Sigma k_t = T$$

$$= \frac{T-1}{T-2} - \frac{\Sigma k_t x_t^2}{(T-2)\Sigma_i x_i^2}$$

which will be less than one if

$$\frac{\Sigma_t k_t x_t^2}{\Sigma_i x_i^2} > 1$$

Now, it is obvious that when using a weighted average, if the larger observations are given more weight and the smaller observations are given less weight, then the weighted average will be greater than the arithmetic average. Since the larger x's have the larger k_t's, and since $\Sigma_t k_t x_t^2$ is a weighted sum of the x_t^2 (by k_t) it follows that

$$\frac{\Sigma\, k_t x_t^2}{\Sigma\, x_i^2} \geq \frac{\Sigma\, 1.x_t^2}{\Sigma\, x_i^2} = 1$$

Thus, the summation term will be less than one, that is

$$E(S_{y.x}^2) \leq \sigma^2$$

i.e. $$R^2 = 1 - \frac{S_{y.x}^2}{S_y^2} > 1 - \frac{\sigma^2}{\sigma_y^2} = P^2$$

Thus, the multiple correlation coefficient is overestimated, which means that it appears as if the regression equation gives a better fit than it actually does if heteroskedasticity (of the form which results in an increase in the variance as x increases) is present.

A P P E N D I X H

THE UNIVERSE OF SECURITIES USED IN CHAPTER SEVEN

1. Durban Roodepoort Deep Ltd
2. East Daggafontein Mines Ltd.
3. East Rand Proprietary Mines Ltd.
4. The Grootvlei Proprietary Mines Ltd.
5. Marievale Consolidated Mines Ltd.
6. The Randfontein Estates Gold Mining Co. Witwatersrand Ltd.
7. The South African Land and Exploration Co. Ltd.
8. South Roodepoort Main Reef Areas Ltd.
9. Vlakfontein Gold Mining Co. Ltd.
10. West Rand Consolidated Mines Ltd.
11. Bracken Mines Ltd.
12. Leslie Gold Mines Ltd.
13. Winkelhaak Mines Ltd.
14. The Afrikander Lease Ltd.
15. Buffelsfontein Gold Mining Co. Ltd.
16. Hartebeesfontein Gold Mining Co. Ltd.
17. Stilfontein Gold Mining Co. Ltd.
18. Vaal Reefs Exploration and Mining Co. Ltd.
19. Zandpan Gold Mining Co. Ltd.
20. Free State Geduld Mines Ltd.
21. Harmony Gold Mining Co. Ltd.
22. Loraine Gold Mines Ltd.
23. President Brand Gold Mining Co. Ltd.
24. President Steyn Gold Mining Co. Ltd.
25. St. Helena Gold Mines Ltd.
26. Welkom Gold Mining Co. Ltd.
27. Western Holdings Ltd.
28. Blyvooruitzicht Gold Mining Co. Ltd.
29. Doornfontein Gold Mining Co. Ltd.
30. Libanon Gold Mining Co. Ltd.
31. Venterspost Gold Mining Co. Ltd.
32. West Driefontein Gold Mining Co. Ltd.
33. Western Areas Gold Mining Co. Ltd.
34. Western Deep Levels Ltd.
35. Amalgamated Collieries of S.A. Ltd. (Taken over by Anglo American Coal Corporation Ltd. on 9/1/76).
36. Anglo-Transvaal Collieries Ltd.
37. Apex Mines Ltd.
38. The Clydesdale (Transvaal) Collieries Ltd.
39. Natal Ammonium Collieries (1946) Ltd.
40. Natal Anthracite Colliery Ltd.
41. Tavistock and South Witbank Collieries Ltd.
42. Vierfontein Colliery Ltd.

43. Vryheid Coronation Ltd.
44. Wankie Colliery Co. Ltd.
45. Welgedacht Exploration Co. Ltd.
46. Witbank Colliery Ltd.
47. De Beers Consolidated Mines Ltd.
48. The Messina (Transvaal) Development Co. Ltd.
49. M.T.D. (Mangula) Ltd.
50. Associated Manganese Mines of South Africa Ltd.
51. South African Manganese Amcor Ltd.
52. Lydenburg Platinum Ltd.
53. Potgietersrust Platinums Ltd.
54. The Rooiberg Minerals Development Co. Ltd.
55. Zaaiplaats Tin Mining Co. Ltd.
56. Consolidated Murchison Co. Ltd.
57. Msauli Asbes Bpk.
58. Anglo American Corporation of South Africa Ltd.
59. Anglo Transvaal Consolidated Investment Co. Ltd.
60. Consolidated Gold Fields Ltd.
61. Federale Mynbou Bpk.
62. Free State Development and Investment Corporation Ltd.
63. General Mining and Finance Corporation Ltd.
64. Johannesburg Consolidated Investment Co. Ltd.
65. Middle Witwatersrand (Western Areas) Ltd.
66. New Central Witwatersrand Areas Ltd.
67. New Witwatersrand Gold Exploration Co. Ltd.
68. Rand Selection Corporation Ltd.
69. Transvaal Consolidated Land and Exploration Co. Ltd.
70. Tweefontein United Collieries Ltd.
71. Union Corporation Ltd.
72. Anglo-Transvaal Industries Ltd.
73. Barlow Rand Ltd.
74. Bonuskor Bpk.
75. De Beers Industrial Corporation Ltd.
76. Hippo Holdings Co. Ltd.
77. South Atlantic Corporation Ltd.
78. Trade and Industry Acceptance Corporation Ltd.
79. W & A Investment Corporation Ltd.
80. The Common Fund Investment Society Ltd.
81. K and L Timbers Ltd.
82. Standard Bank Investment Corporation Ltd.
83. Volkskas Bpk.
84. The South African Breweries Ltd.
85. Anglo-Alpha Cement Ltd.
86. Blue Circle Cement Ltd.
87. Murray and Roberts Holdings Ltd.
88. Plate Glass and Shatterprufe Industries Ltd.
89. Pretoria Portland Cement Co. Ltd.
90. The African Clothing Factory (Ensign) Ltd.
91. Berkshire International (S.A.) Ltd.
92. Delswa Ltd.
93. I.L. Back and Co. Ltd.
94. Rex Trueform Clothing Co. Ltd.
95. The Natal Chemical Syndicate Ltd.
96. Rand Carbide Ltd.
97. Lamberts Bay Canning Co. Ltd.

98. Marine Products Ltd.
99. Sea Products (S.W.A.) Ltd.
100. South West Africa Fishing Industries Ltd.
101. African Products Manufacturing Co. Ltd.
102. Bakers South Africa Ltd.
103. T.W. Beckett and Company Ltd.
104. Imperial Cold Storage and Supply Company Ltd.
105. The Premier Milling Co. Ltd.
106. Stein Brothers (Holdings) Ltd.
107. Tiger Oats and National Milling Co. Ltd.
108. Edworks (1936) Ltd.
109. Bradlow's Stores Ltd.
110. Phil Morkel Ltd. (Assumed investment in Bothners Holdings Pty. Ltd. prior to 1965.)
111. Steel and Barnett Ltd.
112. African Cables Ltd.
113. Cullinan Holdings Ltd.
114. Dorman Long Vanderbijl Corporation Ltd.
115. Edward L. Bateman Ltd.
116. Eveready South Africa Ltd.
117. C.J. Fuchs Ltd.
118. Hubert Davies and Co. Ltd.
119. International Combustion (Africa) Ltd.
120. Massey-Ferguson (South Africa) Ltd.
121. National Bolts Ltd.
122. Reunert and Lenz Ltd.
123. Stewarts and Lloyds of South Africa Ltd.
124. The Union Steel Corporation of South Africa Ltd.
125. Wispeco Holdings Ltd.
126. Bus Bodies (S.A.) Ltd.
127. Eriksen Consolidated Holdings Ltd.
128. The General Tire and Rubber Company (South Africa) Ltd.
129. McCarthy Rodway Ltd.
130. Robbs Holdings Ltd.
131. Williams Hunt South Africa Ltd.
132. Consolidated Glass Works Ltd.
133. The Metal Box Company of South Africa Ltd.
134. Premier Paper Mills Ltd.
135. Sappi Ltd.
136. Argus Printing and Publishing Co. Ltd.
137. C.N.A. Investments Ltd.
138. Hortors Ltd.
139. Edgars Stores Ltd.
140. Greatermans Stores Ltd.
141. John Orr Holdings Ltd.
142. O.K. Bazaars (1929) Ltd.
143. Stuttaford and Co. Ltd.
144. Truworths Ltd.
145. Woolworths Holdings Ltd.
146. Hulett's Corporation Ltd.
147. Illovo Sugar Estates Ltd.
148. Tongaat Group Ltd.
149. Consolidated Textile Mills Investment Corporation Ltd.

150. Natal Canvas Rubber Manufacturers Ltd.
151. Romatex Ltd.
152. The Lion Match Co. Ltd.
153. Rembrandt Beherende Beleggings Bpk.
154. Rembrandt Group Ltd.
155. Utico Holdings Ltd.
156. Gallo (Africa) Ltd.
157. South African Marine Corporation Ltd.

APPENDIX I

THE FIVE SAMPLES USED IN CHAPTER SEVEN

TABLE I.1 : SAMPLE 1

No	Share	No. in Table H
1.	Grootvlei	4
2.	Marievale	5
3.	S.A. Land	7
4.	Vlakfontein	9
5.	Winkelhaak	13
6.	President Steyn	24
7.	St. Helena	25
8.	Welkom	26
9.	Vryheid	43
10.	Wankie	44
11.	De Beers	47
12.	M.T.D. (Mangula)	49
13.	Associated Manganese	50
14.	S.A. Manganese	51
15.	Lydenburg Platinum	52
16.	Rooiberg	54
17.	Zaaiplaats	55
18.	Federale Mynbou	61
19.	Free State Development	62
20.	New Central Wits	66
21.	New Witwatersrand	67
22.	Tweefontein	70
23.	Union Corporation	71
24.	Anglo-Transvaal Ind.	72
25.	De Beers Ind.	75
26.	Common Fund	80
27.	Standard Bank	82
28.	Volkskas	83
29.	Blue Circle	86
30.	Plate Glass	88
31.	African Clothing	90
32.	Delswa	92
33.	Rand Carbide	96
34.	S.W.A. Fishing	100
35.	T.W. Beckett	103
36.	Premier Milling	105
37.	Bradlow's	109
38.	Ed. L. Bateman	115
39.	Hubert Davies	118
40.	Reunert and Lenz	122
41.	Stewarts and Lloyds	123
42.	Wispeco	125

TABLE I.2 : SAMPLE 2

No.	Share	No. in Appendix H
1.	E.R.P.M.	3
2.	Grootvlei	4
3.	South Roodepoort	8
4.	West Rand Cons.	10
5.	Winkelhaak	13
6.	Stilfontein	17
7.	Harmony	21
8.	Welkom	26
9.	Blyvooruitzicht	28
10.	Doornfontein	29
11.	Libanon	30
12.	Amalgamated Collieries	35
13.	Clydesdale	38
14.	Natal Anthracite	40
15.	Welgedacht	45
16.	Messina	48
17.	Rooiberg	54
18.	Cons. Murchison	56
19.	Msauli	57
20.	Anglo American Corp.	58
21.	Anglovaal	59
22.	Johannesburg Consolidated Inv.	64
23.	T.C. Land	69
24.	Barlow Rand	73
25.	De Beers Ind.	75
26.	Trade and Industry	78
27.	Standard Bank	82
28.	Anglo-Alpha	85
29.	Murray and Roberts	87
30.	African Clothing	90
31.	Rex Trueform	94
32.	Bakers	102
33.	Premier Milling	105
34.	African Cable	112
35.	Fuchs	117
36.	National Bolts	121
37.	Reunart and Lenz	122
38.	Union Steel	124
39.	Wispeco	125
40.	Eriksen	127
41.	General Tire	128
42.	McCarthy Rodway	129
43.	Robbs	130
44.	Premier Paper	134
45.	Sappi	135
46.	Hortors	138
47.	Woolworths	145
48.	Consolidated Textile	149
49.	Utico Holdings	155
50.	Gallo	156

TABLE I.3 : SAMPLE 3

No.	Share	No. in Appendix H
1.	Durban Roodepoort Deep	1
2.	East Daggafontein	2
3.	E.R.P.M.	3
4.	Marievale	5
5.	South Roodepoort	8
6.	Vlakfontein	9
7.	West Rand Cons.	10
8.	Leslie	12
9.	Winkelhaak	13
10.	Hartebeesfontein	16
11.	Stilfontein	17
12.	President Brand	23
13.	President Steyn	24
14.	Welkom	26
15.	Blyvooruitzicht	28
16.	Venterspost	31
17.	Western Areas	33
18.	Clydesdale	38
19.	Tavistock	41
20.	General Mining	63
21.	Middle Witwatersrand	65
22.	T.C. Land	69
23.	Tweefontein	70
24.	Anglo-Transvaal Ind.	72
25.	Bonuskor	74
26.	Trade and Industry	78
27.	Common Fund	80
28.	K. & L. Timbers	81
29.	Volkskas	83
30.	Delswa	92
31.	Rex Trueform	94
32.	Marine Products	98
33.	African Products	101
34.	Bakers	102
35.	Imperial Cold Storage	104
36.	Stein Brothers	106
37.	Tiger Oats	107
38.	Bradlow's	109
39.	Eveready	116
40.	Massey-Ferguson	120
41.	General Tire	128
42.	Robbs	130
43.	Metal Box	133
44.	Premier Paper	134
45.	Greatermans	140
46.	Hulett's	146
47.	Illovo	147
48.	Tongaat	148
49.	Romatex	151
50.	Rembrandt Group	154

TABLE I.4 : SAMPLE 4

No.	Share	No. in Appnedix H
1.	Randfontein	6
2.	Bracken	11
3.	Afrikander Lease	14
4.	Buffelsfontein	15
5.	Vaal Reefs	18
6.	Zandpan	19
7.	Free State Geduld	20
8.	Loraine	22
9.	Western Holdings	27
10.	West Driefontein	32
11.	Western Deep Levels	34
12.	Anglo-Transvaal Collieries	36
13.	Apex Mines	37
14.	Natal Ammonium	39
15.	Vierfontein	42
16.	Witbank	46
17.	Potgietersrust Platinums	53
18.	Consolidated Gold Fields	60
19.	Johannesburg Consolidated Investment	64
20.	Rand Selection	68
21.	Anglo-Transvaal Ind.	72
22.	Hippo Holdings	76
23.	South Atlantic	77
24.	W. & A. Investment Corp.	79
25.	S.A. Breweries	84
26.	Pretoria Portland Cement	89
27.	African Clothing	90
28.	Berkshire	91
29.	I.L. Back	93
30.	Natal Chemical	95
31.	Lamberts Bay	97
32.	Sea Products (SWA)	99
33.	Edworks	108
34.	Bradlow's	109
35.	Phil Morkel	110
36.	Steel and Barnett	111
37.	Cullinan	113
38.	Dorman Long	114
39.	International Combustion	119
40.	Union Steel	124
41.	Consolidated Glass	132
42.	Argus	136
43.	C.N.A.	137
44.	Edgars	139
45.	John Orr	141
46.	O.K. Bazaars	142
47.	Stuttaford	143
48.	Natal Canvas	150
49.	Lion Match	152
50.	Rembrandt Beherende	153

TABLE I.5 : SAMPLE 5

No.	Share	No. in Appendix H
1.	Durban Roodepoort Deep	1
2.	Marievale	5
3.	Randfontein	6
4.	Winkelhaak	13
5.	Stilfontein	17
6.	Vaal Reefs	18
7.	Harmony	21
8.	President Brand	23
9.	Western Holdings	27
10.	Blyvooruitzicht	28
11.	Western Deep Levels	34
12.	Amalgamated Collieries	35
13.	Tavistock	41
14.	De Beers	47
15.	Messina	48
16.	S.A. Manganese	51
17.	Lydenburg Platinum	52
18.	Rooiberg	54
19.	Cons. Murchison	56
20.	Federale Mynbou	61
21.	General Mining	63
22.	T.C. Land	69
23.	Barlow Rand	73
24.	Trade and Industry	78
25.	Common Fund	80
26.	K. & L. Timbers	81
27.	Volkskas	83
28.	S.A. Breweries	84
29.	Murray and Roberts	87
30.	African Clothing	90
31.	Natal Chemical	95
32.	Marine Products	98
33.	Imperial Cold Storage	104
34.	Tiger Oats	107
35.	Edworks	108
36.	Phil Morkel	110
37.	Dorman Long	114
38.	Stewarts and Lloyds	123
39.	Wispeco	125
40.	General Tire	128
41.	McCarthy Rodway	129
42.	Premier Paper	134
43.	Hortors	138
44.	Stuttaford	143
45.	Woolworths	145
46.	Hulett's	146
47.	Romatex	151
48.	Rembrandt Group	154
49.	Gallo	156
50.	Safmarine	157

APPENDIX J

THE COMPUTER PROGRAMS FOR THE MARKOWITZ AND SHARPE PORTFOLIO SELECTION MODELS

This Appendix contains a brief description of the computer programs written to solve the original Markowitz (1959) model, and the index models of Sharpe ((1963) and (1970)). Included is a description of the input required for each program as well as a listing of each program.

The algorithms used were those proposed by Sharpe (1970), which basically construct and solve a set of simultaneous equations for each corner portfolio. The actual workings of the algorithms are rather tedious and lengthy, and hence are not presented here. For any information in this regard, the reader is referred to Sharpe (1970).

It should be noted that the algorithms were coded in FORTRAN V and all computations were performed on the University of Cape Town's Univac 1106 computer. The programs use two of the system supplied subroutines; the first being GJR, which provides a solution to a set of simultaneous equations; and the second being SIMPLX, which solves a linear programming problem and is used to obtain the first corner portfolio, that with maximum expected return.

C.1 COMPUTER PROGRAM FOR THE MARKOWITZ MODEL

A listing of the program is given on the following pages, but prior to that a brief description of the input required (card by card) is given.

CARD 1: N, M

N = The number of shares considered.

M = The number of linear constraints (not bounds) imposed.

FORMAT: (2I5)

CARD 2: IFLAG, DLAM

IFLAG = 0 indicates that all corner portfolios are to be printed;

= 1 indicates that only the portfolio corresponding to the λ specified by the particular investor is to be printed.

DLAM = the value of λ chosen by the investor (only needs to be specified if IFLAG = 1).

FORMAT: (I1,F9.6)

CARD 3: E(I), I = 1,2,...,N.

E(I) = The expected return on the I^{th} security.

FORMAT: (F12.9)

<u>CARD 4</u>: COV(I,J); J = 1,2,...,N; I = 1,2,...,N.

COV(I,J) = The covariance (estimated) between the I^{th} and J^{th} securities.

FORMAT: (F12.9)

<u>CARD 5</u>: MASK1, MASK2

MASK1 = 0 implies that no linear constraints are imposed.

$\neq$ 0 implies linear constraints are imposed.

MASK2 = 0 implies that all securities have the same upper and lower bounds.

= 10 implies the upper and lower bounds are 1,0 and 0,0 respectively.

$\neq$ 0 or 10 implies different upper and lower bounds for each security.

FORMAT: (2I5)

<u>CARD 6</u>: (Present only if MASK1 $\neq$ 0)

D(I,J); J = 1,2,...,N+1; I = 1,2,...,M

D(I,J) = the coefficient of the J^{th} variable in the I^{th} linear constraint. It should be noted that the last column (i.e. J = N+1) of this array contains the constant terms.

FORMAT: (F12.9)

<u>CARD 7</u>: 1) If MASK2 $\neq$ 0 or 10:

SMAL(I), UP(I); I = 1,2,...,N.

SMAL(I) = lower bound for the I^{th} security.

UP(I) = upper bound for the I^{th} security.

2) If MASK2 = 0 or 10:

SSS1, SSS2

SSS1 = lower bound imposed on all securities.

SSS2 = upper bound imposed on all securities.

FORMAT: (2F10.5)

<u>CARD 8</u>: IFIX(I); I = 1,2,...,20

Required for SIMPLX subroutine - the reader is referred to the manual "SIMPDX/SIMPLX Linear Programming Subroutines, reference manual for the 1108, University of Wisconsin Computing Centre, May 1970" for any further information.

FORMAT: (10I5)

```
      DIMENSION E( 50),COV( 50, 50),TAB( 70, 70),TABI( 70, 70),WORKA( 70
     1, 70),D( 20, 50),SMAL( 50),UP( 50),IDUM( 50),A(121, 50),RHS(121),T
     2(121),COST( 50),XX(123),JX(121),PI(121),EE(123,123),ERR(4),TOL(4),
     3IFIX(20),IOUT(4),TT(3),Y(123),S( 50),V(2),X( 50),P( 50),SMP( 50),J
     4C( 70)
      DATA (TT(J),J=1,3)/1H ,1H+,1H-/(IDUM(I),I=1, 50)/ 50*1/
      NCNT1 = 70
      NCNT2 = 70
      IIFLG = 0
C
C
C
C
C          DATA INPUT AND INITIALIZATION SECTION  .
C
C
      READ(8,100) N,M
100   FORMAT(2I5)
      READ(8,101) IFLAG,DLAM
101   FORMAT(I1,F9.6)
      IN = N + 1
      K1 = N+M+2
      K2 = N+M+3
      LL = N+M+1
      NM = N+M
      READ(20,110) (E(I),I=1,N)
110   FORMAT(F12.9)
      DO 120 I=1,N
      READ(20,110) (COV(I,J),J=1,N)
120   CONTINUE
      READ(8,100) MASK1,MASK2
      IF (MASK1.EQ.0)  GO TO 140
      DO 130 I=1,M
      READ(8,110) (D(I,J),J=1,IN)
130   CONTINUE
140   IF (MASK2.EQ.0)  GO TO 150
      IF (MASK2.EQ.10)  GO TO 150
      READ(8,145) (SMAL(I),UP(I),I=1,N)
145   FORMAT(2F10.5)
      GO TO 200
150   READ(8,145) SSS1,SSS2
      DO 160 I=1,N
      SMAL(I) = SSS1
      UP(I) = SSS2
160   CONTINUE
C
C
C
C
C          CONSTRUCTION OF TAB .
C
C
200   DO 230 I=1,N
```

```
      DO 210 J=1,N
      TAB(I,J) = 2.0*COV(I,J)
210   CONTINUE
      IF (MASK1.EQ.0)  GO TO 225
      DO 220 J=IN,NM
      K = J-N
      TAB(I,J) = -D(K,I)
220   CONTINUE
225   TAB(I,LL) = -1.0
      TAB(I,K1) =  0.0
      TAB(I,K2) = E(I)
230   CONTINUE
C
C
      IF (MASK1.EQ.0)  GO TO 265
C
      DO 260 I=IN,NM
      K = I-N
      DO 240 J=1,N
      TAB(I,J) = D(K,J)
240   CONTINUE
      DO 250 J=IN,LL
      TAB(I,J) = 0.0
250   CONTINUE
      TAB(I,K1) = D(K,IN)
      TAB(I,K2) =0.0
260   CONTINUE
C
C
265   DO 270 J=1,N
      TAB(LL,J) = 1.0
270   CONTINUE
      DO 280 J=IN,LL
      TAB(LL,J) = 0.0
280   CONTINUE
      TAB(LL,K1) = 1.0
      TAB(LL,K2) = 0.0
      WRITE(5,285)
285   FORMAT(1H1)
      DO 295 I=1,LL
      WRITE(5,289)
289   FORMAT(1H0)
      WRITE(5,290) (TAB(I,J),J=1,K2)
290   FORMAT(10F10.4)
295   CONTINUE
C
C
C
C
C          FINDING THE MAXIMUM E(P) PORTFOLIO .
C
C
      DO 310 J=1,N
      A(1,J) = 1.0
310   CONTINUE
      RHS(1) = 1.0
      T(1) = IT(1)
C
      IF (MASK1.EQ.0)  GO TO 340
C
```

```
      DO 330 I=1,M
      KKK = I+1
      DO 320 J=1,N
      A(KKK,J) = D(I,J)
320   CONTINUE
      RHS(KKK) = D(I,IN)
      T(KKK) = TT(1)
330   CONTINUE
C
340   DO 360 I=1,N
      KKK = I+1+M
      DO 350 J=1,N
      A(KKK,J) = 0.0
350   CONTINUE
      A(KKK,I) = 1.0
      RHS(KKK) = SMAL(I)
      T(KKK) = TT(3)
360   CONTINUE
C
      DO 380 I=1,N
      KKK = M+N+I+1
      DO 370 J=1,N
      A(KKK,J) = 0.0
370   CONTINUE
      A(KKK,I) = 1.0
      RHS(KKK) = UP(I)
      T(KKK) = TT(2)
      COST(I) = -E(I)
380   CONTINUE
C
      READ(8,390) (IFIX(I),I=1,20)
390   FORMAT(10I5)
      IFIX(3) = N+N+M+1
      IFIX(4) = N
      DO 400 I=1,4
      TOL(I) = 0.0
400   CONTINUE
C
      CALL SIMPLX (A,T,RHS,COST,IFIX,TOL,OBJ,XX,JX,PI,EE,ERR,IOUT,Y,S)
C
      DO 410 I=1,N
      X(I) = 0.0
410   CONTINUE
      IFIZ = IFIX(3)
      DO 420 I=1,IFIZ
      II = JX(I)
      IF (II.GT.N)  GO TO 420
      X(II) = XX(I)
420   CONTINUE
C
      WRITE(5,430)
430   FORMAT(1H1,30H THE MAXIMUM E(P) PORTFOLIO IS  )
      WRITE(5,440) (I,X(I),I=1,N)
440   FORMAT(3H X(,I3,3H) =,F6.4)
      EP = 0.0
      VP = 0.0
      SDP = 0.0
      DO 460 I=1,N
      EP = EP + X(I)*E(I)
      DO 450 J=1,N
```

```
      VP = VP + X(I)*X(J)*COV(I,J)
450   CONTINUE
460   CONTINUE
      SDP = SQRT (VP)
      WRITE(5,470) EP,VP,SDP
470   FORMAT(///,7H E(P) =,F8.5,/,7H V(P) =,F8.5,/,7H SD(P)=,F8.5)
      OLDLAM = 1000.00
C
C
C
C
C          ASSIGN STATUS TO EACH VARIABLE AND MODIFY THE SET OF EQUATIONS
C
C
      DO 500 I=1,N
      IF (X(I).LE.SMAL(I))  IDUM(I) = 0
      IF (X(I).GE.UP(I))    IDUM(I) = 2
500   CONTINUE
      EMAX = 1000.0
C
      DO 560 I=1,N
      IF (IDUM(I))   ,,520
      DO 510 J=1,LL
      TABI(I,J) = 0.0
510   CONTINUE
      TABI(I,I) = 1.0
      TABI(I,K1) = SMAL(I)
      TABI(I,K2) = 0.0
      GO TO 560
C
520   IF (IDUM(I)-2) 540,,
      DO 530 J=1,LL
      TABI(I,J) = 0.0
530   CONTINUE
      TABI(I,I) = 1.0
      TABI(I,K1) = UP(I)
      TABI(I,K2) = 0.0
      IF ( E(I) .GT. EMAX )  GO TO 560
      EMAX = E(I)
      IREM = I
      GO TO 560
C
540   DO 550 J=1,K2
      IDUM(I) = 1
      TABI(I,J) = TAB(I,J)
550   CONTINUE
C
560   CONTINUE
C
C
      IF (MASK2 .NE. 10 )  GO TO 565
      IDUM(IREM) = 1
      DO 562 J=1,K2
      TABI(IREM,J) = TAB(IREM,J)
562   CONTINUE
C
565   DO 580 I=IN,LL
      DO 570 J=1,K2
      TABI(I,J) = TAB(I,J)
570   CONTINUE
```

```
560   CONTINUE
C
C
C
C
C          SOLVE THE SET OF EQUATIONS .
C
C
600   DO 610 I=1,LL
      DO 610 J=1,K2
      WORKA(I,J) = TABI(I,J)
610   CONTINUE
      V(1) = 4.0
C
      CALL GJR (WORKA,NCNT1,NCNT2,LL,K2,$620,JC,V)
      GO TO 700
620   WRITE(5,630)
630   FORMAT(1H1,13H ERROR RETURN  )
      GO TO 1000
C
C
C
C
C          CONSTRUCT PARTIAL DERIVATIVE TABLES .
C
C
700   DO 730 I=1,N
      IF (IDUM(I).EQ.1)  GO TO 730
      P(I) = 0.0
      SMP(I) = -E(I)
      DO 710 J=1,N
      P(I) = P(I) + 2.0*COV(I,J)*WORKA(J,K1)
      SMP(I) = SMP(I) + 2.0*COV(I,J)*WORKA(J,K2)
710   CONTINUE
      IF (MASK1.EQ.0)  GO TO 725
      DO 720 J=1,M
      KKK = J+N
      P(I) = P(I) - D(J,I)*WORKA(KKK,K1)
      SMP(I) = SMP(I) - D(J,I)*WORKA(KKK,K2)
720   CONTINUE
725   P(I) = P(I) - WORKA(LL,K1)
      SMP(I) = SMP(I) - WORKA(LL,K2)
730   CONTINUE
C
C
C
C
C          FIND THE LARGEST CRITICAL VALUE OF LAMBDA .
C
C
      ICNT = 0
      PLAMAX = 0.0
      DO 870 I=1,N
      IF (IDUM(I) -1) 810,830,820
C
810   IF (SMP(I).LE.0.0)  GO TO 870
      PLAMC = -P(I) / SMP(I)
      GO TO 860
C
820   IF (SMP(I).GE.0.0)  GO TO 870
```

```
      PLAMC = -P(I) / SMP(I)
      GO TO 860
C
 830  IF (WORKA(I,K2)) 840,870,850
C
 840  PLAMC= ( UP(I) - WORKA(I,K1) ) / WORKA(I,K2)
      GO TO 860
C
 850  PLAMC = ( SMAL(I) - WORKA(I,K1) ) / WORKA(I,K2)
C
 860  IF (PLAMC.LT.PLAMAX)  GO TO 870
      PLAMAX = PLAMC
      ICNT = I
C
 870  CONTINUE
C
      IF ( IFLAG .EQ. 0 )  GO TO 874
      IF ( OLDLAM .GT. DLAM .AND. PLAMAX .LT. DLAM )  GO TO 970
      GO TO 900
 874  WRITE(5,875) PLAMAX,OLDLAM
 875  FORMAT(1H1,26H FOR VALUES OF LAMBDA FROM,F10.5,3H TO,F11.5,16H THE
     1 SOLUTION IS,/)
      WRITE(5,880) (I,WORKA(I,K1),WORKA(I,K2),I=1,N)
 880  FORMAT(3H X(,I3,4H) = ,F9.5,3H + ,F11.5,7H LAMBDA     )
C
 882  ZZ = 0.0
      YY = 0.0
      PP = 0.0
      QQ = 0.0
      RR = 0.0
      DO 890 I=1,N
      ZZ = ZZ + E(I)*WORKA(I,K1)
      YY = YY + E(I)*WORKA(I,K2)
      DO 885 J=1,N
      PP = PP + WORKA(I,K1)*WORKA(J,K1)*COV(I,J)
      QQ = QQ +(WORKA(I,K1)*WORKA(J,K2) + WORKA(J,K1)*WORKA(I,K2) ) *COV
     1(I,J)
      RR = RR + WORKA(I,K2)*WORKA(J,K2)*COV(I,J)
 885  CONTINUE
 890  CONTINUE
C
      WRITE(5,892) ZZ,YY
 892  FORMAT(///,8H E(P) = ,F8.5,3H + ,F8.5,7H LAMBDA)
      WRITE(5,893) PP,QQ,RR
 893  FORMAT(8H V(P) = ,F8.5,3H + ,F8.5,7H LAMBDA,3H + ,F8.5,10H LAMBDA*
     1*2)
C
      EP = ZZ + YY*OLDLAM
      VP = PP + QQ*OLDLAM + RR*(OLDLAM**2)
      SPD = SQRT (VP)
      WRITE(5,895) OLDLAM,EP,VP,SPD
 895  FORMAT(//,13H AT LAMBDA = ,F10.5,5X,8H E(P) = ,F8.5,/,29X,7HV(P) =
     1 ,F8.5,/,26X,10HS.D.(P) = ,F8.5)
C
      EP = ZZ + YY*PLAMAX
      VP = PP + QQ*PLAMAX + RR*(PLAMAX**2)
      SPD = SQRT (VP)
      WRITE(5,895) PLAMAX,EP,VP,SPD
C
      IF ( IIFLG .EQ. 1 )  GO TO 1000
```

```
      IF ( PLAMAX.LE.0.0 )  GO TO 1000
      OLDLAM = PLAMAX
C
C
C
C
C         CHANGE STATUS OF VARIABLES AND MODIFY THE EQUATION .
C
C
900   I = ICNT
      IF (IDUM(I).EQ.1)  GO TO 920
      IDUM(I) = 1
      DO 910 J=1,K2
      TABI(I,J) = TAB(I,J)
910   CONTINUE
      GO TO 960
C
920   RR= WORKA(I,K1) + WORKA(I,K2)*PLAMAX
      IF (RR.LE.SMAL(I) )  GO TO 940
C
      IDUM(I) = 2
      DO 930 J=1,LL
      TABI(I,J) = 0.0
930   CONTINUE
      TABI(I,I) = 1.0
      TABI(I,K1) = UP(I)
      TABI(I,K2) = 0.0
      GO TO 960
C
940   IDUM(I) = 0
      DO 950 J=1,LL
      TABI(I,J) = 0.0
950   CONTINUE
      TABI(I,I) = 1.0
      TABI(I,K1) = SMAL(I)
      TABI(I,K2) = 0.0
C
960   GO TO 600
970   WRITE(5,975)
975   FORMAT(1H1,' THE OPTIMAL SOLUTION IS ',//)
      DO 990 I=1,N
      XXX = WORKA(I,K1) + WORKA(I,K2)*DLAM
      WRITE(5,980) I,XXX
980   FORMAT (3H X(,I2,4H) = ,F7.5 )
990   CONTINUE
      IIFLG = 1
      OLDLAM = DLAM
      GO TO  882
C
C
C
C
1000  STOP
      END
```

C.2 COMPUTER PROGRAM FOR SHARPE'S INDEX MODEL

A listing of the program is given on the following pages, but once again, prior to this a brief description of the input required is given.

CARD 1: NCR, NP

NCR = Logical unit number with which cards 3,4,5, and 6 are read.

NP = Number of shares in the universe.

FORMAT: (2I5)

CARD 2: N,L

N = Actual number of shares considered.

L = Number of indices used.

FORMAT: (2I5)

CARD 3: E(I), I = 1,2,...,N

E(I) = Expected return on the I^{th} security.

FORMAT: (F14.6)

CARD 4: VARU(I), I = 1,2,...,N

VARU(I) = Variance of the residual (that is, $\sigma^2_{u_i}$) in the equation below.

$$R_i = \alpha_i + \beta_{i1}I_1 + \beta_{i2}I_2 + \ldots + \beta_{iL}I_L + u_i \quad (J.1)$$

FORMAT: (F14.6)

CARD 5: COV(J,K); J = 1,2,...,L; K = 1,2,...,L

COV(J,K) = Covariance (estimated) between the J^{th} and K^{th} indices.

FORMAT: (F14.6)

CARD 6: BETA(I,J); I = 1,2,...,N; J = 1,2,...,L

BETA(I,J) = estimator of β_{ij} in equation (J.1).

FORMAT: (F14.6)

CARD 7: MASK2

MASK2 = 0 implies that all shares have the same upper and lower bounds;

= 10 implies that all shares have upper and lower bounds of 1,0 and 0,0 respectively;

≠ 0 or 10 implies that they have differing upper and lower bounds.

FORMAT: (I5)

CARD 8: 1) If MASK2 ≠ 0 or 10

SMAL(I), UP(I): I = 1,2,...,N

SMAL(I) = Lower bound of I^{th} security;

UP(I) = Upper bound of I^{th} security.

2) If MASK2 = 0 or 10

SSS1, SSS2

SSS1 = lower bound imposed on all securities;

SSS2 = upper bound imposed on all securities.

FORMAT: (2F10.4)

<u>CARD 9</u>: IFIX(I); I = 1,2,...,20

As discussed in Section C.1 for the Markowitz Model.

FORMAT: (10I5)

```
      DIMENSION E( 50),COV( 10, 10),BETA( 50, 10),VARU( 50),TAB( 70, 70)
     1,TAB1( 70, 70),WORKA( 70, 70),SMAL( 50),UP( 50),IDUM( 50),A(101, 5
     20),RHS(101),T(101),COST( 50),XX(103),JX(101),PI(101),EE(103,103),E
     3RR(4),TOL(4),IFIX(20),IOUT(4),TT(3),Y(103),S( 50),V(2),X( 50),P( 5
     40),SMP( 50),JC( 70)
C
      DATA (TT(J),J=1,3) /1H ,1H+,1H-/(IDUM(I),I=1, 50)/ 50*1/
      NCNT1 = 70
      NCNT2 = 70
      EMAX = 1000.0
C
C
C
C
C                DATA INPUT AND INITIALIZATION SECTION .
C
C
      READ(8,100) NCR,NP
100   FORMAT(2I5)
      READ(8,100) N,L
      M1 = N+1
      M2 = N+L
      M3 = N+L+1
      M4 = N+2+L
      LL = M4+1
      K1 = LL+1
      K2 = K1+1
C
      READ(NCR,110) (E(I),I=1,N)
110   FORMAT(F14.6)
      READ(NCR,110) (VARU(I),I=1,N)
      DO 120 J=1,L
      READ(NCR,110) (COV(J,K),K=1,L)
120   CONTINUE
      DO 130 J=1,L
      READ(NCR,110) (BETA(I,J),I=1,N)
130   CONTINUE
C
      READ(8,100) MASK2
      IF (MASK2.EQ.0)  GO TO 150
      IF (MASK2.EQ.10)  GO TO 150
      READ(8,140) (SMAL(I),UP(I),I=1,N)
140   FORMAT(2F10.4)
      GO TO 200
C
150   READ(8,140) SSS1,SSS2
      DO 160 I=1,N
      SMAL(I) = SSS1
      UP(I) = SSS2
160   CONTINUE
C
C
C
C
C                CONSTRUCTION OF TAB .
C
C
200   DO 250 I=1,N
      DO 210 J=1,M2
```

```
      TAB(I,J) = 0.0
210   CONTINUE
      TAB(I,I) = 2.0 * VARU(I)
      DO 220 J=M3,M4
      KKK = J-M2
      TAB(I,J) = -1.0 * BETA(I,KKK)
220   CONTINUE
      TAB(I,LL) = -1.0
      TAB(I,K1) = 0.0
      TAB(I,K2) = E(I)
230   CONTINUE
C
      DO 270 I=M1,M2
      III = I-N
      DO 240 J=1,N
      TAB(I,J) = 0.0
240   CONTINUE
      DO 250 J=M1,M2
      JJJ = J-N
      TAB(I,J) = 2.0 * COV(III,JJJ)
250   CONTINUE
      DO 260 J=M3,K2
      TAB(I,J) = 0.0
260   CONTINUE
      TAB(I,M2+III) = 1.0
270   CONTINUE
C
      DO 300 I=M3,M4
      III = I-M2
      DO 280 J=1,N
      TAB(I,J) = -1.0 *BETA(J,III)
280   CONTINUE
      DO 290 J=M1,K2
      TAB(I,J) = 0.0
290   CONTINUE
      KKK = I-L
      TAB(I,KKK) = 1.0
300   CONTINUE
C
      DO 310 J=1,N
      TAB(LL,J) = -1.0
310   CONTINUE
      DO 320 J=M1,LL
      TAB(LL,J) = 0.0
320   CONTINUE
      TAB(LL,K1) = -1.0
      TAB(LL,K2) = 0.0
C
      WRITE(5,330)
330   FORMAT(1H1)
      DO 360 I=1,LL
      WRITE(5,340)
340   FORMAT(1H0)
      WRITE(5,350) (TAB(I,J),J=1,K2)
350   FORMAT(10F10.4)
360   CONTINUE
C
C
C
C
```

```
C                 FINDING THE MAXIMUM E(P) PORTFOLIO .
C
C
      DO 410 J=1,N
      A(1,J) =1.0
410   CONTINUE
      RHS(1) = 1.0
      T(1) = TT(1)
C
      DO 430 I=1,N
      KKK = I+1
      DO 420 J=1,N
      A(KKK,J) = 0.0
420   CONTINUE
      A(KKK,I) = 1.0
      RHS(KKK) = SMAL(I)
      T(KKK) = TT(3)
430   CONTINUE
C
      DO 450 I=1,N
      KKK = N+I+1
      DO 440 J=1,N
      A(KKK,J) = 0.0
440   CONTINUE
      A(KKK,I) = 1.0
      RHS(KKK) = UP(I)
      T(KKK) = TT(2)
      COST(I) = -E(I)
450   CONTINUE
C
      READ(8,460) (IFIX(I),I=1,20)
460   FORMAT(10I5)
      IFIX(3) = N+N+1
      IFIX(4) = N
      DO 470 I=1,4
      TOL(I) = 0.0
470   CONTINUE
C
      CALL SIMPLX (A,T,RHS,COST,IFIX,TOL,OBJ,XX,JX,PI,EE,ERR,IOUT,Y,S)
C
      DO 480 I=1,N
      X(I) = 0.0
480   CONTINUE
      IFIZ = IFIX(3)
      DO 490 I=1,IFIZ
      II = JX(I)
      IF (II.GT.N)  GO TO 490
      X(II) =XX(I)
490   CONTINUE
      OLDLAM = 1000.00
C
C
C
C
C          ASSIGN STATUS TO EACH VARIABLE AND MODIFY THE SET OF EQUATIONS
C
C
      DO 500 I=1,N
      IF (X(I).LE.SMAL(I))  IDUM(I) = 0
      IF (X(I).GE.UP(I) )   IDUM(I) = 2
```

```
500   CONTINUE
C
      DO 560 I=1,N
      IF(IDUM(I)) ,,520
      DO 510 J=1,LL
      TABI(I,J) = 0.0
510   CONTINUE
      TABI(I,I) = 1.0
      TABI(I,K1) = SMAL(I)
      TABI(I,K2) = 0.0
      GO TO 560
C
520   IF (IDUM(I)-2) 540,,
      DO 530 J=1,LL
      TABI(I,J) = 0.0
530   CONTINUE
      TABI(I,I) = 1.0
      TABI(I,K1) = UP(I)
      TABI(I,K2) = 0.0
      IF ( E(I) .GT. EMAX )  GO TO 560
      EMAX = E(I)
      IREM = I
      GO TO 560
C
540   IDUM(I) = 1
      DO 550 J=1,K2
      TABI(I,J) = TAB(I,J)
550   CONTINUE
C
560   CONTINUE
C
C
      IF ( MASK2 .NE. 10 )  GO TO 565
      IDUM(IREM) = 1
      DO 562 J=1,K2
      TABI(IREM,J) = TAB(IREM,J)
562   CONTINUE
C
565   DO 580 I=M1,LL
      DO 570 J=1,K2
      TABI(I,J) = TAB(I,J)
570   CONTINUE
580   CONTINUE
C
C
C
C
C              SOLVE THE SET OF EQUATIONS .
C
C
600   DO 610 I=1,LL
      DO 610 J=1,K2
      WORKA(I,J) = TABI(I,J)
610   CONTINUE
      V(1) = 4.0
C
      CALL GJR (WORKA,NCNT1,NCNT2,LL,K2,$620,JC,V)
C
      GO TO 700
620   WRITE(5,630)
```

```
650   FORMAT(1H1,15H ERROR RETURN )
      GO TO 1000
C
C
C
C
C                CONSTRUCT PARTIAL DERIVATIVE TABLES .
C
C
700   DO 720 I=1,N
      IF (IDUM(I).EQ.1)  GO TO 720
      P(I) = 2*VARU(I)*WORKA(I,K1)
      SMP(I) = 2*VARU(I)*WORKA(I,K2) - E(I)
C
      DO 710 J=1,L
      JJJ = M2+ J
      P(I) = P(I) - WORKA(JJJ,K1)*BETA(I,J)
      SMP(I) = SMP(I) - WORKA(JJJ,K2)*BETA(I,J)
710   CONTINUE
C
      P(I) = P(I) -WORKA(LL,K1)
      SMP(I) = SMP(I) - WORKA(LL,K2)
720   CONTINUE
C
C
C
C
C                FIND THE LARGEST CRITICAL VALUE OF LAMBDA .
C
C
      ICNT = 0
      PLAMAX = 0.0
      DO 870 I=1,N
      IF (IDUM(I)-1) 810,830,820
C
810   IF (SMP(I).LE.0.0)  GO TO 870
      PLAMC = -P(I) / SMP(I)
      GO TO 860
C
820   IF (SMP(I).GE.0.0)  GO TO 870
      PLAMC = -P(I) /SMP(I)
      GO TO 860
C
830   IF (WORKA(I,K2)) 840,870,850
C
840   PLAMC = ( UP(I)-WORKA(I,K1) ) / WORKA(I,K2)
      GO TO 860
C
850   PLAMC = ( SMAL(I)-WORKA(I,K1) ) / WORKA(I,K2)
C
860   IF (PLAMC.LT.PLAMAX)  GO TO 870
      PLAMAX = PLAMC
      ICNT = 1
C
870   CONTINUE
C
C
      WRITE(5,875) PLAMAX,OLDLAM
875   FORMAT(1H1,28H FOR VALUES OF LAMBDA FROM,F10.5,3H TO,F11.5,16H THE
     1 SOLUTION IS,/)
```

```
      WRITE(5,880) (I,WORKA(I,K1),WORKA(I,K2),I=1,N)
  880 FORMAT(3H X(,I3,4H) = ,F9.5,3H + ,F9.5,7H LAMBDA )
C
      ZZ = 0.0
      YY = 0.0
      PP = 0.0
      QQ = 0.0
      RR = 0.0
C
      DO 885 I=1,N
      ZZ = ZZ + E(I)*WORKA(I,K1)
      YY = YY + E(I)*WORKA(I,K2)
      PP = PP + (WORKA(I,K1)**2) * VARU(I)
      QQ = QQ + 2.0*WORKA(I,K1)*WORKA(I,K2)*VARU(I)
      RR = RR + (WORKA(I,K2)**2)*VARU(I)
  885 CONTINUE
C
      DO 890 J=1,L
      JJJ = N+J
      DO 889 K=1,L
      KKK = N+K
      PP = PP + WORKA(JJJ,K1)*WORKA(KKK,K1)*COV(J,K)
      QQ = QQ + ( WORKA(JJJ,K1)*WORKA(KKK,K2) + WORKA(KKK,K1)*WORKA(JJJ,
     1K2) )*COV(J,K)
      RR = RR + WORKA(JJJ,K2)*WORKA(KKK,K2)*COV(J,K)
  889 CONTINUE
  890 CONTINUE
C
C
      WRITE(5,892) ZZ,YY
  892 FORMAT(///,8H E(P) = ,F8.5,3H + ,F8.5,7H LAMBDA )
      WRITE(5,893) PP,QQ,RR
  893 FORMAT(8H V(P) = ,F8.5,3H + ,F8.5,7H LAMBDA,3H + ,F8.5,10H LAMBDA*
     1*2)
C
      DO 897 I=1,2
      EP = ZZ + YY*OLDLAM
      VP = PP + QQ*OLDLAM + RR*(OLDLAM**2)
      SDP = SQRT(VP)
      WRITE(5,895) OLDLAM,EP,VP,SDP
  895 FORMAT(//,13H AT LAMBDA = ,F10.5,5X,8H E(P) = ,F8.5,/,29X,7HV(P) =
     1 ,F8.5,/,26X,10HS.D.(P) = ,F8.5)
C
      OLDLAM = PLAMAX
  897 CONTINUE
      IF (PLAMAX.LE.0.0)  GO TO 1000
C
C
C
C
C          CHANGE STATUS OF VARIABLES AND MODIFY THE SET OF EQUATIONS .
C
C
      I = ICNT
      IF (IDUM(I).EQ.1)  GO TO 920
      IDUM(I) = 1
      DO 910 J=1,K2
      TABI(I,J) = TAB(I,J)
  910 CONTINUE
      GO TO 960
```

```
C
920   R = WORKA(I,K1) + WORKA(I,K2)*PLAMAX
      IF (R.LE.SMAL(I))  GO TO 940
      IDUM(I) = 2
      DO 930 J=1,LL
      TABI(I,J) = 0.0
930   CONTINUE
      TABI(I,1) = 1.0
      TABI(I,K1) = UP(I)
      TABI(I,K2) = 0.0
      GO TO 960
C
940   IDUM(I) = 0
      DO 950 J=1,LL
      TABI(I,J) = 0.0
950   CONTINUE
      TABI(I,1) = 1.0
      TABI(I,K1) = SMAL(I)
      TABI(I ,K2) = 0.0
C
960   GO TO 600
C
C
C
1000  STOP
      END
```

A P P E N D I X K

T A B L E K.I

THE GOLD SHARES USED FOR TRAPEZOIDAL DATA EXAMPLE

1. East Rand Proprietary Mines Ltd.
2. Bracken Mines Ltd.
3. Kinross Mines Ltd.
4. Winkelhaak Mines Ltd.
5. Southvaal Holdings Ltd.
6. Free State Geduld Mines Ltd.
7. President Brand Gold Mining Co. Ltd.
8. President Steyn Gold Mining Co. Ltd.
9. St. Helena Gold Mines Ltd.
10. Welkom Gold Mining Co. Ltd.
11. Western Holdings Ltd.
12. East Driefontein Gold Mining Co. Ltd.
13. Kloof Gold Mining Co. Ltd.
14. West Driefontein Gold Mining Co. Ltd.
15. Western Areas Gold Mining Co. Ltd.
16. Western Deep Levels Ltd.

B I B L I O G R A P H Y

1. AFFLECK-GRAVES, J.F. (1974): Portfolio Selection on the Johannesburg Stock Exchange. Unpublished M.Sc. thesis, University of Cape Town.

2. AFFLECK-GRAVES, J.F. and MONEY, A.H. (1975): A Note on the Random Walk Model and South African Share Prices. The South African Journal of Economics, Vol. 43, pp 382-388.

3. AFFLECK-GRAVES, J.F. and MONEY, A.H. (1976): A Comparison of Two Portfolio Selection Models. The Investment Analysts Journal, Vol. 7, pp 35-40.

4. ALTMAN, E.I., JACQUILLAT, B., and LEVASSEUR, M. (1974): Comparative Analysis of Risk Measures: France and the United States. Journal of Finance, Vol. 29, pp 1495-1511.

5. ANDERBERG, M.R. (1973): Cluster Analysis for Applications. Academic Press, New York.

6. ANDERSON, T.W. (1958): An Introduction to Multivariate Statistical Analysis. John Wiley & Sons, N.Y.

7. BACHELIER, L. (1900): Theorie de la Speculation. Gauthier-Villars, Paris. Reprinted in Cootner (1964).

8. BAESEL, J.B. (1974): On the Assessment of Risk : Some Further Considerations. Journal of Finance, Vol. 29, pp 1491-1494.

9. BAUMOL, W.J. (1963): An Expected Gain-Confidence Limit Criterion for Portfolio Selection. Management Science, Vol. 10, pp 174-182.

10. BEAVER, W., KETTLER, P., and SCHOLES, M. (1970): The Association Between Market Determined and Accounting Determined Risk Measures. The Accounting Review, Vol. 45, pp 654-682.

11. BLUME, M.E. (1971): On the Assessment of Risk. Journal of Finance, Vol. 26, pp 1-10.

12. BROWN, S.J. (1977): Heteroscedasticity in the Market Model: A Comment. Journal of Business, Vol. 50, pp 80-83.

13. CHEN, A.H., JEN, F.C. and ZIONTS, S. (1971): The Optimal Portfolio Revision Policy. Journal of Business, Vol. 44, pp 51-61.

14. COHEN, A.W. (1966): The Chartcraft Method of Point and Figure Trading. Chartcraft Inc.,N.Y.

15. COHEN, K.J. and FITCH, B.P. (1966): The Average Investment Performance Index. Management Science, Vol. 12, pp 195-215.

16. COHEN, K.J. and POGUE, J.A. (1967): An Empirical Evaluation of Alternative Portfolio Selection Models. Journal of Business, Vol. 40, pp 166-193.

17. COOTNER, P.H. (1962): Stock Prices: Randon vs Systematic Changes. Industrial Management Review, Vol. 3, pp 24-45. Reprinted in Cootner (1964).

18. COOTNER, P.H. (1964): The Variation of Certain Speculative Prices. M.I.T. Press, Massachussetts.

19. COWLES, A. (1933): Can Stock Market Forecasters Forecast? Econometrica, Vol. 1, pp 309-324

20. DIXON, W.J. (1975): BMDP Biomedical Computer Programs (Editor). University of California Press, Berkley, L.A.

21. ELTON, E.J. and GRUBER, M.J. (1972): Security Evaluation and Portfolio Analysis. Prentice-Hall Inc., N.J.

22. FAMA, E.F. (1965a): The Behaviour of Stock Market Prices. Journal of Business, Vol. 38, pp 34-105.

23. FAMA, E.F. (1965b): Portfolio Analysis in a Stable Paretian Market. Management Science, Vol. 11, pp 404-419.

24. FAMA, E.F. (1968): Risk, Return and Equilibrium : Some Clarifying Comments. Journal of Finance, Vol. 23, pp 29-40.

25. FAMA, E.F., FISHER, L., JENSEN, M.C., and ROLL, R. (1969): The Adjustment of Stock Prices to New Information. International Economic Review, Vol. 10, pp 1-21. Reprinted in Elton and Gruber (1972).

26. FARRAR, D.E. (1962): The Investment Decision Under Uncertainty. Prentice-Hall Inc., N.J.

27. FEENEY, G.J. and HESTER, D.D. (1967): Stock Market Indices: A Principal Components' Analysis. In Hester and Tobin (1967).

28. FRANCIS, J.C. and ARCHER, S.H. (1971): Portfolio Analysis. Prentice-Hall, Inc., N.J.

29. GOLDBERGER, A.S. (1964): Econometric Theory. John Wiley & Sons, N.Y.

30. GRANGER, C.W.J. and MORGENSTERN, O. (1963): Spectral Analysis of New York Stock Market Prices. Kyklos, Vol. 16, pp 1-27. Reprinted in Cootner (1964).

31. GRANVILLE, J.E. (1963): New Key to Stock Market Profits. Prentice-Hall Inc., N.J.

32. GROSS, R. (1974): Risk Analysis of Ordinary Shares Quoted on the Johannesburg Stock Exchange. The Investment Analysts Journal, Vol 4, pp 30-34.

33. HAKANSSON, N.H. (1971): Multi-Period Mean Variance Analysis: Towards a General Theory of Portfolio Choice. Journal of Finance, Vol. 26, pp 857-884.

34. HARTIGAN, J.A. (1975): Clustering Algorithms. Wiley & Sons, N.Y.

35. HASTIE, K. (1967): The Determination of Optimal Investment Policy. Management Science, Vol. 13, pp 757-774.

36. HAYCOCKS, H.W. and PLYMEN, J. (1964): The Design, Application and Future Development of the Financial Times - Actuaries Index. Journal of the Institute of Actuaries, Vol. 90, pp 267-304.

37. HESTER, D.D. and TOBIN, J. (1967): Risk Aversion and Portfolio Choice. John Wiley & Sons, N.Y.

38. HIMMELBLAU, D.M. (1972): Applied Nonlinear Programming. McGraw-Hill, N.Y.

39. HODGES, S. and SCHAEFER, S. (1974): The Interpretation of the Geometric Mean : A Note. Journal of Financial and Quantitative Analysis, Vol. 9, pp 497-504.

40. HWEI-JU CHEN, GNANADESIKAN, R., and KETTENRING, J.R. (1974): Statistical Methods for Grouping Corporations. Sankhya, Vol. 36, pp 1-28.

41. JOHNSTON, J. (1963): Econometric Methods, Second Edition. McGraw-Hill, N.Y.

42. KANTOR, M. (1972): On Market Indexes. In Szegö and Shell (1972) pp 620-632.

43. KENNEY, J.F. and KEEPING, E.S. (1951): Mathematics of Statistics : Part II. Second Edition. D Van Nostrand, N.J.

44. KING, B.F. (1966): Market and Industry Factors in Stock Price Behaviour. Journal of Business, Vol. 39, pp 139-190.

45. LATANÉ, H.A., TUTTLE, D.L., and YOUNG, W.E. (1971): Market Indexes and Their Implications for Portfolio Management. Financial Analysts Journal, Vol. 27, pp 75-85.

46. LATANÉ, H.A., TUTTLE, D.L., and JONES, C.P. (1975): Security Analysis and Portfolio Management. The Ronald Press Co., N.Y.

47. LEVY, H., and SARNAT, M. (1972): Investment and Portfolio Analysis. John Wiley and Sons.

48. LEVY, R.A. (1968): The Relative Strength Concept of Common Stock Price Forecasting. Investors Intelligence, N.Y.

49. LEVY, R.A. (1971): On the Short Term Stationarity of Beta Coefficients. Financial Analysts Journal, Vol. 27, pp 55-62.

50. LINTNER, J. (1965): Security Prices, Risk, and Maximal Gains from Diversification. Journal of Finance, Vol. 20, pp 587-613.

51. MANDELBROT, B. (1963): The Variation of Certain Speculative Prices. Journal of Business, Vol. 36, 394-419.

52. MARKOWITZ, H.M. (1952): Portfolio Selection. Journal of Finance, Vol. 7, 77-91.

53. MARKOWITZ, H.M. (1959): Portfolio Selection : Efficient Diversification of Investments. Wiley & Sons, N.Y.

54. MARKS, P. and STUART, A. (1971): An Arithmetic Version of the Financial Times Industrial Ordinary Share Index. Journal of the Institute of Actuaries, Vol. 97, pp 297-324.

55. MARKSTEIN, D.L. (1966): How to Chart Your Way to Stock Market Profits. Parker Publishing Co., N.Y.

56. MARTIN, J.D. and KLEMKOSKY, R.C. (1975): Evidence of Heteroscedasticity in the Market Model. Journal of Business, Vol. 48, 81-86.

57. MILLER, R.G. (1966): Simultaneous Statistical Inference. McGraw-Hill, N.Y.

58. MODIGLIANI, F. and POGUE, G.A. (1974): An Introduction to Risk and Return : Concepts and Evidence. Financial Analysts Journal, Vol. 30, pp 69-86.

59. MORRISON, D.F. (1967): Multivariate Statistical Methods. McGraw-Hill, N.Y.

60. MOSSIN, J. (1968): Optimal Multiperiod Portfolio Policies. Journal of Business, Vol. 41, 215-229.

61. NIEDERHOFFER, V. and OSBORNE, M.F.M. (1965): Market Making and Reversal on the Stock Exchange. Journal of the American Statistical Association, Vol. 61, pp 897-916.

62. OSBORNE, M.F.M. (1959): Brownian Motion in the Stock Market. Operations Research, Vol. 7, pp 145-173. Reprinted in Cootner (1964).

63. OSBORNE, M.F.M. (1962): Periodic Structure in the Brownian Motion of Stock Prices. Operations Research, Vol. 10, pp 345-379.

64. PRAETZ, P.D. (1969): Australian Share Prices and the Random Walk Hypothesis. Australian Journal of Statistics, Vol. 11, pp 123-139.

65. ROSENBERG, B. and McKIBBEN, W. (1973): The Prediction of Systematic and Specific Risk in Common Stocks. Journal of Financial and Quantitative Analysis, Vol. 8, 317-333.

66. SCHWARTZ, R.S. and WHITCOMB, D.K. (1977): The Time-Variance Relationship : Evidence on Autocorrelation in Common Stock Returns. Journal of Finance, Vol. 32, No.1, pp 41-55.

67. SHARPE, W.F. (1963): A Simplified Model for Portfolio Analysis. Management Science, Vol. 9, pp 277-293.

68. SHARPE, W.F. (1964): Capital Asset Prices : A Theory of Market Equilibrium under Conditions of Risk. Journal of Finance, Vol. 19, pp 425-442.

69. SHARPE, W.F. (1967): A Linear Programming Algorithm for Mutual Fund Portfolio Selection. Management Science, Vol. 13, pp 499-510.

70. SHARPE, W.F. (1970): Portfolio Theory and Capital Markets. McGraw-Hill, N.Y.

71. SHARPE, W.F. and COOPER, G.M. (1972): Risk-Return Classes of New York Stock Exchange Stocks, 1931-1967. Financial Analysts Journal, Vol. 28, pp 46-54.

72. SIEGEL, S. (1956): Nonparametric Statistics for the Behavioral Sciences. McGraw Hill, N.Y.

73. SMITH, ADAM (1967): The Money Game. Random House, N.Y.

74. SMITH, K.V. (1967): A Transition Model for Portfolio Revision. Journal of Finance, Vol. 22, pp 425-439.

75. SOKAL, R.R. and SNEATH, P.H.A. (1963): Principles of Numerical Taxonomy. Freeman, S.F.

76. SOLNIK, B.H. (1973): Note on the Validity of the Random Walk for European Stock Prices. Journal of Finance, Vol. 28, pp 1151-1159.

77. SPRECHER, C.R. (1975): Introduction to Investment Management. Houghton Mifflin, Boston.

78. SZEGÖ, G.P. and SHELL, K. (1972): Mathematical Methods in Investment and Finance. North Holland, Amsterdam.

79. THEIL, H. (1960): Best Linear Index Numbers of Prices and Quantities. Econometrica, Vol. 28, pp 464-480.

80. TOBIN, J. (1958): Liquidity Preferences as Behaviour Towards Risk. Review of Economic Studies, Vol. 25, pp 65-86. Reprinted in Elton and Guber (1972).

81. TREYNOR, J.L. (1965): How to Rate Management of Investment Funds. Harvard Business Review, Vol. 43, pp 63-75.

82. TROSKIE, C.G. (1970): Principal Components and its Applications. Unpublished paper read at the Conference of the S.A. Statistical Association, October, 1970.

83. TRYON, R.C. and BAILEY, D.E. (1970): Cluster Analysis. McGraw-Hill, N.Y.

84. WALLINGFORD, B.A. (1967): A Survey and Comparison of Portfolio Selection Models. Journal of Financial and Quantitative Analysis, Vol. 2, pp 85-106.

85. WOLFE, P. (1959): The Simplex Method for Quadratic Programming. Econometrica, Vol. 27, pp 382-398.

86. YOUNG, W.E. (1971): Random Walk of Stock Prices : A Test of the Variance-Time Function. Econometrica, Vol, 39, pp 797-812.