On the efficiency of cloud providers: A DEA approach incorporating categorical variables

E. Filiopoulou, P. Mitropoulou, N.Lionis, and C. Michalakelis[1]

***Abstract* — Cloud computing is a growing industry and it has already dominated many IT markets segments. Cloud providers offer numerous equivalent IaaS services aiming to fulfill clients' requirements. In addition, cloud clients need to choose solutions that minimize costs without compromising efficiency though. However, not only the confusion due to the large variety of cloud services but also the uncertainty about the efficiency specifications that their cloud services should have, can make cloud computing services selection a difficult task for the users. Into this context, this paper presents an approach to a multi-attribute decision-making problem that focuses on the calculation of efficiency of IaaS cloud services as a measurable driver for both clients and providers. A DEA input-oriented model is described, which estimates efficiency of cloud services based on functional and non-functional parameters. Furthermore, the contribution of functional and non-functional features to the overall performance is examined. This innovative model urges providers to optimize the efficiency of their services aiming to increase their market share and, at the same time, assists clients in choosing a cost-effective cloud solution.**

***Index Terms*— Cloud Services, Data Envelopment Analysis, Efficiency, Functional Parameters, Non-Functional Parameters.**

I. INTRODUCTION

Cloud computing is an industry in exponential growth and it is anticipated to continue developing at a robust rate for the following years. This general purpose technology, by offering on-demand self-service, broad network access, resource pooling, rapid elasticity and measured service, provides a fundamental contribution to promote productivity, growth and competition among companies [1]. Therefore, more and more companies are switching to these facilities, instead of expanding their own datacenters.

Clients that decide to migrate to cloud environment should be able to understand the differences between IaaS, PaaS and SaaS platforms and also identify the market leaders in each category. Even though cloud computing is not a new technology but a different way of contracting for services and technologies, it can be complex, fragmented and confusing for potential cloud clients. Clients try to overcome all the complexities of the cloud environment by choosing cloud services that fulfill their requirements, but mainly combine advantageous price and high performance.

Cloud performance depicts the cost-effectiveness of cloud services and it helps clients to choose not only the optimal cloud solution but also the most cost-effective one. It evaluates numerous equivalent cloud services from different providers in an objective perspective [2]. This study adopts Data Envelopment Analysis (DEA), a non-parametric method that estimates empirically productive efficiency of decision making units, and proposes a DEA-oriented model methodology to assess cloud performance and rank cloud services by relative efficiency [3-4]. Functional and non-functional requirements together with price describe each cloud bundle of the sample data. Functional requirements define the straightforward cloud services, while non-functional parameters refer to the anticipated quality of services and indicate the constraints under which services should operate.

Into the context of the work performed in this paper, an innovative approach is developed emphasizing on the efficiency of cloud services and indicating the contribution of non-functional parameters to the overall performance. Furthermore, the proposed model can constitute a powerful tool for cloud clients and providers, since it can guide clients through cloud selection process and can also prompt providers to enrich and upgrade their services aiming to performance improvement. This work intends to fill a gap in the literature about the impact of non-functional parameters to relative efficiency. It also highlights the fact that price, despite the fact that it is a quantitative and explicit variable, it is not a decisive indicator for cloud selection.

The rest of the paper is organized as follows: Section II presents the related work, while Section III introduces a general description of DEA technique, emphasizing on the incorporation of non-functional parameters. Section IV

Submitted: July, 6th 2017
E. Filiopoulou is with the Department of Informatics and Telematics, Harokopio University of Athens, Omirou 9 str. 177 78, Athens Greece (e-mail: evangelf@hua.gr)
P. Mitropoulou is with the Department of Informatics and Telematics, Harokopio University of Athens, Omirou 9 str. 177 78, Athens Greece (e-mail: persam@hua.gr)
N. Lionis is with the Department of Informatics and Telecommunications, National and Kapodistrian University of Athens, Panepistimiopolis, Ilisia 157 84, Athens, Greece (e-mail: nlionis@di.uoa.gr).
C. Michalakelis is with the Department of Informatics and Telematics, Harokopio University of Athens, Omirou 9 str. 177 78, Athens Greece (e-mail: michalak@hua.gr).

develops the proposed methodology and section V presents the evaluation results. Finally, Section IV concludes the paper and suggests future work.

II. Related Work

There are several proposed schemes that discuss the cloud service performance, as well as approaches that combine cloud services selection based on performance and employing the DEA methodology. To this end, the body of the relevant literature includes studies that address cloud services selection based on performance and studies that discuss cloud services selection based on DEA.

Brebner and Liu [5] use a suite of cloud testing applications on a variety of cloud infrastructures, including Google App Engine, Amazon EC2, and Microsoft Azure. Their evaluations were combined with a Service Modeling Technology that model performance and scalability of applications, in order to predict the resource requirements in terms of cost and performance.

Ostermann and Losup [6] analyze the performance of cloud computing services for scientific computing workloads of Many -Task Computing (MTC) users. They present an empirical evaluation of the performance of four cloud computing providers, including Amazon EC2 that is the leading cloud industry. Using simulation they compare performance and cost models of each cloud provider.

Furthermore in [7] a comprehensive comparison of four public cloud providers based on performance, is examined. The performance is described by metrics that characterize cloud performance. Performance metrics have to be understandable and fair due to the numerous cloud providers.

Park and Jeong [8] propose a QoS model for SaaS ERP that fulfills six criteria; Functionality, Usability, Business, Reliability, Maintainability, Efficiency. Based on this QoS model, they also propose a Multi Criteria Decision Making (MCDM) that returns the SaaS ERP which meets the above criteria and the user requirements. In this paper Social Network Group was used in order to get opinions by various types of expert groups. Finally, two personalized QoS ranking prediction approaches are presented in [9], which predict the QoS rankings directly. Extensive real-world experiments are conducted employing real-world QoS data, including globally 300 distributed users and 500 real-world Web services. The experimental results show that our approaches outperform other competing ones.

Kumar and Saurabh [10] rank the cloud services based on performance, including price as well. Their model helps customers to choose the cloud solution that fulfills their requirement but it also helps cloud providers to improve their services. For the performance calculation Data Envelopment Model (DEA), Analytic Hierarchy Process (AHP) and Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) model have been used [10].

In addition, a non-parametric method is proposed in the work performed in [11] which evaluates the relative efficiency of cloud services based on Data Envelopment Analysis. It ranks cloud services into different efficiency levels and indicates solutions towards performance improvement. The evaluation of the performance was based on a group of IaaS services.

The Data Envelopment Analysis methodology is also used in [12] as an approach to performance measurement of cloud computing platforms, considering the importance of each resource in a specific application by defining the weight for each resource. Two benchmark suites were used for the evaluation, High-Performance Computing Challenge (HPCC) and Phoronix Test Suite (PTS).

Following the literature review it becomes obvious that the DEA methodology has been applied to cloud IaaS services, with the limitation that these services are described only by functional parameters such as CPU, storage and memory. However, non-functional factors which describe system attributes such as security, reliability, performance are as critical as the functional features, since they define quality and constraints of cloud services. Cloud vendors that provide options described by non-functional parameters enhance their cloud services and become more competitive, efficient and profitable.

The proposed DEA-oriented model estimates the relative efficiency of cloud services and focuses on the impact of non-functional parameters on efficiency. This approach aims to fill an existing gap in literature about the application of DEA, taking non-functional parameters under consideration and studying their key role on performance.

III. Data Envelopment Analysis (DEA)

A. General Description

Data Envelopment Analysis (DEA) is a performance evaluation methodology and a benchmarking technique and has been widely studied, used and analyzed gaining increasing popularity among researchers. The method was originally introduced by Charnes, Cooper and Rhodes [3], in 1978, as a mathematical programming model to evaluate non-profit and public sector organizations. The method has been proven significantly useful to locate ways for assessing the comparative efficiencies of decision-making units (DMUs; e.g. banks, schools, hospitals, factories, etc.), especially when the presence of multiple inputs and outputs makes comparison with other techniques difficult [10, 13, 14].

The usual measure of efficiency is simply defined as the ratio of output to input: the more the output per unit of input achieved, the greater the relative efficiency is. This kind of measure is often inadequate in more complex situations, such as the cloud computing services selection. This is due to the existence of multiple outputs and inputs related to different resources, activities and environmental factors and due to the numerous DMUs being evaluated are only relatively homogeneous and cannot be easily analyzed [13]. This is the reason why DEA is a really powerful tool able to evaluate the efficiency of a number of producers or units and allows efficiency to be measured without having to specify either the form of the production function or the weights for the different inputs and outputs chosen [14]. As a linear based multi-criteria decision making methodology, it compares each unit with only the "best" units, identifying not only the most efficient units or best practice units but also the inefficient ones, in which real efficiency improvements are possible. Finally, the DEA method defines a non-parametric best practice frontier that can be used as a reference for efficiency measures and calculates the numerical coefficients given to each DMU, estimating its relative efficiency [4, 13].

A common measure for relative efficiency of a many input–many output DMU is the following:

$$Efficiency = \frac{weighted\ sum\ of\ outputs}{weighted\ sum\ of\ inputs}, \tag{1}$$

which, after introducing the usual notation for a DMU with m outputs and n inputs, can be written as:

$$Efficiency\ of\ DMU\ j = \frac{u_1 y_{1j} + u_2 y_{2j} + \ldots + u_m y_{mj}}{v_1 x_{1j} + v_2 x_{2j} + \ldots + v_n x_{nj}} \tag{2}$$

where u_i is the weight given to output i, v_i is the weight given to input i, y_{ij} is the amount of output i from *DMU j* and x_{ij} is the amount of input i in *DMU j*.

According to the above equations, the measurement of the relative efficiency of a DMU, with multiple possibly inputs and outputs, is achieved by constructing a hypothetical efficient unit, as a weighted average of efficient units, to act as a comparator for an inefficient unit [15]. Towards this direction, the application of a common set of weights across all DMUs is not a simple action. In some cases it is considered to be very difficult to value the inputs and outputs of a DMU, while at the same time units may value their inputs and outputs in a different way and therefore adopt different weights. Thus, the assumption of setting universally valid weights is unsatisfactory and DEA gives a solution to this problem by determining a set of weights in the most favorable light for each DMU in comparison to other units. Efficiency (h_0) of a target unit (j_0) can be obtained as a solution to the following problem [13-15]:

$$\left.\begin{aligned}
&\max \mathrm{h}_0 = \frac{\sum_r u_r y_{rj_0}}{\sum_i v_i x_{ij_0}} && (3)\\
&\text{subject to:}\\
&\frac{\sum_r u_r y_{rj}}{\sum_i v_i x_{ij}} \le 1, \quad \textit{for each DMU } j = 1,2,\ldots,k && (4)\\
&u_r, v_i \ \ge \varepsilon, \quad r = 1,2,\ldots,m \quad i = 1,2,\ldots,n\,. && (5)
\end{aligned}\right\}\text{(M1)}$$

The variables of the above problem are the weights *(u, v)* which are the most favorable to unit j_0, as compared to the other *k-1* DMUs and bounded to be greater than or equal to some small positive arbitrary quantity ε in order to avoid any input or output being totally ignored. The relative efficiency of each DMU is subject to the constraint that no unit can be more than 100% efficient when the same weights are applied to each DMU, meaning that the efficiency is bounded to be lower than or equal to 1 [3].

Model M1 is fractional linear programming, and it first needs to be converted into a linear form so that the methods of linear programming can be applied. The linearization process is relatively straightforward and the linear version of the constraints of M1 is shown in the following model M2 [3, 13]:

$$\left.\begin{aligned}
&\max \mathrm{h}_0 = \sum_r u_r y_{rj_0} && (6)\\
&\text{subject to:}\\
&\sum_i v_i x_{ij_0} = 1 && (7)\\
&\sum_r u_r y_{rj} - \sum_i v_i x_{ij} \le 0, \quad \textit{for each DMU } j = 1,2,\ldots,k && (8)\\
&u_r, v_i \ \ge \varepsilon, \quad r = 1,2,\ldots,m \quad i = 1,2,\ldots,n\,. && (9)
\end{aligned}\right\}\text{(M2)}$$

As far as the maximization of a fraction is concerned, the most important point is the relative magnitude of the numerator and denominator and not their individual values. In other terms, there would be the same result if the denominator is set equal to a constant and the numerator is maximized. The relative efficiency of the target unit can be obtained by solving model M2. The efficiencies of the entire set of DMUs can be measured by finding the solution to the linear program focusing on each unit in turn [15].

B. Different Models

DEA accords name "envelopment" because of the way it envelops the observations in order to identify "efficiency frontier" that is used to evaluate relative performance of all peer units [10]. There are two types of DEA models. The first is called multiplier model – the primal model, and the other is called envelopment model – the dual model. The primal model was mainly described at the previous section, whereas the dual model is constructed by assigning a variable (dual variable) to each constraint in the primal model and then formulating a model on these variables, resulting in the following model M3, where h_0* is the efficiency score of a target unit (j_0) that would result in the optimum (calculated, e.g., by splitting the sample) [15, 16]:

$$\mathrm{h}_0^* = \min \mathrm{h}_0 \tag{10}$$

subject to:

$$\sum_j \lambda_j x_{ij} \le h_0 x_{ij_0}, \quad i = 1,2,\ldots,n \tag{11}$$

$$\sum_{j} \lambda_j y_{rj} \geq y_{rj_0}, \quad r = 1, 2, \ldots, m \qquad (12) \qquad \text{(M3)}$$

$$\sum_{j} \lambda_j = 1 \qquad (13)$$

$$\lambda_j \geq 0, \quad j = 1, 2, \ldots, k. \qquad (14)$$

The solution to either the original LP (the primal) or the partner (the dual) provides the same information about the problem being modelled. The solution to the dual model is seeking to minimize the efficiency with values of λ_j to form a composite unit with inputs $\sum_{j} \lambda_j x_{ij}$, $i=1, 2, \ldots n$ and outputs $\sum_{j} \lambda_j y_{rj}$, $r=1, 2, \ldots, m$ more efficient than unit j_0 which is being evaluated. More specifically, the weighted sum of the inputs of the other DMUs should be less than or equal to the inputs of unit j_0 and the weighted sum of the outputs of the other DMUs should be greater than or equal to unit j_0. The weights are the λ values. All the other DMUs with non-zero λ values are the units against which each inefficient DMU was found to be most directly inefficient [13, 15].

In addition, there are two approaches to apply the DEA model. The input-oriented approach aims to increase the efficiency of a DMU by minimizing inputs while keeping outputs fixed at the same level. On the other hand, the output-oriented DEA model is used when outputs are maximized while keeping input level as constant as possible [10].

C. Incorporation of Categorical Variables

In a DEA model input and output variables are used as essentially quantitative but sometimes it is necessary for the contribution of non-quantitative variables, such as ordinal or nominal variables, to be evaluated. For example, one might like to compare institutions of the same type, such as public or private schools with different qualitative features. This is accomplished by introducing dummy/categorical variables containing numbers for order or identifiers for names. In the context of DEA method, categorical or control variables refer not only to ordered variables (i.e. of the type “low”, “medium”, “high”) but also to variables that take on only a finite number of values, are inputs or outputs of certain types and cannot be represented by continuous variables [17-19].

DEA categorical variables function as further constraints on making comparisons between subsets of comparable DMUs. There must be some a priori information about the direction of the disadvantage between different categories, so that comparisons for DMUs in the same category or in a more unfavorable/lower category are possible. If categories cannot be comparable, a separate DEA should be performed for each category [1].

Even though methods of incorporating categorical variables to a DEA model have been introduced before, they have rarely been applied from researchers. Splitting the data set of a problem according to the distinct values of the categorical variables is a solution proposed in [20]. However, it is considered to be really tedious and slow for a model with more than one categorical variables, due to the fact that a separate DEA run is carried out for each distinct combination of categorical variables. To this end, Banker and Morey [18] showed that splitting of data into subsamples can be avoided by defining descriptor binary variables for each DMU that can replace categorical variables. However, their approach cannot be solved using standard DEA software. This problem is faced from a simple and straightforward alternative quite similar to the Banker and Morey method which is based on indicator variables [19]. Löber and Staat try to incorporate categorical non-discretionary variables in DEA models regardless of the returns to scale assumption. Splitting the data is not required and their method can be solved by any standard DEA Software and can be applied to discretionary categorical variables and non-hierarchical categoricals as even in the absence of numerical data [19].

For an input-oriented model with non-discretionary inputs/indicators, each input indicator must be greater than 0 for the DMU to be excluded from reference sets of other DMUs and equal to 0 for the DMU to be evaluated. These indicators must lead to constraints, which are fully redundant once inadmissible peers have been removed from the data set. Categorical variables must not change the efficiency score for any DMU, for which the referent point in a model without the additional constraint had already been composed of peers from categories not better than its own [19, 20]. The following LP represents a general solution for an input-oriented model with P inputs x and Q outputs y for N DMUs indexed by n. The true efficiency score θ for a DMU (x_0, y_0) with R categorical variables with C_r categories is:

$$\left.\begin{array}{ll}
\min \theta & (15) \\
\text{subject to:} & \\
\sum_{n=1}^{N} \lambda_n y_{qn} \geq y_{q0}, \quad q = 1, \cdots, Q & (16) \\
\sum_{n=1}^{N} \lambda_n x_{pn} \leq \theta x_{p0}, \quad p = 1, \cdots, P & (17) \\
\sum_{n=1}^{N} \lambda_n i_{rn}^{C_r} \leq \theta i_{r0}^{C_r}, \quad r = 1, \ldots, R, C_r = 1, \ldots, C_r - 1 & (18) \\
\lambda_n, x_{pn}, y_{qn} \geq 0, \quad n = 1, \ldots, N & (19)
\end{array}\right\} \text{(M4)}$$

To exclude DMUs in a better category as peers in referent points for DMUs in a lower category, indicator i_{rn} must be equal to x_{pn}, where x_p can be either one of the P inputs (clearly redundant, whichever one is chosen) for all DMUs in a lower category and must be equal to 0 for all other DMUs.

IV. Methodology for Cloud Computing Services Selection

A. Data Collection

The proposed method is based on IaaS cloud computing, the most straightforward cloud service. Data collection was based on Cloudorado (http://www.cloudorado.com) [7], a cloud computing comparison service that provides pricing bundles for IaaS providers. It accepts specifications on each customer's needs such as memory, storage, processor computing capabilities and operating systems and returns a comparison of different cloud services based on price. The platform also allows applying filters based on non-functional requirements such as security, reliability and cloud management features. According to Cloudorado's founder, Marcin Okraszewski [21], the task of comparing manually all the providers is very hard for the companies and it can take days or even weeks. This is the reason why Cloudorado aims to keep up to date, so that businesses can be able to make even better decisions and finally be successful.

The collected IaaS cloud bundles were categorized into three groups:

- ***Computation Optimized Instances*:** Instances for compute-bound applications that feature high performance processors.
- ***Memory Optimized Instances***: Instances for memory-intensive applications
- ***Storage Optimized Instances***: Instances that are designed for applications that require high sequential read and write access to very large data sets on local storage [22].

The total number of the collected price bundles is 806 out of 23 providers. The number of compute, memory and storage optimized instances is 401, 205, 200 respectively.

As mentioned above, the price instances of DEA method are derived from 23 providers, shown in Table 1. The collection of cloud bundles was based on criteria that were not fulfilled by all cloud providers, thus the number of the collected price bundles of each provider differs.

TABLE 1
CLOUD IAAS PROVIDERS

Providers	
Microsoft azure	Stratogen
Amazon	Eapps
Google	Data dimension
Cloudsigma	Cloudware
Atlantic.net	Zippycloud
M5	Exoscale
Elastichosts	Vps.net
Bitrefinery1	Dreamhost
Storm	Zettagrid
Rackspace	Cloudsolutions
E24cloud.com	Gigenet
Joynet	

Cloudorado also includes non-functional features such as:

- Security compliance and certifications (SSAE 16, HIPPA, FISMA, PCI DSS, etc.),
- Cloud management features
- Support levels
- Service Level Agreement (SLA)
- Elasticity of the offers
- Level of reliability
- Various technical details regarding networking, servers
- Used technologies
- Security design
- Licenses
- Billing details
- Support by third-party tools [21].

The client can easily determine which cloud provider best fits their organization's specific needs in terms of cost and performance.

The cloud instances are determined by functional requirements (CPU, Memory and Storage Capacity) shown in TABLE 2, TABLE 3 and TABLE 4, respectively, together with the considered values.

TABLE 2
COMPUTE OPTIMIZED INSTANCES

vCPU	Memory(GB)	Storage(GB)
1	2	50
2	4	100
4	8	100
4	8	200
8	16	200
8	16	500
16	32	200
16	32	500
16	32	1000
32	64	500
32	64	1000
32	64	2000

TABLE 3
MEMORY OPTIMIZED INSTANCES

vCPU	Mem(GB)	Storage(GB)
2	8	50
2	16	50
2	16	100
4	32	100
8	64	200
16	128	500
32	256	1000

TABLE 4
STORAGE OPTIMIZED INSTANCES

vCPU	Mem(GB)	Storage(GB)
2	16	500
2	16	1000
4	32	1000
4	32	2000
8	64	2000
8	64	5000
16	128	5000

16	128	10000
32	256	10000

B. Categorical Variables Description

This paper proposes an alternative methodology for the selection of IaaS cloud computing services by applying DEA to a multi-attribute decision-making problem, where each performance may depend on a number of functional and non-functional factors. Such a selection is difficult, due to the many qualitative features that are continuously being offered together with IaaS bundles. In this way, cloud bundles of services can be selected, not just according to the price but also based on the importance that each non-functional requirement has for the users [8, 21, 23].

This section focuses mainly on the description of non-functional parameters, which are treated as categorical variables for the input-oriented DEA model used in the context of this paper [19]. There are non-functional aspects which constitute some of the main promises of cloud computing, such as scalability, elasticity, service levels, and others that are the most important user concerns such as security, availability, ease of migration, true reliability levels, and usability [5, 24]. Based on these features, the diverse qualitative attributes of cloud bundles are grouped into corresponding categories, so that selection of cloud providers and services could be easier. A functional requirement about the supported operating system is included in this section as well, since it is considered to be another categorical variable for the DEA model of this study [19, 21].

According to the comparison of several cloud providers given by Cloudorado [7], the IaaS cloud services selection problem consists of 13 non-functional requirements and a functional one, grouped into 4 categories which are described below together with the considered values as shown in TABLE 5:

1) Security

- **Encrypted Storage**: If the storage volume is encrypted.
- **Safe Harbor / EU Directive 95/46/EC**: If the provider is compliant with EU Directive 95/46/EC on the protection of personal data. Regarding US companies, the equivalent is the Safe Harbor principles.

2) Availability/ Reliability

- **Service Level Agreement (SLA) Level:** The uptime SLA level expressed in percentage points of availability. A 5-level scale measuring from 99.90% (Level 1) until 100% (Level 5) of availability is used.
- **Backup Storage:** If storage-based backup is available or not.
- **Free Support:** If support cost is included in the price of the basic plan of each provider; any other additional support beyond the basic plan is paid.

3) Elasticity/ Performance

- **Burstable CPU:** The CPU allocation can be either fixed or can burst to a higher capacity if current conditions allow it. The burstable CPU is favorable for the selection of a cloud bundle of services, since it allows gaining extra CPU power at no additional cost, whereas with fixed CPU allocation there is no hope for a free CPU in case of a spike, but the CPU power is known, almost like with a dedicated server.
- **Auto-scaling:** A 4-level scale is used. Level 1 refers to vertical auto-scaling, meaning when it is possible to scale up a server automatically, by adding more resources, such as disk space, RAM or processing units. Horizontal auto-scaling (Level 2) is about adding more servers, quickly and easily deploy new images based on existing virtual machines depending on the workload. It is possible that none of these 2 types of auto-scaling is supported (Level 0) and on the other hand both of them may be supported from a cloud provider (Level 3). It is favorable to have a horizontal auto-scaling, because it increases capacity of existing hardware by connecting multiple servers, whereas vertical auto-scaling restricts capacity extension to the capacity of the actual server [25].
- **Resource usage Monitoring:** If there are integrated monitoring solutions (i.e. monitoring tools, alerts, indicators, etc.) being offered by cloud companies, so that users can monitor current resource utilization (i.e. CPU, RAM, disk, network etc.) in their cloud servers for no additional cost. This feature is needed for performance and capacity management, since often measurements from within the cloud server might not show the full image of resources usage.

4) Usability/ Portability

- **Web Interface:** If a web management interface is available or not.
- **API:** If an API management is available for automating cloud servers and interacting with them or not.
- **One Account for All Locations:** If there is one account and single interface to manage all different locations or a separate account for each location.
- **Image from Cloud Server:** If a provider supports creating an image from an existing VM and then deploying it to other cloud servers.
- **Limited Free Trial:** If cloud companies offer a free trial of their services for a limited period of time or for a certain amount of credit to be spent on cloud services, so that customers can use it to run tests.
- **Supported Operating System:** There are two different supported operating systems (regardless of version) available as pre-configured images from cloud providers, Linux and Windows. The choice of Linux is considered to be favorable for a cloud computing bundle of services, as it has no additional cost for the bundle

TABLE 5.
FUNCTIONAL AND NON-FUNCTIONAL FEATURES OF CLOUD BUNDLES OF SERVICES TREATED AS DEA CATEGORICAL VARIABLES WITH CORRESPONDING VALUES

CLOUD FEATURES	NUMBER OF CATEGORIES	VALUES
	SECURITY	
Encrypted Storage	2	YES/ NO
Safe Harbor / EU Directive 95/46/EC	2	YES/ NO
	AVAILABILITY/ RELIABILITY	
SLA Level	5	1:99.90%, 2:99.95%, 3:99.98%, 4:99.99%, 5:100%
Backup Storage	2	YES/ NO
Free Support	2	YES/ NO
	ELASTICITY/ PERFORMANCE	
Burstable CPU	2	Burstable/ Fixed
Auto-scaling	4	0:None, 1:Vertical, 2:Horizontal, 3:Both
Resource usage Monitoring	2	YES/ NO
	USABILITY/ PORTABILITY	
Web Interface	2	YES/ NO
API	2	YES/ NO
One Account for All Locations	2	YES/ NO
Image from Cloud Server	2	YES/ NO
Limited Free Trial	2	YES/ NO
Supported Operating System	2	LINUX/WINDOWS

At present, different types of cloud services with diverse resources, qualitative requirements, functional features and prices are available in the market and from different cloud providers. The decision maker needs to identify and select the best suited bundle in order to achieve the desired output with minimum cost and maximum performance.

C. DEA Evaluation

Following the collection of data from cloud computing providers, the proposed methodology proceeds to calculate the relative efficiency rates, according to the DEA. DMUs, input and output parameters, as well as categorical variables are specified in an input-oriented DEA model, in the way as it was previously described. The essence of such a model in measuring the efficiency of cloud bundles lies in maximizing the efficiency rates by reducing inputs and keeping outputs at the current level.

Cloud bundles are designated as Decision Making Units (DMUs). Each DMU is denoted by the provider's name and a serial number. For example, and in the context of Memory Optimized Instances group, DMU 'Google5' describes the fifth cloud instance of Google, as derived from the Cloudorado platform.

Price for an annual subscription is selected as an input parameter and is determined from Cloudorado. The price is a multidimensional factor because a provider who is considered to be the cheapest for one cloud instances group might be the most expensive for another one. The DEA model aims to increase the efficiency of each bundle by decreasing its price, while keeping outputs at the same level.

Memory (GB), Storage capacity (GB) and Compute Power (CPU cores) that specify each cloud service are defined as the outputs of the model. There are a few more characteristics, such as Time On, Transfer In, Transfer Out, and the option that the CPUs, the RAM and the storage can be distributed among more than one physical server, participating in the price bundling of Cloudorado which are not considered as output parameters into this study. The Transfer In (the number of bytes received by server from the internet per month) does not contribute at a substantial level to the shaping of the pricing bundles, since many providers (i.e. Amazon, ecloud24 etc.) charge customers only for the outgoing traffic and the others include it as a small amount in the total price of services. Furthermore, the Transfer Out (the number of bytes sent by server to Internet per month) does not seem to affect much pricing [26]. Therefore, with no loss of generality, the Transfer In attribute was considered to be at 1GB and the Transfer Out at 10GB per month. As far as Time On is concerned, it was set at a level of 100% availability per day and the default offered value of non-distributed resources was also considered.

In addition, all of the categorical variables described at the previous section are used as input indicators, since the DEA model of this paper is an input-oriented model with P inputs x and Q outputs y for n DMUs indexed by i. Therefore, each categorical variable, from R categoricals in total, has a number of categories Cr, as shown in

. A corresponding number of binary dummies of the type proposed by Banker and Morey [18] is generated in this way, Cr-1. Hence, these variables are necessary and are designed to be multiplied with the price of each bundle that constitutes the input of the DEA model, thereby creating Cr-1 input indicators in order to achieve the desired result [19]. The proposed DEA model with all the input and output parameters is depicted in Fig. 1.

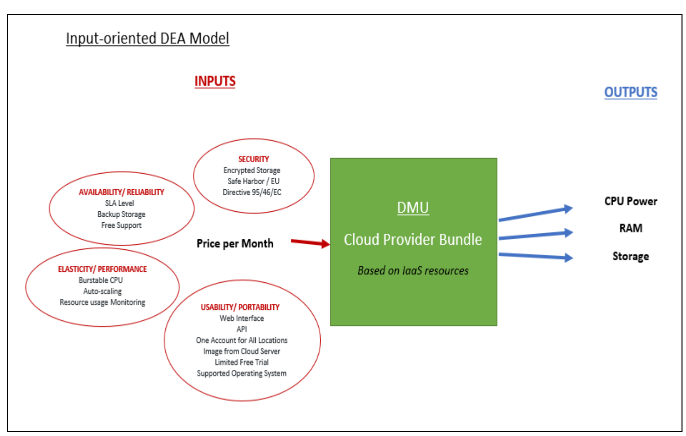

Fig. 1. The proposed input-oriented DEA model with all input and output parameters.

It should be also taken into account that all these additional variables lead to more constraints in the LP in the envelopment form in order to be able to use standard DEA code, as explained in section 3.3 of this paper. The use of 0-1 dummy variables guarantees that only DMUs from the worst category are admissible as peers for DMUs from this category [19]. The objective function of the input-oriented model subject to the constraint on the indicators z_{ri} is given below, where θ^* is the efficiency score that would result in the optimum:

$$\left.\begin{aligned}
&\theta^* = \min \theta && (20)\\
&\text{subject to:} && \\
&\sum_{i=1}^{n} \lambda_i y_{qi} \geq y_{q0}, \quad q = 1, \cdots, Q && (21)\\
&\sum_{i=1}^{N} \lambda_i x_{pi} \leq \theta x_{p0}, \quad p = 1, \cdots, P && (22)\\
&\sum_{i=1}^{N} \lambda_i z_{ri}^{C} \leq \theta z_{r0}^{C}, \quad r = 1, \ldots, R,\ C = 1, \ldots, Cr - 1 && (23)\\
&\lambda_i, x_{pi}, y_{qi} \geq 0, \quad i = 1, \ldots, N && (24)
\end{aligned}\right\} \quad (M5)$$

According to the previous categorization of cloud bundles, services have been characterized as compute, memory and storage oriented instances and the proposed model is applied over each group returning the most efficient services. In the case of compute optimized instances there are 401 DMUs, 205 DMUs for the memory-oriented group and, finally, 200 DMUs correspond to the storage optimized instances.

Evaluation was based on MaxDEA software [16], a user-friendly and easy to use software with no limitation on the number of DMUs. MaxDEA returns the efficient cloud bundles as well as the rate of the inefficient cloud bundles. In addition, DEA calculates the slacks that are defined as the additional improvement needed for an inefficient DMU to become efficient [11]. At the proposed input-oriented CCR model the slack variables correspond to price reduction. Furthermore projection values, which are the efficient targets, are estimated for improving the performance of a service which does not lie into the efficient frontier [10].

A simple example of 3 different DMUs (cloud computing provider bundles) of the compute-oriented category with 5 categorical variables using the DEA indicator approach is shown in Fig. 2.

Variable	DMU - Cloud Provider Bundle		
	CloudSolutions5	Google2	e24cloud.com2
CPU (cores)	2	2	1
RAM (GB)	16	13	4
Storage (GB)	200	100	140
Price ($)	119	50	60
Auto-scaling	None	Horizontal	Vertical
$d^{AS(V)}$	0	1	1
$d^{AS(H)}$	0	1	0
$d^{AS(B)}$	0	0	0
$i^{AS(V)}$	0	50	60
$i^{AS(H)}$	0	50	0
$i^{AS(B)}$	0	0	0
Resource usage monitoring	NO	YES	YES
$d^{RU(Y)}$	0	1	1
$i^{RU(Y)}$	0	50	60
Web Interface	YES	YES	YES
$d^{WEB(Y)}$	1	1	1
$i^{WEB(Y)}$	119	50	60
Supported OS	Linux	Linux	Linux
$d^{OS(W)}$	1	1	1
$i^{OS(W)}$	119	50	60
Burstable CPU	Fixed	Fixed	Fixed
$d^{CPU(Y)}$	0	0	0
$i^{CPU(Y)}$	0	0	0

Fig. 2. A sample of dataset with some categorical variables using the DEA indicator approach.

The dummy variables that were created are labeled as $d^{a(b)}$ and the input indicators as $i^{a(b)}$. The dummies were multiplied with the input of the monthly price, resulting in the indicators. DMUs "CloudSolutions5" and "Google2" were found to be efficient, whereas the efficiency score of DMU "e24cloud.com2" was estimated to be 75.83% having "CloudSolutions5" in its reference set with λ equal to 0.152174. Both "CloudSolutions5" and "Google2" lie into the efficient frontier but "Google2" needs to reduce its price 14.50$ (final bundle price: 45.50$), in order to be efficient as well.

V. Results and Discussion

A. Model Results Without Including Categorical Variables

The proposed model is initially applied over a dataset of cloud bundles including only functional parameters. The implementation of DEA to a rather simple dataset, i.e. without including the non-functional factors, highlights the contribution of these parameters to the performance. In the group of IaaS services each cloud bundle is described by Compute Power (vCPUs), Memory (GB), Storage (GB) and Price ($). The only functional parameter, which participates as a categorical variable in the DEA model, is the supported operating system. Linux is chosen because it is freeware, thus the impact on the price of cloud instance is negligible.

The efficient DMUs for the compute, memory and storage optimized groups are 3 out of 204 DMUs, 8 out of 103 DMUs and 5 out of 101 DMUs respectively. Therefore, only 1.47% of compute cloud bundles are efficient, while 7.76% memory bundles and 4.95% storage bundles lie on the efficient frontier. Fig. 3, Fig. 4 and Fig. 5 illustrate the results of the model.

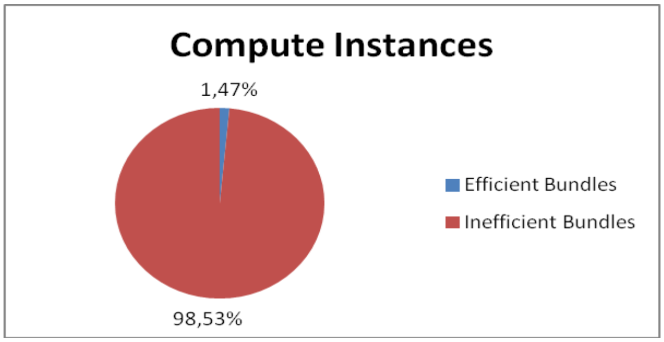

Fig. 3. Overall evaluation of compute DMUs.

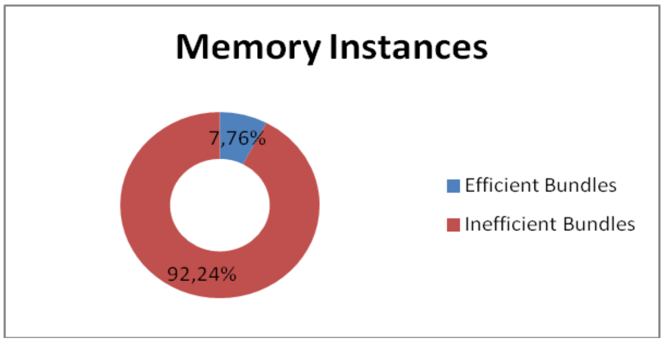

Fig. 4. Overall evaluation of memory DMUs.

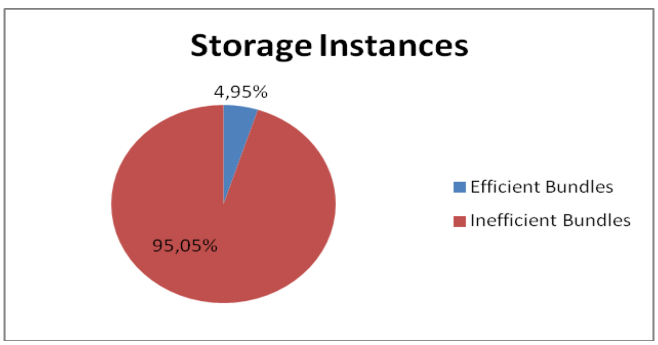

Fig. 5. Overall evaluation of storage DMUs.

In the "compute optimized" group only 3 out of 23 providers offer efficient cloud bundles. Furthermore, in the memory and storage instances the providers that demonstrate efficient bundles are 4 out of 23 and 3 out of 23 respectively. TABLE 6, Table 7 and TABLE 8 present the above results.

TABLE 6
NUMBER OF EFFICIENT COMPUTE INSTANCES PER CLOUD PROVIDER

Providers	Number of efficient cloud bundles
Google	1
DreamHost	1
Vps.Net	1

TABLE 7
NUMBER OF EFFICIENT MEMORY INSTANCES PER CLOUD PROVIDER

Providers	Number of efficient cloud bundles
Microsoft Azure	1
CloudSigma	3
DreamHost	1

TABLE 8
NUMBER OF EFFICIENT STORAGE INSTANCES PER CLOUD PROVIDER

Providers	Number of efficient cloud bundles
Microsoft Azure	3
Google	1
DreamHost	1

Fig. 6 illustrates an overall ranking of the above providers. It is obvious that the most efficient provider is Microsoft Azure while Vps.Net presents the smallest number of efficient bundles.

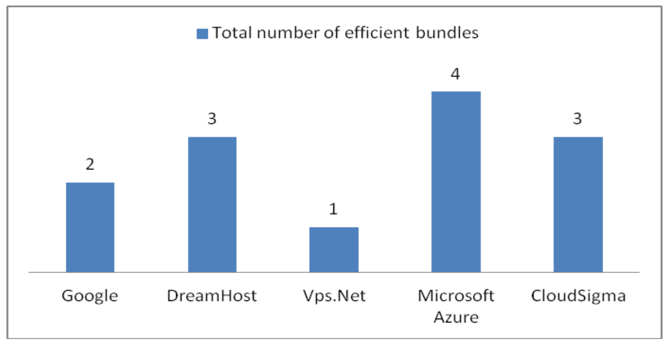

Fig. 6. Efficient Providers

The model moves inefficient bundles to the efficient point by reducing the input (price) and preserving the outputs (CPU, Memory, Storage) at their current levels. Therefore, the model focuses on the inefficient bundles and calculates the average reduction and the standard deviation of the price. As calculated, compute optimized inefficient bundles need a 286.08$ price reduction on average, with a standard deviation s=408.95$. In the memory optimized group the average reduction of the price is 311.31$ and standard deviation is 497,47$. Finally, in the storage group, the average reduction of the price is 578.23$ and the standard deviation 753.20$. The values of standard deviations reveal the lack of homogeneity among bundles, since they are rather widely spread around mean values.

The proposed methodology is based on the benchmarking of efficient bundles. Each efficient bundle is used as a benchmark for other bundles in order to estimate relative efficiency. The more times each efficient bundle is applied as a benchmark, the more significant the benchmark is. In the compute group the most important bundle is Google2, which was used as a benchmark 189 times. The most significant bundle in the memory group is CloudSigma2, used as a benchmark 51 times and, finally, in the storage group Microsoft Azure1 is the most essential bundle, used as a benchmark 64 times. The characteristics of the above bundles are summarized in TABLE 9.

TABLE 9
IAAS CHARACTERISTICS OF THE MOST IMPORTANT CLOUD

BUNDLES

DMU	CPU	RAM	Storage	Price	Operating System
Google2	2	13	100	50	Linux
CloudSigma2	1	16	50	62	Linux
Microsoft Azure1	2	16	500	96	Linux

B. *Model Results Including Categorical Variables*

The proposed DEA-model is also applied to a more complex group of cloud bundles, including the non-functional parameters. However, in this case the sample data has been enlarged because the different values of the categorical variables generate new bundles and consequently affect the price. In addition to the previous case, cloud bundles are described by Compute Power (vCPUs), Memory (GB), Storage (GB) and Price ($), as well as by 14 categorical parameters.

The efficient DMUs are 39 out of 401, 48 out of 204 and 35 out of 200 DMUs for the compute, memory and storage optimized instances, respectively. Thus, 9.72% of compute bundles and 23,5% of memory bundles are efficient. Finally, in the storage optimized group the rate of the efficient bundles is 17.5%. Fig. 7, Fig. 8 and Fig. 9 illustrate the rates of the efficient and inefficient DMUs.

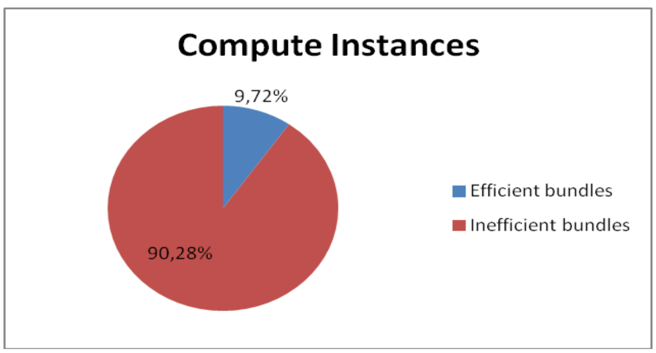

Fig. 7. Overall evaluation of compute DMUs.

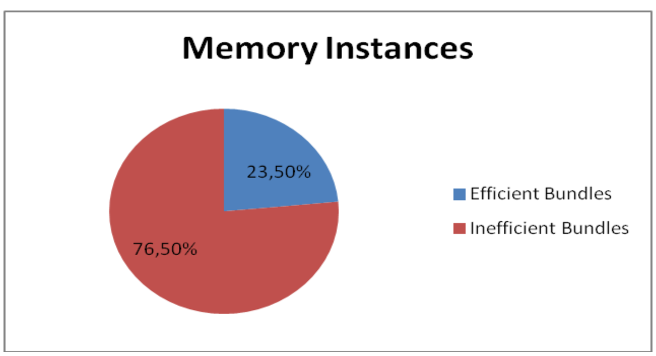

Fig. 8. Overall evaluation of memory DMUs.

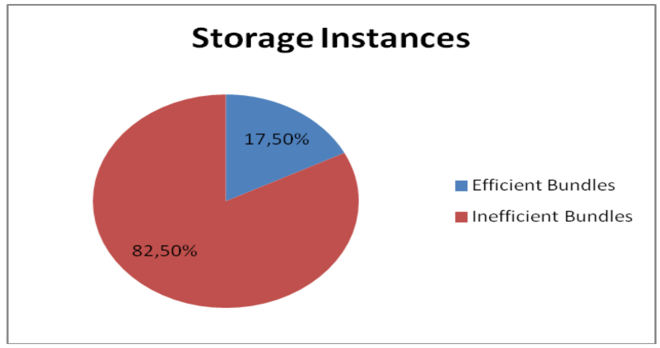

Fig. 9. Overall evaluation of storage DMUs.

Despite the fact that the sample data includes 23 providers, only a proportion of them produce efficient bundles. The providers offering bundles with exceptional relative efficiency are presented in TABLE 10, TABLE 11 and Table 12**Σφάλμα! Το αρχείο προέλευσης της αναφοράς δεν βρέθηκε.**, together with the corresponding number of efficient bundles.

TABLE 10
NUMBER OF EFFICIENT COMPUTE INSTANCES PER CLOUD PROVIDER

Compute Optimized Group	
Providers	Number of efficient cloud bundles
Amazon	2
Atlantic.net	1
Cloudsigma	1
Cloudsolutions	4
Cloudware	4
Dreamhost	3
e24cloud.com	1
Exoscale	8
Gigenet	1
Google	1
M5	1
Microsoft Azure	2
Storm	3
VP.NET	1
Zippycloud	6

TABLE 11
NUMBER OF EFFICIENT MEMORY INSTANCES PER CLOUD PROVIDER

Memory Optimized Group	
Providers	Number of efficient cloud bundles
Amazon	5
Atlantic.net	1
CloudSigma	5
CloudSolutions	2
Cloudware	3

Dreamhost	3
e24cloud.com	1
eApps	1
Exoscale	6
Gigenet	2
M5	3
Microsoft Azure	3
Storm	3
VPS.NET	2
Zippycloud	8

TABLE 12
NUMBER OF EFFICIENT STORAGE INSTANCES PER CLOUD PROVIDER

Storage Optimized Group	
Providers	Number of efficient cloud bundles
Amazon	3
Atlantic.net	1
Microsoft Azure	3
CloudSigma	4
Cloudsolutions	2
Cloudware	1
Dreamhost	2
e24cloud.com	3
eApps	1
Google	1
M5	2
Storm	4
VPS.NET	1
Zettagrid	1
Zippycloud	6

In the compute as well as in the memory group, CloudSolutions5 is the most important efficient bundle, since it was used as a benchmark for other bundles more times. In addition, and regarding the storage group, CloudSolutions1 is the most significant bundle. The functional and the non-functional attributes of the most important bundles are summarized in TABLE 13 and TABLE 14.

TABLE 13
FUNCTIONAL ATTRIBUTES

DMU	CPU	Memory	Storage	Price	Times as benchmark
CloudSolutions1	1	2	50	31$	27
CloudSolutions5 (compute)	2	16	200	119$	100
CloudSolutions5 (memory)	2	16	200	119$	27

TABLE 14
NON-FUNCTIONAL ATTRIBUTES

CloudSolutions1 CloudSolutions5	Encrypted Storage	No
	Safe Harbor / EU Directive 95/46/EC	No
	SLA Level	99.95%
	Backup Storage	No
	Free Support	No
	Burstable CPU	Fixed
	Auto-scaling	No
	Resource usage Monitoring	No
	Web Interface	Yes
	API	Yes
	One Account for All Locations	No
	Image from Cloud Server	Yes
	Limited Free Trial	No
	Supported OS	Linux

Furthermore, the proposed approach indicates the average reduction and the standard deviation of the price required in order to move inefficient bundles to the efficient point. In the compute optimized instances the average reduction of the price is 521.96$ and standard deviation is 2163$. In addition, in the memory group the average reduction of the price is estimated to be 461.9$ and standard deviation is 1019.5$. Finally, in the storage group the average reduction is 802.3$ and standard deviation is s = 2098.3$.

C. Comparison

It is notable that the proportion of the bundles which include the non-functional requirements and lie on the efficient frontier is substantially increased as compared to the bundles with functional requirements only, indicating that non-functional requirements affect performance. This is graphically depicted in Fig. 10, Fig. 11 and Fig. 12.

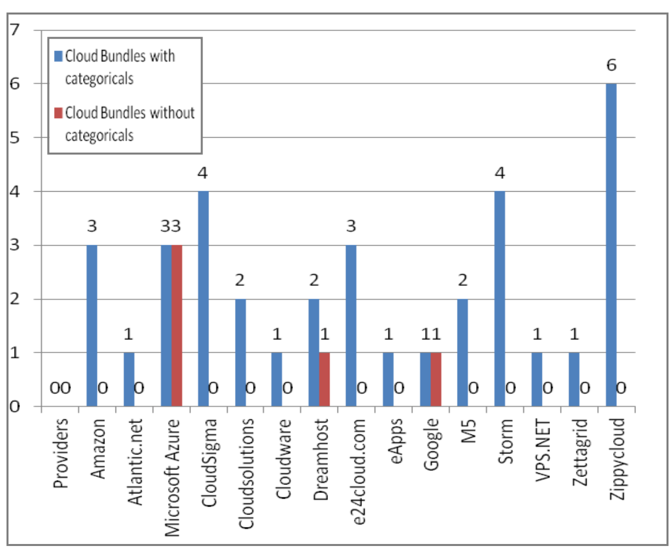

Fig. 10. Comparison between compute efficient instances with and without categorical variables.

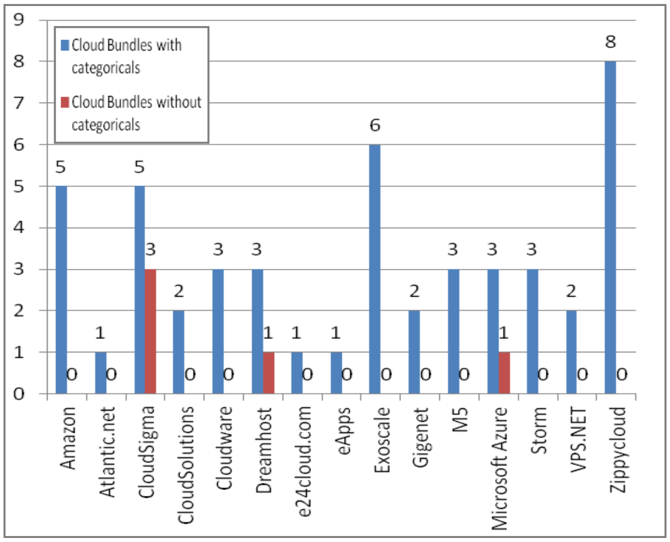

Fig. 11. Comparison between memory efficient instances with and without categorical variables.

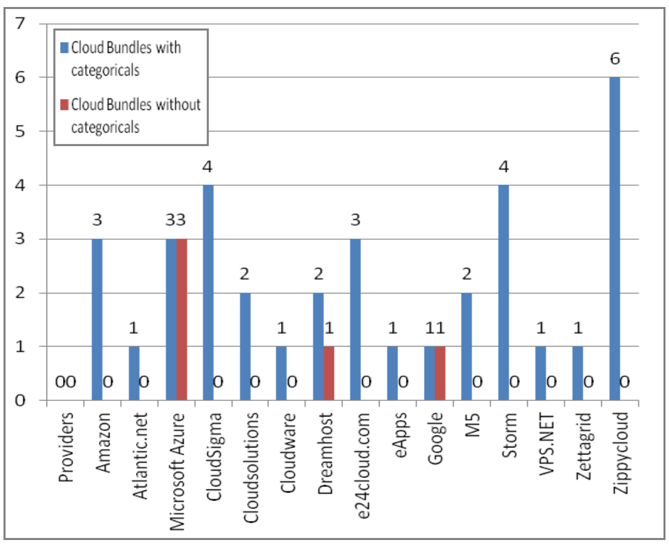

Fig. 12. Comparison between storage efficient instances with and without categorical variables.

Since the proposed model is input oriented, inefficient DMUs become efficient through the proportional reduction of their input [27] which corresponds to the reduction of price. An overall proportionate movement of the price, for each group (compute, memory, storage) is depicted in the figures below making obvious that non-functional parameters affect the proportionate price reduction.

Fig. 13 and Fig. 14 illustrate distributions of price reduction for compute inefficient bundles. They highlight that categorical variables succeed in moving a larger number of inefficient bundles to efficient frontier with less price reduction. The majority of inefficient computed bundles, at a level of approximately 40%, can move to the efficient frontier with a maximum price reduction up to 100$ if categorical variables are not taken into account. If categorical variables are included, the 41% of inefficient bundles become efficient with a price reduction up to 70$.

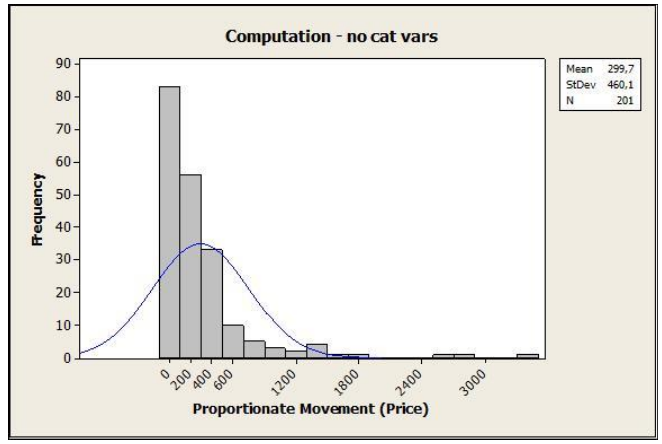

Fig. 13. Distribution of price reduction for compute inefficient bundles without categorical variables.

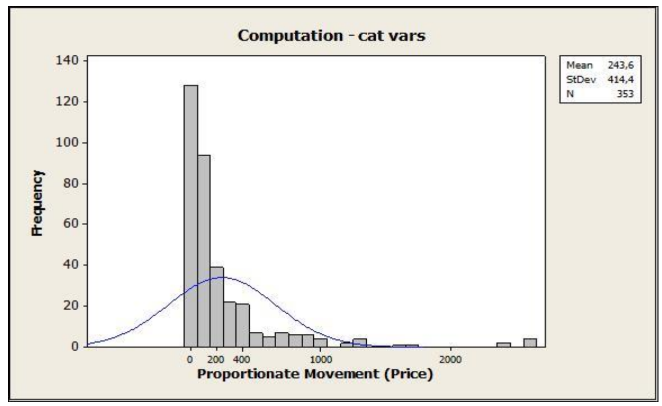

Fig. 14. Distribution of price reduction for compute inefficient bundles with categorical variables.

The same results regarding distributions of price reduction for memory group are illustrated in Fig. 15 and Fig. 16. The results show that the largest share of inefficient bundles (37%) without the non-functional parameters requires a price reduction of about 300$ in order to be efficient. Instead, 46% of cloud bundles that support non-functional factors require price reduction up to 150$.

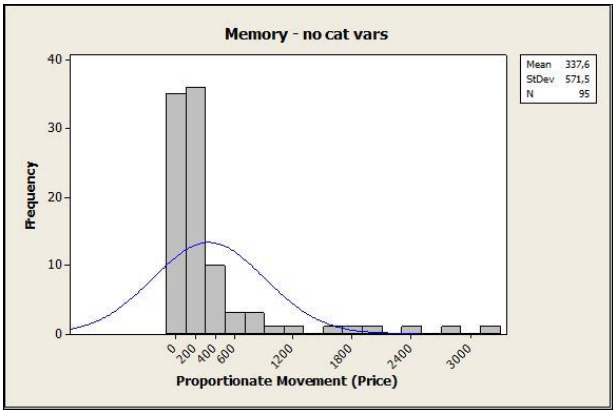

Fig. 15. Distribution of price reduction for memory inefficient bundles without categorical variables.

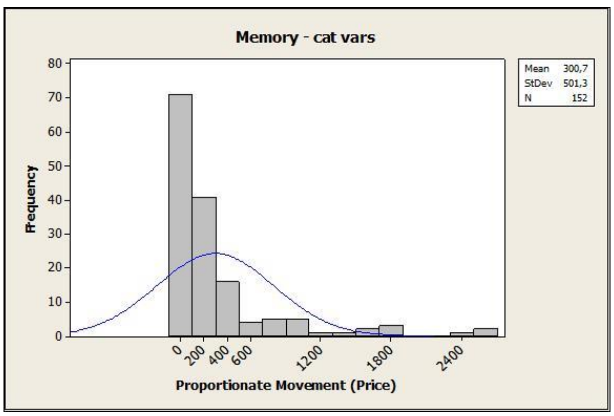

Fig. 16. Distribution of price reduction for memory inefficient bundles with categorical variables.

In addition, in the storage group and without considering non-functional parameters, the greater part of inefficient bundles, almost 32.5%, becomes efficient by reducing the mean price at about 250$, as shown in Fig. 17. The incorporation of non-functional parameters is depicted in Fig. 18, where 38.2% of inefficient bundles reach efficient frontier by reducing price up to 100$. All of the considered cases reveal the importance of the non-functional parameters to the shaping of efficient pricing schemes.

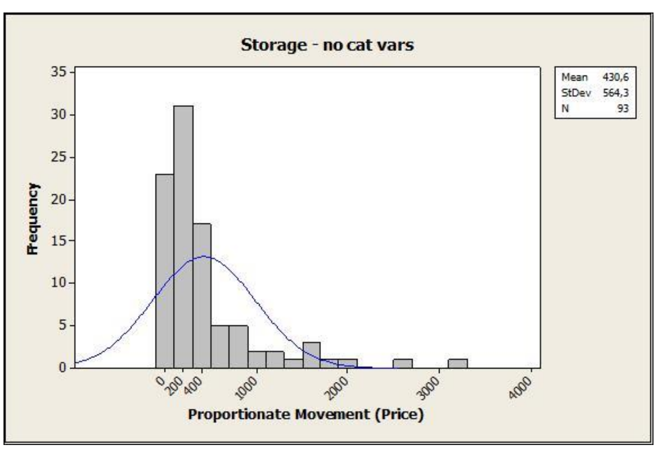

Fig. 17. Distribution of price reduction for storage inefficient bundles without categorical variables.

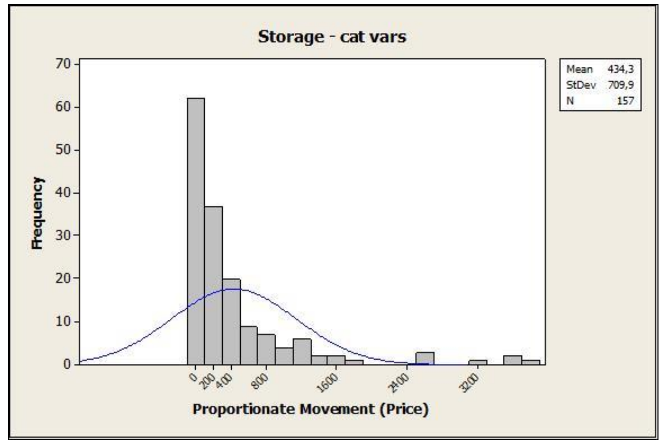

Fig. 18. Distribution of price reduction for storage inefficient bundles with categorical variables.

VI. CONCLUSIONS

Cloud market is constantly growing and cloud providers offer numerous services with different pricing schemes, aiming to fulfill client requirements. Even though, clients need to choose the optimal cloud solution that meets with their requirements, the variety of cloud services in different pricing schemes causes confusion and sharpens clients' doubt of adopting cloud services. Although, pricing strategies are considered to be a powerful tool that classifies services, their complexity discourages clients, especially the non-experienced ones.

Therefore, relative efficiency is introduced to this paper, contributing to ranking cloud services based on the combination of performance and price. Cloud bundles were collected by the platform of Cloudorado for a number of 23 providers and are described not only by functional parameters, such as CPU, memory and storage, but also by non-functional requirements being represented by security, reliability, elasticity, availability, portability and usability features. These characteristics mainly define the qualitative attributes of cloud bundles, which are expected to be very important to clients for the cloud computing services selection, but they have not been examined thoroughly yet so that their contribution is measured in some way.

Towards this direction, the proposed methodology consists of a dual DEA model, which was applied to the whole dataset of collected cloud bundles and finally estimates their relative efficiency. It highlights the importance of non-functional attributes of cloud computing services to the performance, pointing out at the same time that high prices do not necessary signal high quality. According to the results, cloud providers that offer IaaS bundles including exclusively functional parameters present a smaller rate of efficient bundles comparing to providers that enrich their bundles with non-functional attributes. Thus, non-functional parameters boost efficiency of cloud bundles and make vendors more competitive and profitable.

Furthermore, the DEA methodology is used to provide decision makers with a valuable techno-economic analysis tool that focuses on different competing cloud computing services available in the market. The proposed model examines inefficient bundles as well and can direct cloud providers how to be efficient and competitive. It estimates price reduction that is necessary in order to move inefficient bundles to efficient frontier. Inefficient bundles that include non-functional attributes require a smaller price reduction than IaaS bundles that are based only on functional parameters (CPU, Storage, Memory).

The extension of the DEA model by using and adopting efficiency indices (restricted weights) for input variables related to cloud providers is a way to incorporate subjective judgements about the degree of importance of non-functional parameters and how they individually affect relative performance of cloud services. More benefits will be also derived from this approach, since the DEA model can adapt easily to different estimations and various pricing strategies. Another interesting research direction will be to apply to the same dataset of IaaS bundles of services some other decision making methodologies and alternative efficiency measures, such as AHP, TOPSIS, cross-efficiency DEA etc. and then compare the results with those obtained from the DEA model, so that customers are offered further support for their cloud computing services selection. Finally, it is highly recommended to analyze and apply these methods in the context of specific scenarios of use.

References

1. Charnes, A., et al., *Data envelopment analysis: Theory, methodology, and applications*. 2013: Springer Science & Business Media.

2. Zhang, A.L.X.Y.S.K.M., *Comparing Public-Cloud Providers.* IEEE Internet Computing, 2011. **15**(2): p. 50-53.

3. Charnes, A., W.W. Cooper, and E. Rhodes, *Measuring the efficiency of decision making units.* European journal of operational research, 1978. **2**(6): p. 429-444.

4. Li, H., J. Liu, and G. Tang. A pricing algorithm for cloud computing resources. in Network Computing and Information Security (NCIS), 2011 International Conference on. 2011. IEEE.

5. Rodrigues, T. *Side-by-side comparisons of IaaS service providers*. 2012 August 7, 2012, 7:18 AM PST]; Thoran Rodrigues compares the major players and some of the newcomers in the increasingly competitive IaaS space, examining both their service offerings and how well they meet user cloud concerns. Available from: http://www.techrepublic.com/blog/the-enterprise-cloud/side-by-side-comparisons-of-iaas-service-providers/.

6. Iosup, A., et al., *Performance Analysis of Cloud Computing Services for Many-Tasks Scientific Computing.* IEEE Transactions on Parallel and Distributed Systems, 2011. **22**(6): p. 931-945.

7. Okraszewski, M. *Cloud Computing Price Comparison Engine*. 2016; Available from: http:/www.cloudorado.com.

8. Jeong, J.J.H.P.H.-Y., *The QoS-based MCDM system for SaaS ERP applications with Social Network.* The Journal of Supercomputing, 2012. **2**(66): p. 614-632.

9. Wang, Z.Z.X.W.Y.Z.M.R.L.J., *QoS Ranking Prediction for Cloud Services.* IEEE Transactions on Parallel and Distributed Systems, 2012. **24**(6): p. 1213-1222.

10. Kumar, S., *Evaluating Cloud Services based on DEA, AHP and TOPSIS*, D.G.R. Gangadharan, Editor 2014: Institute of Development and Research in Banking Technology, Hyderabad.

11. Wang, C.X.Y.M.X., A Non-Parametric Data Envelopment Analysis Approach for Cloud Services Evaluation, 2015.

12. de Souza, L.M. and M.P. Fernandez, Performance Evaluation Methodology for Cloud Computing using Data Envelopment Analysis. ICN 2015, 2015: p. 70.

13. Sherman, H.D. and J. Zhu, *Data Envelopment Analysis Explained.* Service Productivity Management: Improving Service Performance using Data Envelopment Analysis (DEA), 2006: p. 49-89.

14. Braglia, M. and A. Petroni, *Evaluating and selecting investments in industrial robots.* International Journal of Production Research, 1999. **37**(18): p. 4157-4178.

15. Emrouznejad, A. *DEA Tutorial*. 2012; This is a tutorial on Data Envelopment Analysis. Available from: http://deazone.com/en/resources/tutorial.

16. LtdCompany, B.R.S. 2009; Available from: http://maxdea.com/.

17. Førsund, F.R., CATEGORICAL VARIABLES IN DEA1. 2001.

18. Banker, R.D. and R.C. Morey, *The use of categorical variables in data envelopment analysis.* Management science, 1986. **32**(12): p. 1613-1627.

19. Löber, G. and M. Staat, Integrating categorical variables in Data Envelopment Analysis models: A simple solution technique. European Journal of Operational Research, 2010. **202**(3): p. 810-818.

20. Charnes, A., W.W. Cooper, and E. Rhodes, Evaluating program and managerial efficiency: an application of data envelopment analysis to program follow through. Management science, 1981. **27**(6): p. 668-697.

21. Okraszewski, M. *Cloudorado Launches New Cloud Computing Comparison Service*. 2015 [cited 2017 25/1/2017]; Available from: http://www.prweb.com/releases/cloud/computing/prweb12455967.htm.

22. *Amazon EC2 Instance Store*. 2017 [cited 2017 25/1/2017]; Available from: http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/InstanceStorage.htm.

23. Arias, D.C.B., Resource Provisioning in Clouds via Non-Functional Requirements, 2013, The University of Melbourne.

24. Glinz, M. On non-functional requirements. in 15th IEEE International Requirements Engineering Conference (RE 2007). 2007. IEEE.

25. Beaumont, D. *How to explain vertical and horizontal scaling in the cloud*. 2014 [cited 2017; Available from: https://www.ibm.com/blogs/cloud-computing/2014/04/explain-vertical-horizontal-scaling-cloud/.

26. Mitropoulou, P., et al. A Hedonic Price Index for Cloud Computing Services. in CLOSER 2015, 5th International Conference on Cloud Computing and Services Science. 2015. Lisbon, Portugal.

27. Cooper, W.W., L.M. Seiford, and K. Tone, Introduction to data envelopment analysis and its uses: with DEA-solver software and references. 2006: Springer Science & Business Media.