# Sensitivity Analysis in OLAP Databases

Emiel Caron and Hennie Daniels
*Department of Management, Tilburg University, Warandelaan 2, Tilburg, The Netherlands*

Keywords: OLAP Databases, Business Analytics, Explanatory Analytics, Sensitivity Analysis, Decision-support Systems.

Abstract: The theoretical underpinnings under which sensitivity analysis is valid in OLAP databases are dealt with in this paper. Sensitivity analysis is considered to be the reverse of explanation generation in diagnostic reasoning. Our exposition differentiates between sensitivity analysis in systems of purely drill-down equation and mixed systems of equations with also business model equations. It is proven that there is an unique additive drill-down measure defined on all cubes of the aggregation lattice. This proof is the basis for sensitivity analysis in OLAP databases, where a change in some base cell in the lattice is propagated to all descendants in its upset. For sensitivity analysis in mixed systems of equations a matrix notation is presented and the conditions for solvability are discussed. Due to the fact that such systems are typically overdetermined in OLAP databases, the implicit function theorem cannot be applied. Therefore, we proposed a method to reduce the number of equations in the system and apply the implicit function theorem on a subsystem of the original system. We conclude with an alternative method for what-if analysis in mixed systems of equations.

## 1 INTRODUCTION

The main goal of our research is "to extend the functionality of multi-dimensional (or OLAP) business databases with diagnostic capabilities to support managerial decision-making" (Caron, 2013). In this paper, the OLAP database is extended with functionality for *sensitivity analysis*. The purpose of the methods and algorithms presented here, is to provide OLAP databases with more powerful explanatory analytics and reporting functions. In this paper, we describe how sensitivity analysis can be implemented in a multi-dimensional database.

Sensitivity analysis in multi-dimensional databases is related to the notion of comparative statics in economics. Where the central issue is to determine how changes in independent variables affect dependent variables in an economic model. Comparative statics is defined as the comparison of two different equilibrium states solutions, before and after a change in one of the independent variables, keeping the other variables unchanged (Samuelson, 1941). It is one of the primary analytical methods used in economics, where it is commonly used, for example, in the study of changes in supply and demand when analyzing a market. Instead of repeating the phrase "keeping the other variables unchanged", economists use the more compact Latin equivalent *ceteris paribus* (c.p.). The underlying model for comparative statics is a set of equations that define the vector of dependent variables $y_1, y_2, \ldots, y_m$ as functions of the vector of independent variables $x_1, x_2, \ldots, x_n$, i.e.

$$\mathbf{y} = f_l(\mathbf{x}), l = 1, 2, \ldots, m. \quad (1)$$

On the one hand, this corresponds to a *system of business model equations* between measures in an OLAP database. Relations between measures are denoted by

$$y^{\mathbf{i}}(C) = f(\mathbf{x}^{\mathbf{i}}(C)), \quad (2)$$

where $y$ and $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ are measures on the same cube $C = [i_1, i_2, \ldots, i_n]$. On the other , the application of a specific aggregation function $f$, i.e. SUM(), Count(), AVG(), etc., on the measure values of each cube $y(C)$ in $L$ creates a *system of drill-down equations*, given by

$$y^{i_1 \ldots i_q \ldots i_n}(C) = f(y^{i_1 \ldots (i_q - 1) \ldots i_n}(R_q^{-1}(C))). \quad (3)$$

In Equation (2), the function $f$ might be non-linear, in Equation (3), the function $f$ is linear for SUM() and Count() aggregations, where $R$ is a drill-down operator in dimension $q$. In the latter situation we use the terms non-base variables and base variables, for dependent and independent variables, respectively. To implement sensitivity analysis in OLAP, we define a new cube operator that supports



Caron, E. and Daniels, H.
Sensitivity Analysis in OLAP Databases.
DOI: 10.5220/0006791702210228
In *Proceedings of the 20th International Conference on Enterprise Information Systems (ICEIS 2018)*, pages 221-228
ISBN: 978-989-758-298-1

the analyst in answering typical managerial what-if questions, while navigating the cube. We distinguish between two types of what-if questions:

- Questions related to a system of drill-down equations. For example, “How is the profit in the year 2010 affected when the profit for a certain product is changed with one percent in the first quarter in The Netherlands, c.p.?”
- Questions related to a system of business model equations. For example, “How is the profit in the year 2010 for a certain product affected when its unit price is changed with one additional unit in the sales model, c.p.?”

This paper is structured as follows. In Subsection 1.1 we discuss related work. In Section 2 we discuss sensitivity analysis in systems that consist of purely drill-down equations. In Section 3 we elaborate on sensitivity analysis in systems that consist of purely business model equations and mixed systems of equations. Finally, in Section 4 we draw some conclusions.

### 1.1 Related Work

The variables, parameter values, and assumptions of any business or economic model are subject to change. Sensitivity analysis, generally defined, is the investigation of these potential changes and their impacts on conclusions to be drawn from the model (e.g. (Baird, 1990)). There are many possible applications of sensitivity analysis, described here within the categories of decision support, communication, increased understanding or quantification of the system, and model development (Pannell, 1997). There is a very large literature on procedures and techniques for sensitivity analysis (Clemson et al., 1995). Two general classes of techniques for sensitivity analysis are the *implicit function theorem* (Currier, 2000; Heckman, 2000) and *monotone comparative statics* (Milgrom and Shannon, 1994). These are methods for characterizing whether an increase in a parameter causes the dependent variable to increase or decrease. Historically the implicit function theorem was used for this purpose and the implicit function theorem not only tells you whether the dependent variable increases or decreases but also the magnitude of change. In contrast, monotone comparative statics tells you only “up” or “down”, i.e., it gives an ordinal rather than cardinal answer. In our research, we focused solely on quantitative what-if analysis within the multi-dimensional database.

To the best of our knowledge, (Balmin et al., 2000) and (Lakshmanan et al., 2007) are the only published research works that address sensitivity analysis in OLAP databases in a significant way. In (Balmin et al., 2000), the authors have developed the SESAME system for the processing of hypothetical queries. For this system query algebra operators are proposed that are suitable for spreadsheet-style what-if computations. In the system hypothetical queries are modeled as a list of hypothetical modifications on the data in the fact table. A shortcoming of their approach is that it lacks a good mathematical underpinning, to decide whether a certain change is allowed in the model or not, as opposed to our approach. In (Lakshmanan et al., 2007), a different perspective is taken on what-if analysis. They focus on what-if analysis related to changes in dimensions and their hierarchical structure. However, our focus is on data-driven what-if scenarios, as opposed to structural ones.

In many OLAP software products, sensitivity analysis is not possible at the moment. If one wants to do sensitivity analysis in these products one has to copy the data to a reporting environment, for example MS Excel, to compute manually the impact of changes in certain cells of the data cube. An exception is the software product Clickview (Cliqview Corporation, 2017), where a fixed change in a base variable can be induced in a system of additive drill-down measures, to determine its impact on non-base variables. The difference with our approach is that we can induce variable changes in systems of additive and average drill-down measures and under certain conditions in non-linear systems of business equations. For this purpose we have designed a prototype application for sensitivity analysis in MS Excel with Pivot tables, with additional features implemented in Visual Basic.

## 2 SENSITIVITY ANALYSIS IN A SYSTEM OF DRILL-DOWN EQUATIONS

In this section we investigate the influence of a change in a measure value of a cell in any OLAP cube, on a higher level value of the same measure in the aggregation lattice. Or in formal notation, what is the effect of changing $y(c')$ to $y(c')+\delta$ on a dependent variable $y(c)$ in the upset of $c'$. To solve this consider the lattice $L'$ with top cube $C_p = [i_1, i_2, \ldots, i_n]$ and base cube $C_q = [j_1, j_2, \ldots, j_n]$. Notice that $L'$ is a sublattice of $L$ and $L' = \{\downarrow c\} \cap \{\uparrow c'\}$. The values of the measure $y$ in the cube $C_q$ are denoted by $x(c'_i)$, and are called the base variables where $i = 1, 2, \ldots, |C_q|$, and the values of the measure $y$ in $\{\uparrow C_q\}$ are denoted by $y(c)$, and are called the non-base variables. We distinguish between the original values of a measure without change $x^r(C_q)$ and $y^r(C_p)$, and the values of the changed measure: $x^a(C_q)$ and $y^a(C_p)$, where $x^a(C_q) = x^r(C_q)$ except for one cell $c'_i$ in the cube $C_q$, for which $x^a(c'_i) - x^r(c'_i) = \delta$.

The following theorem shows how the values of $y$ change in the lattice $L'$.

**Theorem 1.** There is an unique additive drill-down measure $y^a(c)$ defined on all cube cells in the sublattice $L'$ such that:

$$y^a(c) = y^r(c) + \beta(c) \cdot (x^a(c'_i) - x^r(c'_i)), \tag{4}$$

where:

$\beta(c) = 1$ if $c \in \{\uparrow c'_i\}$, and
$\beta(c) = 0$ if $c \notin \{\uparrow c'_i\}$.

**Proof.** To show that $y^a(c)$ is additive it is sufficient to show that $\beta(c) \cdot (x^a(c'_i) - x^r(c'_i))$ is additive, because the sum of additive measures is also additive and $y^r(c)$ is additive by the consistency assumption. Hence, we must show that:

$$\beta(c) = \sum_q \beta(R_q^{-1}(c)), \tag{5}$$

where $R_q^{-1}$ is the drill-down operation defined on a cell $c$ in the lattice $L$. Now there are two cases:

1. $c \in \{\uparrow c'_i\}$, i.e. $c$ is an ancestor of $c'_i$. In that case $c'_i$ is also a descendant of *one* of the cells in $R_q^{-1}(c)$, $c'_i \in \{\downarrow R_q^{-1}(c)\}$, which is a child of $c$ in dimension $q$. This property does not depend on dimension $q$. So both sides of Equation (5) are equal to 1.
2. $c \notin \{\uparrow c'_i\}$, i.e. $c$ is not an ancestor of $c'_i$. In that case, $c'_i$ is also not a descendant of *one* of the children of $c$. Hence, both sides of Equation (5) are zero. □

Notice that the drill-down measure $y^a(c)$ is unique. This follows from the general proposition that every additive measure with given values on the base cube is unique. This follows immediately from Theorem on OLAP equations described in (Caron, 2013) (page 40), and the fact that $L'$ is a lattice of cubes.

In the case that $c \in \{\uparrow c'_i\}$, we can rewrite Equation (4) as follows

$$y^a(c) = y^r(c) + \inf(y^a(c'), y^a(c)). \tag{6}$$

If $y(c)$ is an additive drill-down measure then we use

$$\inf(y^{a;\mathbf{j}}(c'), y^{a;\mathbf{i}}(c)) = y^{a;\mathbf{j}}(c') - y^{r;\mathbf{j}}(c'), \tag{7}$$

for the computation of $\inf(y^a(c'_i), y^a(c))$ in Equation (6) and if the variable $x^r(c')$ is changed with $\delta$ in sensitivity analysis then $y^a(c)$ is computed as $y^a(c) = y^r(c) + (x^a(c') - x^r(c'))$. This result follows immediately from the Theorem described in (Caron, 2013) (page 40).

Moreover, in the case that $y^r(c)$ is an average drill-down measure we use specific influence measure for the computation of $\inf(y^a(c'_i), y^a(c))$ in Equation (6) and if the variable $x^r(c')$ is changed with $\delta$ in sensitivity analysis then $y^a(c)$ is computed as $y^a(c) = y^r(c) + \frac{1}{|C_q|}(x^a(c') - x^r(c'))$, where $C_q$ is the context cube under consideration. This result is not proven here but the proof is similar to the proof of Theorem described in (Caron, 2013) (page 40), with the difference that the RHS of the drill-down equation is divided by the number of cells in the context cube.

Here we present a numeric example of a what-if analysis in the cube $C$ = Store $\times$ Products for the measure sales, aggregated by the average function. The data of the cube is depicted in Table 1. We want to analyse a change $\delta$ in the cell $(A, P_1)$ on its upset $\{\uparrow (A, P_1)\}$. The reference value of the cell is given by $\text{sales}^r(A, P_1) = 1$ and the actual value is given by $\text{sales}^a(A, P_1) = 1 + \delta$. By applying Equation (6) we compute the effect of this change on $\{\uparrow (A, P_1)\}$; these effects are given by,

$$\begin{aligned} \text{sales}^a(\text{All}, P_1) = &\quad \text{sales}^r(\text{All}, P_1) + \tfrac{1}{3}\delta \text{ where } |R^{+1}_{\text{Stores}}(C)| = 3, \\ \text{sales}^a(A, \text{All}) = &\quad \text{sales}^r(A, \text{All}) + \tfrac{1}{4}\delta \text{ where } |R^{+1}_{\text{Products}}(C)| = 4, \\ \text{sales}^a(\text{All}, \text{All}) = &\quad \text{sales}^r(\text{All}, \text{All}) + \tfrac{1}{12}\delta \text{ where } |C| = 12. \end{aligned}$$

For example, here $R^{+1}_{\text{Stores}}(C)$, represent a roll-up $(+1)$ in the Stores's dimension hierarchy, and $|R^{+1}_{\text{Stores}}(C)| = 3$ represents the number of stores.

Table 1: Sensitivity analysis in the example cube Store $\times$ Products for the average drill-down measure sales. Here the value of the cell $(A, P_1)$ is changed with $\delta$ and this change is propagated in the cell's upset.

| AVG(sales) | | Stores | | | |
|---|---|---|---|---|---|
| | | $A$ | $B$ | $C$ | All |
| Products | $P_1$ | $1+\delta$ | 2 | 3 | $2+\frac{1}{3}\delta$ |
| | $P_2$ | 4 | 5 | 6 | 5 |
| | $P_3$ | 7 | 8 | 9 | 8 |
| | $P_4$ | 10 | 11 | 12 | 11 |
| | All | $5.5+\frac{1}{4}\delta$ | 6.5 | 7.5 | $6.5+\frac{1}{12}\delta$ |

The subsystem of drill-down equations that corresponds with $\{\uparrow c'\}$ has an unique solution, after a change in $y(c')$ with some $\delta$, as a result of the theorem (Caron, 2013) . However, the complete system of equations becomes inconsistent because Equation (3) does not hold in that case:

$$y^{\max_1 \max_2 \ldots \max_n}(c) + \delta(c') \neq \sum_{c_n \in R_n^{-\max_n} \circ \ldots \circ R_2^{-\max_2} \circ R_1^{-\max_1}(c)} y^{00\ldots0}(c_n).$$

In other words, when the change in what-if analysis is not induced by a variable in the base cube, but by a (non-base) variable on some intermediate level in the lattice $L$, the complete system of equations will become inconsistent. For analysis restricted to $\{\uparrow c'\}$ this does not matter, however analysis in the complete system is obviously not useful anymore. The inconsistencies in the complete system of drill-down equations, can be corrected by a straightforward procedure, that repairs the OLAP database (Caron and Daniels, 2008).

# 3 SENSITIVITY ANALYSIS IN A SYSTEM OF BUSINESS EQUATIONS

In this section we discuss managerial what-if questions related to a system of business model equations and a mixed system of drill-down and business model equations.

Multiple related measures in the business model and associated dimensions, result in a mixed, often non-linear, system of equations. For example, consider Table 2 with the equations of Figure 1. In 13 these equations are given in shorthand notation. The system of equations in (13) are represented as a graph in Figure 2. In this system we want to change an independent variable $x_i$, e.g. $x_4$ (= Volume(2005.Q2)) and/or $x_5$ (= Unit Price(2005.Q2)), and study the impact on its upset, in particular, the dependent root variable $y_1$ (= Revenues(2005)). Notice that (13) is overdetermined, because we have 4 independent variables and 5 equations.

In general, for a mixed system of equations

- the equations are linear and non-linear, and
- the system of equations is overdetermined.

Similarly, a system of solely drill-down equations is also overdetermined in the case of multiple dimensions. Equation (13) can be written as

$$f_l(\mathbf{y},\mathbf{x}) = \mathbf{0}. \tag{8}$$

The linearization of (8) in a neighborhood of a solution $(y_0, x_0)$ reads:

$$A_1\mathbf{y} + A_2\mathbf{x} = \mathbf{0}. \tag{9}$$

The matrix $A_1$ is the $l \times m$ coefficient submatrix for dependent variables and $A_2$ is the $l \times n$ coefficient submatrix for independent variables. Here the matrix of the first derivatives of $\mathbf{f}$ with respect to $\mathbf{y}$ is represented by $A_1 = \mathbf{D_y f}(\mathbf{y},\mathbf{x})$ and the matrix of first derivatives of $\mathbf{f}$ with respect to $\mathbf{x}$ is represented by $A_2 = \mathbf{D_x f}(\mathbf{y},\mathbf{x})$.

With (9) we can examine the impact of a change in one or more independent variables c.p., given by $\Delta\mathbf{x}$, on the dependent variables, given by $\Delta\mathbf{y}$, where equation (8) has to be satisfied. In the next section, we investigate the conditions for consistency and solvability of (9), which is a necessary condition for solvability of (8).

## 3.1 Conditions for Solvability

A necessary condition for solvability in a system of linear equations is the *rank criterium*. A system of linear equations (9), of $A_1\mathbf{y} + A_2\mathbf{x} = \mathbf{0}$, is solvable if and only if $\text{rank}(A_1| - A_2\mathbf{x}) = \text{rank}(A_1)$. The proof of this theorem is, for example, given in (Schott, 1997). In words, the rank criterium says that the vector $-A_2\mathbf{x}$ must be in the column space (range) of $A_1$ for the system to be solvable.

To investigate the solvability of (8), we assume that

$$(\mathbf{y}_0, \mathbf{x}_0) = (y_1^0, y_2^0, \ldots, y_m^0, x_1^0, x_2^0, \ldots, x_n^0)$$

is a solution of (8). We substitute this solution in the derivative matrices $A_1$ and $A_2$ to obtain the linearized matrix $[A_1 A_2]$ at the solution $(\mathbf{y}_0, \mathbf{x}_0)$. The linearized system of equations $A_1\Delta\mathbf{y} + A_2\Delta\mathbf{x} = \mathbf{0}$ is solvable if and only if $\text{rank}(A_1) = \text{rank}(A_1| - A_2\Delta\mathbf{x})$. Similarly, the linearized system of equations is solvable for an independent variable $\partial x_i$, if and only if, $\text{rank}(A_1) = \text{rank}(A_1|\text{column } x_i \text{ from } A_2)$. A column vector $x_i$ of the submatrix $A_2$ is represented by $a_2(i)$. Accordingly, the rank criterium can be used to determine whether an independent variable $\partial x_i$ qualifies for what-if analysis in a system of business model equations. However, in the next section it is shown, that this criterium is a necessary but not sufficient condition for the solvability of a non-linear system of equations.

When the submatrix $A_1$ is nonsingular then the solution of $A_1\Delta\mathbf{y} + A_2\Delta\mathbf{x} = \mathbf{0}$ is unique and given by

$$\Delta\mathbf{y} = -A_1^{-1}A_2\Delta\mathbf{x}.$$

Notice that the rank criterium is a necessary but not sufficient condition for the solvability of a *non-linear system of equations*. Practically, this means that in such models the number of equations must be equal to the number of dependent variables to produce a square submatrix $A_1$ ($l = m$).

Now suppose that we are given an overdetermined system of equations as in (8) and a solution $(\mathbf{y}_0, \mathbf{x}_0)$ to this system such that all the equations are satisfied. The first derivatives of the equations can be written in matrix form as in (9). If the rank criterium for consistency holds for a certain independent variable $x_i$, considered for what-if analysis, then the solution $\mathbf{f}(\mathbf{y}_0, \mathbf{x}_0) = \mathbf{0}$ is filled in Equation (9). Subsequently,

$$\alpha_1 \cdot \text{eq. } 1 + \alpha_2 \cdot \text{eq. } 2 + \ldots + \alpha_l \cdot \text{eq. } l = 0, \tag{10}$$

holds if all the $\alpha_i$'s exist. If the $\alpha_i$'s exist we remove $(l-m)$ dependent equations from the system of equations and derive a $(m \times m)$ submatrix $A_1$. If the remaining system of equations in $A_1$ is nonsingular the implicit function theorem can be applied and the $\alpha_i$'s determined. In that case the removed equations are satisfied too, because Equation (10) holds and the general solution for $x_i$ can be determined.

### 3.2 What-if Analysis Example

In this example we want to change an independent variable $x_i$ and study the impact on elements in its upset. The Jacobian of the system of equations in (13) is given in (14). Observe that the vector

$$(\mathbf{y_0}\ \mathbf{x_0}) = (48\ 16\ 15\ 3.2 | 13\ 12\ 7\ 4\ 4\ 6\ 3\ 2\ 2.75\ 3\ 3.25),$$

is a solution to the system of equations. The Jacobian at $(x_0, y_0)$ is given in (15) The rank criterium for solvability in this system is satisfied for the variables $x_4$ (= Volume(2005.Q2)) and $x_5$ (= Unit Price(2005.Q2)): $\mathrm{rank}(A_1|a_{2(4)}) = \mathrm{rank}(A_1) = 4$ and $\mathrm{rank}(A_1|a_{2(5)})$ $=\mathrm{rank}(A_1) = 4$. It can easily be verified that the rank criterium is not satisfied for the other independent variables. For example, for variable $x_1$ it can be concluded that $\mathrm{rank}(A_1|a_{2(1)}) > \mathrm{rank}(A_1)$. Therefore, the only candidate independent variables for what-if analysis in this example are $x_4$ and $x_5$.

As we saw, the rank criterium is a necessary but not sufficient condition for solvability. We cannot apply the implicit function theorem to verify solvability here, because the submatrix $A_1$ is non-square $(5 \times 4)$. But in this case we may eliminate one of the equations because we can find $\alpha_i$ such that:

$$\begin{aligned} \alpha_1 \cdot \text{ eq. } 1 + \alpha_2 \cdot \text{ eq. } 2 + \alpha_3 \cdot \text{ eq. } 3 + \\ \alpha_4 \cdot \text{ eq. } 4 + \alpha_5 \cdot \text{ eq. } 5 = 0. \end{aligned} \tag{11}$$

These $\alpha_i$'s are given by

$$\begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \\ \alpha_5 \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \\ -1 \\ 0 \\ y_3 \end{pmatrix}.$$

Now we proceed as follows. In the system of equations in (13) all independent variables are replaced by the solution $(\mathbf{y}_0, \mathbf{x}_0)$ except the independent variables $x_4$ and $x_5$, that are under consideration for what-if analysis. From the original system of equations, one dependent equation is removed and we derive a reduced system of equations, where the matrix $A_1$ is square. Removing eq. 2 yields

$$f(y_1, y_2, y_3, y_4, x_4, x_5) = \begin{cases} -y_1 + 32 + y_2 = 0 \\ -y_2 + x_4 x_5 = 0 \\ -y_3 + 11 + x_4 = 0 \\ -y_4 + (32 + x_4 x_5)/y_3 = 0. \end{cases} \tag{12}$$

$(\mathbf{y}_0, \mathbf{x}_0) = (48, 16, 15, 3.2, 4, 4)$ is a solution of (12). The $4 \times 4$ derivative submatrix $A_1$ of $\mathbf{f}$ with respect to $\mathbf{y}$ in $(48, 16, 15, 3.2, 4, 4)$ is

$$\mathbf{D_y f}(48, 16, 15, 3.2, 4, 4) = \begin{pmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & -\frac{48}{225} & -1 \end{pmatrix} = A_1.$$

It can easily be verified that

$$A_1 A_1^{-1} = \begin{pmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & -\frac{48}{225} & -1 \end{pmatrix} \begin{pmatrix} -1 & -1 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & \frac{48}{225} & -1 \end{pmatrix} = I_4.$$

By the implicit function theorem we can find continuous differentiable functions $\varphi_i(x_4, x_5) : \mathcal{B} \to \mathbb{R}$, where $\mathcal{B} = \mathcal{B}_r(48, 16, 15, 3.2, 4, 4)$, such that

$$\begin{cases} y_1 = \varphi_1(x_4, x_5) \\ y_2 = \varphi_2(x_4, x_5) \\ y_3 = \varphi_3(x_4, x_5) \\ y_4 = \varphi_4(x_4, x_5), \end{cases}$$

is a solution of the system of equations (12). Moreover, also the removed equation $-y_1 + y_3 y_4 = 0$ (eq. 2) is satisfied because of (11). Computation gives:

$$\begin{cases} y_1 = 32 + x_4 x_5 \\ y_1 = (11 + x_4)(\frac{32 + x_4 x_5}{11 + x_4}) = 32 + x_4 x_5 \\ y_2 = x_4 x_5 \\ y_3 = 11 + x_4 \\ y_4 = \frac{32 + x_4 x_5}{11 + x_4}. \end{cases}$$

## 4 CONCLUSIONS

In this paper, we stated the theoretical underpinnings under which sensitivity analysis is allowed in multidimensional databases. We also discussed some theoretical issues and procedures related to sensitivity analysis in OLAP databases.

For sensitivity analysis in systems of additive drill-down measures we proved Theorem 1, and showed that there is an unique additive drill-down measure $y^a(c)$ defined on all cubes of the aggregation lattice. This theorem is the basis for sensitivity analysis here, where a change in some base cell in the lattice is propagated to all descendants in its upset. For the average drill-down measure a similar expression is determined. Moreover, sensitivity analysis might cause the multi-dimensional database to become corrupted, if the analysis is not carried out on cells in the base cube. To overcome this problem we proposed a correction procedure.

For sensitivity analysis in mixed systems of equations we introduced a matrix notation and we discussed the conditions for solvability. Because mixed systems are typically overdetermined the implicit function theorem cannot be applied. Therefore, we proposed a method to reduce the number of equations in the system and apply the implicit function theorem on a subsystem.

## REFERENCES


Baird, B. (1990). *Managerial Decisions Under Uncertainty: An Introduction to the Analysis of Decision Making*. John Wiley & Sons, Inc., New York.

Balmin, A., Papakonstantinou, Y., and Papadimitriou, T. (2000). Optimization of hypothetical queries in an olap environment. *Data Engineering, International Conference on*, 0:311.

Caron, E. (2013). *Explanation of Exceptional Values in Multi-dimensional Business Databases*. PhD thesis, Erasmus University.

Caron, E. A. M. and Daniels, H. A. M. (2008). Extensions to the olap framework for business analysis. In Cordeiro, J., Shishkov, B., Ranchordas, A., and Helfert, M., editors, *ICSOFT (ISDM/ABF)*, pages 240–247. INSTICC Press.

Clemson, B., Yongming, T., Pyne, J., and Unal, R. (1995). Efficient methods for sensitivity analysis. *System Dynamics Review*, 11(1):31–49.

Cliqview Corporation (2017). Cliqview.

Currier, K. (2000). *Comparative Statics Analysis in Economics*. World Scientific Publishing Co, Singapore.

Heckman, J. (2000). Causal parameters and policy analysis in economics: A twentieth century retrospective. *The Quarterly Journal of Economics*, 115(1):45–97.

IBM Cognos Software (2017). Ibm cognos business intelligence, powerplay.

Lakshmanan, L., Russakovsky, A., and Sashikanth, V. (2007). What if olap queries with changing dimensions perspectives are everything. In *VLDB '07: Proceedings of the 33th International Conference on Very Large Data Bases*, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

Milgrom, P. and Shannon, C. (1994). Monotone comparative statics. *Econometrica*, 62(1):157–80.

Pannell, D. J. (1997). Sensitivity analysis of normative economic models: theoretical framework and practical strategies. *Agricultural Economics*, 16(2):139–152.

Samuelson, P. A. (1941). The stability of equilibrium: Comparative statics and dynamics. *Econometrica*, 9(2):97–120.

Schott, J. (1997). *Matrix Analysis for Statistics*. Wiley, New York.

## APPENDIX

A star model representing a multi-dimensional financial database is shown in Figure 1 and is used as an illustrative example in this paper. This database, called GoSales, contains the financial figures from a generic fictitious company that sells sports equipment, obtained from the Cognos OLAP product PowerPlay (IBM Cognos Software, 2017). Figure 1 depicts a central *fact table* and five *dimensions tables*. The central fact table represents the financial data set. It lists the measures of the data, like profit, revenues, costs, etc. The financial data set has five *dimensions tables*: Time (T), Product (P), Location (L), Customer (C), and Vendor (V), and all dimensions have a 2-4 level hierarchy.

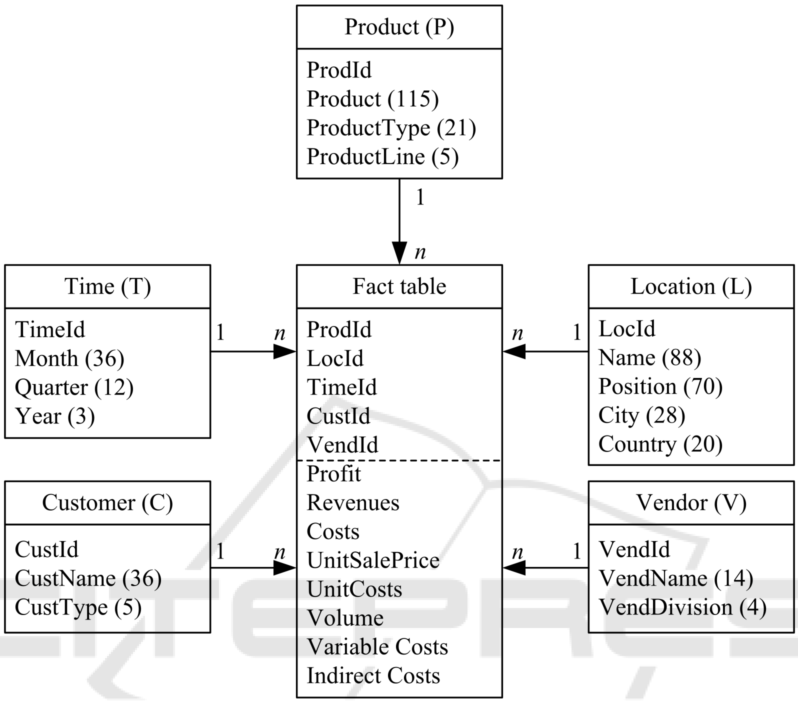


Figure 1: Star model with five dimension tables and a central fact table representing the financial data set.



Table 2: Subsystem of business model and drill-down equations derived from a multi-dimensional financial database.

| | |
|---|---|
| 1. | Rev.(2005) = Rev.(2005.Q1) + Rev.(2005.Q2) + Rev.(2005.Q3) + Rev.(2005.Q4) |
| 2. | Rev.(2005) = Vol.(2005) × Unit Pr.(2005) |
| 3. | Rev.(2005.Q2) = Vol.(2005.Q2) × Unit Pr.(2005.Q2) |
| 4. | Vol.(2005) = Vol.(2005.Q1) + Vol.(2005.Q2) + Vol.(2005.Q3) + Vol.(2005.Q4) |
| 5. | Unit Pr.(2005) = ((Vol.(*.Q1) × Unit Pr.(*.Q1)) + (Vol.(*.Q2) × Unit Pr.(*.Q2)) + (Vol.(*.Q3) × Unit Pr.(*.Q3)) + (Vol.(*.Q4) × Unit Pr.(*.Q4))) / Unit Pr.(*) |

Table 2 in shorthand notation

$$\begin{cases} -y_1 + x_1 + y_2 + x_2 + x_3 = 0 \\ -y_1 + y_3 \times y_4 = 0 \\ -y_2 + x_4 \times x_5 = 0 \\ -y_3 + x_6 + x_4 + x_7 + x_8 = 0 \\ -y_4 + ((x_6 \times x_9) + (x_4 \times x_5) + (x_7 \times x_{10}) + (x_8 \times x_{11}))/y_3 = 0, \end{cases} \tag{13}$$

where $y_i$ with $i = 1,2,3,4$ are the dependent variables and $x_i$ with $i = 1,2,\ldots,11$ are the independent variables.

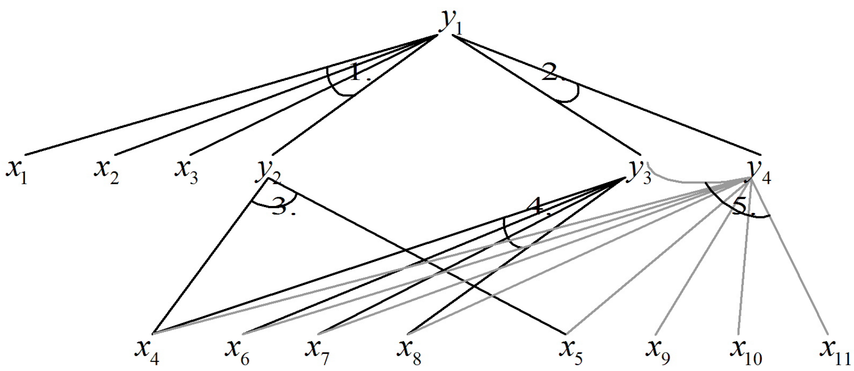


Figure 2: Graph representation of the implicit system of equations.

$$A = [A_1\ A_2] =$$
$$\left(\begin{array}{rrrr|rrrrrrrrrrr} -1 & 1 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & y_4 & y_3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 & x_5 & x_4 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & )^* & -1 & 0 & 0 & 0 & \frac{x_5}{y_3} & \frac{x_4}{y_3} & \frac{x_9}{y_3} & \frac{x_{10}}{y_3} & \frac{x_{11}}{y_3} & \frac{x_6}{y_3} & \frac{x_7}{y_3} & \frac{x_8}{y_3} \end{array}\right). \tag{14}$$
$$)^* = -\frac{x_6 x_9 + x_4 x_5 + x_7 x_{10} + x_8 x_{11}}{(y_3)^2}$$

$$A_0 = [A_1\ A_2] =$$
$$\left(\begin{array}{rrrr|rrrrrrrrrrr} -1 & 1 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 3.2 & 15 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 & 4 & 4 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & -\frac{48}{225} & -1 & 0 & 0 & 0 & \frac{4}{15} & \frac{4}{15} & \frac{2.75}{15} & \frac{3}{15} & \frac{3.25}{15} & \frac{6}{15} & \frac{3}{15} & \frac{2}{15} \end{array}\right). \tag{15}$$