General Model for Automated Diagnosis of Business Performance

Emiel Caron and Hennie Daniels

ERIM REPORT SERIES *RESEARCH IN MANAGEMENT*	
ERIM Report Series reference number	ERS-2005-058-LIS
Publication	October 2005
Number of pages	33
Persistent paper URL	
Email address corresponding author	ecaron@rsm.nl
Address	Erasmus Research Institute of Management (ERIM) RSM Erasmus University / Erasmus School of Economics Erasmus Universiteit Rotterdam P.O.Box 1738 3000 DR Rotterdam, The Netherlands Phone: + 31 10 408 1182 Fax: + 31 10 408 9640 Email: info@erim.eur.nl Internet: www.erim.eur.nl

Bibliographic data and classifications of all the ERIM reports are also available on the ERIM website:
www.erim.eur.nl

ERASMUS RESEARCH INSTITUTE OF MANAGEMENT

REPORT SERIES

RESEARCH IN MANAGEMENT

ABSTRACT AND KEYWORDS	
Abstract	In this paper, we describe an extension of the methodology for explanation generation in financial knowledge-based systems, offering the possibility to automatically generate explanations and diagnostics to support business decision tasks. The central goal is the identification of specific knowledge structures and reasoning methods required to construct computerized explanations from financial data and business models. A multi-step look-ahead algorithm is proposed that deals with so-called calling-out effects, which are a common phenomenon in financial data sets. The extended methodology was tested on a case-study conducted for Statistics Netherlands involving the comparison of financial figures of firms in the Dutch retail branch. The analyses are performed with a diagnostic software application which implements our theory of explanation. Comparison of results of the classic explanation methodology with the results of the extended methodology shows significant improvements in the analyses when cancelling-out effects are present in the data.
Free Keywords	Decision support systems, Finance, Production statistics, Artificial intelligence, Explanation
Availability	The ERIM Report Series is distributed through the following platforms: Academic Repository at Erasmus University (DEAR), DEAR ERIM Series Portal Social Science Research Network (SSRN), SSRN ERIM Series Webpage Research Papers in Economics (REPEC), REPEC ERIM Series Webpage
Classifications	The electronic versions of the papers in the ERIM report Series contain bibliographic metadata by the following classification systems: Library of Congress Classification, (LCC) LCC Webpage Journal of Economic Literature, (JEL), JEL Webpage ACM Computing Classification System CCS Webpage Inspec Classification scheme (ICS), ICS Webpage

General model for automated diagnosis of business performance

Automated interfirm comparison: Case-study for Statistics Netherlands

Emiel Caron Hennie Daniels

Erasmus Research Institute in Management (ERIM), Erasmus University Rotterdam, PO Box 90153,
3000 DR Rotterdam, phone +31 010 4082574, e-mail: ecaron@rsm.nl
The Netherlands

Abstract

In this paper, we describe an extension of the methodology for explanation generation in financial knowledge-based systems, offering the possibility to automatically generate explanations and diagnostics to support business decision tasks. The central goal is the identification of specific knowledge structures and reasoning methods required to construct computerized explanations from financial data and business models. A multi-step look-ahead algorithm is proposed that deals with so-called calling-out effects, which are a common phenomenon in financial data sets. The extended methodology was tested on a case-study conducted for Statistics Netherlands involving the comparison of financial figures of firms in the Dutch retail branch. The analyses are performed with a diagnostic software application which implements our theory of explanation. Comparison of results of the classic explanation methodology with the results of the extended methodology shows significant improvements in the analyses when cancelling-out effects are present in the data.

Keywords: Decision support systems; Finance; Production statistics; Artificial intelligence; Explanation

1 Introduction

The diagnostic process for interfirm comparison (IFC) or competition benchmarking is now carried out manually by (business) analysts, where the analyst has to explore large data sets in the domain of business and finance to spot firms that expose exceptional behaviour compared to some norm behaviour. After that the analyst has to find the reasons, the set of financial variables responsible, behind the firm's exceptional behaviour. This diagnostic process is fully automated and implemented in a computer program, based on an explanation formalism described in this paper, to support human decision makers.

Here diagnosis is defined as *finding the best explanation of observed symptoms of a system under study*. This definition assumes that we know which behaviour we may expect from a correctly working system. The expected behaviour or norm model can be derived from some statistical model or can be expert knowledge from financial analysts. Two important consecutive phases in a diagnostic process are *problem identification* (or symptom detection) and *explanation generation* [11]. When a discrepancy between actual and norm behaviour is discovered, and is qualified as unacceptable with respect to some specified norm, the next step is to explain this discrepancy using our "understanding" of the system.

The rationale behind this paper is to improve and extend the methodology for automated business diagnosis and its software implementation, and to deal specifically with some tricky problems of the classic explanation methodology. Therefore, we propose a number of improvements and further develop the methodology. Firstly, a more sophisticated method for symptom detection is presented in this paper that takes into account the probability distribution of the variable under consideration. The detection of symptoms for computerized diagnosis in financial data is not fully developed in earlier methods [1, 2], where it is described as a rather "crude" method to filter out symptoms. In this method symptom detection is the simple process where the difference value is taken between the actual and norm value of each variable. If this value is below or above some specified threshold, a symptom is added to the list of symptoms.

Secondly, in this paper we extend the explanation methodology as described in [1, 2], with a procedure to deal with so-called *cancelling-out* or *neutralisation effects* in data sets. A neutralisation effect is the phenomenon that the effects of two or more lower-level variables may cancel each other out in a system of equations, so that their (joint) influence on a higher-level variable is partially or fully neutralized. For example, the first half-year positive financial results could partially cancel out the negative financial results of the next six months. If one starts diagnosis with the method described in [1, 2] on the aggregated year level these effects are not identified. However, these effects are quite common in financial data and other data sets and could lead to erroneous results in the form of incorrect or incomplete explanation trees. Therefore, a substitution method is introduced that can take possible cancelling-out effect into account in all levels of the underlying business model.

Finally, an older version of the diagnostic program was implemented in a constraint logic programming language called CHIP, similar to the well-known PROLOG language [2]. This type of implementation has some advantages in terms of (declarative) knowledge representation, but it also has some clear disadvantages in terms of applicability in an office environment and presentation of program output. To deal with these disadvantages we have implemented the extended explanation model in MS Excel in combination with Visual Basic (VB), a software package which is present in most offices. In the older implementation only text generation of output results is possible and is rather inflexible. Therefore, we present the output as *trees of causes*: representing symptoms and causes graphically in a so-called

Windows “folder structure”. In this way the analyst can intuitively access the results of the explanation generation process.

1.1 Explanation model

Today’s systems for automated financial diagnosis and interfirm comparison have little explanation or diagnostic capabilities. Such functionality can be provided by extending these systems with an explanation formalism, which mimics the work of human analysts in diagnostic processes. Our exposition on diagnostic reasoning and causal explanation is largely based on Feelders and Daniels’ notion of explanations in [1, 2], which is essentially based on Humpreys’ notion of aleatory explanations [10] and the theory of explaining differences by Hesslow [8]. Where causal influences can appear in two forms: *contributing* and *counteracting*. The canonical format for causal explanations is taken from [1, 2]:

$$\langle a, F, r \rangle \text{ because } C^{+}\text{, despite } C^{-}\text{.}$$

where $\langle a, F, r \rangle$ is the event to be explained, C^{+} is non-empty set of contributing causes, and C^{-} a (possibly empty) set of counteracting causes. The explanation itself consists of the causes to which C^{+} jointly refers. C^{-} is not part of the explanation, but gives a clearer notion of how the members of C^{+} actually brought about the event. In words, the explanandum is a three-place relation between an object *a* (e.g. the ABC-company), a property *F* (e.g. having a low profit) and a reference class *r* (e.g. other companies in the same branch or industry). The task is not to explain why *a* has property *F*, but rather to explain why *a* has property *F when the members of r do not*. This general formalism for explanation constitutes the basis of our extended framework for diagnosis in financial models developed in this paper.

1.2 Related work

A lot of literature is available about medical diagnosis and diagnosis of technical devices, see [2, 11] for an overview. However the literature discussing explanation and diagnosis in quantitative financial or accountancy models is not very large. In [11], a comparison is made between automated financial diagnosis and the other diagnostic domains. Several approaches have been proposed for the automatic generation of explanations based on financial models [1, 2, 6, 10]. To position this paper we discuss some related work.

Daniels and Feelders [1, 2] describe a formal framework for explanation and diagnosis of company performance with both qualitative and quantitative information. For the construction of explanations, the canonical format of explanations (see paragraph 1.1) is adapted to the requirements of the business domain. Their method reduces the sets of contributing and counteracting causes to parsimonious sets, to avoid the inclusion of insignificant causes. Furthermore, they developed an influence measure, to determine contributing and counteracting causes, which embodies a kind of ceteris paribus reasoning. As stated in the introduction their method does have some fundamental problems. However, because of its importance for this research the theory and methodology of Daniels and Feelders is elaborated upon in section 2.

Kosy and Wise [10] describe a general system for generating explanations in financial models, not directed specifically at diagnostic problem solving. Their algorithm explains any difference between two values of a variable, as long as these values have been generated by the same equation. However, no strict separation is made between contributing and counteracting causes, which leads to counterintuitive results in some cases and it may cause the system to leave out significant causes from the explanation.

Courtney et al. [6] describe a decision support system directed specifically at managerial problem diagnosis. Functional relations that are allowed to sustain explanations are restricted to linear functions however. The restriction to linear relationships is not very realistic however in a financial context. A clear distinction is made in their system between contributing and counteracting influences. The system is not fully automatic however, in the sense that the user has tot decide which of the complete list of influences presented in considered relevant.

This paper is organised as follows. In the next section we first review the explanation model as described in literature and subsequently introduce extensions for the explanation model. In section 3 the extensions are illustrated by an extensive case study on interfirm comparison with financial data collected at Statistics Netherlands. In the case study we compare two explanations, in the form of trees of causes, for detected symptoms derived from companies in the Dutch retail industry. One tree is generated with the classic explanation methodology and the other tree with our new extended methodology. In section 4 we describe the software implementation and architecture of the diagnostic program. Finally, we draw a number of conclusions in section 5. In Appendix A the complete list of variables used in the case study is included and in Appendix B the actual data for interfirm comparison is added. In addition, Appendix C contains additional figures referenced in the paper.

2 Explanation model

Diagnosis of business performance is defined in [1, 2] as explaining the difference between the actual performance of a company, and its norm performance. In the explanation model, *a* refers to the object that shows the actual behaviour and *r* refers to the object that shows the norm behaviour. Two principal knowledge structures for diagnosis of business performance are identified:

- Knowledge of general laws, relating variables pertaining to business performance: *the business model*.
- Knowledge of normal behaviour: the *normative model*.

In this section we present a summary on the general theory and methodology for automated business diagnosis. Furthermore, an extension is made to the theory on symptom detection.

2.1 The business model

Feelders and Daniels state that explanations are usually based on general laws expressing relations between events, such as cause effect relations or constraints between variables. These laws are represented in a business model *M*. The model *M*, which is a form of domain knowledge, can be derived from many domains. The business model *M* represents quantitative financial and operating variables by means of mathematical equations of the form:

$$y = f(\mathbf{x}) \text{ where } \mathbf{x} = (x_1, \ldots, x_n) .$$

In paragraph 3.2, an example is given of a business model used by Statistics Netherlands for gathering production statistics in the retail branch. The quantitative model is used to propagate both deviating and non-deviating values.

A directed graph, the *explanatory graph* $E(M) = (\mathcal{V}, \mathcal{E})$, is associated with the business model *M*. The vertex set $\mathcal{V}$ contains as elements all variables appearing in the model. The edge set $\mathcal{E}$ contains a directed edge from vertex x_i to x_j iff: $x_j = f(\ldots, x_i, \ldots) \in M$. A restriction is placed on the model *M* to exclude cycles in the explanatory graph $E(M)$. In Fig. 1 the explanatory graph of the business model in paragraph 3.2 is depicted.

[Please insert Figure 1 about here]

The arcs between the nodes, which represent the variables in the business model, indicate the direction of influence, or causal direction. Interpreting the = in the equations of model *M* as a $\leftarrow$ gives the causal direction as used by economists, accountants or financial analysts. Thus, in the model the effects appear on the left-hand side (LHS) of the equations and the causes on the right-hand side (RHS). However, as we shall see, the diagnostic reasoning direction is the reverse of the causal direction. In other words, the explanation generation process takes part from the whole (the LHS variables) to the parts (the RHS variables). For example, in the reasoning process the LHS of equation 1 (results before taxation) in model *M* is broken down into its constituent result parts.

2.2 The normative model

The normative model specifies which reference object(s) should be used to compare. It also specifies the variables with respect to which the comparison should be made. The most common "reference objects" to diagnose business performance are [2]:

- Historical reference values
- Industry averages as reference values

Historical reference values
In this case the reference value for a particular variable is its (average) value in one or more previous time periods. The number of historical periods considered should not be too large, in view of the possibility of "structural changes", such as a change in the macro-economic climate. Such a structural change would lead to "comparing the incomparable". In historical comparisons the only judgment that can be made is *better* or *worse* than in the previous period. It does not enable one to say that some property is *good* or *bad* in an absolute sense. It might be the case that a company has a declining return on total assets, but that the industry on average is doing even worse.

Industry averages as reference values
The industry average of companies operating within the same industry is often used as a reference for the company. Such industry averages can usually only be usefully compared for *ratios*, not for *nominal* variables. The problem with any comparison between different firms is that factors such as accounting methods or the size of the firm may have a considerable influence on the results. Since these factors might not be constant among firms in the same branch or industry, this diminishes the comparability of the firms. To take into account the size of the firm we normalize the data by dividing each variable in the business model by the total number of employees expressed in FTE (Full-Time Employees) for each individual firm.

Another drawback is that firms are compared to "mediocrity", i.e. industry average, and not for example to the best in their line of business. For companies that are currently below industry average, this norm may be a good objective for aim for. Firms that are above average will probably set different goals. Despite these possible objections, the industry average is often viewed as the normal case when comparing several firms in one line of business; significant deviations from this norm are viewed as a signal to look for underlying causes. Reference values in the case study for Statistics Netherlands could, for example, be averages of financial variables for industries, like transportation, wholesale, retail, building, and service industry.

2.3 Symptom detection

A diagnosis in a financial model is an explanation for observed exceptional behaviour of a company. The first step in diagnostic process is problem or symptom identification, the detection of abnormal behaviour. The central question in problem identification for business diagnosis is: "Which firms deviate significantly from their branch average or historic average?" Suppose the normative model contains a reference value for variable y. And the data set may contain several reference values, besides the actual values for business variables. For diagnosis of company performance the event to be explained is specified as in [1, 2]:

1. a = the actual behaviour of a company,
2. F = a particular variable deviates from its norm value,
3. r = the norm behaviour for the company involved.

Because the actual object *a* and reference object *r* will always be clear from the context, the explanation formalism is simplified to:

$$\partial y = q \text{ occurred because } C^{+}, \text{ despite } C^{-}.$$

In this expression, $\partial y = q$ where $q \in \{\text{low, normal, high}\}$, specifies an event in the financial data set, i.e. the occurrence of a quantitative difference between the *actual* and the *reference value* of y, denoted by y^{a} and y^{r}, respectively.

Symptom detection in [1, 2] is described as a fairly simple process where a value $g(y^{a}, y^{r})$ is computed for each variable, where *g* is some user-defined function such as percentage difference or absolute difference. And if this value is below (above) some specified threshold, a symptom $\partial y = low\ (high)$ is added to the list of symptoms. The result of symptom identification is a set of symptoms $S = \{\partial y_1 = q_1, \ldots, \partial y_n = q_n\}$ where $q \in \{low, high\}$. Note that for the purpose of diagnosis, it is not interesting to explain symptoms with the label $\partial y = "normal"$, since it is only required to explain why a variable deviates from its reference value [1, 2].

The above-mentioned symptom detection process is a rather “crude” method to filter symptoms out of the underlying business model and data set. A more sophisticated method is developed that takes into account the probability distribution, e.g. the normal distribution, of the business variable under consideration. In this method first the average value for each variable is estimated based on a statistical model. When a statistical model is used as a normative model then $y^{r} = \hat{y}$. The residual of the model is simply defined as: $\Delta y = y - y^{r} = y - \hat{y}$. Furthermore, the larger the absolute value of the residual, the more exceptional the variable is. If we now normalize the residual of the model by the standard deviation of the variable in the sample, we get the normalized residual $\Delta y / \sigma$. The exact population parameters of the distribution are usually unknown; therefore they are estimated and replaced by the sample mean and sample variance. Correspondingly, the problem of looking for exceptional company behaviour is equivalent to the problem of looking for exceptional normalized residuals. Statistically defined, a variable is a *symptom* or exceptional value if it is higher (lower) than some user-defined threshold δ ($-\delta$). Usually, we select $\delta = 1.645$ corresponding to a probability of 95% in the normal distribution. Automatically, the following series of statistic tests is performed on each variable in the business model, by the diagnostic program, to detect symptoms in the data set under consideration:

- if $\Delta y / s > \delta$ (one-tailed test) then the symptom is labelled $\partial y = "high"$,
- if $\Delta y / s < -\delta$ (one-tailed test) then the symptom is labelled $\partial y = "low"$ and
- if $-\delta \le \Delta y / s \le \delta$ then the symptom is labelled $\partial y = "normal"$.

2.4 Diagnosis and explanation

If $\partial y = q$ is identified as a symptom, we want to explain the difference $\partial y = y^{a} - y^{r}$ where y^{r} is a reference value of the variable under study. An explanation is given based on the financial equations of the business model. To determine the contributing and counteracting causes that explain the quantitative difference, between the actual and reference value of *y,* a *measure of influence* is defined in literature [1, 2, 10] as follows:

$$\inf(x_i, y) = f(\mathbf{x}^r_{-i}, x^a_i) - y^r,$$

where $f(\mathbf{x}^r_{-i}, x^a_i)$ denotes the value of $f(\mathbf{x})$ with all variables evaluated at their reference values, except x_i. In words, $\inf(x_i, y)$ indicates what the difference between the actual and reference value of *y* would have been if *only* x_i would have deviated from its reference value [1]. Here it is explicitly assumed that $y^a = f(x^a_1, x^a_2, \ldots, x^a_n)$ and $y^r = f(x^r_1, x^r_2, \ldots, x^r_n)$.

Furthermore, the correct interpretation of the influence measure depends on the form of the function *f*; the function has to satisfy the so-called *conjunctiveness constraint* [1, 2]. This constraint captures the intuitive notion that the influence of a single variable should not turn around when it is considered in conjunction with the influence of other variables. Two classes of functions satisfy the conjunctiveness constraint, namely *additive* (or difference) functions and *monotonic* functions, this is proven in [1], which frequently occur in business model relations. By monotonicity we mean the monotonicity in all variables separately, on the domain under consideration. For correct explanation generation the conjunctiveness constraint has to hold for the actual function as well as for the reference function. The form of the reference function depends on the type of statistical model applied.

In the situation that actual and reference function are both additive, then $\inf(x_i, y)$ is correctly interpreted as a quantitative specification of the change in *y* that is explained by a change in x_i.

Proposition: if $f = \sum_{i=1}^{n} x_i$ and $y^r = \sum_{i=1}^{n} x^r_i$ then

$$\Delta y = y^a - y^r = \sum_{i=1}^{n} \inf(x_i, y).$$

Proof: $\Delta y = y^a - y^r = \sum_{i=1}^{n} \inf(x_i, y)$.

$$\Delta y = \sum_{i=1}^{n} \inf(x_i, y) = \sum_{i=1}^{n} \left\{ f(\mathbf{x}^r_{-i}, x^a_i) - y^r \right\} =$$

$$\sum_{i=1}^{n} \left\{ \sum_{i=1}^{n} (\mathbf{x}^r_{-i}, x^a_i) - \sum_{i=1}^{n} x^r \right\} = \sum_{i=1}^{n} (x^a_i - x^r_i) = \sum_{i=1}^{n} x^a_i - \sum_{i=1}^{n} x^r_i = y^a - y^r.$$

Furthermore, if f is non-additive but differentiable, $y^r = f(\mathbf{x}^r)$ and $\delta_i = x^a_i - x^r_i$ is small then $\Delta y \approx \sum_{i=1}^{n} \inf(x_i, y)$. However in general Δy is not necessarily equal to $\Delta y = \sum_{i=1}^{n} \inf(x_i, y)$. This occurs when $y^r \neq f(\mathbf{x}^r)$, or when *f* is non-additive and $\delta_i = x^a_i - x^r_i$ is large. For monotonic functions, the interpretation of $\inf(x_i, y)$ becomes more difficult and context-dependent, but the sign of $\inf(x_i, y)$ is not context-dependent. Therefore, reference values are made *internally consistent* [1, 2] in this situation to maintain the assumption of $y = f(\mathbf{x})$.

The definition of the influence-measure makes it possible to operationalize the concepts of contributing and counteracting causes. When explanation is supported by a business model equation the set of contributing (counteracting) causes C^+ (C^-) consists of measures x_i of $\mathbf{x}$ with $\inf(x_i, y) \times \Delta y > 0 \;\; (< 0)$. In words, the contributing causes are those variables whose influence values have the same sign as ∂y, and the counteracting causes are those variables whose influence values have the opposite sign.

In the explanation method, insignificant influences are left out of the explanation by a filter measure. In [1, 2] the set of causes is reduced to the so-called *parsimonious set of causes*. The *parsimonious set of contributing (counteracting) causes* C_p^+ (C_p^-) is the smallest subset of the set of contributing causes, such that its influence on *y* exceeds a particular fraction (T^+) of the influence of the complete set. The fractions T^+ and T^- are numbers between 0 and 1, and will typically 0.85 or so.

In [1, 2] the *maximal explanation* method is described. Where the main idea is that for $\partial y = q$, explanation generation is continued only for its parsimonious contributing causes, whereas non-parsimonious causes and counteracting causes are not explained any further. This process is continued until a contributing cause is encountered that cannot be explained within the business model *M*, because the business model does not contain a relation in which this contributing cause appears on the LHS. Maximal explanation extends the idea of *one-level* explanations, that is based on only one relation from the business model, to *multi-level* explanations. The maximal explanation process results in a so-called *tree of causes* (or explanation tree), where *y* is the root of the tree and its children, grandchildren, great-grandchildren and so on are parsimonious contributing and counteracting causes. In this way explanations are chained together and a tree of causes is formed.

Moreover, there is a natural way to construct reference objects for the RHS variables of business model equations. The basic idea is that the context and statistical model selected as reference object for the LHS determines the reference objects for the variables on the RHS. In addition, if some RHS variable has again a relation in which it appears on the LHS, the construction of reference objects can be continued for the RHS variables of this equation following the same principle. In this way, a *chain of reference objects* is constructed in the explanation generation process.

2.5 Making cancelling-out effects visible by substitution

The explanation/diagnosis methodology as described in literature has a serious short-coming, because it cannot deal with so-called *cancelling-out* or *neutralisation effects.* Cancelling-out is the phenomenon that the effects of two or more lower-level variables in the business model may cancel each other out so that their joint influence on a higher-level variable in the business model is partly or fully neutralized. These effects are quite common in financial models as we shall see in the upcoming case study. The problems with these effects were first mentioned by Kosy and Wise in [10], however no solution was presented in their article. For the top-down explanation generation process this means that in some data sets possible significant parsimonious causes for a symptom will not be detected when cancelling-out effects are present. These not-detected parsimonious causes by multi-level explanation are called *hidden causes*. Hidden causes are significant causes that are not visible at first due to the *neutralisation of a higher level variable* in the business model. In theory, cancelling-out effects may occur on every level in the business model. Therefore, one does not have a clue a priori on what level in the business model detection for these effects should start and whether these effects are significant or not. Of course, financial analysts would like to be informed about significant hidden causes, and would consider an explanation tree without mentioning these causes as incomplete and not accurate. Therefore, we developed a flexible method that takes into account cancelling-out effects.

Because hidden causes and neutralised variables are not reported with the classic explanation methodology, we introduce a *multi-step look-ahead method* for finding these causes. Basically, this method is an extension of the *maximal explanation method* as described in [1, 2] and can be switched on or off. In short, the look-ahead method is composed out of two consecutive steps and begins the explanation process for a symptom as before with the first available equation in the business model by determining parsimonious causes. In the first step, all non-parsimonious contributing and counteracting causes (if there are any) are subjected to the *look-ahead test* for detection of cancelling-out effects at a specific level in the business model. When cancelling-out effects are detected, the second step makes candidate hidden causes visible by means of *function substitution*. Where the lower-level equations at level *j* in the business model are substituted into the higher-level equation under consideration for explanation. Subsequently, the substituted function is added to the business model and considered for explanation generation. When the substituted equation in combination with the data under consideration yields different sets of parsimonious causes compared with the initial equation we have detected significant hidden causes. Now we will elaborate on these two steps in the look-ahead method and present an algorithm for it. Furthermore, the method will be illustrated in the case study of section 3.

Suppose that we are explaining a symptom in a part of business model *M* with the following two equations:

eq. 1 $y = f(x_1, x_2, \ldots, x_n) \in M$

eq. 2 $x_i = g(z_1, z_2, \ldots, z_n) \in M$

Furthermore, suppose that maximal explanation results in a set of parsimonious causes where the variable x_i is not part of, thus $x_i \notin C_p^+$ or $x_i \notin C_p^-$. An extreme situation occurs when $\inf(x_i, y) = 0$ then the variable x_i has no influence on Δy. To make sure that the explanation is correct, and cancelling-out effects are taken into account, the variable x_i is subjected to the look-ahead test one level ahead in the business model. Here the *depth of the business model* (*n*) is defined as the number of levels in the business model or associated directed graph. Where the root of the tree is on level 0, the children of the root are on level 1, the children of the children of the root are on level 2, and so on. The indicator for cancelling-out effects is the presence of so-called *neutralized variables* on the RHS of eq. 1.

Definition 1 (*Neutralized variable*).
A neutralized variable is a RHS variable x_i, that has marginal or no influence on the above symptom (Δy), because its influence is cancelled out by the effect of lower level variables in the business model *M*.

The definition states that the neutralisation of variable x_i is caused, in the extreme situation, by the fact that the influence of the set of contributing causes $\inf(C^+, x_i)$ is just as large as the influence of the set of counteracting causes $\inf(C^-, x_i)$. However in a lot of situations the effect of neutralisation will be partial, therefore we apply a *look-ahead test to determine the degree of neutralisation*. The level of neutralisation is the ratio r ($0 \le r \le 1$) between the influence of the complete set of contributing causes and the influence of the complete set of counteracting causes. For variable x_i the look-ahead tests are defined in Table 1.

Table 1
Definition of look-ahead tests

	Look-ahead test
$\Delta x_i > 0$	$\rho \leq \| \inf(C^-, x_i) \| / \inf(C^+, x_i) < 1$
$\Delta x_i = 0$	$\| \inf(C^-, x_i) / \inf(C^+, x_i) \| = r = 1$
$\Delta x_i < 0$	$\rho \leq \inf(C^-, x_i) / \| \inf(C^+, x_i) \| < 1$

For $r = 1$ the level of neutralisation is maximal (100%). When the ratio is between some user-defined threshold (for example, $\rho = 0.60$ (60% neutralisation) or so) and 1, the variable x_i is considered to be neutralized. Naturally, for $r = 0$ there is no cancelling-out effect in variable x_i on the level in the business model under consideration. The latter means that when $\Delta x_i > 0$ or $\Delta x_i < 0$ all RHS-variables of eq. 2 are contributing causes, thus $z_1, z_2, \ldots, z_n \in C^+$. Moreover, there is a relation between the magnitudes of the fractions T^+ and T^-, and the number of times the look-ahead test needs to be performed. By choosing a relatively high value for the fractions T^+ and T^- more causes will be labelled parsimonious, because less insignificant influences are filtered out of the explanation. Therefore, more causes are considered for maximal explanation and less non-parsimonious causes have to be subjected to the look-ahead test.

However the presence of a neutralized variable is a necessary but not sufficient condition for the existence of hidden causes somewhere in the business model.

Definition 2 (*Hidden causes*).
Parsimonious contributing or counteracting causes C_p^+ (C_p^-), not visible immediately due to the effect of a neutralized variable x_i in the business model *M* on level *j*, that become visible due to explanation generation with the substituted function $f^{sub[j]}$.

When there is a neutralisation effect in variable x_i, *all* (lower-level) children on level *j* in the business model are *substituted* into the higher-level function, and subsequently the *substituted function* $f^{sub[j]}$ is added to the business model. Here all related functions are substituted in eq. 1, because all RHS variables of $f^{sub[2]}$ (eq. 2) have to be evaluated *jointly* to take into account the absolute magnitudes of their influence values. With reference objects for all RHS variables of $f^{sub[2]}$, based on the canonical way of forming reference objects, we compute the influences of the variables on Δy and again determine sets of parsimonious contributing and counteracting causes. In addition, new detected causes are added to the explanation tree. As a general remark we state that the functional form of the substituted equation $f^{sub[j]}$ has to satisfy the conjunctiveness constraint [1, 2].

The existence of hidden causes, on level *j* in the business model for the data under consideration, is determined by making comparison between sets of parsimonious causes generated by explanation generation sustained by equations f and $f^{sub[j]}$. A number of typical situations are encountered in making comparison:

- When the sets of parsimonious causes sustained by equations f and $f^{sub[j]}$ are identical, then there are *no hidden causes*. Here no new parsimonious causes are identified by application of the look-ahead method.

- When the sets of parsimonious causes sustained by $f^{sub[j]}$ only contain children of parsimonious causes sustained by f, then there are *no hidden causes*. Here the look-ahead method does not detect parsimonious causes that would not have been detected by maximal explanation.
- When the sets of parsimonious causes sustained by $f^{sub[j]}$ contain children of neutralized variables then these (lower-level) variables are considered to be *hidden causes*. Here the look-ahead method detects new parsimonious causes that would be missed by maximal explanation.
- When the sets of parsimonious causes sustained by $f^{sub[j]}$ *only* contain children of neutralized variables and no children of parsimonious causes sustained by f, then we have the special case of *dominating hidden causes*. Here the look-ahead method takes over maximal explanation in constructing the explanation tree.

A multi-step look-ahead algorithm is proposed that is composed out of two repetitive tasks, the application of the look-ahead tests and the substitution of lower level functions into higher level functions to make hidden causes visible, as described earlier in this section. For each required look-ahead the two tasks are simply repeated, in this way we can do: one step look-ahead, two step look-ahead, and so on in the business model. The maximum number of look-ahead steps that can be carried out in a business model is $n-2$. For example, for $n=2$ obviously no look-ahead needs to be carried out, for $n=3$ one step look-ahead is the maximum, for $n=4$ two step look-ahead is the maximum, and so on. In addition, the number of look-ahead steps (the horizon) in the business model is user-defined and is based on the domain knowledge of the analyst.

The look-ahead algorithm, when activated, is executed each time after parsimonious contributing and counteracting causes have been determined with maximal explanation. More specific the algorithm is invoked after step 2, 3 and 4 of the maximal explanation tree definition [1, 2]. This means that we designed the algorithm as an extension of the maximal explanation method. Explanation generation is always initialised with maximal explanation, however after the determination of causes with the equation under consideration, cancelling-out effects are investigated with the look-ahead algorithm in the business model with a certain horizon. When no cancelling-out effects are present the algorithm continues with maximal explanation. However when cancelling-out effect are present the look-ahead algorithm takes over and makes hidden causes visible by substitution up to a specified horizon in the business model. The multi-step look-ahead method has the following pseudo-code:

```
For the number (1 to $n$) of specified look-ahead steps do
  All non-parsimonious causes in the equation under consideration are subjected to the look-ahead test
  If there is a neutralized variable then
    Substitute all children of the RHS variables on level $j$ into equation under consideration
    Add $f^{sub[j]}$ to business model and derive parsimonious contributing and counteracting causes for it
    Update the explanation tree
      Add new parsimonious causes
      Remove non-parsimonious causes
  Else
    Substitute all children of the RHS variables on level $j$ into equation under consideration and derive
    $f^{sub[j]}$
Next look-ahead step ($j$ + 1)
Continue with maximal explanation
```

3 Case-study: Interfirm analysis at Statistics Netherlands

3.1 Interfirm comparison at Statistics Netherlands

Intra-company benchmarking or interfirm comparison can be defined as the regular measuring and comparing of a company's performance against its competitors, against industry leaders or industry averages. The aim is often to learn how the company can improve its own performance. By comparing the financial variables of a company with those of other companies, the company can assess its performance against objective standards and see where the company is strong or weak. The data for interfirm comparison in this case study is obtained from Statistics Netherlands [12]. Statistics Netherlands is responsible for collecting, processing, and publishing statistics used in practice, by policymakers and for scientific research. The information Statistics Netherlands publishes incorporates a multitude of societal aspects, form macro-economic indicators such as economic growth and consumer prices, to the incomes of individual people and households.

3.2 Case description

The business model we present in this section has been taken from surveys for gathering production statistics for companies in the Dutch *retail and wholesale trade*. Statistics Netherlands uses different surveys for different types of industry. The business model relations used in the case study are:

(1) $r_1 = r_2 + r_3 + r_4 + r_5$

(2) $r_2 = r_6 - r_7$

(3) $r_3 = r_8 - r_9$

(4) $r_4 = r_{10} - r_{11}$

(5) $r_5 = r_{12} - r_{13}$

(6) $r_6 = r_{14} + r_{15}$

(7) $r_{14} = r_{16} + r_{17} + r_{18} + r_{19} + r_{20}$

(8) $r_{15} = r_{21} + r_{22}$

(9) $r_7 = r_{23} + r_{24} + r_{25} + r_{26} + r_{27} + r_{28} + r_{29} + r_{30} + r_{31} + r_{32} + r_{33} + r_{34}$

(10) $r_{23} = r_{35} + r_{36}$

(11) $r_{24} = r_{37} + r_{38} + r_{39} + r_{40}$

(12) $r_{25} = r_{41} + r_{42} + r_{43} + r_{44}$

(13) $r_{26} = r_{45} + r_{46} + r_{47} + r_{48} + r_{49} + r_{50}$

(14) $r_{27} = r_{51} + r_{52} + r_{53}$

(15) $r_{28} = r_{54} + r_{55} + r_{56} + r_{57} + r_{58} + r_{59} + r_{60}$

(16) $r_{29} = r_{61} + r_{62} + r_{63}$

(17) $r_{30} = r_{64} + r_{65} + r_{66} + r_{67} + r_{68}$

(18) $r_{32} = r_{69} + r_{70} + r_{71} + r_{72} + r_{73} + r_{74}$

(19) $r_{33} = r_{75} + r_{76} + r_{77} + r_{78} + r_{79} + r_{80} + r_{81}$

We will now elaborate on the meaning of the left-hand side (LHS) variables in the equations of the business model M (with depth $n = 4$), which Statistics Netherlands applies for companies in the Dutch retail and wholesale trade. See Appendix A for a complete overview of model variables and their meaning. Three types of business equations are identified in the business model namely: *result* (eq. 1 through 5), *revenue* (eq. 6 through 8), and *cost* (eq. 9 through 19) *equations*. The variable (r_1) in the root result equation gives the company's total result before taxation. This variable is split up into four types of results namely: total operating results (r_2), total financial results (r_3), total results allowances (r_4), and total extraordinary results (r_5). These result variables are the difference between a revenues component like: total revenues (r_6), financial gains (r_8), deductions from allowances (r_{10}) and extraordinary gains (r_{12}), and a costs component like: total costs (r_7), financial losses (r_9), additions to allowances (r_{11}) and extraordinary losses (r_{13}). Here the variable financial gain is the collection of interests received, gains from participations, payments of dividends, and profits from investments and other financial gains. The additions to allowances (r_{11}) are the sum of additions to internal provident funds, like initial expenses, funds for business restructuring and maintenance. And extraordinary profits are all gains that do not result from normal business management, like profits made on disposal of subsidiaries, fixed assets, and foreign business units. Because Statistics Netherlands is very interested in the structure of the variables r_6 and r_7 for retail businesses these variables play an important part of their surveys for business statistics. Therefore, these variables are further decomposed into revenues and costs equations.

The variables total net sales (r_{15}) and total additional revenues (r_{14}) provide a break-down of r_6. Total net revenue is the sum of net sales from the main activities of the company (r_{21}) and the net sales from other activities (r_{22}). And the total additional revenues are the sum of the variables: allowances for secondment (r_{16}), activated production for the company (r_{17}), subsidies and restitutions (r_{18}), received payments of damages (r_{19}), and other additional revenues that are not mentioned in another category (r_{20}).

Moreover the cost variables: costs of goods sold (r_{23}), total costs of labour (r_{24}), total additional personnel expenses (r_{25}), total costs of transportation (r_{26}), total costs of energy (r_{27}), total housing costs (r_{28}), total cost of production machines, equipment, installations, and office equipment (r_{29}), total selling expenses (r_{30}), total costs of communication (r_{31}), total costs of third party professional services (r_{32}), total other operating cost not mentioned elsewhere (r_{33}), and deprecations on tangible and intangible fixed assets (r_{34}) give a breakdown of r_7. Some of the RHS (right-hand side) variables of equation 9 are again decomposed in more specific cost components (see equations 10 through 19).

Statistics Netherlands distinguishes between two types of questionnaires (surveys) to gather production statistics for the retail branch, namely one for small companies and a more extended one for medium-size and large companies. The difference is that for larger companies additional variables are collected, thus the top structure for the business model is identical. Moreover, both surveys include more variables than presented in the business model like, for example, the number of employees, the type of retail company (for example, supermarkets, do-it-yourself stores, bakeries, florists, etc.) and complete specifications of net sales in terms of product lines. However the latter variables usually do not have a great filling in the data set, therefore these variables are excluded from the analysis. In addition, we use production statistics from two consecutive years, namely the year 2001 and 2002. In both data sets more than 5000 different retail and wholesale companies are present classified into branch sections. Every record in the data set represents a unique company. Because we only have the production statistics of two consecutive years at our disposal, namely the production statistics of the year 2001 and 2002, limitations are placed on the number of historic business comparisons we can perform.

3.3 Reference objects for interfirm comparison

We have to determine which reference (norm) objects are appropriate for diagnosis of business performance in the case study for Statistics Netherlands. Because we do not want to "compare the incomparable", we have to take several aspects into account. In computing norm values for diagnosis, we therefore take into consideration the Standard Industry Classification (SIC) for the retail and wholesale industry and the size of the company. Therefore, we make computerized selections on the original data set and perform diagnosis on *subsets of the data*, for example, we make selections on: supermarkets, liquor stores, do-it-yourself shops, etc. And within these subsets we make a further selection on the size class (we distinguish between small, medium and large) of the companies. The company size classes are based on the number of employees of the firm in FTE's (full-time employees). The intervals for the different size classes are:

- small: 1 through 9 employees,
- medium: 10 through 99 employees,
- and large: as from 100 employees and more.

Moreover, we can optionally make further selections in the size classes by making new size classes in the old ones. In this way we make homogeneous subsets of the data for analysis. For the analysis we *normalized the data* by dividing all variables in the business model by the total number of FTE's of each individual company. And in the case of *missing data* for a business model variable, for example, because a certain retail business left a cell blank on the questionnaire, we assign the value zero to that variable. In this way it is enforced that the business model equations always "fit" the data. In conclusion, these reduced and normalized homogeneous subsets of the data constitute the basis for computation of the norm values.

Furthermore, the original source data obtained from the questionnaires suffers from *obvious mistakes*, like "1000-errors" (e.g. € 125,235 should be noted as € 125), erroneous negative values, and empty sub(totals). Because of these mistakes the diagnostic process could result in incorrect explanations for detected symptoms, because this data will not always fit the equations. Therefore, we use the data (records) that results from the *statistical editing process* [7]. In this process the data is automatically corrected for these mistakes, for example, empty sub(totals) are filled with the sum of the constituent variables or the absolute value is taken for erroneous negative values. In addition, records with possibly less obvious mistakes are selected and edited interactively by statistical analysts, while the remaining records are edited by the computer.

The original data set has two dimensions, namely a company and a time dimension. Therefore, two different types of comparison analysis can be performed on the retail production statistics, namely comparisons against industry or branch averages and historic comparisons. For both types of analysis, reference values (usually based on averages) are determined in similar procedures and added to the original data set, firstly for the purpose of symptom detection and secondly for the purpose of explanation generation. For interfirm comparisons, the *sample mean* (industry average) is computed by taking the mean value of all the companies in the selected normalized sample of a specific year for all variables (r_1 through r_{81}) in the business model. And for historic comparisons the reference values for the business model variables are the values in one or more previous time periods, for example we can compare the results for the actual year with the results of the previous year for a certain company. In addition, the sample mean can also be computed as an historic average by taking the mean value of a single company over a number of periods again for all variables in the business model.

The correct interpretation of the inf-measure depends on the form of the function f. As stated before, the conjunctiveness constraint must hold for the actual function as for the reference function. In the situation when both functions are additive this constraint is satisfied (see proof). Furthermore, when both functions are additive, $\inf(x_i, y)$ is correctly interpreted as a quantitative specification of the change in y that is explained by the change in x_i. If we assume that industry averages serve as reference values, them we are effectively explaining a change of y between industry averages and the actual values for business model variables. In the context of the case study this the reference function is additive $y^r = \sum_{i=1}^{n} x_i^r$, because it can easily be seen that the following relation holds $\overline{y} = \sum_{i=1}^{n} \overline{x}_i$. It has to be remarked that in the situation of missing data the latter relation only holds when we apply our procedure for missing values and replace empty cells with zeros. Otherwise the sample size may vary over the business model variables. Because there are only additive functions in the business model *internal consistency* [1] is guaranteed in the case study. In addition, if we assume that previous period values serve as norm values, we are effectively explaining a change of y between periods $t-1$ and t. Here the reference function is additive by definition, thus the conjunctiveness constraint is again satisfied.

3.4 Symptom detection

For the case study we perform our analyses on a specific homogeneous sample selected out of the original data set with production statistics for the year 2001. The selected sample is composed out of n = 69 *fashion shops* out of the size class “medium”. Problem identification in the data set starts with the variable results for taxation (r_1) on the root level of the business model. This variable is normally distributed in the sample with mean 11.30 (the industry average) and standard deviation 28.85. The exact population parameters of the distribution are unknown; therefore they are estimated and replaced by the sample mean and sample variance. The central question in problem identification for this case study is: “Which firms deviate significantly from their branch average?” The symptom detection module of the diagnosis application finds 9 firms that are higher (or lower) than the specified threshold value in the sample data set (see Table 2 for a full specification of the norm model). Here we select $\delta = 1.645$ ($\delta = -1.645$) corresponding to a probability of 95% in the normal distribution. With these test specifications we derive the following distribution of the number of firms over the three symptom types:

- high symptom: 5 firms,
- normal symptom: 60 firms,
- and low symptom: 4 firms.

The firms with symptoms high and low are placed in a separate file/spreadsheet and are candidates for explanation generation. In the spreadsheet these firms are marked with a color, so that the analysts can immediately spot exceptional firms.

Table 2: Specification of norm model for diagnostic example

Slot name	Slot entry
variable	results before taxation ($r1$)
norm object	industry average
industry	fashion shops
size class	medium
year	2001
number of firms	69
distribution	normal distribution $r_1 \sim N(11.30, 832.17)$
threshold	$\alpha = .05$ (two one-tailed tests)

For one of the fashion shops in the spreadsheet we present complete diagnostics with symptom detection and explanation generation. The diagnosis will be performed for the ABC-company; the selected fashion shop in the data set. Moreover, the data is anonimized because Statistics Netherlands does not allow the exposure of data on the micro level; therefore we do not mention the real names of the companies involved in the case study. However we will present data on the aggregated level, for example, industry averages, and specific micro data we will present anonimized. However all data and statistics are available at Statistics Netherlands. For the ABC-company the detected symptom is "*high*" when comparing the actual result before taxation of the company with the branch average, because the one-tailed test $(61.75-11.30)/28.85 > 1.645$ is above the threshold value. Furthermore, the relative difference between the actual value and industry average for r_1 is $(61.75-11.30)/11.30 = 4.996$. Thus, the ABC-company is doing particularly good compared to its industry average, almost 5 times better.

3.5 Example explanation generations for ABC-company

In this paragraph comparison is made between the results of two explanations for the event/symptom: $\langle \text{ABC- company(2001)}, \partial r_1 = high, \text{ branch\_average(2001)} \rangle$. Here we will try to answer the following question: "Why did the results before tax go up for the ABC-company?" Here the branch average for (large) fashion shops in the year 2001 in the Netherlands is taken as a reference object for both explanations. The first explanation for this event is generated by the "classic" explanation method as described in the literature [1, 2]. We shall show for the data under consideration that this method will not give the optimal explanation in the case of cancelling-out effects. Moreover, comparison of human analysis and the result of the classic explanation method shows noticeable differences when these effects occur. Therefore, we present a second explanation generated with our extended explanation method, with detection for hidden causes switched on, as described in the theory section. The latter explanation will illustrate the theory and extra capabilities of the extended model to automatically detect hidden causes that are not detected by the classic methodology for automated business comparison. The two explanations and additional explanatory trees are both generated automatically by our prototype computer program named the "*Statistics Netherlands diagnosis application*".

Explanation generation with maximal explanation

The classic explanation method yields the following results, taking $T^+ = T^- = 0.85$ for the fraction. In Table 3 comparison is made between the actual results before taxation of the ABC-company and the branch average in the year 2001. From the data in Table 3 we infer that $C_p^+ = \{r_2\}$ and $C_p^- = \{\ \}$. The variable r_2 (total operating results) explains 90.44% of the difference ∂r_1, and is therefore identified as the single parsimonious contributing cause because its value exceeds the fraction. Thus, the result variables r_3, r_4 and r_5 are filtered out of the explanation because their influences are considered to be too small. Therefore, the variable r_2 is the single child node of its parent (root node) r_1 in the explanation tree.

Table 3
Actual and norm values for $r_1 = r_2 + r_3 + r_4 + r_5$

	Actual	Norm	$\inf(x_i, y)$	diff. %
r_1	61.75	11.30		466.31
r_2	60.42	14.79	45.62	308.38
r_3	1.33	–2.55	3.88	152.30
r_4	0.00	–0.15	0.15	100.00
r_5	0.00	–0.79	0.79	100.00

The diagnostic process is now continued for the parsimonious contributing cause, supported for further explanation by equation 2, to explain for the initial difference in ∂r_1. Here one clearly sees that the classic explanation method assumes that there are no significant cancelling-out effects in the data for lower-level equations 3, 4 and 5. In the second part of this section we shall see that this assumption does not hold for this particular financial data set. Under this assumption the new event (analogous to the previous example) to be explained is specified as: $\langle \text{ABC-company(2001)}, \partial r_2 = \mathit{high}, \text{ branch\_average(2001)}\rangle$. Table 4 summarizes the results for the explanation of the ABC-company's relative high total operating result. From the data in the table it follows that $C_p^+ = \{r_6, r_7\}$, since both r_6 (explains 45.73%) and r_7 (explains 54.73%) contributed to the difference between norm value and the actual value, and are both needed to explain the desired fraction of $\inf(C^+, r_2)$. In words, the total operating results for the ABC-company are relatively high, because of the fact that the total operating revenues (r_6) are high and the total operating costs (r_7) are low in comparison with their branch averages. Obviously, $C_p^- = \{\ \}$. Thus, the variable r_2 has two successor children in the explanation tree. Both children correspond to equations (eq. 6 and 9) in the business model and can therefore be explained further.

Table 4
Actual and norm values for $r_2 = r_6 - r_7$

	Actual	Norm	$\inf(x_i, y)$	diff. %
r_2	60.42	14.79		308.38
r_6	329.50	308.64	20.86	6.76
r_7	269.09	293.84	24.76	–8.43

Analogous to the previous example, the events to be explained are specified as follows: $\langle \text{ABC- company(2001)}, \partial r_6 = \mathit{high}, \text{branch\_average(2001)}\rangle$ and $\langle \text{ABC-company(2001)}, \partial r_7 =$ $\mathit{low}, \text{ branch\_average(2001)}\rangle$. In other words, we want to determine which lower level revenues and costs variables in the business model contributed significantly to these events. Firstly, Table 5 summarizes the model results for the explanation of the ABC-company's relatively high total operating revenues. From the data in the table it follows that $C_p^+ = \{r_{14},$ $r_{15}\}$ and $C_p^- = \{\ \}$. Although, both r_{14} and r_{15} are needed to explain the desired fraction, the variable total net sales (r_{15}) is clearly a more important cause, explaining 83.80%, of ∂r_6. Secondly, from Table 6 it can be concluded that $C_p^+ = \{r_{25}, r_{28}, r_{30}, r_{33}, r_{34}\}$ and $C_p^- = \{r_{23}, r_{24}\}$. Notice that the model does not mention that the contributing cause total cost of transportation

(r_{26}) is below average. The reason behind is that the contributions to the overall contributing influence $(\inf(C^{+}, r_2) = -35.50)$ on the total operating costs is marginal. The same reasoning applies for the total cost of energy (r_{27}) and the total cost of third party professional services (r_{32}). This shows that the model filters insignificant influences out of the explanations. Furthermore, the parsimonious counteracting causes are mentioned by the explanation model, since $C_p^{-} = \{r_{23}, r_{24}\}$. However the marginal counteracting cause total costs of communication $(\inf(r_{31}, r_7) = 0.02)$ is omitted from this set of causes.

Table 5
Actual and norm values for $r_6 = r_{14} + r_{15}$

	Actual	Norm	$\inf(x_i, y)$	diff. %
r_6	329.50	308.64		6.76
r_{14}	4.92	1.54	3.38	220.06
r_{15}	324.58	307.10	17.48	5.69

Table 6
Actual and norm values for $r_7 = r_{23} + r_{24} + r_{25} + r_{26} + r_{27} + r_{28} + r_{29} + r_{30} + r_{31} + r_{32} + r_{33} + r_{34}$

	Actual	Norm	$\inf(x_i, y)$	diff. %
r_7	269.09	293.84		–8.43
r_{23}	181.42	178.30	3.12	1.75
r_{24}	64.00	56.42	7.58	13.43
r_{25}	0.42	3.61	–3.19	–88.45
r_{26}	0.50	1.71	–1.21	–70.72
r_{27}	1.92	2.27	–0.36	–15.64
r_{28}	2.17	18.47	–16.31	–88.27
r_{29}	0.33	0.67	–0.34	–50.43
r_{30}	8.42	11.99	–3.57	–29.79
r_{31}	1.00	0.98	0.02	2.46
r_{32}	3.50	4.39	–0.89	–20.19
r_{33}	1.42	5.00	–3.59	–71.68
r_{34}	4.00	10.04	–6.04	–60.16

Explanation generation proceeds with the identified parsimonious contributing causes in Table 5 and Table 6 that are supported with additional equations in the business model *M*. The variables (causes) r_{14} and r_{15} are sustained by the revenue equations 7 and 8. And the variables (causes) r_{25}, r_{28}, r_{30} and r_{33} are sustained by the cost equations 12, 15, 17 and 19. For these equations the influence values are presented in Appendix B and are omitted here because of space limitations. Notice that variable r_{34} has no equation in the business model, so for this variable the explanation process is terminated.

The previous examples of different one-level explanations are now combined to a complete tree of causes. Fig. 2 summarizes the results of the complete diagnostic process, where dashed lines indicate counteracting causes. Since there is only one symptom to be explained, the diagnosis contains one maximal explanation. Thus, Fig. 2 actually depicts the maximal explanation, as specified in paragraph 2.4, for $\partial r_1 = "high"$.

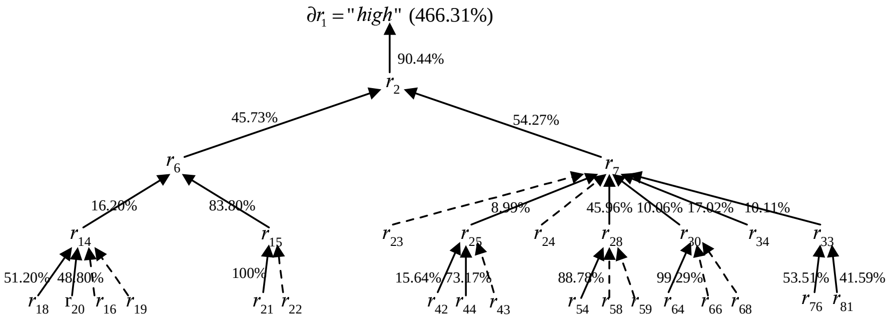


Fig. 2 Diagnosis for S = {∂r_1 = low} at ABC-company

Explanation generation with multi-step look-ahead algorithm

In this section, we explain the initial event for the ABC-company with the multi-step look-ahead algorithm activated, because we want to test for cancelling-out effects in the data set. In other words, we drop the assumption of *no cancelling-out effects* in the data set under consideration. Therefore, we initialize the procedure for detection of hidden causes in our diagnostic program. The procedure will be configured initially for *one-step look-ahead* and we take again $T^+ = T^- = 0.85$ for the fractions.

Explanation generation starts again with the root equation in the business model and from the data in Table 3 we derive obviously the same set of causes as in the classic method. However, instead of proceeding with purely explanation of the parsimonious contributing causes, we now first test for potential cancelling-out effects, one step ahead in the business model. Here the look-ahead procedure takes into account the effects of all not-identified variables one level deep, thus the effects of the RHS-variables in equations 3, 4 and 5.

First the procedure subjects all variables that are excluded from the parsimonious sets to the look-ahead test with lower boundary $\rho = 0.30$. Thus the degree of neutralisation is computed for all LHS's of equations 3, 4, and 5. Based on Table 7, the degree of neutralisation for equation 3 is computed as (|5.44| / 9.33) * 100% = 58.31%, because $\inf(C^+, r_3) \geq \inf(C^-, \ r_3)$. This means, in words, that more than half of the effect of the contributing cause (financial revenues) is neutralized by the counteracting cause (financial expenses). Therefore, the effect of total financial results (r_3) on ∂r_1 is only marginal, and is excluded from the parsimonious contributing set in the classic explanation methodology. The variable r_3 is clearly a partially neutralized variable. However analysts would like to take these secondary "long-distance" effects (neutralized variables) into account because they are important for an accurate and complete explanation. In addition, for equation 4 and 5 the degrees of neutralisation are somewhat lower, respectively 6.25% and 28.18%. Fig. 3 shows the one step look-ahead tests with arrows "stepping over" the intermediate nodes, and pointing at the RHS variables of equation 3, 4 and 5, in the partial explanation tree. In this figure, the straight black lines indicate the parsimonious causes that were actually detected.

Table 7
Actual and norm values for $r_3 = r_8 - r_9$

	Actual	Norm	$\inf(x_i, y)$	Diff. %
r_3	1.33	–2.55		152.30
r_8	11.17	1.84	9.33	506.59
r_9	9.83	4.39	–5.44	123.98

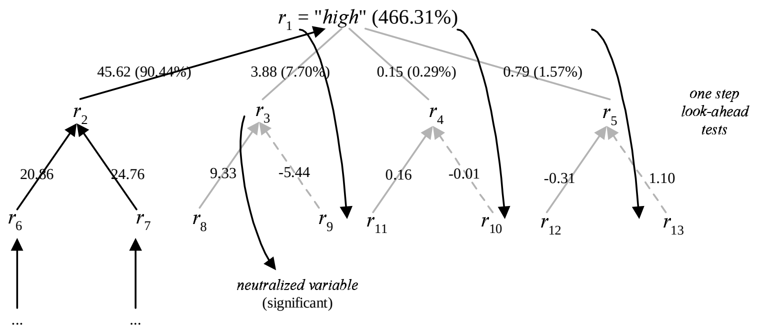


Fig. 3 Illustration of look-ahead method

In the first part of the procedure the look-ahead test found some cancelling-out effects. However we still do not know whether these effects are significant or not. Therefore, in the second phase of the procedure the *substitution method* is applied to find parsimonious contributing and counteracting causes, which were missed in the local explanation of differences by standard multi-level explanation. Equations 2 through 5 are substituted into the root equation and the following new equation for explanation generation is derived: $r_1^{sub[2]} = (r_6 - r_7) + (r_8 - r_9) + (r_{10} - r_{11}) + (r_{12} - r_{13})$. This substituted equation is added to the set of business model equations (equation 20), *changing* the original business model *M*. Because the substituted function is again additive the conjunctiveness constraint is satisfied. The specification of the event to explain ∂r_1 remains the same, however now we apply equation 20 to explain for the difference. Table 8 summarizes the results of our extended model of the ABC-company's relatively high results before taxation. From the data in Table 8 it follows that $C_p^+ = \{r_6, r_7, r_8\}$ and $C_p^- = \{r_9\}$, and we conclude that the effects of causes r_8 and r_9 are significant at the specified fractions for parsimonious sets.

Table 8
Actual and norm values for $r_1 = (r_6 - r_7) + (r_8 - r_9) + (r_{10} - r_{11}) + (r_{12} - r_{13})$

	Actual	Norm	$\inf(x_i, y)$	diff. %
r_1	61.75	11.30		466.31
r_6	269.09	293.84	24.76	6.76
r_7	329.50	308.64	20.86	–8.43
r_8	11.17	1.84	9.33	506.59
r_9	9.83	4.39	–5.44	123.98
r_{10}	0.00	0.16	0.16	–100.00
r_{11}	0.00	0.01	–0.01	–100.00
r_{12}	0.00	0.31	–0.31	–100.00
r_{13}	0.00	1.10	1.10	–100.00

Furthermore, these *hidden causes* were missed in the classic way of analysing differences because their joint effect was neutralized in variable r_3 (see again Fig. 3). For the tree of causes this means that two new children are added to the root node: a parsimonious contributing child for r_8 and a parsimonious counteracting child for r_9. As a result the top branches of the original tree are updated, as can be seen in Fig. 4. Notice that the neutralized variable r_3 is not part of the tree of causes (grey line).

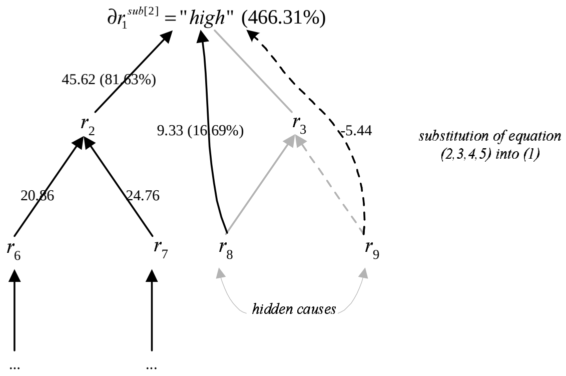


Fig. 4 Diagnosis with detection of hidden causes for $S=\left\{\partial r_1^{sub[2]}="high"\right\}$ at ABC-company

The analysis continues as follows. For r_8 explanation generation stops because there are no supportive equations for it in the business model. Again for causes r_6 and r_7 explanation is sustained by equation 6 and equation 9. From the data in Table 5 and 6 parsimonious causes are determined as before. Because all RHS variables in Table 5 are identified as parsimonious contributing causes the look-ahead procedure is not applied here. In this case, potential cancelling-out effects do occur due to the fact that all contributing causes are explored further in the business model. In addition, it can easily be seen that there are no hidden causes under node r_6 in the tree of causes represented by Fig. 2, because all its children and grandchildren are identified.

Based on the data from Table 6 the look-ahead test needs to be applied for all not identified variables in equation 9 one step ahead in the business model. For most of these variables the degree of neutralisation for the sets of causes is relatively low. However the degree of neutralisation for the RHS variables of equation 18 is more than 40% ((0.65/|-1.53|)*100%), showing the possibility of the existence of significant hidden causes. For the second time function substitution is needed to exclude the phenomenon of hidden causes in the underlying data set. Substituting equations 10 through 19 into equation 9 yields the following additive equation: $r_7^{sub[4]}=r_{31}+r_{34}+r_{35}+\ldots+r_{80}+r_{81}$ (added as equation 21 to the business model). From the data in Appendix B it follows that $C_p^+=\{r_{34,}r_{40},r_{44},r_{54},r_{60},$ $r_{64},r_{74},r_{76},r_{81}\}$ and $C_p^-=\{\ r_{35},r_{37},r_{38},r_{59},r_{66},r_{68},r_{69}\}$. From these parsimonious sets it is concluded that r_{32} is a neutralized variable and that r_{69} and r_{74} are significant hidden causes that were not mentioned in the first explanation tree (see Fig. 2). Now explanation generation continuous again for all parsimonious contributing causes. Eventually, the tree of causes is updated for the substituted equation and new branches are added; resulting in the (sub-)tree of causes represented by Fig. 5. Notice the differences in comparing the successors of variable r_7 in Fig. 5 with the successors of the same variable in Fig. 2. In Fig. 5, clearly important additional information is given. For example, now the large counteracting cause r_{35} under node r_{23} is made visible.

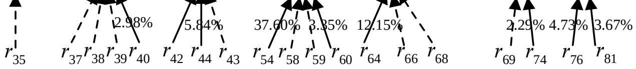


Fig. 5 Diagnosis with detection of hidden causes for $S=\left\{\partial r_7^{sub[4]}="low"\right\}$ at ABC-company

4 Software implementation

In this section we shortly present the most important concepts of the software implementation of the prototype diagnosis application in MS Excel in combination with Visual Basic. This application is initially programmed to perform the experiments and analyses for the case study at Statistics Netherlands. However the prototype software could handle data and business models from multiple domains. Fig. 6 depicts the overview architecture of the program for diagnosis.

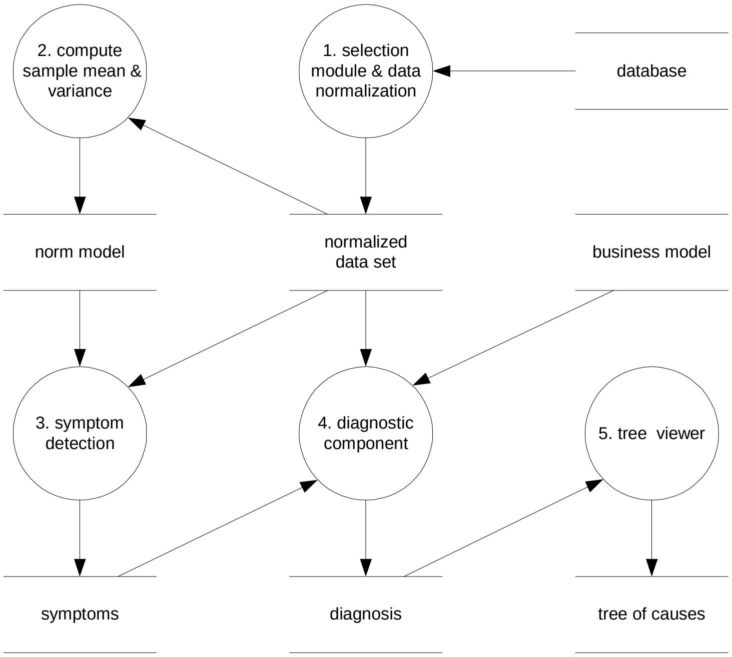


Fig. 6 Data flow diagram (DFD) of diagnosis application

Most elements of the program are discussed in the previous parts of this paper. However the procedure the *diagnostic component*) was not discussed earlier. This procedure contains both the method for maximal explanation as the multi-step look-ahead algorithm. For the implementation of the procedure we applied *tree programming* to generate the tree of causes

Furthermore, we include some screenshots of the software in Appendix C. In Fig. 7, the main user interface screen is depicted with buttons for symptom detection and explanation generation. One immediately notices the textboxes where user-defined fields (e.g. threshold values, fractions, and number of look-aheads) have to be filled out.

[Please insert Figure 7 about here]

The tree-viewer interface of the program is depicted in Fig. 8. In the viewer the whole explanatory graph can be made visible my manipulating the tree. In addition, the tree of causes is projected on the explanatory graph by highlighting parsimonious causes. By clicking on the cause under consideration the details for the cause become visible in the right panel of the screen.

[Please insert Figure 8 about here]

Naturally, the software is available upon request.

5 Summary and conclusions

In this paper, we extended the formal model for automated business diagnosis and improved its software implementation. Therefore, the explanation model is extended in two ways: in the symptom detection phase the probability distribution of business model variables is taken under consideration and in the explanation generation phase hidden causes can be made visible by function substitution. The problem of looking for exceptional company behaviour in financial data sets is translated into the problem of looking for exceptional normalized residuals. In this way a statistical definition for a symptom is derived. Furthermore, the multi-level look-ahead algorithm is proposed to enhance the explanation methodology so that it can deal with cancelling-out effects; the common effect that variables cancel each other out somewhere in the business model with the result that their effect on a higher level in the business model is partially or fully neutralized. The extended model is implemented as a prototype in MS Excel in combination with Visual Basic. Within the software implementation special attention is given to presentation of the program output, where symptoms and causes are presented graphically as a tree of causes. In this manner, a manager or financial analyst can view and access the results of the explanation process for diagnosis of company performance as a compact tree. This tree can be navigated with simple mouse-clicks.

The applicability of the extended method is illustrated by a case study on interfirm/historic comparison in the Dutch retail and wholesale trade, based on production statistics obtained from Statistics Netherlands. In the case study it is shown that in the presence of cancelling-out effects the extended model with the multi-level look-ahead procedure makes significant causes visible that would be missed by the explanation methodology of maximal explanation. In addition, the fully automated diagnostic process makes it possible to detect and explain abnormal company behaviour in large data sets. We believe that this enhanced framework could assist analysts and improve the decision-making process, by automatically generating explanations for exceptional values in various data sets and business models.

Acknowledgments. We want to thank Jeffrey Hoogland for his support at Statistics Netherlands.

References

[1] H.A.M. Daniels, A.J. Feelders, "*Theory and methodology: a general model for automated business diagnosis*", European Journal of Operational Research, 130: 623-637, (2001).

[2] A.J. Feelders, "*Diagnostic reasoning and explanation in financial models of the firm*", PhD thesis, University of Tilburg, (1993).

[3] E.A.M. Caron , H.A.M. Daniels, "*Diagnosis in the OLAP context*", ERIM working paper, http://hdl.handle.net/1765/1492, (2004).

[5] E.A.M. Caron, H.A.M. Daniels, "*Automated business diagnosis in the OLAP Context*", In H. Fleuren, P. Kort, and D. den Hertog (Eds.), Operations Research Proceedings 2004, pp. 425-433, Springer, (2005).

[6] J.F. Courtney, D.B. Paradice and N.H. Ata Mohammed, A knowledge-based DSS for managerial diagnosis, Decision Sciences, 18, 373-399, (1987).

[7] A. de Jong, "*Uni-edit: Standardized processing of structural business statistics in the Netherlands*", Conference on European Statistics, UNECE work session on statistical data editing, Helsinki, working paper no. 27, (2002).

[8] G. Hesslow: "*Explaining differences and weighting causes*", Theoria, 49:87-111, (1984).

[9] P.W. Humphreys: "*The chances of explanation*", Princeton University Press, Princeton, New Jersey, (1989).

[10] D.W. Kosy, B.P. Wise: "*Self-explanatory financial planning models*", in Proceedings of AAAI-84, Los Altos, CA, Morgan Kaufmann, pages 176-181, (1984).

[11] W. Verkooijen: "*Automated financial diagnosis: a comparison with other diagnostic domains*", Journal of Information Science, 19, pages 125-135, (1993).

[12] Statistics Netherlands, http://www.cbs.nl, (2005).

Appendix A: List of variables

Because of space limitations only the meaning of the variables identified by the explanation model are described here in detail. The variable descriptions have been translated from the original Dutch questionnaires. The complete list of variables and their definitions are available upon request.

Result variables:

r_1: results before taxation

r_2: total operating results

r_3: total financial results

r_4: total results allowances

r_5: total extraordinary results

r_6: total operating revenues

r_7: total operating costs

r_8: financial revenues

r_9: financial expenses

r_{10}: additions to allowances

r_{11}: deductions from allowances and provisions released

r_{12}: extraordinary profits

r_{13}: extraordinary losses

Revenue variables:

r_{14}: total additional revenues

r_{15}: total net sales

r_{16}: allowances for secondment

r_{17}: activated production for own company

r_{18}: subsidies and restitutions

r_{19}: received payments of damages

r_{20}: other additional revenues (not mentioned elsewhere)

r_{21}: net sales main activity of company

r_{22}: net sales other activities

Cost variables:

r_{23}: costs of goods sold

r_{24}: total costs of labour

r_{25}: total additonal personnel expenses

r_{26}: total costs of transportation

r_{27}: total costs of energy

r_{28}: total housing costs

r_{29}: total cost of production machines, equipment, installations, and office equipment

r_{30}: total selling expenses
r_{31}: total costs of communication
r_{32}: total cost of third party professional services
r_{33}: total other operations costs (not mentioned elsewhere)
r_{34}: depreciations on tangible en intangible fixed assets

r_{35}: costs of commodity goods sold
r_{36}: other costs of goods sold

r_{37}: gross wages and salaries
r_{38}: employer part´s social security insurance
r_{39}: pensions
r_{40}: other social security contributions

r_{41}: payments to temporary workers
r_{42}: payments to other temporary workers
r_{43}: training costs
r_{44}: other personnel expenses

r_{45}: costs of leasing/renting means of transportation
r_{46}: costs of maintenance for means of conveyance
r_{47}: costs of fuel
r_{48}: ownership tax
r_{49}: insurance premiums for means of conveyance
r_{50}: other costs of transportation

r_{51}: natural gas
r_{52}: costs of electricity
r_{53}: other costs of energy (excluding fuels)

r_{54}: costs of leasing/renting land and buildings
r_{55}: maintenance/repairs land and buildings
r_{56}: costs of cleaning land and buildings
r_{57}: environment tax
r_{58}: property tax
r_{59}: insurance premium for building and contents assurances
r_{60}: other housing costs

r_{61}: renting/leasing machines, equipment, installations, and office equipment
r_{62}: maintenance/reparation machines, equipment, installations, and office equipment
r_{63}: other costs machines, equipment, installations, and office equipment

r_{64}: advertizing and promotion expenses
r_{65}: commissions for agents
r_{66}: travelling, accommodation and representation costs
r_{67}: research and development costs
r_{68}: other selling expenses

r_{69}: banking business
r_{70}: other insurance premiums (not mentioned elsewhere)
r_{71}: accountancy, juridical, economical, tax advice
r_{72}: third-party services for automation and computerization
r_{73}: refuse and waste processing
r_{74}: other third-party costs for professional services

r_{75}: licenses, royalties, copyright
r_{76}: intra concern/administrative costs
r_{77}: stationary, contributions, subscriptions, specialist literature
r_{78}: other costs for renting/leasing (not mentioned elsewhere)
r_{79}: other maintenance/reparation costs (not mentioned elsewhere)
r_{80}: other cost price increasing taxes (not mentioned elsewhere)
r_{81}: other general costs (not mentioned elsewhere)

Appendix B: Data for comparison ABC-company

The values for the lower level cost variables are omitted because of space limitations.

variables	ABC-company			
	actual	reference	diff. %	influence
r_{1}	61.75	11.30	466.31	50.45
r_{2}	60.42	14.79	308.38	45.62
r_{3}	1.33	–2.55	152.30	3.88
r_{4}	0.00	–0.15	100.00	0.15
r_{5}	0.00	–0.79	100.00	0.79
r_{6}	269.09	293.84	6.76	24.76
r_{7}	329.50	308.64	–8.43	20.86
r_{8}	11.17	1.84	506.59	9.33
r_{9}	9.83	4.39	123.98	–5.44
r_{10}	0.00	0.16	–100.00	0.16
r_{11}	0.00	0.01	–100.00	–0.01
r_{12}	0.00	0.31	–100.00	–0.31
r_{13}	0.00	1.10	–100.00	1.10
r_{14}	4.92	1.54	220.06	3.38
r_{15}	324.58	307.10	5.69	17.48
r_{16}	0.00	0.22	–100.00	-0.22
r_{17}	0.00	0.00	0.00	0.00
r_{18}	2.33	0.35	559.19	1.98
r_{19}	0.00	0.26	–100.00	–0.26
r_{20}	2.58	0.70	270.91	1.89
r_{21}	324.58	304.42	6.62	20.16
r_{22}	0.00	2.68	–100.00	–2.68
r_{23}	181.42	178.30	1.75	3.12
r_{24}	64.00	56.42	13.43	7.58
r_{25}	0.42	3.61	–88.45	–3.19
r_{26}	0.50	1.71	–70.72	–1.21
r_{27}	1.92	2.27	–15.64	–0.36
r_{28}	2.17	18.47	–88.27	–16.31
r_{29}	0.33	0.67	–50.43	–0.34
r_{30}	8.42	11.99	–29.79	–3.57
r_{31}	1.00	0.98	2.46	0.02
r_{32}	3.50	4.39	–20.19	–0.89
r_{33}	1.42	5.00	–71.68	–3.59
r_{34}	4.00	10.04	–60.16	–6.04

Data for cost variables

variables	ABC-company			
	actual	reference	diff. %	influence
r_{35}	181.42	177.69	2.10	3.73
r_{36}	0.00	0.61	–100.00	–0.61
r_{37}	53.50	45.93	16.49	7.57
r_{38}	6.83	6.17	10.76	0.66
r_{39}	3.50	2.95	18.78	0.55
r_{40}	0.17	1.38	–87.93	–1.21
r_{41}	0.00	0.36	–100.00	–0.36
r_{42}	0.00	0.51	–100.00	–0.51
r_{43}	0.17	0.12	44.29	0.05
r_{44}	0.25	2.62	–90.47	–2.37
r_{45}	0.00	0.62	–100.00	–0.62
r_{46}	0.00	0.16	–100.00	–0.16
r_{47}	0.00	0.33	–100.00	–0.33
r_{48}	0.00	0.06	–100.00	–0.06
r_{49}	0.00	0.12	–100.00	–0.12
r_{50}	0.50	0.42	19.40	0.08
r_{51}	0.67	0.51	29.74	0.15
r_{52}	1.17	1.38	–15.45	–0.21
r_{53}	0.08	0.38	–77.97	–0.29
r_{54}	0.00	15.26	–100.00	–15.26
r_{55}	0.50	0.86	–41.73	–0.36
r_{56}	0.00	0.21	–100.00	–0.21
r_{57}	0.08	0.05	51.78	0.03
r_{58}	0.58	0.24	147.92	0.35
r_{59}	1.00	0.49	104.18	0.51
r_{60}	0.00	1.36	–100.00	–1.36
r_{61}	0.00	0.20	–100.00	–0.20
r_{62}	0.17	0.37	–54.92	–0.20
r_{63}	0.17	0.10	69.54	0.07
r_{64}	1.83	6.76	–72.89	–4.93
r_{65}	0.00	0.03	–100.00	–0.03
r_{66}	1.00	0.48	106.96	0.52
r_{67}	0.00	0.01	–100.00	–0.01
r_{68}	5.58	4.71	18.63	0.88
r_{69}	1.17	0.64	82.85	0.53
r_{70}	0.67	0.54	22.87	0.12
r_{71}	1.33	1.81	–26.24	–0.47
r_{72}	0.33	0.43	–21.67	–0.09
r_{73}	0.00	0.04	–100.00	–0.04
r_{74}	0.00	0.93	–100.00	–0.93
r_{75}	0.00	0.00		0.00
r_{76}	0.00	1.92	–100.00	–1.92
r_{77}	0.67	0.69	–3.91	–0.03
r_{78}	0.00	0.01	–100.00	–0.01
r_{79}	0.00	0.14	–100.00	–0.14
r_{80}	0.00	0.00		0.00
r_{81}	0.75	2.24	–66.53	–1.49

Appendix C: Additional figures

Figure 1: Explanatory graph for business model Statistics Netherlands

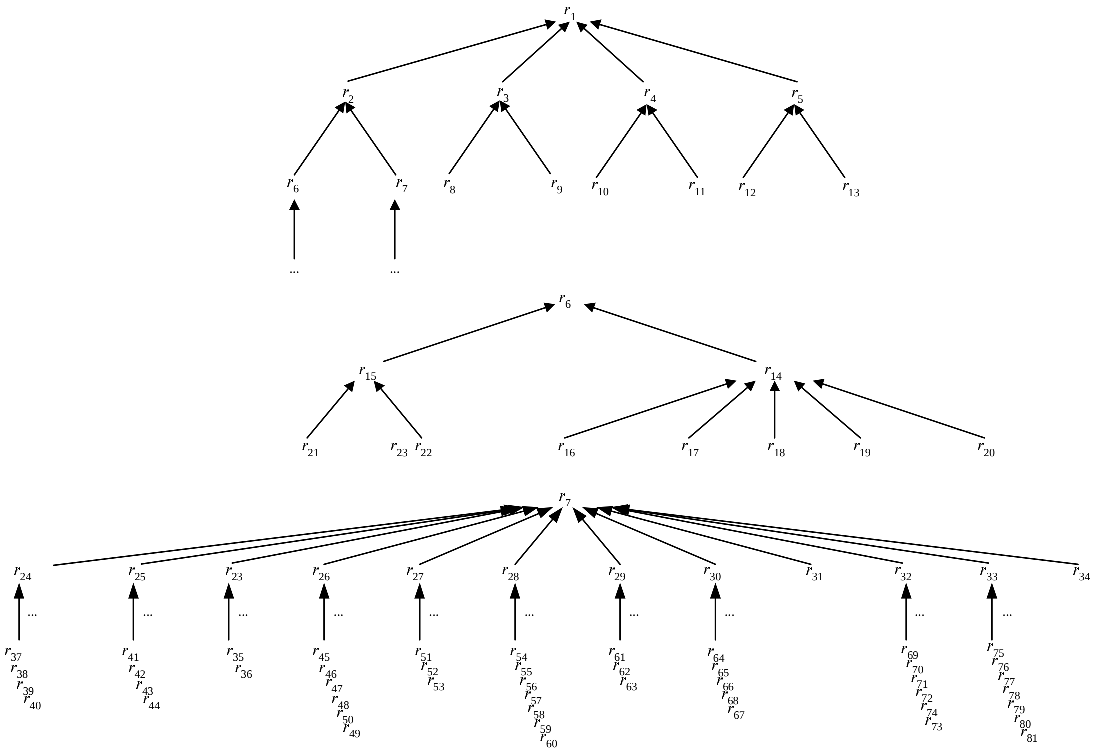
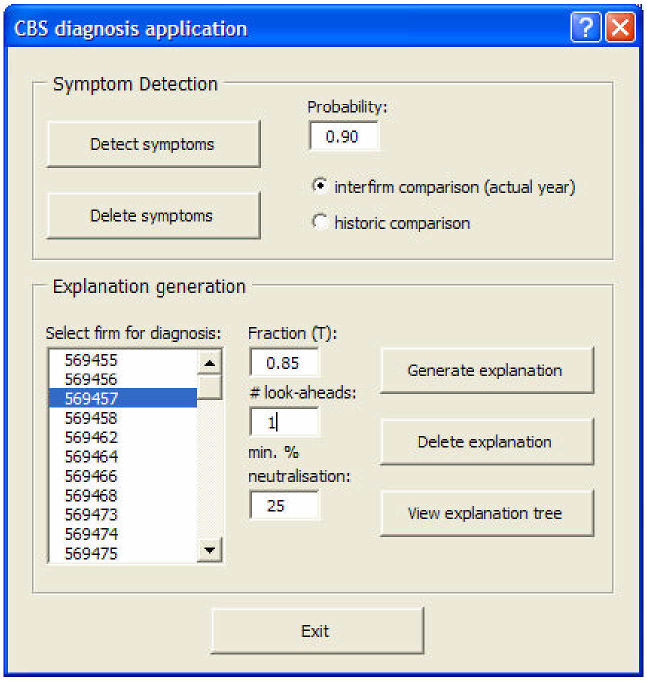

r_{1}
r_{2}
r_{3}
r_{4}
r_{5}
r_{6}
r_{7}
r_{8}
r_{9}
r_{10}
r_{11}
r_{12}
r_{13}
...
...
r_{6}
r_{15}
r_{14}
r_{21}
r_{23} r_{22}
r_{16}
r_{17}
r_{18}
r_{19}
r_{20}
r_{7}
r_{24}
r_{25}
r_{23}
r_{26}
r_{27}
r_{28}
r_{29}
r_{30}
r_{31}
r_{32}
r_{33}
r_{34}
...
r_{37} r_{38} r_{39} r_{40}
r_{41} r_{42} r_{43} r_{44}
r_{35} r_{36}
r_{45} r_{46} r_{47} r_{48} r_{50} r_{49}
r_{51} r_{52} r_{53}
r_{54} r_{55} r_{56} r_{57} r_{58} r_{59} r_{60}
r_{61} r_{62} r_{63}
r_{64} r_{65} r_{66} r_{68} r_{67}
r_{69} r_{70} r_{71} r_{72} r_{74} r_{73}
r_{75} r_{76} r_{77} r_{78} r_{79} r_{80} r_{81}

Figure 7: CBS diagnosis application; Main user interface



CBS diagnosis application

Symptom Detection

Probability: 0.90

● interfirm comparison (actual year)

○ historic comparison

Explanation generation

Select firm for diagnosis:

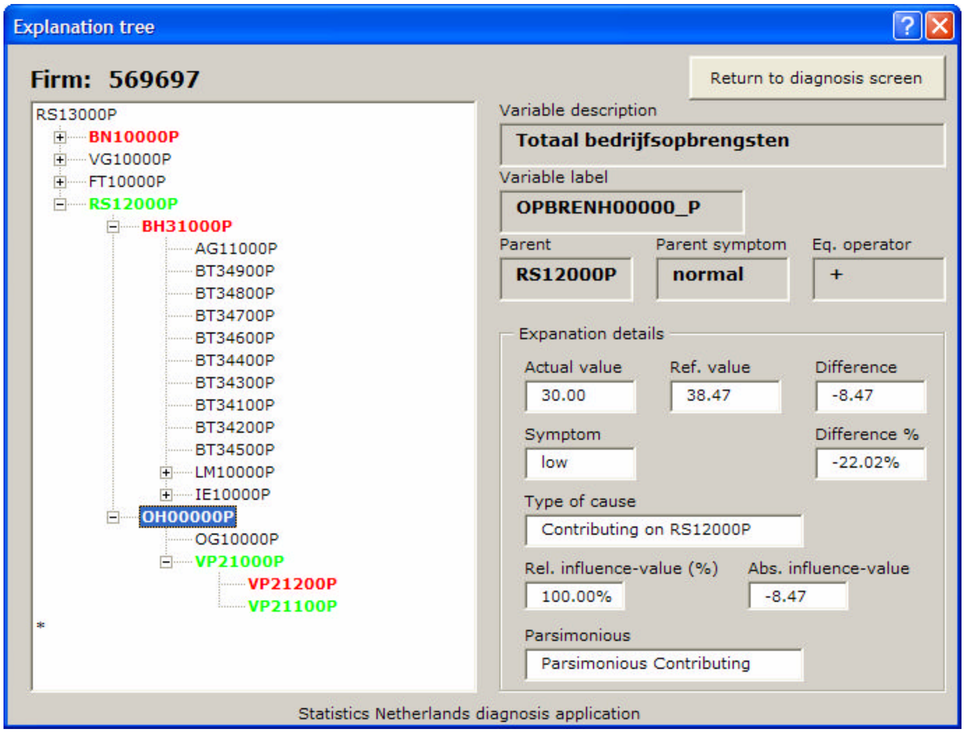
569455
569456
569457
569458
569462
569464
569466
569468
569473
569474
569475

Fraction (T): 0.85

look-aheads: 1

min. % neutralisation: 25

Figure 8: CBS diagnosis application; Explanation tree viewer



Explanation tree

Firm: 569697

Return to diagnosis screen

```
RS13000P
  ⊞ BN10000P
  ⊞ VG10000P
  ⊞ FT10000P
  ⊟ RS12000P
      ⊟ BH31000P
            AG11000P
            BT34900P
            BT34800P
            BT34700P
            BT34600P
            BT34400P
            BT34300P
            BT34100P
            BT34200P
            BT34500P
          ⊞ LM10000P
          ⊞ IE10000P
      ⊟ OH00000P
            OG10000P
          ⊟ VP21000P
                VP21200P
                VP21100P
*
```

Variable description
Totaal bedrijfsopbrengsten

Variable label
OPBRENH00000_P

Parent	Parent symptom	Eq. operator
RS12000P	normal	+

Expanation details

Actual value	Ref. value	Difference
30.00	38.47	-8.47

Symptom	Difference %
low	-22.02%

Type of cause
Contributing on RS12000P

Rel. influence-value (%)	Abs. influence-value
100.00%	-8.47

Parsimonious
Parsimonious Contributing

Statistics Netherlands diagnosis application

Publications in the Report Series Research* in Management

ERIM Research Program: “Business Processes, Logistics and Information Systems”

2005

On The Design Of Artificial Stock Markets
Katalin Boer, Arie De Bruin And Uzay Kaymak
ERS-2005-001-LIS
http://hdl.handle.net/1765/1882

Knowledge sharing in an Emerging Network of Practice: The Role of a Knowledge Portal
Peter van Baalen, Jacqueline Bloemhof-Ruwaard, Eric van Heck
ERS-2005-003-LIS
http://hdl.handle.net/1765/1906

A note on the paper Fractional Programming with convex quadratic forms and functions by H.P.Benson
J.B.G.Frenk
ERS-2005-004-LIS
http://hdl.handle.net/1765/1928

A note on the dual of an unconstrained (generalized) geometric programming problem
J.B.G.Frenk and G.J.Still
ERS-2005-006-LIS
http://hdl.handle.net/1765/1927

Privacy Metrics And Boundaries
L-F Pau
ERS-2005-013-LIS
http://hdl.handle.net/1765/1935

Privacy Management Contracts And Economics, Using Service Level Agreements (Sla)
L-F Pau
ERS-2005-014-LIS
http://hdl.handle.net/1765/1938

A Modular Agent-Based Environment for Studying Stock Markets
Katalin Boer, Uzay Kaymak and Arie de Bruin
ERS-2005-017-LIS
http://hdl.handle.net/1765/1929

Lagrangian duality, cone convexlike functions
J.B.G. Frenk and G. Kassay
ERS-2005-019-LIS
http://hdl.handle.net/1765/1931

Operations Research in Passenger Railway Transportation
Dennis Huisman, Leo G. Kroon, Ramon M. Lentink and Michiel J.C.M. Vromans
ERS-2005-023-LIS
http://hdl.handle.net/1765/2012

Agent Technology Supports Inter-Organizational Planning in the Port
Hans Moonen, Bastiaan van de Rakt, Ian Miller, Jo van Nunen and Jos van Hillegersberg
ERS-2005-027-LIS
http://hdl.handle.net/1765/6636

Faculty Retention factors at European Business Schools
Lars Moratis, Peter van Baalen, Linda Teunter and Paul Verhaegen
ERS-2005-028-LIS
http://hdl.handle.net/1765/6559

Determining Number of Zones in a Pick-and-pack Orderpicking System
Tho Le-Duc and Rene de Koster
ERS-2005-029-LIS
http://hdl.handle.net/1765/6555

Integration of Environmental Management and SCM
Jacqueline Bloemhof and Jo van Nunen
ERS-2005-030-LIS
http://hdl.handle.net/1765/6556

On Noncooperative Games and Minimax Theory
J.B.G. Frenk and G.Kassay
ERS-2005-036-LIS
http://hdl.handle.net/1765/6558

Optimal Storage Rack Design for a 3-dimensional Compact AS/RS
Tho Le-Duc and René B.M. de Koster
ERS-2005-041-LIS
http://hdl.handle.net/1765/6730

Strategies for Dealing with Drift During Implementation of ERP Systems
P.C. van Fenema and P.J. van Baalen
ERS-2005-043-LIS
http://hdl.handle.net/1765/6769

Modeling Industrial Lot Sizing Problems: A Review
Raf Jans and Zeger Degraeve
ERS-2005-049-LIS
http://hdl.handle.net/1765/6912

Cyclic Railway Timetabling: a Stochastic Optimization Approach
Leo G. Kroon, Rommert Dekker and Michiel J.C.M. Vromans
ERS-2005-051-LIS
http://hdl.handle.net/1765/6957

Linear Parametric Sensitivity Analysis of the Constraint Coefficient Matrix in Linear Programs
Rob A. Zuidwijk
ERS-2005-055-LIS

Diffusion of Mobile Phones in China
Sunanda Sangwan and Louis-Francois Pau
ERS-2005-056-LIS

An Elementary Proof of the Fritz-John and Karush-Kuhn-Tucker Conditions in Nonlinear Programming
S.I. Birbil, J. B. G. Frenk and G. J. Still
ERS-2005-057-LIS

General model for automated diagnosis of business performance
Emiel Caron and Hennie Daniels
ERS-2005-058-LIS

[*] A complete overview of the ERIM Report Series Research in Management:
https://ep.eur.nl/handle/1765/1

ERIM Research Programs:

LIS	Business Processes, Logistics and Information Systems
ORG	Organizing for Performance
MKT	Marketing
F&A	Finance and Accounting
STR	Strategy and Entrepreneurship