International Institute for
Applied Systems Analysis
www.iiasa.ac.at

Identifying Regularities in Stock Portfolio Tilting

Susmaga, R., Michalowski, W. and Slowinski, R.

IIASA

INTERIM REPORT   IR-97-66/September

Identifying Regularities in Stock Portfolio Tilting

Robert Susmaga (robert.susmaga@cs.put.poznan.pl)
Wojtek Michalowski (michalow@iiasa.ac.at)
Roman Slowinski (slowinski@sol.put.poznan.pl)
.

Approved by
Pekka Korhonen (korhonen@iiasa.ac.at)
Leader, *Decision Analysis and Support Project (DAS)*

Contents

Abstract

The paper deals with the issues associated with identification of stocks generating abnormal returns. Following the findings of a finance theory regarding portfolio tilting, a set of price-related stocks' attributes was analyzed. The analysis was conducted with the help of rough sets methodology which allows to distinguish "important" attributes for problem description, and to generate decision rules which can be later used to predict stocks' performance. Validity of the approach was tested on the Toronto Stock Exchange data.

Keywords: rough sets, decision rules, reducts, portfolio tilting, anomalies theory

About the Authors

Roman Slowinski is Professor of Decision and Computer Sciences and Head of the Laboratory of Intelligent Decision Support Systems, Institute of Computer Science, Poznan University of Technology, Poznan, Poland.

Robert Susmaga is on the research and teaching staff of the Institute of Computer Science at Poznan University of Technology, Poznan, Poland.

Wojtek Michalowski is a Senior Research Scholar with the Decision Analysis and Support Project of IIASA.

Identifying Regularities in Stock Portfolio Tilting

Robert Susmaga
Wojtek Michalowski
Roman Slowinski

1. Introduction

If a stock market's behavior would exactly follow predictions of the efficient market hypothesis and a single factor capital asset pricing model, there might not be a possibility to outperform the market in a systematic fashion. However, an increasing number of studies reports the departures in terms of stocks' returns from what would be predicted by classical finance methodology. Apart from enhancing our knowledge and understanding of the risk/return trade-off, these studies also suggest that it is possible to design certain investment strategies with potential to generate "abnormal" returns. Most of the researchers and practitioners agree that these strategies should be based on considering price-related stock attributes while making investment decisions. The purpose of the research described in this paper is to determine relative importance of well established price-related attributes in a so called "portfolio tilting" process. Our approach is different from that reported in classical finance literature which essentially considers a cross-sectional multivariate regression of stock or portfolio returns on a set of predetermined price-related variables besides the systematic risk (Jog and Li, 1995). Such an analysis does not provide important information regarding reduction of the original set of the attributes and regarding performance of such a reduced set. In order to generate this information, a stock performance should be studied from a different perspective, such as, for example those provided by a broad range of knowledge discovery methods. One such method, tested in a variety of applications, is a rough sets (RS) methodology (Pawlak, 1991; Slowinski, 1992). We are proposing to apply the RS to the evaluation of price-related attributes for portfolio tilting, and to the identification of a predictive value of a reduced set.

The paper is organized as follows. In the next section we describe the main issues of portfolio tilting. In Section 3 we present fundamental notions of the RS theory and describe its software implementation. In Section 4 we discuss application of the RS to analyze stock data from the Toronto Stock Exchange, and present the results of this analysis. The paper concludes with the discussion.

2. Portfolio Tilting

Since the advent of the Modern Portfolio Theory (MPT) arising from the work of Markowitz (1959), the notion of diversified portfolios has become one of the most fundamental concepts of portfolio management. While developed as a financial economic theory in a conditional-normative framework, the MPT has spawned a variety of applications and further theoretic constructs. For the purpose of this paper, three aspects of the research arising from MPT are most relevant.

First, the MTP was derived using a representative investor belonging to the von Neumann and Morgenstern family of functions, which manifested in portfolio optimization techniques based on the mean-variance framework. There were many attempts to evaluate the impact of the assumptions about specific types of utility functions on portfolio optimization. However, the original Markowitz mean-variance framework proved to be sufficiently rich to provide the main theoretical background for the analysis of importance of diversification.

Second, MPT gave rise to a variety of asset pricing models for stock pricing, the most known among them being the Capital Asset Pricing Model (CAPM) (Mossin 1966) and the Arbitrage Pricing Model (APM) (Roll and Ross, 1984; King, 1966).

Third, it gave rise to performance measurement techniques designed to evaluate the performance of professional portfolio managers against a benchmark portfolio (Jensen, 1968).

All these three aspects of the MPT led to the notion that the best managed portfolio is the one which is most widely diversified and the sufficient value of such a portfolio may be created through passive buy-and-hold investment strategy. This led to a belief that investors should not expect professional portfolio managers to outperform a well-diversified benchmark, such as the S & P500 in the U.S. or the TSE300 in Canada. However, somewhat restrictive assumptions behind the MTP and CAPM have also resulted in a line of research which has criticized this way of reasoning for portfolio management and performance evaluation. The main thrust of this research was a notion that reliance on the MTP and CAPM for active portfolio management may result in ignoring both the firm (or stock) specific attributes as well as the multi-attribute nature of the investor's preferences. Consequently, the notion of 'tilting' was proposed, which defines a systematic approach to constructing a portfolio which has a higher (or lower) value of a particular attribute than this which can be found in a benchmark portfolio.

Professional managers, as well as sophisticated investors, began evaluating the impact of tilting the portfolio based on attributes other than the mean-variance and systematic risk, believing that the resulting portfolio will allow for superior returns without sacrificing risk. Thus, tilting results in over- or under-investing in stocks which have the same (expected) mean-variance but which have certain values of other attributes deemed as desirable by an investor. In addition, a variety of anomalies in stocks' returns reported in the literature allowed to further question the viability of sole reliance on the MPT and CAPM for portfolio decisions. Researches started also to investigate the impact of investor's characteristics on portfolio decisions. As it is beyond the scope of this paper to summarize the vast amount of literature in this area, we recommend an interesting review by Fama (1991). Nevertheless, it is important to stress that all these studies point to the fact that by choosing stocks with certain values of their attributes, an investor may be able to achieve superior performance compared to the buy and hold

portfolio strategy. The main thrust of the frameworks for such portfolio construction has been to investigate the potential models for and consequences of moving away from the benchmark portfolios or their proxies as represented by various stock indices. The intent has been to provide tools and techniques to allow for deviating from passive portfolios which are considered to be optimal portfolios in the long run. Such deviation is identical to 'tilting' the portfolio for certain attributes in expectation to generate excess returns, or to be more consistent with the investor's preferences. For example, if the investor believes that, all else being equal, he/she would prefer a low dividend paying stock over high dividend paying stock, it implies tilting the portfolio in such a way that it would either include a higher number of low dividend paying stocks or would have a higher investment in the few low dividend paying stocks.

In our analysis we are interested in tilting the portfolios according to some well established price-related stock attributes. The information provided to an investor through an analysis of these attributes can sometimes be confusing, so it is important to acquire a clear picture regarding importance and validity of the stocks' attributes in a tilting process. This can be accomplished by analyzing the contribution of specific price-related attributes to the identification of top performing stocks in terms of their returns. As a by-product of such an analysis it maybe also possible to determine predictive power of specific subsets of price-related attributes. All this required information can be generated by analyzing stock data with the help of the RS methodology which introduces reducts and decision rules – the notions which are very useful in determining attributes' relative importance and predictive power.

3. Rough Set Theory

3.1 Main Notions

The rough set (RS) theory proposed by Pawlak (1991) is concerned with qualitative data analysis. The data set being analyzed consists of *objects* (also *examples*, *cases*) which may represent states, clients, stocks, *etc*., described by values of the *attributes*, representing features, parameters, *etc*. of the objects. The set of attributes is divided into two disjoint subsets, called *condition* and *decision* attributes. The important distinction between these two sets is that the condition attributes express some descriptive information about the objects, while the decision attributes express some decisions or conclusions made about the objects. The set of objects described by attributes and represented in a table form is called *a decision table*.

One of the fundamental notions of the RS theory is the *indiscernibility relation*. Indiscernibility identifies groups of objects which are indiscernible from one another when only some of the attributes are taken into account. The relation is the exemplification of the fact that the values of attributes are the sole source of knowledge about the objects. Indiscernibility is a equivalence relation, so it defines a *partition* of objects into disjoint subsets. The main concern of the RS theory is to examine different partitions induced by sets of condition and decision attributes and the relationship between these two partitions.

Two particular partitions of the objects are of special interest. One of them is the partition induced by the set of all the decision attributes. This partition defines *classes* of objects – namely the sets of objects described by the same values of the decision

attributes. Because often there is only one decision attribute in the table, the classes are sets of objects featuring the same value of the decision attribute. The other partition of interest is induced by the set of all the condition attributes. The elements of this partition, called *atoms*, contain objects which are indiscernible from one another with regard to all of the condition attributes. The name 'atom' is intended to stress the fact that it is possibly the smallest and unsplittable 'granule' of information. Any two objects belonging to the same atom are consequently described by the same set of values and should be treated in exactly the same way in further analysis.

Defining the atoms and classes allows for the next step of the RS analysis in which the equivalence sets induced by the condition attributes (atoms) and the decision attributes (classes) are matched to check if the classes can be expressed by the atoms. Any subset of objects (called a *concept*) is *definable* by some attributes if it may be represented as a union of elements of the partition generated by this set of attributes. In that sense the concept is built of atoms. A very interesting situation arises in case of inconsistency, that is when a concept is not definable by a subset of attributes, and thus it cannot be constructed from the atoms, because there are atoms in the decision table which contain objects both within the concept's definition and outside it. Because the atoms cannot be split any further, the RS theory introduces a notion of approximation and instead of representing the concepts it represents their approximations.

Lower approximation of a concept is a set of objects belonging to atoms which are completely included in the concept's definition. If an atom contains objects of which only some belong to the concept's definition, such an atom constitutes the *upper approximation* of the concept. Final notion of interest to the RS analysis is a boundary region which is defined as the difference between the upper and lower approximations. In the case when inconsistency does not occur and the concepts are definable in terms of atoms, then both the lower and upper approximations of the concept are simply equal to each other. This implies that a boundary region is empty.

The RS theory introduces a measure of inconsistency, called the *quality of approximation*, which is defined as the ratio of the number of all objects belonging to lower approximations of all classes to the number of all objects in the decision table. Maximum value of this measure, being *1.0*, indicates that all the classes may be fully defined using the condition attributes. If the quality of approximation is satisfactory, then it may be interesting to see if there are some subsets of condition attributes which are sufficient to generate the same quality of approximation as a whole set. This notion leads directly to the idea of attributes' reduction.

3.2 Reducts and Their Computation

A *reduct* is defined as a subset of attributes which ensures the same quality of approximation as the original set. In general, it is possible that there is more than one reduct for a given decision table. In that case the set called the *core of attributes* is defined as the intersection of all reducts. Removal of any single attribute included in the core always leads to the drop of the quality of approximation.

Generating reducts is a computationally complex task and normally two classes of reduct generating algorithms are used:

- *Exact algorithms*: designed for generating all reducts from a decision table. Their main disadvantage is the computational complexity that may grow non-polynomially with the number of attributes in the table.
- *Approximate algorithms*: designed for generating single reducts, with the aim of reducing the computing time. This allows for avoiding the non-polynomial computational load but it produces approximate solutions[1] which are not true reducts.

3.3 Decision Rules

A very important part of the RS analysis concerns expressing the dependencies between the values of the condition and the decision attributes. The main task is to find these dependencies and represent them in an easily interpretable manner. One possible way of doing this is through the construction of *decision rules*. A decision rule is expressed as a logical statement consisting of the condition and decision parts. The interpretation of the rule is as follows:

'IF the condition attributes assume values indicated in the condition part of the rule THEN the value of decision attribute(s) is one of those indicated by the decision part'.

Decision rule may be categorized as *consistent* if all objects matching its condition part have the value which is indicated by the decision part of the rule, otherwise a rule is *inconsistent*; and it may be categorized as *exact* if the decision part of the rule contains only one elementary condition, otherwise a rule is *approximate*. Approximate rule is a result of an approximate description of concepts in terms of atoms. It implies that using the available knowledge, it is not possible to decide whether some objects (from the boundary region) belong to a given decision class or not.

Each decision rule is characterized by its *strength*, which is defined as the number of objects satisfying the condition part of the rule (or, in other words, which are *covered* by a rule) and belonging to a given decision class. In the case of approximate rules, the strength is calculated separately for each possible decision class. Stronger rules are often more general, i.e. their condition parts are shorter and less specialized, and they are usually more interesting from an analyst's point of view.

Procedures for generating decision rules from decision tables are mostly based on induction. The existing inductive algorithms use one of the following strategies to produce the set of decision rules:

- generation of a minimal set of rules covering all objects,
- generation of an exhaustive set consisting of all possible rules,
- generation of a set of 'strong' decision rules, consistent or not, each covering relatively many objects, but not necessarily all of them.

An important feature of the RS methodology is that the rules for a given decision class are actually generated not from the concept definition but from its approximations. Lower approximation serves first to produce exact rules for the class, while the

---

[1] An approximate solution is a set of attributes that achieves or exceeds the required degree of discrimination but is not minimal (i. e. it could be further reduced by removing some attributes without violating the discrimination degree) or a set of attributes that does not achieve the required discrimination degree (but its degree is high enough).

boundary regions are subsequently used to generate approximate rules that cover the inconsistencies.

3.4 Classification of New Objects

Decision rules derived from a decision table can be used for recommendations regarding the classes of new objects (*classification*). Specifically, the classification of a new object can be supported by matching its description to one of the decision rules. This matching may lead to one of three situations (Slowinski and Stefanowski ,1994a):

(a) the new object matches exact rules indicating one decision class,

(b) the new object matches exact or approximate rules indicating different decision classes,

(c) the new object does not match any of the rules.

In case of situation (a) the recommendation is univocal. In the case of ambiguous matching (situation (b)), the final classification is made on a basis of rule strengths. For each class which is indicated by any of the rules, a coefficient called *class support* defined as the total strength of all the rules supporting this class is computed. The decision class featuring the highest value of the class support is assigned as a class for the new object.

In the case when no rule matches a new object (situation (c)), an object may remain unclassified. However, some other method (such as the *default class classifier*) for predicting an object's class may be used in this case. The default classifier assigns an object to a selected class, so called *default class* (each object is assigned the same class). Usually the most frequently occurring class in a decision table is selected to be a default class, in which case a classifier is referred to as *majority class classifier*.

In summary, the process of classifying objects may actually consist of two phases, the first of them being classification using the rules, while the second involves handling the still unclassified objects. These objects may be assigned class membership according to, for example, a majority class classifier.

3.5 Validation Tests

Proper validation of the decision rules generated for a set of objects should be provided by the domain expert. However, it could also be done by classifying new objects for which the classes are already known. When no new objects are available, the rules may be still evaluated using *validation tests*. In a validation test, a subset of objects (*testing sample*) is selected randomly and removed temporarily from the set of all objects while the decision rules are being generated for this reduced set of objects (*learning sample*). Subsequently, objects from a testing sample are classified, and their real classes are compared to those established by the decision rules. As a result, a percentage of incorrectly classified objects (*error ratio*) is computed. It should be stressed that the error ratio obtained through validation tests may be only considered as an approximation of the real error ratio of the rules.

Validation tests may be employed with any classifying method, but if used with the decision rules, apart from generating ratios of correctly classified objects (*accuracy of classification*) and incorrectly classified objects (error ratio), they also generate ratio of unclassified objects.

The major validation tests can be categorized as:

- *Random Tests*. The testing sample is selected and classified only once, with its size being set arbitrarily to a constant percentage of all objects (e.g. to 33%), or generated randomly. Generally, random validation tests should be used with large data sets.

- *N-fold Cross-Validation Tests*. The set of all available objects is initially divided into *N* disjoint subsets, called *folds*. The learning and classifying phases are then conducted *N* times, with each of the folds acting successively as the testing sample and all the remaining folds as the learning sample. The main idea behind the cross-validation test is that every object is classified exactly once and it serves to generate the decision rules *N-1* times. The final results of the *N* fold cross-validation test are given as averages of the *N* individual tests. This type of test should be used with medium-sized data sets.

- *Leaving-1-out Tests*. It is a variation of the *N*-fold cross-validation test in which *N* is set to the number of all objects. That means that in every iteration only 1 object is excluded from the learning sample and this object is subsequently classified by the generated classifier (at each time it is a different object). This test should be applied to small data sets (below one hundred of objects).

3.6 Discretization of Continuous Attributes

The discernibility relation of the RS may be applied only if the values of the attributes are discrete, as even very small differences in continuous values might affect calculation of the atoms. To prevent this from happening, the continuous attributes in the decision table should be discretized. As a result of discretizing, precision of the original data is decreased (in the sense that the original values of the attributes cannot be reconstructed on a basis of the discrete values), but its generality is increased.

It should be stressed also that discretization of continuous values is embedded in human reasoning. For example, a doctor evaluating body temperature rarely operates on actual value of the measurement, but instead considers discretized values such as 'low', 'medium' or 'high'.

Most typical discretization information (called a hard discretization) consists of a finite set of numbered subintervals defined over the range of values of continuous attributes. This type of discretization is also referred to as *norms*, because the subintervals are frequently defined following some norms in the subject domain. The subintervals are used to discretize the continuous values by substituting an original value with a number of an interval to which it belongs. A more advanced form of discretization involves subintervals represented as fuzzy numbers with overlapping bounds. This fuzzy form of discretization requires different, usually more advanced techniques for processing the discretized decision tables (Slowinski and Stefanowski, 1994a).

When the subintervals for the discretization are specified by a domain expert following his/her judgement or using some norms established in the subject domain, then they are called *expert discretizations*. On the other hand, when they are defined automatically,

then they are called *automatic discretizations* (for review of the automatic discretization procedures see for example, Chmielewski and Grzymala-Busse (1995), Dougherty *et al*. (1995)).

3.7 Software Implementation of the RS Methodology

All the computations reported in this paper were conducted using the *ProFIT* computer system (Mienko *et al*., 1996) which implements the generalized RS methodology (Slowinski and Stefanowski, 1994a). The functions incorporated in the *ProFIT* system include:

- discretization of continuous attributes,
- building approximations,
- generating reducts,
- generating decision rules,
- classifying new objects,
- conducting cross-validation tests.

3.7.1 Input Data, Discretization and Creation of Approximations

The *ProFIT* system accepts both discrete and continuous input data. To handle the continuous attributes the system implements a way to store discretizations in either hard or fuzzy form. In the latter case the discretizing process usually leads to non-unique assignment of continuous values to intervals (this is because the fuzzy intervals may overlap and a continuous value of an object may be positioned within the overlapping region, which results in this value being assigned simultaneously to two subintervals). The generalized RS methodology implements a technique of splitting the objects to overcome the problem of multiple assignment and allows for imperfect input data.

ProFIT includes an automatic method of generating discretizations according to a supervised hard discretization algorithm first introduced by Nguyen and Skowron (1995). The method conducts a breadth-first search for the cutting points in the domains of the attributes to discriminate globally objects of different classes. At every step the best cut on all the attributes is chosen. The original algorithm terminates after all possible pairs of objects from different classes have been discriminated, which corresponds to the quality of classification equal to *1.0*. The actual procedure implemented in *ProFIT* has been modified to be quality driven, which means that it is terminated after an arbitrarily specified value of quality is reached. Another important modification concerns the breadth-first search combined with some randomness in choosing the cuts. This modification deprives the original procedure of its strictly "greedy" character and allows for more diversified discretization to be accomplished.

After all continuous condition attributes are discretized, *ProFIT* allows for creating the approximations of classes including: computing atoms, computing lower and upper approximations, and calculating quality of approximation.

3.7.2 Reduct Generation

ProFIT implements two exact and one heuristic method of generating reducts.

- An exact method for generating all reducts introduced by Skowron and Rauszer (1992). The algorithm is based on the notion of the discernibility matrix that is computed for all pairs of objects in the decision table. The reduct computing procedure utilizes the fact that the discernibility matrix, after being created, may be substantially reduced in size using the law of absorption. As a result, the computing time may drop considerably, thus making the reduct computation very efficient even with non-trivial data sets.

- An exact method for generating all reducts based on scanning all supersets of the core of the attributes. This method should normally be used for smaller data sets, but may also prove effective in case of decision tables in which the core includes relatively many attributes. The method may also be applied for searching for pseudo-reducts, that is subsets of attributes that guarantee the quality of classification equal to or higher than an arbitrarily specified value.

- A heuristic method for generating single reducts based on the forward selection principle. The method creates a candidate reduct using core of the attributes in a repetitive manner. In every iteration, the quality of approximation for the candidate reduct is calculated, and, if it is still too low, each of the attributes not included in candidate reduct is appended to it temporarily and the quality is recalculated. Finally, the attribute which gave the highest increase of the quality is added to the candidate reduct and the calculations are repeated. The method terminates as soon as the quality of approximation reaches the required value.

3.7.3 Rule Generation

ProFIT implements two procedures for rule generation.

- 'LEM2' by Grzymala-Busse (1992) is a machine learning based heuristic which seeks to generate the minimal number of rules that cover all the objects. In each iteration, a rule which can cover the biggest number of objects is produced and the objects being covered by this rule are removed from a learning sample. The iterations continue until the learning sample is empty, which signifies that all of the objects are covered by at least one of the rules.

- 'All Rules' by Stefanowski and Vanderpooten (1994) is a knowledge acquisition based algorithm. In its standard version it generates all possible rules for a given data set. However, its computing effectiveness increases rapidly when some constraints are imposed on the rules which are to be generated. The first one involves a rule length upper limit, so as a result only the rules which are shorter than the specified limit are generated. The second one is a rule strength lower limit, so only the rules which are stronger than the specified limit are generated. It is also possible to generate rules satisfying both constraints.

3.7.4 Classifying New Objects

The following automatic classification procedures are employed in *ProFIT* for classification of new objects:

- classification using the rules,

- classification using the rules and the Valued Closeness Relation (VCR) (Slowinski and Stefanowski, 1994b),

- classification using the rules and the default class classifier.

Classifying by the rules remains the main classification method, however, it is possible to use the VCR or the default class classifier at the second stage for the remaining unclassified objects. To test the quality of rules, *ProFIT* implements three types of cross-validation tests discussed in Section 3.5.

4. Case Study: Analyzing Stocks on the Toronto Stock Exchange

The data set analyzed in this study includes all publicly traded companies listed on the Toronto Stock Exchange (TSE) available in version 4.6 of the Stock Guide database. The rate of return for the stocks under consideration and the market indices were drawn from the return and index files of the TSE/Western database. To qualify for inclusion in the analysis for a given year, a firm had to meet the following requirements:

1) the firm must belong to an industry group other than "Gold and Precious Metal"; "Mining Exploration without Production"; "Utilities"; "Financial Services including Insurance",

2) the firm must be listed on the TSE,

3) the firm's common stock must be traded at a yearly closing price equal to or above CAD $1.50 in 1985 constant dollars,

4) the firm must have a total annual revenue higher than CAD $1.5 million in 1985 constant dollars,

5) the firm must have a year-end total asset base larger than CAD $3.0 million in 1985 constant dollars,

6) the firm must have defined total annual revenue, year-end price, year-end total assets and other relevant data available in the Stock Guide database as per requirements (2), (3), and (4) above and for calculating price-related stock attributes,

Requirement (1) excludes firms in those industry groups where price-related attributes may not be easily interpreted. Requirements (2), (3), and (4) impose size constraint on the firms, to ensure that the considered companies are large enough to be frequently researched and considered by investment professionals. In addition, these requirements control potential biases that characterize very small firms. The requirements (5) and (6) ensure that meaningful data are available. Although a number of stocks under consideration are reduced by enforcing the above requirements, a net benefit associated with clarity of the data set is considerable.

4.1 Experimental Design

The data set used in the experiment includes stocks' information of the TSE traded firms for the period 1989-93. Each stock was described using 11 attributes, including annual rate of return. The list of the attributes is given in Table 1.

Attribute name	Attribute description
ANR	Annual rate of return
RE_Q4	Relative earnings past 4 quarters
RE8_Q4	Relative earnings past 8 quarters
RP8_Q4	Relative price past 8 quarters
PMR_Q4	Price momentum rank
UE_Q4	Unexpected quarterly earnings over the same quarter last year
QPR_Q4	Price at the end of 4th quarter
PBR_Q4	Price to book ratio
MC_Q4	Market capitalization
PSR_Q4	Price to sales ratio
QEM_Q4	Quarterly earnings momentum

Table 1. Attributes and their description

Since financial theory does not provide strong guidance for the identification or construction of the specific price-related attributes, the selection given in Table 1 was directed by existing evidence for the US market and was influenced by its popularity among practitioners of "attribute investing". The annual rate of return (ANR) was treated in our study as a decision attribute used to determine *ex post* evaluation of stocks' performance. All remaining attributes were treated as condition attributes for which data was collected for the fourth quarter of every year. Table 2 includes information on the number of stocks (objects) meeting the requirements for inclusion in the analysis, for which data was available for every year during a five year period of the analysis.

Year	Number of stocks
1989	346
1990	345
1991	364
1992	362
1993	405

Table 2. Number of stocks analyzed in consecutive years

4.1.1 Discretization

The condition attributes were discretized using standard norms which were specified by a financial expert. A decision attribute (ANR) was discretized by dividing stocks into performance groups depending on the values of the ANR. The top 25% performers were categorized as TOP group, bottom 25% performers were categorized as BOTTOM group, and all remaining stocks were assigned to a MIDDLE group. Determination of the TOP or BOTTOM groups as a sharp 25% cut of all stocks proved to be too categorical, hence a technique of adjusting the groupings was introduced. Adjusting was especially required when the sharp 25% border of a group happened to appear in the middle of a streak of ordered stocks featuring the same value of the ANR. In this case the actual border of a group needed to be shifted so as to prevent splitting objects with the same value of ANR into two different groups.

4.2 The Analysis

4.2.1 Establishing the Class Definitions

The stocks outperforming the others in terms of the ANR (TOP group) are of particular interest to the investors, therefore the class definitions used in the RS analysis should be carefully scrutinized in order to examine their influence on the results produced by the *ProFIT* system. Four schemes of class definitions were considered:

- 3-class cut: TOP, MIDDLE and BOTTOM,
- 2-class cut: TOP and REST,
- 4-class cut: TOP, BUFFER, MIDDLE and BOTTOM,
- 3-class cut: TOP, BUFFER and REST,

where class REST includes stocks normally belonging to the MIDDLE and BOTTOM groups, while BUFFER class contains stocks from the TOP group which form an immediate boundary of that group with a MIDDLE group. With the distinction of the BUFFER class, the TOP class was further contracted to include just top stock performers. Table 3 presents exemplary results of the analysis of impact of class definitions on rule parameters and the accuracy of classification, applied to the 1989 data set. The column 'Rule strength' presents the average strength for rules generated by the LEM2 algorithm, and the column '10-fold CV' shows the results of applying the 10-fold Cross-Validation test.

Class definitions				Rule strength	10-fold CV
Top	Middle	Bottom	----------	3.0	54.9
Top	Rest	----------	----------	5.7	69.9
Top	Buffer	Middle	Bottom	3.5	49.7
Top	Buffer	Rest	----------	5.3	72.3

Table 3. Exemplary results of computations designed to examine the impact of class definitions

Analysis of the above results suggests that the most appropriate description of the decision table is accomplished when the classification scheme involving TOP, BUFFER and REST classes is used.

4.2.2 Final Results

Table 4 presents the final results derived at for a whole time period under consideration. The columns 'Rule strength' and '10-fold CV' are the same as defined in Table 3, while the column 'Self-test' gives the percentage of correctly classified stocks when rules are applied to a learning sample. In the columns '# rules' and '# reducts' the numbers of rules and reducts generated from the data sets respectively, are given.

Year	# rules	Rule strength	Self-test	10-fold CV	# reducts
1989	96	5.3	96.5	72.3	2
1990	72	8.4	99.7	79.1	3
1991	105	5.2	97.5	68.1	1
1992	85	7.0	98.9	72.9	1
1993	113	6.0	100.0	66.4	1

Table 4. Final results of computations for the selected classification

The results in Table 4 confirm good performance of the rules in classifying the stocks. This is particularly evident while looking at the very good results of the self-test and the 10-fold CV test (the average of self-test results for consecutive years is 98.5% and the average of cross-validation tests is 71.8%). The results of a self-test should be considered despite the fact that the rules are applied to a learning sample. In reality the rules that would have been unable to correctly classify stocks (objects) from which they were generated would be assigned only very limited credibility by the experts, making the results of self-test the first plausible step of rules' verification.

The results of the 10-fold CV tests are probably the most important in assessing quality of the rules as classifiers. In cross-validation tests, the rules are applied to classify objects not belonging to a learning sample, so these results are considered to be one of the best approximations of the real-life performance of the rules. Again, the results reported here are very promising.

The low number of reducts (see column "# reducts") for each year indicates that there is a distinct disproportion in the importance of condition attributes while considering their ability to approximate the classes. This observation supports the idea of attributes' reduction, indicating that some of the attributes may be discarded from further considerations without much impact on the quality of classification. The attributes to be discarded are those which do not appear in the reducts.

4.2 Idea of a Common Reduct

In this study we are interested in finding out if it is possible to tilt the portfolio using as few of the stocks' attributes as possible. In that sense, information provided by the reducts becomes very important. Reducts calculated for each year include the subsets of attributes that ensure the same quality of classification as the original set. However, deciding about the number of attributes to be used in a general problem of portfolio tilting just on a basis of the reducts is not a very easy task due to the following reasons:

- usually there are more than one reduct per year,
- there is no consistency among the reducts for different years.

It is seldom possible to generate a single reduct that would be the same for all years under consideration. In that case an artificial single reduct needs to be created with the help of some heuristic. We call a reduct created in such a way a "Common Reduct (CR)".

One possible method of creating the CR is to consider jointly all of the reducts identified by the *ProFIT* system, and computing for every attribute a frequency of occurrence in those reducts. Subsequently, a histogram may be used to analyze the frequencies, and the attributes most frequently occurring in the reducts should be included in the CR.

Table 5 gives the frequencies of all the condition attributes in the reducts and Figure 1 presents the histogram of those values.

Attributes' name	Frequency
Relative earnings past 4 quarters	6
Relative earnings past 8 quarters	7
Relative price past 8 quarters	4
Price momentum rank	8
Unexpected quarterly earnings over the same quarter last year	8
Price at the end of 4th quarter	5
Price to book ratio	8
Market capitalization	7
Price to sales ratio	8
Quarterly earnings momentum	5

Table 5. Frequencies of the attributes in reducts

An analysis of the histogram suggests that the attributes that occur at least 7 times in the reducts should be included in the CR. Thus, according to this heuristic, the CR (denoted further as CR1) consists of the following condition attributes: relative earnings in past 8 quarters (RE8), price momentum rank (PMR), unexpected quarterly earnings over the same quarter last year (UE), price to book ratio (PBR), price to sales ratio (PSR), and market capitalization (MC).

A potential problem with the application of heuristic described above to calculating the CR is that a number of reducts in different years is uneven, thus contributing to overweighing years with bigger number of reducts. To account for that situation, an alternative, two-phase method of establishing the CR was tested. In the first phase the reducts generated for each year were analyzed using an entropy measure and one for every year was selected. The second phase involved application of the heuristic described above, but only to the five reducts identified in phase one. Table 6 and Figure

2 give frequency information and a histogram of these frequencies for the attributes analyzed according to the alternative procedure.
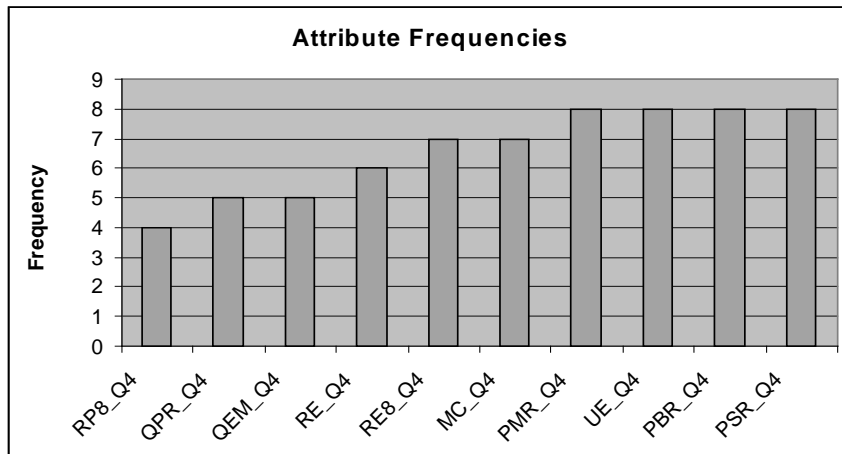

Figure 1. Histogram of attributes' frequencies

Attributes' name	Frequency
Relative earnings past 4 quarters	3
Relative earnings past 8 quarters	5
Relative price past 8 quarters	3
Price momentum rank	5
Unexpected quarterly earnings over the same quarter last year	5
Price at the end of 4th quarter	3
Price to book ratio	5
Market capitalization	4
Price to sales ratio	5
Quarterly earnings momentum	4

Table 6. Frequencies of attributes according to the alternative procedure

The above results suggest that the following attributes should be included in the CR (labelled as CR2 to distinguish it from the CR1 formed according to the first heuristic): relative earnings in past 8 quarters (RE8), price momentum rank (PMR), unexpected quarterly earnings over the same quarter last year (UE), price to book ratio (PBR), price to sales ratio (PSR), market capitalization (MC), and quarterly earnings momentum (QEM).

Both CR1 and CR2 are very similar, the only difference being the attribute QEM (which is contained in CR2 and not in CR1).
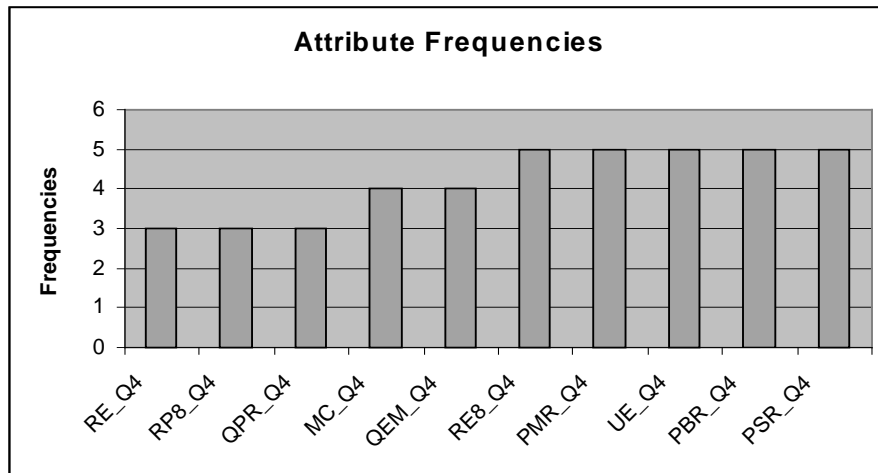
Figure 2. Histogram of attributes' frequencies after applying alternative procedure

4.3 Application of the CR to Describe Tilting

After creating the CR, its classification performance needs to be verified by comparing it with the situation when all the attributes are used. Tables 7 and 8 give results of computations when only the CR1 and the CR2 are used to describe portfolio tilting.

Year	# rules	Rule strength	Self-test	10-fold CV
1989	101	4.3	94.2	71.1
1990	87	6.3	98.6	79.4
1991	111	4.2	93.1	69.2
1992	103	5.2	95.9	72.9
1993	123	5.1	96.8	71.1

Table 7. Computational verification of the CR1

Information about the attributes is collected separately for every year. Therefore, classifying stocks one or two years ahead using rules generated for a previous time period may provide a good verification mechanism of the predictive quality of the CR attributes. The results of such a classification are given in Tables 9 and 10.

Year	# rules	Rule strength	Self-test	10-fold CV
1989	97	4.6	94.8	69.1
1990	86	7.3	99.7	75.7
1991	116	4.6	95.3	70.1
1992	98	6.0	98.1	73.5
1993	124	4.8	97.8	69.1

Table 8. Computational verification of the CR2

Year	Year + 1	Year + 2
1989	75.4	74.0
1990	68.4	66.0
1991	63.8	59.0
1992	69.6	-----

Table 9. Verification of the CR1 looking 1 and 2 years ahead

Year	Year + 1	Year + 2
1989	72.5	67.6
1990	61.0	55.5
1991	66.6	66.4
1992	72.3	-----

Table 10. Verification of the CR2 looking 1 and 2 years ahead

The results presented in Tables 9 and 10 indicate good predictive quality of the decision rules generated using only the CR attributes. When applied to predict stock membership in the top performers group in the next year, a correct assignment was made on average in almost 70% of cases. It should be stressed that this result is achieved with a reduced attributes' set, thus allowing potential investors to collect and analyze smaller amounts of data. Moreover, results in Tables 9 and 10 confirm that predictive quality of the decision rules diminishes as a time horizon extends, and suggests using a well defined time frame in the analysis.

5. Discussion

The purpose of this research was to analyze price-related attributes used normally in portfolio tilting. Due to a possible number of the attributes involved in a process, it is always useful to determine their contribution to identification of top stock performers.

The RS analysis provides tools for such identification, allowing also for a construction of the decision rules which might be applied to the evaluation of new stocks. Quality of these rules indirectly validates the usefulness of reducing the original set of stocks' attributes. Although results of cross-validation tests served as one of the main measures in the analysis, our study was not concerned with establishing the best classifying system. Instead, the positive results of cross-validation suggest that the portfolio tilting may be conducted with equal efficiency when a smaller number of price-related attributes is evaluated. This in turn provides information to potential investors regarding which price-related attributes are more significant in designing successful investment strategies.

The methodology applied by us is not unique for reducing a number of attributes for portfolio tilting, however, the results support continuation of the research on using the RS and machine learning techniques for describing and analyzing complex decision problems, such as stock portfolio tilting decisions.

References

Chmielewski M. and Grzymala-Busse J. (1995). 'Global Discretization of Continuous Attributes as Preprocessing for Machine Learning'. In: Lin T.Y. and Wildberger A.M. (eds.), *Soft Computing*, Society for Computer Simulation, San Diego, pp. 294–301.

Dougherty J. Kohavi R. and Sahami M. (1995). 'Supervised and Unsupervised Discretizations of Continuous Features'. In: *Proceedings of the 12th International Conference on Machine Learning*, Morgan Kaufmann Publishers, New York, pp. 194–202.

Fama E.F. (1991). 'Efficient Capital Markets II'. *Journal of Finance*, Vol. 46, No. 5, pp. 1575–1617.

Grzymala-Busse J. (1992). 'LERS - A System for Learning from Examples Based on Rough Sets'. In: (Slowinski, 1992), pp. 3–18.

Jensen M.C. (1968). 'The Performance of Mutual Funds in the Period 1945-1964'. *Journal of Finance*, Vol. 23, pp. 389–416.

Jog V. and Li B. (1995). "Price Related Anomalies on the Toronto Stock Exchange". Working Paper WP95-12, School of Business, Carleton University.

King B.F. (1966). 'Market and Industry Factors in Stock Price Behaviour'. *Journal of Business*, Vol. 39, pp. 139–190.

Markowitz H.M. (1959). *Portfolio Selection*. J. Wiley, New York.

Mienko R., Slowinski R., Stefanowski J. and Susmaga R. (1996). 'Rough Family — Software Implementation of Rough Set Based Data Analysis and Rule Discovery Techniques'. In: Tsumoto S., Kobayashi S., Yokomori T., Tanaka H. and Nakamura A. (eds.), *Proceedings of the 4th International Workshop on Rough Sets, Fuzzy Sets and Machine Discovery*, Tokyo University Press, Tokyo, pp. 284–291.

Mossin J. (1966). 'Equilibrium in a Capital Asset Market'. *Econometrica*, Vol. 34, pp. 768–783.

Nguyen S.H. and Skowron A. (1995). 'Quantization of Real Value Attributes: Rough Set and Boolean Reasoning Approach'. In: *Proceedings of the Second Joint Annual Conference on Information Sciences*, Society for Information Processing, Wrightsville Beach, pp. 34–37.

Pawlak Z. (1991). *Rough Sets. Theoretical Aspects of Reasoning About Data*. Kluwer Academic Publishers, Dordrecht.

Roll R. and Ross S. (1984). 'The Arbitrage Pricing Theory Approach to Strategic Portfolio Planning'. *Financial Analyst Journal*, No. 3, pp. 14–26.

Skowron A. and Rauszer C. (1992). 'The Discernibility Functions Matrices and Functions in Information Systems'. In: (Slowinski, 1992), pp. 331–362.

Slowinski R. (ed.) (1992). *Intelligent Decision Support. Handbook of Applications and Advances of the Rough Set Theory*. Kluwer Academic Publishers, Dordrecht.

Slowinski R. and Stefanowski J. (1994a). 'Handling Various Types of Uncertainty in the Rough Set Approach'. In: Ziarko W. (ed.), *Rough Sets, Fuzzy Sets and Knowledge Discovery*, Springer Verlag, London, pp. 366–376.

Slowinski R. and Stefanowski J. (1994b). 'Rough Classification with Valued Closeness Relation'. ICS Research Report 25/94, Institute of Computer Science, Warsaw University of Technology.

Stefanowski J. and Vanderpooten D. (1994). 'A General, Two-Stage Approach to Inducing Rules from Examples'. In: Ziarko W., (ed.), *Rough Sets, Fuzzy Sets and Knowledge Discovery*, Springer Verlag, London, pp. 317–325.