Bus Inf Syst Eng 60(6):543–562 (2018)
https://doi.org/10.1007/s12599-018-0533-5

RESEARCH PAPER

EM-OLAP Framework

Econometric Model Transformation Method for OLAP Design in Intelligence Systems

Jan Tyrychtr · Martin Pelikán · Hana Štiková · Ivan Vrana

Received: 24 April 2017 / Accepted: 15 January 2018 / Published online: 14 March 2018

Abstract Econometrics is currently one of the most popular approaches to economic analysis. To better support advances in these areas as much as possible, it is necessary to apply econometric problems to econometric intelligent systems. The article describes an econometric OLAP framework that supports the design of a multidimensional database to secure econometric analyses to increase the effectiveness of the development of econometric intelligent systems. The first part of the article consists of the creation of formal rules for the new transformation of the econometric model (TEM) method for the econometric model transformation of multidimensional schema through the use of mathematical notation. In the proposed TEM method, the authors pay attention to the measurement of quality and understandability of the multidimensional schema, and compare the proposed method with the original TEM-CM method. In the second part of the article, the authors create a multidimensional database prototype according to the new TEM method and design an OLAP application for econometric analysis.

Accepted after two revisions by Prof. Dr. Jarke.

J. Tyrychtr (✉)
Department of Information Technologies, Faculty of Economics and Management, Czech University of Life Sciencesin Prague, Kamýcká 129, 165 00 Prague, Czech Republic
e-mail: tyrychtr@pef.czu.cz

M. Pelikán · H. Štiková · I. Vrana
Department of Information Engineering, Faculty of Economics and Management, Czech University of Life Sciencesin Prague, Kamýcká 129, 165 00 Prague, Czech Republic

Keywords Information system design · Decision support system · Econometric system · Analytical system · Multidimensional design · OLAP

1 Introduction

The field of econometrics has developed rapidly in the last three decades, and its applications can be found in several areas, such as determination of the level of interest rates, estimation of the price elasticity of oil demand, and the production analysis of business. Econometrics has become an interesting tool that enables the extraction of useful information regarding important business matters related to a company and the economy. However, the application of econometric models is a nontrivial process that requires a good understanding of mathematics and statistics.

Currently, several econometric software programs and tools exist. Renfro (2004) and Belsley and Kontoghiorghes (2009) described the characteristics of the most widespread econometric software programs, e.g., AREMOS, MODLER, TROLL or WinSolve. Greene (2015) mentioned that econometric instruments and methods have gradually changed from an initial emphasis on linear models with one or more equations to the present utilization of many nonlinear techniques. Another study (Küsters et al. 2006) introduced several specific problems related to present econometric and prognostic software. The authors mentioned that it is not possible to solve these problems using methodological innovations; rather, a correct and appropriate construction of the database is required. Software and database designers should establish an environment in which information, along with time series to be predicted, can be stored to enable a consequent analysis and methodological improvements. The authors noted that a stable, robust and fast interface should be available for transactional databases or online data marts. The quality of an analytical database is vital for future econometric systems.

To support decision making as much as possible in the areas of economic analysis, it is necessary to treat econometric problems using intelligent decision support systems for managers, professionals and expert staff at various management levels. To cope with this problem, several attempts have been made in the last two decades. One of the first studies (Dolk and Kridel 1991) examined the feasibility of developing an artificially intelligent econometrician as an active decision support system. The next study (Brown et al. 1995) described an econometric-based system to estimate daily cotton market prices. Another study (Brandl et al. 2006) applied the genetic algorithm to an automated econometric decision support system for a foreign exchange market. Various approaches were extensively used, mainly for the automation of econometric methods (e.g., Assaf and Dugan 2007; Yu et al. 2008; Recio et al. 2010).

1.1 Problem Statement and Previous Research

Econometrics uses panel data to analyze company behavior and employee salaries over a certain period, for instance. Numerous econometric applications use large data panels, such as financial econometrics, where the evolution of stock prices with a minute price change can be analyzed. Such models work with a large number of observations that are not available in conventional time series. Panel data is usually not over aggregated as are typical data in time series, so it is possible to analyze and test more complex hypotheses of dynamics and behavior. The powerful OLAP technology is needed to analyze dozens of variables and a large amount of data of econometric models efficiently and fast. The current econometric tools also do not offer the possibility to analyze the theoretical combinations of values that can occur in the economic reality as well as to support the essential what-if questions (see Sect. 7.3).

Previous studies did not focus on the actual design of the databases of these econometric systems and even on the use of new approaches to online analytical processing (OLAP) concepts. The first such effort involved the development of the transformation of the econometric model into the conceptual model (TEM-CM) method (Tyrychtr and Vasilenko 2015), which has been used to formally transform econometric models into the conceptual model of a multidimensional database as the basis for an econometric system based on OLAP.

OLAP offers a new solution that has not been considered in the context of econometric analysis until now. Specialized econometric software tools do not provide an intuitive analysis of econometric models in a form comprehensible to professionals and managers, whose decision-making needs are related to or based on econometric analysis but who have limited knowledge regarding the development of econometric models. From this perspective, the use of OLAP for the econometrics presented in our paper is unmatched.

Our paper represents a new methodology for developing OLAP solutions for econometric analyses. We do not try to improve upon the OLAP technology. Instead, we utilize its advantages to create a proper framework for designers of econometric or other intelligent systems. We introduce an upgraded transformation of the econometric model (TEM) method, which is based on our original TEM-CM method, for conceptual database design based on econometric models. Several shortcomings of the TEM-CM method exist, which we describe later in this article. Progress in this domain rests in upgrading TEM-CM to enable it to transform econometric models and in designing a new methodology aimed at representing econometric requirements using OLAP solutions in the new decision support systems of an enterprise.

Our research addresses improvements to the method of transformation, which can better support the design of the multidimensional databases of econometric-based systems. Application of this method is illustrated by an agricultural case study. Authors addressing the design of an analytical system for agriculture (e.g., Karmakar et al. 2007; Rai et al. 2008; Schulze et al. 2007; Han and Ju 2008; Nilakanta et al. 2008; Abdullah 2009; Bimonte et al. 2013; Uyan et al. 2013; Fountas et al. 2015) have already considered the utilization of econometric functions. However, econometric analysis (e.g., Bravo-Ureta et al. 2007; Čechura 2014; Nowak et al. 2015) for agricultural businesses has presented great potential for the improvement of their production and technical efficiency. Thus, research faces the challenge of developing appropriate analytical methods that can create an econometric OLAP solution.

1.2 Research Question and Methods

To close the abovementioned research gap, we address the following research question: How can we develop OLAP, which is an online analytical solution, for econometric analyses? This article makes the following contributions: (1) we show how to transform econometric models into conceptual models for analytic database design; (2) we create a new design framework that supports these econometric decision-making processes.

In the first part of our article, we create an innovative TEM method using a formal notation. Accrued rules serve as a methodological framework for designing the conceptual and logical schemas of an analytical database.

In the second part of the article, we create an econometric model for online analytical processing (EM-OLAP) framework by utilizing the TEM method. We create prototypes of OLAP solutions from econometric models and search for convenient ways to design them. This process results in a complex methodological framework of OLAP design for econometric analyses support, which makes implementation of the econometric decision-making principles easier.

1.3 Structure of the Article

The article is organized as follows. Section 2 presents a review of the literature and theoretical background of the OLAP field, econometrics and the description of the TEM-CM method. Section 3 describes the research approach for defining and obtaining the EM-OLAP framework. Section 4 presents the goals and hypothesis. Section 5 describes the creation phase of the new TEM method, including a comparison of the original TEM-CM method to the new TEM method and measurement of the understandability (a quality sub-characteristic) of the conceptual schemas. Section 6 explains the rules of the new TEM method. Section 7 presents the creation of the prototype of a multidimensional database created via application of the TEM method. This section also describes a systematic experiment on the created prototype and the progress achieved by the design. Section 8 presents the EM-OLAP framework for the creation of econometric OLAP systems. Section 9 presents the acceptance of the EM-OLAP framework in a real business. Section 10 discusses the methodology and results of the article. Finally, Sect. 11 concludes the article.

2 Theoretical Background

2.1 OLAP

OLAP is an approach offering decision support that aims to gain information from a data warehouse or data marts (Abelló and Romero 2009). OLAP allows the aggregation of data and inspection of indicators from different points of view. OLAP gains aggregated data by grouping various analytical data from a multidimensional database. Multidimensional data analysis is based on the fact that decision makers need aggregate data related to a particular topic, which will also be assessed according to certain factors. Aggregated data are typically modeled as a generalized data cube, which is the default model for OLAP. A data cube is a data structure used to store and analyze large amounts of multidimensional data (Pedersen 2009). A data cube allows utilizing the benefits of a multidimensional view of data and processing OLAP questions using OLAP operators such as roll-up, drill-down, slice-dice, and pivoting.

Many approaches to formally defining operator data cubes exist (a comprehensive overview can be found in Vassiliadis and Sellis 1999). Generally, a data cube consists of dimensions and measures. Dimensions represent the concepts based on which the analysis of summarized data is carried out. Analysts must often group data together and therefore must assess each dimension at different levels of detail. Hence, it is important to organize data into multidimensional hierarchies. Hierarchies of dimensions specify aggregation levels and granularity. For example, the time dimension can be defined as the following multilevel hierarchy: day → month → quarter → year. Measures (monitored indicators) of the cubes are mainly quantitative data that can be analyzed. Common examples include sales, profit, revenue and costs.

Several technologies can be used for the physical storage of multidimensional data (and the implementation of OLAP applications). The two main ways to store data are the so-called multidimensional OLAP (MOLAP) and the relational OLAP (ROLAP). The multidimensional data model based on the relational model distinguishes two basic types of relations: dimension tables and fact tables. These relation types can be used to create a star schema (e.g., Wu and Buchmann 1997; Chaudhuri and Dayal 1997; Ballard et al. 1998; Boehnlein and Ulbrich-vom Ende 1999), various forms of a snowflake schema, (e.g., Chaudhuri and Dayal 1997; Ballard et al. 1998; Boehnlein and Ulbrich-vom Ende 1999) and a constellation schema (e.g., Abdelhédi and Zurfluh 2013). The problem of choosing an appropriate structure/schema is solved in another work (Levene and Loizou 2003).

2.2 Econometric Models

An *econometric model* (EM) is a mathematical model that is a mathematical-statistical formulation of economic hypotheses. It expresses the dependence of economic variables on the variables that explain the hypothesis. The Cobb–Douglas production function is most often used in the economic literature and can be characterized by constant elasticity of the production factors, invariability in the economies of scale among businesses and a convexity isoquant function towards the beginning. The Cobb–Douglas production function has the following general form (e.g., Felipe and Adams 2005):

$$y = \alpha x_l^{\beta_l} x_p^{\beta_p} x_k^{\beta_k} \tag{1}$$

where y is the amount of output, $x_{l,p,k}$ is the amount of lth, pth and kth input, α, β is the parameters of production function.

An EM may be composed of more than one equation in the enterprise environment. Stochastic equations with random variables and identity equations exist in the model.

In a standard linear model, mathematically (Tvrdoň 2006):

$$\begin{aligned} y_{1t} &= \gamma_{11}x_{1t} + \gamma_{12}x_{2t} + \cdots + \gamma_{1g}x_{gt} + u_{1t} \\ y_{2t} &= \beta_{21}y_{1t} + \gamma_{21}x_{1t} + \cdots + \gamma_{2g}x_{gt} + u_{2t} \\ y_{3t} &= y_{1t} + y_{2t} \end{aligned} \tag{2}$$

y_s is an endogenous *s-type* variable. Its value in the period t is y_{st}, with $s = (1, 2, \ldots g)$, $t = (1, \ldots, n)$. x_r is the rth exogenous variable, with a value in the period t of x_{rt}, where the number of exogenous variables is equal to k. Thus, $r = (1, 2, \ldots, k)$. The time-delayed endogenous variable z expresses the effects of variables for period t, where $z = (1, 2, \ldots, t - z)$. u_{st} is a random variable in the sth equation of explained endogenous variables in period t. β_{is} is a structural parameter in the ith equation of the sth model undelayed endogenous variable, and γ_{ir} in the ith equation of the model of the rth predetermined variable.

The construction phases of simultaneous EMs are as follows (Čechura et al. 2017):

1. The creation of a matrix model and the content of the various matrices and vectors is as follows:
   - matrix B contains parameters of the endogenous variables of the model,
   - matrix Γ contains parameters of the predetermined variables of the model,
   - vector y_t contains endogenous variables of the model,
   - vector x_t contains predetermined variables of the model, and
   - vector u_t includes stochastic variables of the model.
2. The identification of the model is based on the following condition:
   $k_{**} \geq g_{\Delta} - 1$, where g is the total number of endogenous variables in the model, k is the total number of predetermined variables in the model, and Δ indicates that the corresponding variable is included in the equation. If it is identified, ** indicates that the variable in the equation for which the identification is made is not included in other equations of the model.

2.3 The TEM-CM Method

TEM-CM, developed by (Tyrychtr and Vasilenko 2015), is a simple method for creating multidimensional schemas for econometric OLAP design. This method involves several rules for the creation of the constellation schema. The first phase involves the transformation of econometric variables into dimensions and fact tables. The second phase involves the formation of a relationship between the dimensions and fact tables. This procedure is carried out as follows:

Phase 1: Creation of the primary constellation schema
Rule 1.1: Creation of a fact table in an empty schema for each endogenous variable from the EM.
Rule 1.2: Creation of dimensions of the schema for each exogenous variable from the EM.
Rule 1.3: Creation of a time dimension in the schema (if a time variable exists in the EM).
Phase 2: Creation of relationships in the schema
Rule 2.1: If there is a relationship between the exogenous and endogenous variables in the EM, create table-related associations between facts and dimensions in the schema.

3 Research Approach

We present our research approach, which consists of nine main phases for the creation of our innovative econometric OLAP framework (EM-OLAP). The methodological approach of this article is illustrated in Fig. 1.

Rectangles denote the individual phases of the creation process of the EM-OLAP framework. Rectangles with rounded corners denote the general scientific methods used to achieve relevant goals. The arrows show the sequence of methodological solutions. The new EM-OLAP framework is developed based on the following:

Identification: In this section, we formulate the roles of participants, which specify the utilization of the proposed EM-OLAP framework. We also define goals and formulate the hypothesis. All subsequent phases are based on these objectives and hypothesis.

Design of TEM: In this stage, we create analogies of the EM with a multidimensional schema and thus enable the description of the transformation of an EM into the conceptual and logical schemas of the multidimensional data model. The proposed transformation is performed according to the original TEM-CM methodology and compared to a new method simply called TEM, which we present later in this article.

Selection of variants of the TEM designs: Based on the proposed approaches to transforming an EM into a multidimensional schema, considering the measurement of their quality, we use the quality measurement presented by (Serrano et al. 2008; Gupta and Gosain 2010). Both methodological approaches are based on the measurement of the complexity of data warehousing (Calero et al. 2001). Within this phase, we compare the schema quality of the proposed TEM method with that of the original TEM-CM method. This phase results in a decision regarding which

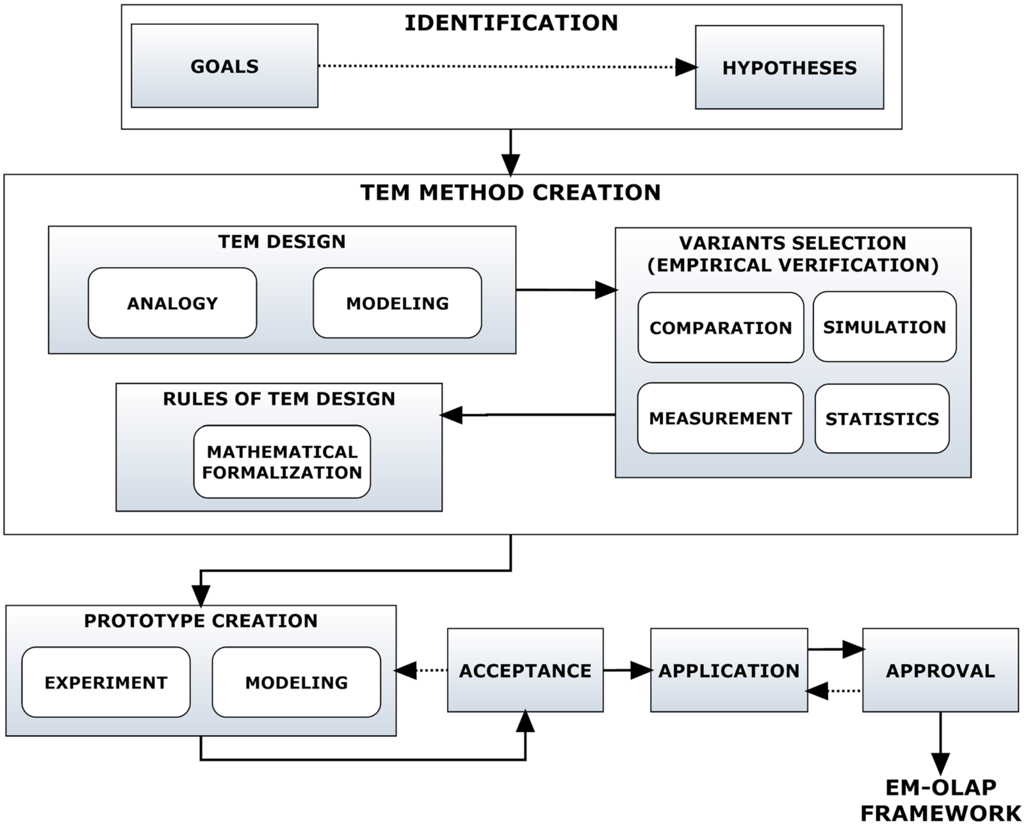


Fig. 1 Research methods for the creation of the EM-OLAP framework

approach is more suitable for transforming EMs into conceptual schemas.

Creation of rules for the TEM method: In this phase, we create formal rules for the new TEM method through the use of mathematical notation.

Creation of the prototype: We create a prototype of conceptual and logical multidimensional schema according to the TEM method. We also create a prototype of multidimensional OLAP databases and applications to ensure econometric analyses. We use Microsoft PowerPivot with Microsoft Excel 2013. The prototype allows us to gain advantages and constraints for creating an EM-OLAP framework to design the physical schema of the multidimensional database. We experiment with various forms of integrated data. To create the prototype, we use a simple production function within the conventional agriculture and then acquire the power forms of (Kroupová 2010):

$$y_{kt} = 205.113L_{kt}^{0.249}WU_{kt}^{0.525}K_{kt}^{0.143}. \qquad (3)$$

The chosen production function is applied to the final output *y*, which is estimated based on constant 2005 prices (measured in thousands of Czech crowns) for the comprehensive analysis of the impact of the fundamental factors of production. The explanatory variables are the following factors of production: land (L) is a hectare of utilized land, work (WU) is the average number of workers, and capital (K) is expressed as the sum of tangible and intangible fixed assets (in thousands of Czech crowns).

Design of the EM-OLAP framework: As a result of the abovementioned phases, we develop a new EM-OLAP framework to create econometrically based OLAP systems.

Acceptance. In this phase, we conduct a systematic experiment using the achieved process design of the EM-OLAP framework. This method is used in conjunction with the creation of a prototype.

Application. The accepted EM-OLAP framework is applied in real cases.

Approval. This final phase of our research coincides with the application phases. The aim of this phase is a new correction of the EM-OLAP framework to adapt it to changes in the application environment.

4 Identification Phase

The aim of the EM-OLAP framework is to provide system designers with methodological guidelines to facilitate the design of systems for decision-making support in econometric analyses. To improve clarity, we present the

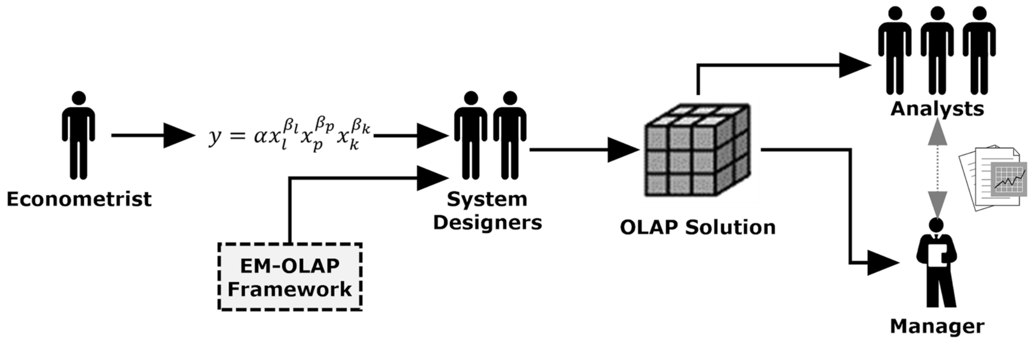


Fig. 2 Role of workers within EM-OLAP framework

following roles, which are directly or indirectly related to the utilization of the EM-OLAP framework (see Fig. 2):

Econometrist – a scientist or analyst that models economic reality using statistic, mathematic or economics instruments. He/she is an expert in economy and statistics and utilizes economic and statistic software. His/her work results in EM equations. He/she does not use an EM-OLAP framework. Rather, he/she only formally identifies the economic reality to design econometric systems. OLAP is not the substantial instrument of an econometrist.

System designer – someone that proposes a system design or architecture. He/she proposes an optimum balance between business needs and technological constraints. Econometric intelligent systems are a rather special part of system design but currently lack methodical guidelines. Thus, the EM-OLAP framework is directly designed to meet the needs of a system designer, enabling him/her to design econometric systems based on OLAP concepts.

Analyst – someone that directly works with the created OLAP solution. He/she performs econometric analyses (e.g., analyses of production factors, consumption changes or unit and marginal costs), creates key performance indicators (KPIs) and develops reports for decision makers (managers).

Decision maker (manager) – someone who evaluates the econometric analyses and proposes further steps.

Because no comprehensive approach to creating econometric systems using OLAP exists, our goal is to suggest a new EM-OLAP framework to improve the design of econometrics-based intelligent systems. We differentiate the main goal from the following subgoals:

1. Creation of the TEM method for the transformation of an EM into a multidimensional paradigm:
    - to perform a comparison of multidimensional schemas via measurements of data mart quality and
    - to create formal rules for the transformation of an EM.
2. Creation of the OLAP prototype allowing econometric analysis:
    - design of conceptual and logical schemas of a multidimensional database and
    - creation and implementation of the physical design of the OLAP prototype.

Table 1 Working hypothesis

Hypothesis
H1: A *possible transformation* of the EM into the physical schema for OLAP exists

To meet the first subgoal of creating the TEM method, we formulate the following working hypothesis (see Table 1).

5 Creation of the TEM Method

We now illustrate the proposed TEM method for a multidimensional database. We first describe the transformation method of TEM-CM and then that of the innovative TEM method. Then, we compare the quality of the resulting schemas and choose the appropriate transformation process. Finally, we use mathematical notation to describe the new method.

5.1 Proposal of TEM

First, we consider the EM with one equation:

$$y_t = \gamma_1 x_{1t} + \gamma_2 x_{2t} + \gamma_3 x_{3t}. \quad (4)$$

Equation (4) may represent a production, cost or any other function. Exogenous variables can represent, e.g., the amount of personnel, material conditions, and the number of aid grants. At this stage, understanding the meaning of each variable is not essential. Now, we use the original

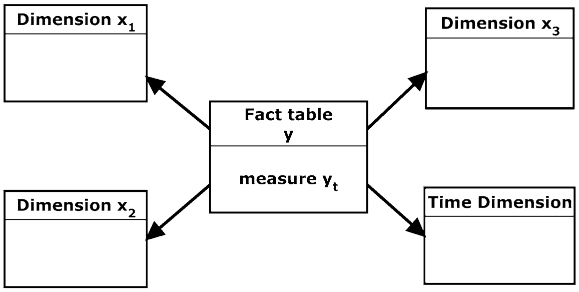


Fig. 3 Conceptual schema for the EM with 1 equation, according to the TEM-CM method

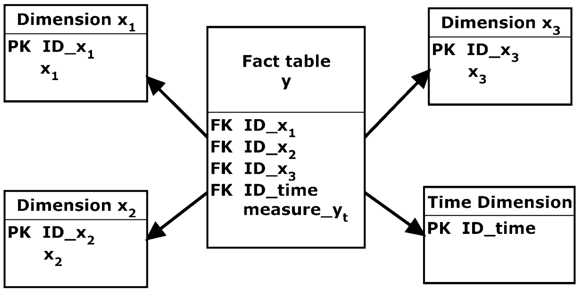


Fig. 4 Logical schema for the EM with 1 equation, according to the TEM-CM method

TEM-CM method to transform the EM (4) into the conceptual schema.

In the first phase of the conceptual design, we create a multidimensional database fact table in an empty conceptual schema (created according to the original TEM-CM method). Based on the EM Eq. (4), we can consider the value of the endogenous variable y as fact. Therefore, y represents the fact table. Exogenous variables x_1, x_2, and x_3 represent dimensions. Since the model contains a time variable t, we add the dimension of time to the schema. The fact table is associated with the roll-up relationship for all relevant dimensions, i.e., the variables on the right side of the equation. All notations of the equation represent the measure and thus serve as observed indicators, which will be part of the fact table. Thus, the created conceptual diagram is as shown in Fig. 3.

For the transformation into the logical schema, we provide each dimension with a numerical primary key and associate each with the fact table via a foreign key. The result of this step is demonstrated in Fig. 4.

The above transformation corresponds to the model with one equation. It is therefore appropriate to consider a more complex model, such as the model with three equations (i.e., model 2) described in Sect. 2. For this EM, we apply the following transformation.

In the first phase, we create the fact table in an empty schema for y_1, y_2 and y_3. Subsequently, we create a dimension in the schema for each exogenous variable in our EM (x_1, x_2, x_3, x_4 and x_5). Since the model contains a time variable t, we also create the dimension of time. We create roll-up table associations between fact tables and dimensions. Thus, for example, the equation $y_{2t} = \beta_{21} y_{1t} + \gamma_{21} x_{1t} + \gamma_{25} x_{5t} + u_{2t}$ indicates that the dimensions x_1 and x_5 are related to a fact table y_2. However, an endogenous variable y_1 appears in this second equation. A roll-up of the association between the fact table y_1 and the fact table y_2 must be created. For the transformation into the logical schema, each dimension is provided with a numerical primary key and associated with the fact table by a foreign key. It is necessary to monitor the measures that will be part of each fact table for each of the three equations. Random variables u_{1t}, u_{2t} are not illustrated in any conceptual or logical schema. The created logical schema is illustrated in Fig. 5.

The resulting schema is physically realizable only when we use ROLAP technology. For an MOLAP implementation, it would not be possible to connect each fact table, or in the MOLAP terminology, the data cubes. Therefore, in the case of simultaneous EMs, it is necessary to convert the model to a reduced form. In the case of model (2), the reduced form of the second equation would be as follows:

$$y_{1t} = \gamma_{11} x_{1t} + \gamma_{12} x_{2t} + \gamma_{13} x_{3t} + \gamma_{14} x_{4t} + u_{1t}$$
$$y_{2t} = \beta_{21} (\gamma_{11} x_{1t} + \gamma_{12} x_{2t} + \gamma_{13} x_{3t} + \gamma_{14} x_{4t}) + \gamma_{21} x_{1t} + \gamma_{25} x_{5t} + u_{2t}. \quad (5)$$

A simple substitution can be expressed using the equation without endogenous variables on the right side of the above equation. The consequence is that in the conceptual (logical) schema, the fact tables are not connected to each other. This model can be implemented for the MOLAP data store but at the cost of a high increase in interconnections between fact tables and dimensions. The model with one equation typically expresses a star schema, while the model with more equations corresponds to a constellation schema (in the case of simultaneous models) or a galaxy. Clearly, the generated logical schema of an EM with three equations (Fig. 5) is more complex than that of an EM with 1 equation (Fig. 4).

Fact tables in the TEM-CM should not be mutually interconnected. Instead, they should be connected by means of the shared dimensions. This could yield constellation or galaxy types of design, which would be more logical for such a design. The disadvantage of TEM-CM in this approach could be the technical design problems of the concrete OLAP platform. Some OLAP tools do not allow connecting several fact tables with the shared dimensions, while for others, it is difficult to do so. Generally, it is possible to connect fact tables with a shared table, but a problem occurs in several shared dimensions. Because of

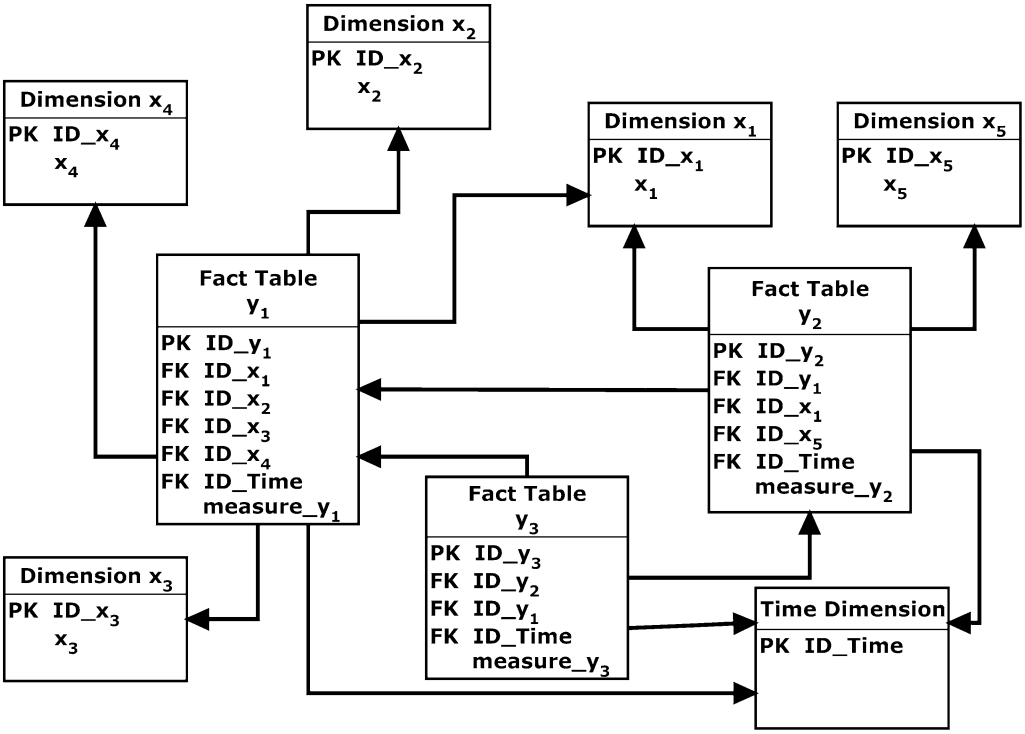


Fig. 5 Logical schema for EM with 3 equations, according to TEM-CM

this technical problem with shared dimensions, we consider other modeling techniques.

Given the above shortcomings of the original transformation methods, we now consider other structural designs of conceptual schemas. Thus, the above claim that y_t is a fact table can be replaced as follows. Consider only one fact table in the schema. Individual endogenous variables will not be expressed in individual fact tables in this scenario. Instead, each equation y_t represents one measure (indicator) of the fact table. We thus proceed in a similar manner. We create a dimension in the schema for each exogenous variable x_t of the EM and create a dimension of time. We create roll-up table associations between facts and dimensions. The result of this transformation approach is illustrated in Fig. 6.

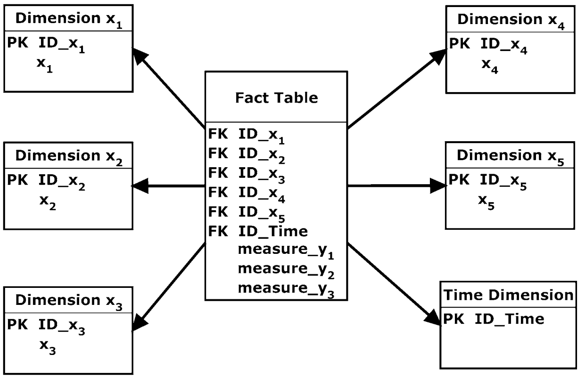


Fig. 6 Logical schema for EM with 3 equations, according to new version of transformation

This schema enables us to record the same econometric variable, as in the originally considered approach (Fig. 5). We focus on the comparison of these two variants of transformation of the schema arising from the above procedures and select the most suitable one to design a multidimensional database in the next part of this article.

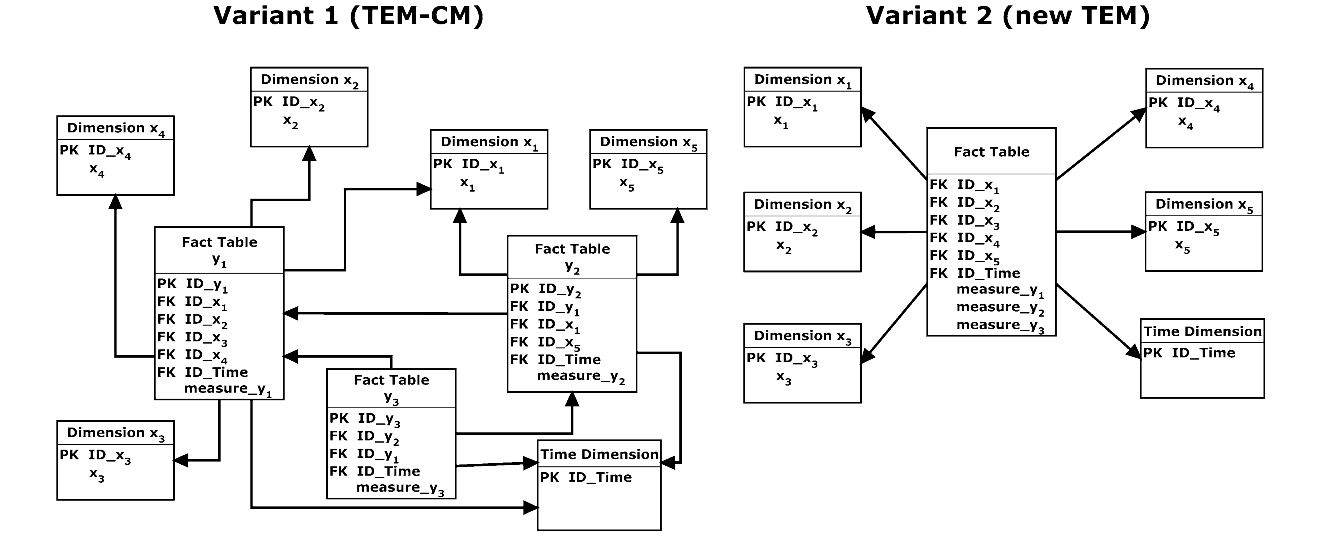
5.2 Comparison of Multidimensional Schemas

The two abovementioned variants of EM transformation are possible for the design of multidimensional schemas.

Given the generally increasing complexity of analytical databases, we should pay attention to the evaluation of their quality during their development. In this part of our work, we verify whether the results of measuring the quality and understandability of the multidimensional schema for the two abovementioned variants of the presented transformation are significant. This verification is important for determining which of the approaches described above should be selected to transform the EM.

5.2.1 Quantitative Comparison

We use a measurement of the quality of data marts developed by (Serrano et al. 2008; Gupta and Gosain 2010) for the quality assessment of the schemas. The first and


Variant 1 (TEM-CM)

Variant 2 (new TEM)

Fig. 7 TEM logical schema for EM with 3 equations, variants 1 and 2

Table 2 Result of quality assessment of the schemas

Variant 1 (TEM-CM)		Variant 2 (TEM)	
Measure	Value of measurement	Measure	Value of measurement
NFT(Sc)	3	*NFT(Sc)*	1
NSDT(Sc)	2	*NSDT(Sc)*	0
Sum	5	Sum	1

Table 3 Result of understandability assessment of the schemas

Variant 1 (TEM-CM)		Variant 2 (TEM)	
Measure	Value of measurement	Measure	Value of measurement
NFT(Sc)	3	NFT(Sc)	1
NDT(Sc)	6	NDT(Sc)	6
NFK(Sc)	12	NFK(Sc)	6
NMFT(Sc)	14	NMFT(Sc)	11
Sum	35	Sum	24

second variant of the transformation are evaluated. We illustrate both resulting schemas in Fig. 7.

The evaluation results are shown in Table 2. We measured several fact tables of the NFT schema and a number of shared dimension tables of the NSDT schema. Clearly, the schema created according to the first variant is structurally more complex; the total value of the measurements is greater than that of the second design type.

Another possible indicator of the quality of the resulting schemas is the measurement of understandability. Evaluation is performed again for the first and second transformation variant. We measured several fact tables of the NFT schema, several dimension tables of the NDT schema, and a number of foreign keys from all the fact tables of the NFK schema. $NFK(Sc) = \sum_{i=1}^{NFT} NFK(FT_i)$, where $NFK(FT_i)$ is the number of foreign keys in the fact table i of the schema Sc. The number of facts in the fact tables of the NMFT schema is determined using $NMFT(Sc) = NA(Sc) - NFK(Sc)$, where $NA(Sc)$ is the number of attributes in the fact tables of the schema Sc. The evaluation results are shown in Table 3.

The measurement results (Table 3) show that the first variant is significantly worse in terms of understandability. Again, the second alternative is more suitable for transforming the EM. Comparison of the proposed TEM method with our original TEM method was the main reason for testing the quality of the designed prototypes of the conceptual schemas. The results of this comparison clearly demonstrate that our proposed method can be used to design schemas with higher quality compared to those of the original method. This outcome occurs because the design type is changed from the original snowflake schema to the star schema, which generally offers designers better intuition.

6 The TEM Method

Based on the results of the quantitative comparison, we create a formalized new TEM method, which is based on the second variant of the multidimensional schema design.

6.1 Formal Representation

To formally define the rules of the TEM method, let us consider a set Y and set X, where:

$Y = \{y_s\} \cup \{y_{st}\}$ is a finite set of endogenous variables,
$X = \{x_r\} \cup \{x_{rt}\}$ is a finite set of exogenous variables and
$Rel \subseteq (X \times Y) \cup (Y \times Y)$ is a set of structural relations in the EM.

The star schema is any set with five elements (*Ent*, *Key*, *Att*, *Ass*, *getKey*), where:

Ent is a non-empty finite set of entities in the schema,
Key is a finite non-empty set of keys in the schema,
Att is a finite non-empty set of attributes in the schema,
$Fact \subseteq Ent$ is a finite set of facts in the schema,
$Dim \subseteq Ent$ is a finite set of dimensions in the schema, and
$Measure \subseteq Fact$ is a finite set of measures in the schema.

Each entity $e \in Ent$ is described by the collection of keys and attributes $\forall e \in Ent : \exists(\{k \in Key\} \cup \{a \in Att\})$.

getKey is a function that returns the *Key* entities in the star schema: $getKey(e) : Ent \rightarrow Key_e \subseteq Key$.

$Ass \subseteq (Dim \times Fact)$ is a finite set of relationships of the entities.

6.2 Design of Rules for the TEM Method

Phase 1: Creation of the basic star schema.
Rule 1.1: Creation of measures in an empty star schema for each endogenous variable of the EM, which is defined by:

$\forall y_s \in Y : m_s \in Measure$ and $\forall y_{st} \in Y : m_{st} \in Measure$.

Rule 1.2: Creation of the dimension in the star schema for each exogenous variable in the EM, which is defined by:

$\forall x_r \in X : d_s \in Dim$ and $\forall x_{rt} \in X : d_{rt} \in Dim$.

Rule 1.3: If there is a time variable in the EM, create the time dimension:

$\forall x_{rt} \in X : d_{rt} \in Dim_{time}$.

Phase 2: Creation of relations between entities in the star schema.
Rule 2.1: If there is a relationship between exogenous variable x, endogenous variable y and function *getKey* that returns a set of keys to these variables, then we create associations between the corresponding fact and the corresponding dimension:

$$\begin{aligned}\forall(x, y) \in Rel : (d, c, K) | (d \in Dim) \wedge (c \in Fact) \\ \wedge ((d, c) \in Ass) \wedge (K \subseteq K_d \cup K_c | \\ \times (K_d = getKey(d)) \wedge (K_c = getKey(c)))\end{aligned}$$

6.3 Application of the Rules of the TEM Method

To verify the rules, we consider EM (1) and the simplified semantic context of the example, where y_{1t} denotes industry production during the period t, y_{2t} denotes other production during the period t, y_{3t} is the total production during the period t, x_{1t} is quantity, x_{2t} is price, x_{3t} is market demand, x_{4t} is supply, x_{5t} is firm-specific information, and u_{1t}, u_{2t} are random components of the period t.

The example describes a situation in which the total production depends on industry production and other production. We should observe the different measures for each of these three endogenous variables. In the first phase, we create measures in the fact table in an empty star schema for y_{1t}, y_{2t} and y_{3t} (rule 1.1). Subsequently, in accordance with rule 1.2, we create a dimension in the star schema for each exogenous variable in our EM: quantity, price, market demand, supply and firm-specific information (e.g., product characteristics). Since model (1) includes a time variable t, the time dimension is created. In the last phase (rule 2.1), we form an association via the generated keys between the fact table and dimensions. Thus, for example, the equation $y_{2t} = \beta_{21}y_{1t} + \gamma_{21}x_{1t} + \gamma_{25}x_{5t} + u_{2t}$ indicates that the level quantity of products and firm-specific information have a relationship with other production (i.e., with the measure y_{1t} in the fact table). In the application context, the equation may be expressed as follows:

$$y_{1t} = 3.45x_{1t} + 1.32x_{2t} + 1.07x_{3t} + 0.43x_{4t} + 284.36$$

Thus, random components u_{1t}, u_{2t} and parameters β, γ are already expressed numerically. Therefore, random components u_{1t}, u_{2t} (or other variables that are not listed in the rules of the TEM method) are not depicted in the schema. The entire schema is thus shaped like a star in Fig. 6.

7 The Creation of the Prototype

To obtain an accurate preview of the future econometric OLAP solution, we create a prototype of a multidimensional database via the TEM method. Creation of the prototype will allow us to obtain benefits and limitations for the creation of the final form of the OLAP framework. To create the prototype, we follow the design of data marts of (Rizzi et al. 2006). To create the prototype, we use conventional production function (1).

7.1 Conceptual Design of the Prototype

To create a conceptual schema, we apply the rules of the TEM method. Application of the TEM method to production function (3) leads to the identification of measures and dimensions. Only one fact table exists for the entire schema. The measures identified by rule 1.1 are therefore a subset of the fact table.

The results of the applied rules (Table 4) allow the creation of the conceptual schema, which we further complete using a logical design.

7.2 Logical Design of the Prototype

A conceptual model is helpful for multidimensional database design, as it facilitates communication between OLAP users and a database designer. However, conceptual models should be converted into logical models for implementation in a database system. The data structure is already described in detail in the logical model regardless of its physical implementation in a database system. To this end, we should perform the following:

7.2.1 Find Relationships Between Different Sets of Entities

First, we apply rule 2.1 of the TEM method, which allocates an appropriate association with the fact table to each identified dimension in the logical design.

Table 4 Description of the results of the TEM

Rule 1.1	Measure	$y_{kt} = 205.113L_{kt}^{0.249}WU_{kt}^{0.525}K_{kt}^{0.143}$
Rule 1.2	Dimension	Land (L)
		Work (WU)
		Capital (K)
Rule 1.3	Dimension	Time

7.2.2 Specify Primary Keys for All Sets of Entities

We add primary keys ID_Land, ID_Work_ ID_Capital and ID_Time to each dimension table. We do not consider surrogate keys for our purposes despite the common use of the surrogate key design pattern to manage entities across disparate source systems.

7.2.3 Find All Attributes for Each Set of Entities

Since the TEM method does not affect the creation of attributes relative to the semantics of variables in the model, it is advisable to add other possible dimension attributes to the schema after using the TEM method. For the land dimension, we include the acreage attribute, which contains data regarding the hectare acreage of land. For the work dimension, we create the number attribute, which includes the number of workers. To the capital dimension, we add the size attribute, which is expressed as the sum of tangible and intangible fixed assets (in thousands of crowns).

7.2.4 Specify the Hierarchy of the Time Dimension

We add attributes to the time dimension that will allow econometric analysis in the long term. From an economic perspective, it is irrelevant to conduct an analysis in the short term, as most of the factors in the equation remain unchanged during the short term.

7.2.5 Identify Granularity and Approach of Slowly Changing Dimension

The type of data granularity must also be chosen. The snapshot granularity is suitable for an econometric analysis. Data are entered into a database with the same time intervals (e.g., every quarter). Thus, the time dimension considers both the year and the quarter. In these intervals, it is possible to identify changes in various dimensions. Generally, a need to follow changes in dimension attributes in the data mart to report historical data also exists. Two situations can occur in the context of econometric analyses:

1. An incorrect concrete variable value needs to be corrected. This can be done by overwriting the old value method. No history of dimension changes is stored in the database in this case. The old dimension value is simply overwritten with a new dimension. This alternative is easily maintainable for an econometric OLAP.
2. The measure calculation needs to be changed. This problem can be split according to two possible situations:

- The first corresponds to the situation in which an econometrist has changed parameters in the econometric equation. The original measure should be preserved, while a new one should be created. This process enables EM-OLAP users to see differences in the calculation of the old and new econometric equations. This situation has no influence on dimension changes.
- The second results from a need to add or remove a variable to/from the EM (to add or remove a relationship with a dimension to/from a measure). This need leads to principal difficulties with a granularity, which is more broadly discussed in Sect. 7.4.2). This problem can be solved by creating a new data model with a new fact table. A high data redundancy is the disadvantage of this solution.

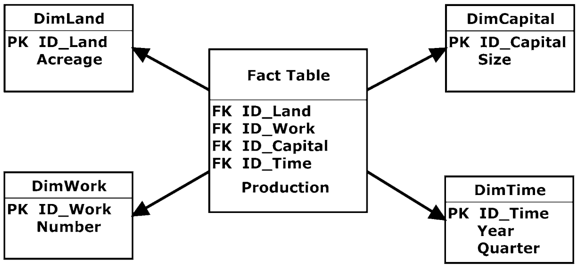


Fig. 8 Logical schema of the prototype

Figure 8 illustrates the resulting final logical schema.

7.3 Physical Design of the Prototype

After creating the logical schema, our next step is to design the physical schema. Essentially, we supplement the logical model with physical characteristics that are typical for OLAP technology and specific database systems. However, at this stage of acceptance of our solution, the optimum specific setting of the proposed database solution is not important, but the opportunity to examine the proposed logical model is. Therefore, we use Microsoft Excel 2013 and PowerPivot to develop the physical design of the prototype, which is sufficient to build our prototype.

First, we integrate data into the fact table and each dimension and create the relationship proposed by the logical schema (Fig. 9) using PowerPivot. Integrated data do not represent specific data of a single company. Data are averaged to represent a medium-sized firm in the period of 2010–2012 with an average number of workers in the interval $< 3;\ 6 >$, capital (millions CZK) in the range of $< 2.5;\ 4 >$ and land area (ha) in the interval $< 90;\ 120 >$. Integrated tables contain attributes designed using the logical schema. Other attributes are not included in the prototype, especially those that could add dimensions to an individual hierarchy.

The physical approach of the prototype design allows us to verify that the logical model proposed by the TEM

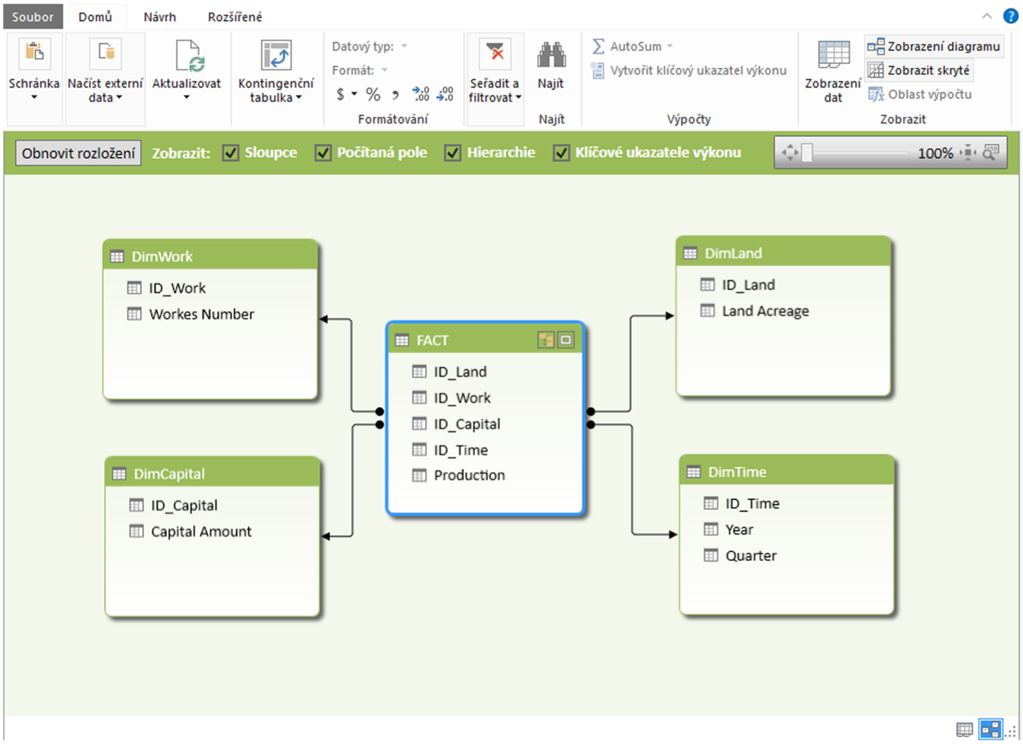


Fig. 9 Diagram of the prototype, developed using PowerPivot

method is feasible and has practical importance in the design of econometric-based intelligence systems. Depending on the needs of econometric analyses, it is particularly necessary to consider the form of the integrated data. Therefore, to realize a physical model of the prototype, we assume several variants of integrated data (referred to as A, B and C), which have an effect on the interpretation of the outputs of the OLAP system.

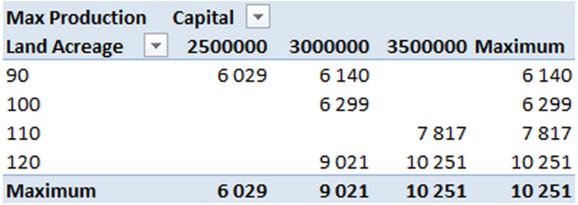
7.3.1 *Integrated Data – A*

The variant denoted as A represents the integrated data that reflect the current state of the factors of the company. For example, in the first quarter of 2016, the amount of land acreage was 125 ha. In the second quarter of 2016, the amount of acreage was 128 ha. Thus, the acquired data in each dimension reflect the current states of the real factors of the company at the time of acquisition.

For practical verification of the prototype of a multidimensional database, we create a PivotTable using PowerPivot (Fig. 10). This output is usually supported by all client applications for OLAP. Within the PivotTable design, we choose a production indicator that is calculated via a dimension in rows and columns (land and time). Due to the characteristics of the production resources, it is possible to perform an aggregation by taking the sum. However, for example, the summation of values of the resulting production for the first and second quarters of 2014 (6029 + 6140) cannot be interpreted correctly. The reason is that the outcome reflects the current status of production factors available for the period. Therefore, it is not possible to interpret the result such that the size of the production in the first two quarters is 12,169 CZK (s. c.). Hence, instead of summation, we apply maximization in the following form of the DAX language of the simplified example:

$$= \mathrm{MAX}\,('\mathrm{FACT}'\,[\mathrm{Production}])$$

This approach enables a clear view of each production size according to the selected factor and the progress of time. However, for econometric analysis, it is appropriate that the OLAP solution allows a factor analysis. In this approach, this process is possible for the actual combination recorded in the fact table. For example, Fig. 11 presents a pivot table with a sparse matrix. During the factor–factor analysis, it is possible to monitor the amount of the factor used for the creation of a specific production. For example, a production value of CZK 6029 (s. c.) is created using 90 ha of land and a capital value of 2.5 million CZK. However, it is not possible to determine the size of production, which could be achieved by using, for example, 100 ha of land.



Max Production	Capital ▼			
Land Acreage ▼	2500000	3000000	3500000	Maximum
90	6 029	6 140		6 140
100		6 299		6 299
110			7 817	7 817
120		9 021	10 251	10 251
Maximum	6 029	9 021	10 251	10 251

Fig. 11 PivotTable: factor–factor

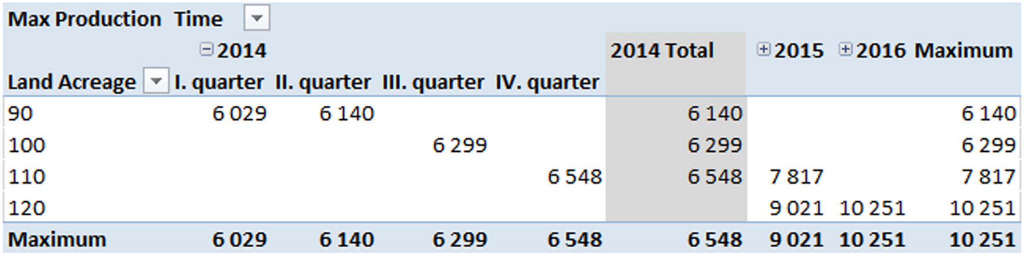
This variant of integrated data is practically feasible, but it is limited within the factor–factor economic analysis, which is important for most businesses.

7.3.2 *Integrated Data – B*

The data acquired in different dimensions reflect changes with respect to the previous state of the factors in the company. For example, the first data acquired is the acreage of the land (90 ha) in the first quarter of 2014, and the next data acquired is the amount of acreage (100 ha) in the third quarter of 2014. These data imply a change of + 10 ha from the first to third period.

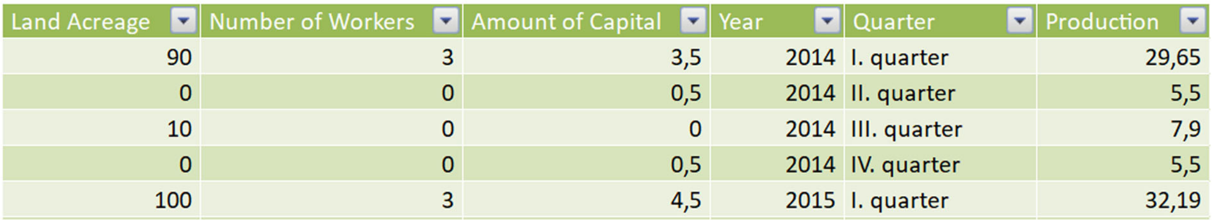
Although the data appear to be fully additive, the prototype implementation suggests the opposite, i.e., that data are non-additive. Because of the mathematical nature of the EM, it is not possible to aggregate data for the period from the first to fourth quarter due to a constant that is included in model (3). Such aggregation includes increased production of a constant for each quarter. Summation would correspond to 48.6 instead of 32.2 (Fig. 12).

It is also necessary to calculate the production using a linearized production function. The reason is that in the case of no change in the factor amount, it would be impossible to mathematically calculate the power function.



Max Production	Time ▼							
	⊟2014				2014 Total	⊞2015	⊞2016	Maximum
Land Acreage ▼	I. quarter	II. quarter	III. quarter	IV. quarter				
90	6 029	6 140			6 140			6 140
100			6 299		6 299			6 299
110				6 548	6 548	7 817		7 817
120						9 021	10 251	10 251
Maximum	6 029	6 140	6 299	6 548	6 548	9 021	10 251	10 251

Fig. 10 PivotTable: factor-time, showing the current state



Land Acreage	Number of Workers	Amount of Capital	Year	Quarter	Production
90	3	3,5	2014	I. quarter	29,65
0	0	0,5	2014	II. quarter	5,5
10	0	0	2014	III. quarter	7,9
0	0	0,5	2014	IV. quarter	5,5
100	3	4,5	2015	I. quarter	32,19

Fig. 12 Fact table in PowerPivot, showing the differences

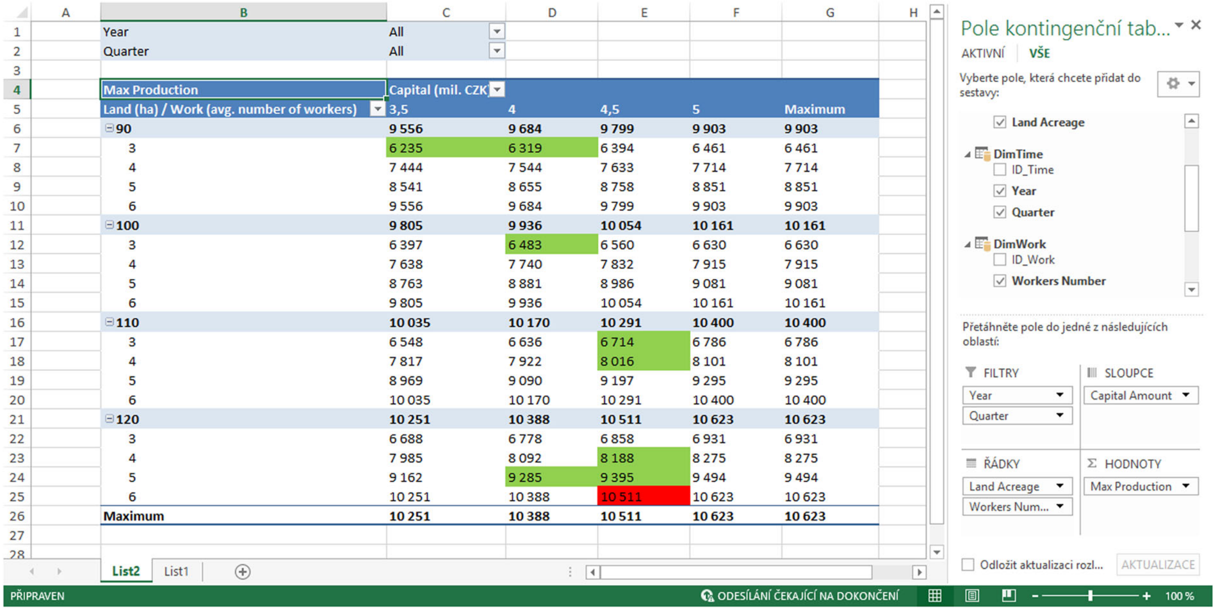
For the above reasons, we reject this data acquisition approach (in the form of differences).

7.3.3 *Integrated Data – C*

For the variant C, we integrate the data such that the data in the dimensions reflect *the current state of the factors* in the company and are created in the fact table *with all possible combinations of factors* that may theoretically occur. As does variant A, variant C reflects the various dimensions of the current state of the company at the time of data acquisition. However, unlike variant A, this type of integrated data also allows us to perform a factor–factor econometric analysis of the theoretical combinations of factors. The basis of the solution is the fact table, into which we record not only the actual combination of factors and corresponding calculated production but also all possible combinations of factors that the company can achieve.

We create the labels of the PivotTable rows based on the acreage and number of employees. The captions of the columns are created based on the size attribute of capital. The formed indicator represents the peak of production. The entire multidimensional data model can perform cuts by year and by quarter (Fig. 13). A rule that requires labeling the results of the factor combinations used at the company is created in the PivotTable. Dark grey represents the last calculated value based on real production of the number of factors (the value CZK 10.511). Grey indicates values from previous production periods. Unmarked items (without colors) represent the theoretical value of production. For example, the production value 6.235 CZK is achieved by using an acreage of 90 ha, an average number of workers equal to 3 and a capital of 3.5 million CZK. The management of this a company can deduce that the purchase of 10 ha of land while other factors remain constant makes it possible to achieve a production of 6.397 million CZK. A production of 7.444 million CZK can be achieved by increasing the average number of employees to 4.

This approach will allow company management to do the following:



Year All
Quarter All

Max Production	Capital (mil. CZK)				
Land (ha) / Work (avg. number of workers)	3,5	4	4,5	5	Maximum
⊟90	**9 556**	**9 684**	**9 799**	**9 903**	**9 903**
3	6 235	6 319	6 394	6 461	6 461
4	7 444	7 544	7 633	7 714	7 714
5	8 541	8 655	8 758	8 851	8 851
6	9 556	9 684	9 799	9 903	9 903
⊟100	**9 805**	**9 936**	**10 054**	**10 161**	**10 161**
3	6 397	6 483	6 560	6 630	6 630
4	7 638	7 740	7 832	7 915	7 915
5	8 763	8 881	8 986	9 081	9 081
6	9 805	9 936	10 054	10 161	10 161
⊟110	**10 035**	**10 170**	**10 291**	**10 400**	**10 400**
3	6 548	6 636	6 714	6 786	6 786
4	7 817	7 922	8 016	8 101	8 101
5	8 969	9 090	9 197	9 295	9 295
6	10 035	10 170	10 291	10 400	10 400
⊟120	**10 251**	**10 388**	**10 511**	**10 623**	**10 623**
3	6 688	6 778	6 858	6 931	6 931
4	7 985	8 092	8 188	8 275	8 275
5	9 162	9 285	9 395	9 494	9 494
6	10 251	10 388	10 511	10 623	10 623
Maximum	**10 251**	**10 388**	**10 511**	**10 623**	**10 623**

Pole kontingenční tab...
Vyberte pole, která chcete přidat do sestavy:
☑ Land Acreage
DimTime
□ ID_Time
☑ Year
☑ Quarter
DimWork
□ ID_Work
☑ Workers Number
Přetáhněte pole do jedné z následujících oblastí:
FILTRY: Year, Quarter
SLOUPCE: Capital Amount
ŘÁDKY: Land Acreage, Workers Num...
HODNOTY: Max Production
□ Odložit aktualizaci rozl...

Fig. 13 PivotTable: factor–factor, corresponding to the current state with combinations

- find a combination of several factors that leads to roughly the same level of production;
- identify the maximum value of production in the reporting period;
- derive the percentage change in the value of one factor during the change in value of the second factor and at a constant level of production.

7.4 Result of Prototyping

7.4.1 Existence of a Solution

Based on the conceptual, logical and physical design of the prototype and multidimensional database, *we accept hypothesis H1* regarding the existence of an allowable transformation of the EM into a physical schema for OLAP.

All mentioned variants of data integration store econometric equations in the form of multidimensional structures. We select variant C of the data integration to design the EM-OLAP framework, which also enables us to store theoretical values.

Storing econometric equations in the form of measures or calculating the column instead of using only the time series is advantageous, as fixed, stored time series are incompatible with theoretical values, which are important for planning changes in individual factors (variables, e.g., queries regarding the influence of an increased number of employees, stored goods, and/or capital on the overall production). This situation represents a difference with respect to classical OLAP solutions, which offer only a current view of the enterprise data and for which data mining instruments are needed to carry out further analytical works.

7.4.2 Limitations and Constraints of the Application

The prototyping result has some design constraints:

- Generally, OLAP focuses on a more effective analysis of a large number of events, which are related to combinations of a limited number of dimensions. Aggregation mechanisms are the advantage of this solution, yielding a better understanding of the observed process or event. Thus, several concepts in different dimensions must be limited. One should ensure that the dimension tables are somehow related to the fact tables.
- It is always necessary to set the range of dimension values according to a concrete economic reality and to predict these ranges. When these ranges are large, the data should be rounded or categorized. In our considered context (agriculture), it is easy to set the ranges of dimensions such as land acreage and number of employees. However, capital is a continuous quantity, and its concrete values should be constrained via rounding or categorization.
- Multiequation models should be treated with care when considering a granularity problem. Several variants should be considered for this design: (1) select only one endogenous variable as a measure and solve the remaining equations as a calculated column. (2) Create a measure for the selected variable as an aggregation function (e.g., average, maximum) and solve the remaining equations as a calculated column. (3) Convert the EM into a reduced form (one equation). (4) Create a separate data cube for each endogenous variable, with the relationships among individual variables being lost. The selected variant will depend on the econometric requirements of the created OLAP solution.

8 EM-OLAP Framework

The presented TEM method is a fundamental method for designing a multidimensional database for econometric analyses. However, this method itself is insufficient for a system designer if he/she does not know which constraints and associations this method applies in the design of a final data mart and OLAP.

Considering all the results from Sects. 6 and 7, we present the EM-OLAP framework. This framework is developed to support the design of a multidimensional database to secure econometric analyses to increase the effectiveness of the development of econometric intelligent systems. This framework is focused on the design of a multidimensional structure from production, cost or demand functions and supports the realization of OLAP via multidimensional databases. The components of the framework are shown in Fig. 14.

We propose the following specific framework processes:

1. *Analysis of requirements*. In this stage, the needs of the end users are examined within the context of the econometric analyses. We identify the type of econometric analysis required by the company (analysis of production costs or demand). We identify the requirement to analyze the relations between factors of production or results of production (e.g., factor–factor and product-factor relations). We classify the requirements for declaring the characteristics of the progress of functions (unit, marginal) and the corresponding flexibility. This stage can be refined during the life cycle of the design and ends with the physical design of a multidimensional database.

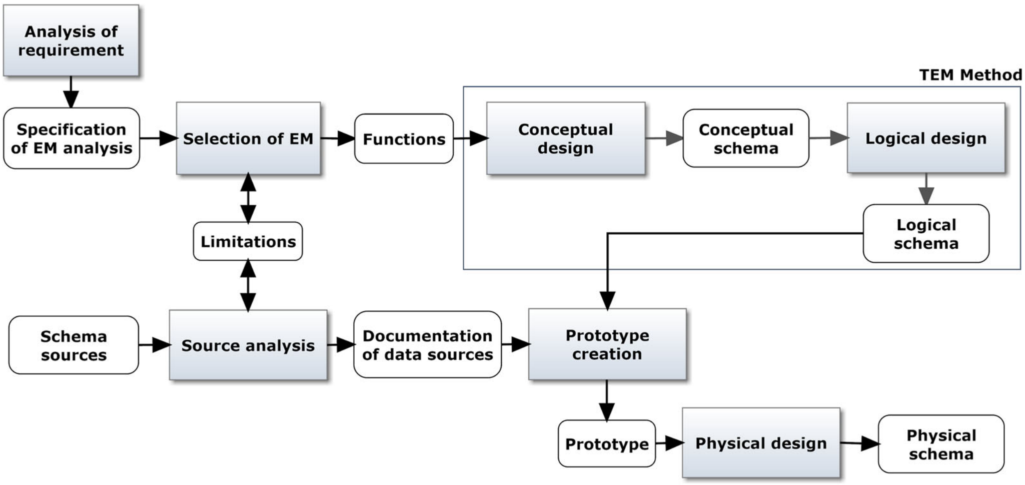


Fig. 14 EM-OLAP framework

2. *Selection of EM.* The aim of the EM selection phase is the selection of the EM type for subsequent econometric analysis. The model can be expressed in a structural or reduced form. However, it must already be assessed and usable for economic interpretation in the company. All variables in the model must be clearly interpretable, and the requirement that the values of these variables must be obtainable from the production resources of the enterprise must be satisfied (e.g., operational databases). The model may represent production or costs. For example, the demand function of the company can be expressed in a power form if none of the variables is equal to zero. If this condition is not met, then the form of a linear function must be chosen. The outcome is an overview of econometric analyses functions along with detailed documentation of the significance of variables and characteristics of their progress.
3. *Analysis of sources*. In this phase, the various schemas of data sources must be analyzed, and they must be aligned to obtain documentation of the data sources for integration. This phase occurs in conjunction with the EM selection stage. Knowledge of the econometric function for which the data will be integrated and knowledge of the available data sources are necessary to select appropriate econometric functions. The result of this phase is the documentation of data sources for integration into the prototype of a multidimensional database. The data to be integrated must satisfy the following conditions:

   (a) The data to be integrated into dimensions must reflect the current state of the factors of the company.
   (b) The fact table must be created using the actual combination of factors that yield the resulting value of the company.
   (c) Meanwhile, in the fact table, all possible combinations of factors that can theoretically occur must be created.
   (d) The number of concepts in each dimension should be limited. According to the economic reality, ranges should be set for continuous variables to ensure the correct function of the aggregation mechanisms. The data should be rounded or categorized when the ranges are broad.

4. *Creation of a data model.* The aim of this phase is to develop conceptual and logical schemas (according to the TEM method). To create a *conceptual schema,* the following steps must be performed:

   (a) Create measures for each endogenous variable from the EM *(Rule 1.1)*. Constraints given granularity should be considered in this step, as mentioned in Sect. 7.4.2).
   (b) Create dimensions for each exogenous variable from the EM *(Rule 1.2)*.
   (c) If there is a time variable in the EM, create the time dimension *(Rule 1.3)*.

5. *Creation of the prototype.* In this phase, the prototype for the validation of a proposed logical schema is created. We integrate data into the fact table and into each dimension. The next step is to create relations between the fact table and the dimensions. Then, the PivotTables, charts and other outputs can be generated for the OLAP. When calculating the indicators, the EM function type must be considered. For example, if the modeled function is linear and does not contain any constant, it is possible to perform summation in the calculation of measures. In other cases, only the aggregation of the maximum or average type makes

economic sense. The resulting display of data in a PivotTable must allow a factor–factor analysis. If any of the above activities ends in failure or if the result is not valid according to the requirements of the analysis, then the conceptual or logical schema must be remodeled.

6. *Physical design.* Acceptance of the prototype ends the process of designing the logical schema. In this stage, the physical properties of the database are determined based on the specific functions provided by the database system, such as indexing and partitioning.

9 Application and Acceptance of the EM-OLAP Framework

Our EM-OLAP framework was applied in the real environment of the AgroKonzulta Žamberk s.r.o. agricultural company, which has operated in the field of agriculture for over 20 years and is engaged not only in agricultural production but also in consultancy and software development for farmers. The requirement of management was to conduct econometric analyses of the production functions of organic farming, taking into account the impact of subsidies. The entire application of the EM-OLAP framework to the company was done using the Pentaho Business Analytics open source software solution, which is compatible with the MySQL relational database. The client application enabled the company management to obtain information regarding the total production over time in the form of PivotTables and graphs. To carry out the factor–factor analysis, which was implemented so that the user could select different combinations of factors (e.g., labor, land, capital, direct payments, and price), the entire dashboard system was set up to highlight key indicators that represent a combination of factors that yield approximately the same level of production. We created additional special outputs that allowed for individual factors to be used to obtain information regarding the production unit, marginal production and production flexibility. The calculations allowed us to model possible variants of the economic evolution and significantly helped the management to make adequate decisions to balance the economics of the company.

10 Discussion

In our work, a new EM-OLAP framework that supports the development of econometric intelligent systems was introduced. Below, we judge the validity of the results achieved during individual phases of the methodology presented in this article:

Design of the TEM method. Given the nature of the used method of analogy as a thought process, the conclusions of the analogy clearly lack the characteristic of irrefutable claims. Therefore, other permissible transformations of the EM into conceptual and logical schemas may exist.

Selection of variants of TEM designs. The quality of the proposed schemas was measured according to scientific methods for measuring data marts presented by (Serrano et al. 2008; Gupta and Gosain 2010). In this area, new ways to measure the quality of multidimensional schemas are continually being developed. Therefore, we cannot evaluate the use of other approaches.

Creation of rules for the TEM method. The formal notation of the TEM method was created via a mathematical apparatus gradually derived step-by-step instead of via the formulation of definitions, theorems and mathematical proofs. The TEM method was successfully presented at the 9th European Computing Conference (Tyrychtr and Vrana 2016).

Creation of the prototype. The creation of the prototype of conceptual and logical schemas (according to the TEM method) and the subsequent creation of the physical schema of a multidimensional database allowed us to accept hypothesis H1. To design a physical schema, we experimented with different variants of integrated data. All variants were based only on data suitable for the analysis of production functions. Evidently, the physical design demonstrated the ability to identify different approaches when different types of econometric context are proposed. In future research, the proposal of physical access (e.g., in the context of cost and demand functions) should be considered.

Acceptance. Several potential problems hindering the adoption of the TEM method exist. According to the design principles of a multidimensional database, fact table measures should be connected to only the combinations of dimensions that determine their values. Thus, only measures sharing all dimensions should be incorporated into the fact table. As a result, the following two possible situations can occur in the design of a *multi-equation model*:

A. Entries in the fact table have a relationship with the NULL element for each dimension, which is not related to the measure within this entry. Introducing a calculated column and selecting only one equation as a measure may be one solution to this problem. For example, y_1 or y_2 from (5) should be solved as a calculated column, and a measure should be defined, e.g., as an average of the production y_2.

B. Entries in the fact table contain measures related to all dimensions. This situation leads to incorrect aggregated results and incorrect semantics. This problem does not occur when a reduced form of the EM exists (i.e., an EM with one equation). For example, y_2 in (5) can be the only measure available in the OLAP model.

The abovementioned problems address multiequation models, which are rare in the current econometrics. The TEM method can be used without these constraints for one-equation models depending on the concrete EM used (which can be based on various methods, e.g., deterministic frontier models, stochastic frontier model, panel data models, and estimation of the technical inefficiency), which may or may not allow transformation into a reduced form.

OLAP is sometimes carried out in a non-standard way:

- Only measures related to all dimensions are stored in a standard way (if possible) in the fact table. In practice, data cubes are created with respect to various areas.
- However, the occurrence of non-standard solutions in which measures are not related to all dimensions is not an exception. These solutions are built using one fact table and multiple dimensions. This, however, leads to (1) many NULL elements in a data cube and (2) the user knowing the correct combinations and when to use a certain measure with a particular dimension.

Thus, we limited our article to existential design solutions, i.e., to the idea that some solutions for the econometric OLAP analyses exist.

Application. In our work, we applied the EM-OLAP framework to and verified it by considering only one company. One primary obstacle to the application of the framework to other types of businesses was the high cost of implementation of the overall EM-OLAP solution. Another obstacle to the validation of the EM-OLAP framework was the rather large complexity of the application of OLAP approaches. In future research, the understandability of the entire solution, its cohesion, its economic impact and the efficiency of testing the applied technology using the final solution must be measured. The progress of the design of EM-OLAP also motivates further research on its integration into existing design methods to allow parallel design of the classic OLAP systems and the proposed EM-OLAP systems. The effects of other approaches (e.g., data mining and competitive intelligence) were not considered in the framework. Despite some setbacks regarding validation of the framework, the resulting EM-OLAP framework allows a company to conduct econometric analysis without company management possessing in-depth knowledge of it. In our work, we placed great emphasis on the application of the production function for OLAP.

11 Conclusions

The motivation of our research was to enable companies to reach full production power and technical efficiency. In this article, we sought approaches capable of facilitating the development of econometric intelligent systems. These systems can easily help company management to interpret an econometric analysis, from which it is possible to obtain relevant knowledge regarding their economic performance. In this article, we presented the EM-OLAP framework. The goal of this framework is the transformation of EMs into multidimensional schemas. The input consists solely of econometric equations that are transformed, via our innovative TEM method, into conceptual and logical schemas of analytic databases. Along with the analysis of data sources, we searched for a prototype that can enable the implementation of the requirements of econometric analysis. The output of the EM-OLAP framework is a physical schema of a multidimensional database based on econometric models, which is applicable to online analytical processing in intelligent systems.

The proposed framework provides system engineers a methodological framework for designing the structures of multidimensional databases. This approach, based on OLAP client applications, enables us to obtain analytical data and present it using dashboards in the form of PivotTables, graphs and other special outputs. Information regarding total production costs or consumption can be developed in real time for the whole company or its parts. The key benefit is the ability to perform factor–factor analysis, which can be implemented in a manner that allows the user to select different combinations of factors (e.g., number of employees, price, quantity of input factors, and population in the region). We can determine combinations of factors that lead to approximately the same level of production and create additional special outputs that provide other economic information for individual factors (such as unit production, marginal production or production elasticity). Finally, we can seek the best combination of factors to maximize production or, conversely, to reduce costs and to help improve company efficiency. Because similar research on econometric analysis via OLAP has not been carried out, the results and benefits presented in this article offer new insights into the development of econometric intelligent systems.

Acknowledgements This work was conducted as part of the following projects: DEPIES – Decision Processes in Intelligent Environments (Grant no. 15-11724S) and ADEPINE – Advanced Decision Problems in Intelligent Environments, funded by the Czech Science Foundation, Czech Republic, and within the Economic value of analytical systems in agriculture (Grant no. 20161019) founded by the IGA of the Faculty of Economics and Management, Czech University of Life Sciences in Prague.

References

Abdelhédi F, Zurfluh G (2013) User support system for designing decisional database. In: ACHI 2013: the sixth international conference on advances in computer-human interactions, Nice, 24 February–1 March 2013

Abdullah A (2009) Analysis of mealybug incidence on the cotton crop using ADSS-OLAP (Online Analytical Processing) tool. Comput Electron Agric 69(1):59–72. https://doi.org/10.1016/j.compag.2009.07.003

Abelló A, Romero O (2009) On-line analytical processing. In: Liu L, Özsu MT (eds) Encyclopedia of database systems. Springer, Heidelberg, pp 1949–1954. https://doi.org/10.1007/978-0-387-39940-9_252

Assaf T, Dugan JB (2007) Decision automation for predictive analysis models. In: RAMS'07: reliability and maintainability symposium, Orlando, January 2007. Annual. IEEE, pp 335–340. https://doi.org/10.1109/rams.2007.328136

Ballard C, Herreman D, Schau D, Bell R, Kim E, Valencic A (1998) Data modeling techniques for data warehousing. IBM Corporation International Technical Support Organization, p 25 (SG24-2238-00)

Belsley DA, Kontoghiorghes E (eds) (2009) Handbook of computational econometrics. Wiley, Hoboken, p 514, ISBN 978-0-470-74385-0

Bimonte S, Pradel M, Boffety D, Tailleur A, André G, Bzikha R, Chanet JP (2013) A new sensor-based spatial OLAP architecture centered on an agricultural farm energy-use diagnosis tool. Int J Decis Support Syst Technol (IJDSST) 5(4):1–20. https://doi.org/10.4018/ijdsst.2013100101

Boehnlein M, Ulbrich-vom Ende A (1999) Deriving initial data warehouse structures from the conceptual data models of the underlying operational information systems. In: Proceedings of the 2nd ACM international workshop on data warehousing and OLAP, Kansas City, 2–6 November 1999, ACM, pp 15–21. https://doi.org/10.1145/319757.319780

Brandl B, Keber C, Schuster MG (2006) An automated econometric decision support system: forecasts for foreign exchange trades. Cent Eur J Oper Res 14(4):401–415. https://doi.org/10.1007/s10100-006-0013-8

Bravo-Ureta BE, Solís D, López VHM, Maripani JF, Thiam A, Rivas T (2007) Technical efficiency in farming: a meta-regression analysis. J Product Anal 27(1):57–72. https://doi.org/10.1007/s11123-006-0025-3

Brown JE, Ethridge DE, Hudson D, Engels C (1995) An automated econometric approach for estimating and reporting daily cotton market prices. J Agric Appl Econ 27(2):409–422. https://doi.org/10.1017/S1074070800028467

Calero C, Piattini M, Pascual C, Serrano MA, Piattini M, Genero M, Calero C, Polo M, Ruiz F (2001) Towards DW quality metrics. In: Proceedings of the international workshop on design and management of data warehouses (DMDW 2001), Interlaken, 4 June 2001, ISSN 1613-0073

Čechura L (2014) Analysis of the technical and scale efficiency of farms operating in LFA. AGRIS On-line Pap Econ Inf 6(4):33–44 (ISSN: 1804-1930)

Čechura L, Hálová P, Kroupová Z, Malý M, Peterová J, Šobrová L (2017) Cvičení z ekonometrie (The Practice of Econometrics). Česká zemědělská univerzita, Provozně ekonomická fakulta. ISBN 978-80-213-2405-3

Chaudhuri S, Dayal U (1997) An overview of data warehousing and OLAP technology. ACM Sigmod Rec 26(1):65–74. https://doi.org/10.1145/248603.248616

Dolk DR, Kridel DJ (1991) An active modeling system for econometric analysis. Decis Support Syst 7(4):315–328. https://doi.org/10.1016/0167-9236(91)90061-F

Felipe J, Adams FG (2005) "A theory of production" the estimation of the Cobb–Douglas function: a retrospective view. East Econ J 31(3):427–445

Fountas S, Carli G, Sørensen CG, Tsiropoulos Z, Cavalaris C, Vatsanidou A, Liakos B, Canavari M, Wiebensohn J, Tisserye B (2015) Farm management information systems: current situation and future perspectives. Comput Electron Agric 115:40–50. https://doi.org/10.1016/j.compag.2015.05.011

Greene WH (2015) Econometric software. In: Wright JD (eds) International encyclopedia of the social & behavioral sciences, 2nd edn. Elsevier, Oxford, pp 1–7. https://doi.org/10.1016/b978-0-08-097086-8.71006-3

Gupta R, Gosain A (2010) Analysis of data warehouse quality metrics using LR. In: Information and communication technologies, ICT 2010, Kochi, Kerala, India, 7–10 September 2010. Springer, Heidelberg, pp 384–388. https://doi.org/10.1007/978-3-642-15766-0_60

Han M, Ju C (2008) Research and application on OLAP-based farm products examination model. In: International symposium on electronic commerce and security, August 2008, IEEE, pp 858–861. https://doi.org/10.1109/isecs.2008.156

Karmakar S, Laguë C, Agnew J, Landry H (2007) Integrated decision support system (DSS) for manure management: a review and perspective. Comput Electron Agric 57(2):190–201. https://doi.org/10.1016/j.compag.2007.03.006

Kroupová Z (2010) Technická efektivnost ekologického zemědělství České republiky. Ekonomická revue 13:61–73. https://doi.org/10.7327/cerei.2010.06.01

Küsters U, McCullough BD, Bell M (2006) Forecasting software: past, present and future. Int J Forecast 22(3):599–615. https://doi.org/10.1016/j.ijforecast.2006.03.004

Levene M, Loizou G (2003) Why is the snowflake schema a good data warehouse design? Inf Syst 28(3):225–240. https://doi.org/10.1016/S0306-4379(02)00021-2

Nilakanta S, Scheibe K, Rai A (2008) Dimensional issues in agricultural data warehouse designs. Comput Electron Agric 60(2):263–278. https://doi.org/10.1016/j.compag.2007.09.009

Nowak A, Kijek T, Domanska K (2015) Technical efficiency and its determinants in the European Union agriculture. Agric Econ 61(6):275–283. https://doi.org/10.17221/200/2014-AGRICECON

Pedersen TB (2009) Cube. In: Liu L, Özsu MT (eds) Encyclopedia of database systems. Springer, Heidelberg, pp 538–539. https://doi.org/10.1007/978-0-387-39940-9_884

Rai A, Dubey V, Chaturvedi KK, Malhotra PK (2008) Design and development of data mart for animal resources. Comput Electron Agric 64(2):111–119. https://doi.org/10.1016/j.compag.2008.04.009

Recio B, García-Mouton E, Castellanos MT, Morató MC, Ibáñez J (2010) An econometric system to assess the economic impact of water restriction policies in Spain. Span J Agric Res 8(3):526–537. https://doi.org/10.5424/sjar/2010083-1248

Renfro CG (2004) A compendium of existing econometric software packages. J Econ Soc Meas 29(1–3):359–409 (ISSN 07479662)

Rizzi S, Abelló A, Lechtenbörger J, Trujillo J (2006) Research in data warehouse modeling and design: dead or alive? In: DOLAP '06 proceedings of the 9th ACM international workshop on data warehousing and OLAP, Arlington, 10 November 2006. ACM, pp 3–10. https://doi.org/10.1145/1183512.1183515

Schulze C, Spilke J, Lehner W (2007) Data modeling for precision dairy farming within the competitive field of operational and analytical tasks. Comput Electron Agric 59(1–2):39–55. https://doi.org/10.1016/j.compag.2007.05.001

Serrano MA, Calero C, Sahraoui HA, Piattini M (2008) Empirical studies to assess the understandability of data warehouse schemas using structural metrics. Softw Qual J 16(1):79–106. https://doi.org/10.1007/s11219-007-9030-7

Tvrdoň J (2006) Ekonometrie (Econometrics). Česká zemědělská univerzita v Praze, Praha. ISBN 80-213-0819-2

Tyrychtr J, Vasilenko A (2015) Transformation econometric model to multidimensional databases to support the analytical systems in agriculture. AGRIS On-line Pap Econ Inf 7(3):71–77 (ISSN 18041930)

Tyrychtr J, Vrana I (2016) Towards econometric OLAP design in the intelligence system. WSEAS Transact Bus Econ 13(59):627–633 (E-ISSN 2224-2899)

Uyan M, Cay T, Akcakaya O (2013) A spatial decision support system design for land reallocation: a case study in Turkey. Comput Electron Agric 98:8–16. https://doi.org/10.1016/j.compag.2013.07.010

Vassiliadis P, Sellis T (1999) A survey of logical models for OLAP databases. ACM Sigmod Rec 28(4):64–69. https://doi.org/10.1145/344816.344869

Wu MC, Buchmann AP (1997) Research issues in data warehousing. In: Datenbanksysteme in Büro, Technik und Wissenschaft. Springer, Heidelberg, pp 61–82. https://doi.org/10.1007/978-3-642-60730-1_5

Yu L, Wang S, Lai KK (2008) Forecasting China’s foreign trade volume with a kernel-based hybrid econometric-AI ensemble learning approach. J Syst Sci Complex 21(1):1–19. https://doi.org/10.1007/s11424-008-9062-5