DLSU Business & Economics Review 25.2 (2016), pp. 166-196

A DEA-based Performance Measurement Mathematical Model and Software Application System Applied to Public Hospitals in the Philippines

Richard C. Li, Jazmin C. Tangsoc, Solomon L. See, Victor John M. Cantor, Martha Lauren L. Tan, and Rachelle Joy S. Yu
De La Salle University, Manila, Philippines
richard.li@dlsu.edu.ph

Evaluation of performance is an important activity in identifying shortcomings in managerial efficiency and devising goals for improvement. However, measuring performance is not an easy task; more so in making sure that it captures a holistic view of performance. This study identified four existing performance measurement issues that organizations face often. These issues were (a) existence of missing data during data collection, (b) accounting undesirable or non-value adding outputs as opposed to desirable or marketable outputs, (c) inclusion of exogenous or environmental factors that affects the organization performance, and (d) arriving with resource allocation decisions that will help improve organizational performance. Linear Programming (LP) and Data Envelopment Analysis (DEA) were used to develop a performance measurement tool that addresses the aforementioned issues. Subsequently, this tool was used to develop the DEA-based performance measurement and reallocation software to aid managers in analyzing organizational performance. A case study on 14 NCR public hospitals was conducted to validate the logic and usefulness of the software. Software results showed that there were two inefficient hospitals and corresponding decisions to increase performance were identified in terms of inputs and outputs. An economic interpretation was then provided to realize the significance of the performance measurement results.

JEL Classifications: C61, C88, I12

Keywords: Data envelopment analysis, Exogenous input, Missing data, Performance measurement, Reallocation decision, Undesirable output

Performance measurement is a fundamental building block of any organization that aims to improve service and business performance. Historically, organizations have always measured performance through their financial performance (Gerschewski & Xiao, 2015). However, traditional performance measures, based on cost accounting information, provide little to support organizations on their performance as a whole. This is because they do not map process performance with the consideration of all inputs used to attain organizational goals (Colledani & Tolio, 2009). It is through performance measurement where organizations should be able to identify and track progress against organizational goals that are often not easily measurable using financial indicators (Tung, Baird, & Schoch, 2011). Performance measurement also plays an important role in identifying opportunities for improvement and comparing performance against standards (both internal and external) (de Lima, da Costa, & de Faria, 2009; Verbeeten & Boons, 2009). Clearly, measuring performance in a holistic view is warranted for an organization to prosper.

Performance measurement is data intensive and requires the organization to have a data collection system in place. However, though procedures are developed in ensuring that important information needed for performance measurement is collected, there are instances wherein data are missing. Missing data pertains to performance-related data that are unavailable due to several reasons that include, but are not limited to, administrative fault in which the staff failed to collect or record the data, malfunctioning of equipment that resulted to data corruption, and refusal of respondent to answer the questions from a survey, among others (Zha, Song, Xu, & Yang, 2013). Missing data exists and are inevitable in all organizations, thus should be accounted for (Kao & Liu, 2000). In any case wherein data are missing, exclusion or elimination of the cases or categories that contain missing data produces a distorted result of performance as some inefficient systems may be considered efficient in the absence of considering a significant input or output in evaluation (Chen, Li, Xie, An, & Liang, 2014). Moreover, the impact of missing data is detrimental not only through its potential hidden biases of the results but also in its practical impact on the sample size available for analysis.

Aside from data collection inadequacies that result to missing data, organizations also need to decipher what kind of data they need to collect and consider for performance measurement. In the performance of any operation, there are many aspects to look into in order to maximize efficiency including: reducing costs, cycle time, waste, material usage, and so forth while maximizing throughput, quality, and so forth. There is multidimensionality in the goals that organizations want to achieve. In optimizing efficiency, organizations would want to use the least amount of inputs to achieve the most amount of output. However, it must be noted that in using inputs, it does not only produce desirable or marketable outputs. Undesirable outputs may also be produced in the process. These outputs are classified as waste or non-value adding for the organization but are jointly produced (Seiford & Zhu, 2002). Hence, it makes sense for a performance measurement system to credit an organization for its provision of desirable outputs and to penalize it for its production of undesirable outputs. For instance, banks increase their number of deposits and loans but incur overdue debts as well. Likewise, hospitals aim to maximize the total number of patients served and treated but incur failed operations and diagnoses that result to deaths. These are some examples of simultaneous occurrence of desirable and undesirable outputs. Zanella, Camanho, and Dias (2015) and Fare, Grosskopf, Lovell, and Pasurka (1989) claimed that ignoring undesirable outputs might produce misleading performance results. Thus, performance measurement should be able to simultaneously consider the decrease of undesirable outputs and the increase of the desirable outputs.

Furthermore, organizations measure their performance in order to compare themselves both with their own standards and the industry standards. Performance measurement results are used to benchmark if the current performance of an organization is lower, higher, or at par with other organizations. However, organizations tend to measure performance and do benchmarking without considering external factors that are non-controllable and yet have an effect on their performance. Inefficiencies in any organization should not be constrained only to managerial and operational inadequacies. Rather, the inefficiencies caused by its operating environment must also be included in the analysis of performance. These are called exogenous inputs that characterize the operating environment within which the production or the organization is taking place or situated, respectively (Macpherson, Principe, & Shao, 2013). For instance, as described by Avkiran and Rowlands (2008), educational attainment of parents could be considered as an exogenous input in measuring literacy and numeracy in primary schools for the reason that educated populations are likely to show higher rating on these measures due to additional resources available to children. Smith and Street (2004) concluded that in whatever way operating environment is defined, usually some organizations operate in more adverse environments than others in the sense that the external circumstances make the achievement of a given level of attainment more or less difficult, thereby leading to imprecise and unreliable assessment of an organization's inefficiency. The benefit of accounting for these exogenous inputs in decision making lies on the idea that it makes the comparison between organizations more realistic by taking all influences (both internal or external) that contribute to performance rating into account and also providing possible sources that can explain the performance behavior of the organization.

Having said that performance measurement is important in any organization and given that missing data, undesirable outputs, and exogenous inputs may conspire to distort the measurement and analysis of performance, a reliable performance measurement system is therefore necessary to aid in effective decision-making while considering all of the issues mentioned.

Data Envelopment Analysis (DEA) is used as a foundation in this paper to come up with a more reliable performance measurement system that addresses all the aforementioned issues. It is a performance measurement and benchmarking tool with the ability to simultaneously consider all inputs and outputs that may be of interest in arriving at an overall performance or efficiency score (Charnes, Cooper, & Rhodes, 1978; Sherman & Zhu, 2006). The model developed in this paper stems from the basic DEA model. The basic DEA model was modified to consider both exogenous inputs and undesirable outputs while a predictive LP model was formulated alongside the modified DEA model in order to account for missing data before comparative analysis among organizations or Decision Making Units (DMUs) is performed. The improved performance measurement system also includes the reallocation of resources to organizations that can realize the most potential in terms of output. An economic interpretation of reallocation decisions made provides an indication of progress towards specific defined organizational objectives and whether expected results are being achieved.

The objective of this research was to develop a performance measurement tool in the form of software that addresses the aforementioned issues. This software is aimed to help managers in performance measurement analysis by calculating efficiencies of units and identifying what levels of inputs and outputs should be achieved to improve performance.

An overview of the general methodology of the study is discussed in Part II. Development of a Linear Programming (LP) model and modification of the basic DEA model to consider the aforementioned issues is then discussed in Part III. The development of the software based on the LP and DEA models and the selection of inputs and outputs for analysis are discussed in Part IV and V, respectively. Discussion of the results will follow in Part VI and conclusions based on all the insights gathered from the study will be presented in Part VII.

METHODOLOGY

Several phases were undertaken in the completion of this research study as shown in Figure 1. The base methodology of the study involves development of the mathematical model and software, the selection and preparation of data to be examined, measurement of DMU efficiency using DEA analysis, reallocation of resources among DMUs for the maximization of overall system output using available resources, and the economic interpretation of the results. The succeeding sub-sections discuss each of the phases aforementioned.

A. Model Development

For the development of the model, existing performance measurement tools were reviewed to identify which tool best suits the objective of the study. DEA was then selected as the foundation tool to be used because of its advantages over other performance measurement tools (e.g., single dimension performance indicators, ratio analysis, regression analysis) such as the ability to simultaneously consider multiple inputs

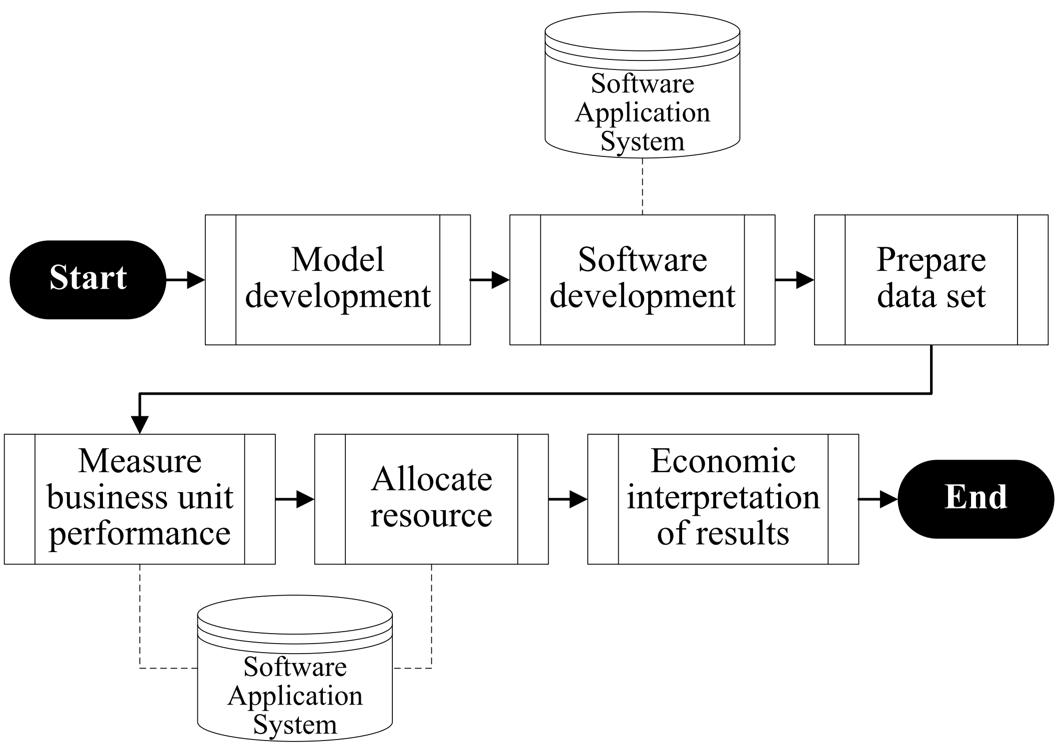

Figure 1. Methodology flowchart.

and outputs and it being a benchmarking tool. Limitations of existing DEA models were then identified, which resulted to three key points to be addressed in this study.

First, existing DEA models do not allow missing cases in the data set. Second, existing DEA models account for undesirable outputs and exogenous inputs separately, which should be taken simultaneously. Third, existing DEA models do not provide insights on how to reallocate excess inputs within the set of DMUs in order to maximize overall system output.

A modified additive DEA model was developed to address the aforementioned concerns and test data from case analyses used in a previous study about libraries (Hoissenzadeh Lofti, Jahanshahloo, & Esmaeli, 2007) to verify if the model gives logical results. Sensitivity analysis was performed to investigate the robustness of the model by considering the effects of parameter adjustments on the model behavior. Model behavior should showcase that inefficient DMUs can be identified if number of DMUs used are within 2*(input + output factors); also that improvement areas are identified for inefficient DMUs; and that expected increase in desirable output has a basis for movement, such that there is either increase in input combinations and/or in undesirable. Upon acceptance of model behavior, software development was then commenced.

B. Software Development

The software development started by identifying user requirements and these were as follows: data setup in the software, needed functions based on the modified additive DEA model, and expected output (numerically and graphically) of the software. To address the user requirements, the software development was divided into phases: (1) data input interface, (2) missing data estimation, (3) DEA efficiency calculation, (4) reallocation of resources, and (5) usability evaluation and enhancements. User testing was conducted on the preliminary software application where comments received from a pool of potential users were used to modify and enhance the software application.

C. Selection and Preparation of Data Set for Case Study

The preparation of the data set began with the selection of DMUs, inputs, and outputs involved in the study as well as the industry for the case study. The healthcare industry was chosen to be the industry for the case study of this research. Typically, inputs considered in analysis are critical resources that were used in the production of the product or service of a DMU. For this purpose, inputs may be, but not limited to, budget allocations, existing capital, buildings, or labor employed (Cantor, Tan, & Yu, 2008). In addition, if there are factors that influence the outcome of outputs and are not in the control of the DMU, these factors should be considered in the analysis as exogenous inputs. Outputs to be considered in analyses, on the other hand, should be the major output, services or products that the company or organization is offering to consumers. If there are waste or non-value adding output, these should minimize and can be considered as undesirable output in the analyses.

The total number of inputs and outputs identified were then checked if compliant with the rule of thumb of DEA analysis that the number of DMUs should be greater than or equal to two times the sum of the number of inputs and outputs. As DEA is data intensive, analysis cannot be conducted if there are missing data. As such, an LP model was developed that will estimate values for missing data.

D. Measurement of DMU Efficiency

A modified additive DEA model was developed and was used to compute for efficiency scores of DMUs under evaluation. Aside from the normal controllable inputs and desirable outputs that existing DEA models consider, the undesirable (bad) outputs, exogenous inputs, and point estimates for missing data were also taken into consideration in the modified additive DEA model.

DEA is a benchmarking tool where DMUs are evaluated relative to other business units by ranking. This is done in order to see the relationship between the set of DMUs such that, for any two units, the first is either “ranked higher than”, “ranked lower than” or “ranked equal to” the second. Rankings make it possible to evaluate complex information according to certain criteria. Computation of the DEA efficiency will generate input slacks and surplus to be used for resource reallocation decisions.

E. Reallocation of Resources

Upon efficiency measurement of DMUs, resource distribution of surplus inputs among DMUs was performed with the goal of improving overall system output. The objective of the reallocation model was to maximize the opportunity to increase the output of the entire system whilst maximizing the use of existing resources within the system. It must be noted that the reallocation model only considers reallocation of the controllable inputs excluding exogenous inputs.

F. Economic Interpretation of Results

Results were analyzed and insights were drawn for possible enhancement of system performance. Initially excess inputs were identified within the system of units being assessed. These excesses were indications of inefficiencies in the original allocation and use of resources. Excesses should not be interpreted as immediate removal of a resource. Rather, it is an indication that measures should be done to reduce cost in that specific resource area. Efficient DMUs can also be identified so that the inefficient DMUs may benchmark their operations against these DMUs.

When inefficiencies are addressed and savings are realized, these savings may be tapped to allocate to DMUs towards increasing overall output of the system. Generally, resource reallocation can be looked into as a means to enhance overall system performance.

MODEL FORMULATION

There were three main phases that correspond to the main functionalities of the DEA benchmarking and reallocation tool. First was the estimation utility for cases of missing data on the data set. Second was the facility for efficiency comparison alongside with the identification of excess inputs or resources of DMUs. Finally, the third functionality was a reallocation model wherein the identified excess inputs were reallocated to DMUs to maximize overall system output. Succeeding sub-sections discuss the details of each of the functionalities mentioned.

A. Estimation of Missing Data

Problem on missing data arises frequently when an organization conducts performance measurement; and the most common method to deal with this problem is through the deletion of categories or cases with missing data. However, such method can seriously affect the number of cases left for analysis, which then can lead to bias and inaccurate findings. To address the problem of deleting categories or cases with missing data, the DEA-based performance measurement and reallocation software allows users to run an analysis with a data set containing missing data through data estimation. That is, the value of the missing data is estimated from the data set instead of deleting. An LP predictive model (see Appendix A) was recommended to estimate the value of missing data. LP was utilized because it does not require assumption of data normality. Statistical regression techniques hold true with assumptions of data normality. LP-based estimates, however, can be used whether the data is normal or non-normal.

The LP predictive model uses pair-wise combination to find the best input or output category that can be used as basis in estimating missing data. Standard error (S_e) is computed from the pair-wise combination and the category that gives the least S_e becomes the basis for estimating the values of missing data. It must be noted that different missing data from different categories may have different bases.

It is usually true using statistical regression that approximately 68% of the estimated values will be within one S_e, and approximately 95% of estimated values will be within two S_e (Winston, 2004). Validation of the LP predictive model was conducted through the data set retrieved from a DEA study on libraries (Hoissenzadeh Lofti et al., 2007) and it was found that approximately 83% of the estimated values were within one S_e.

B. Efficiency Comparison Using the Modified Additive DEA Model With Undesirable Output and Exogenous Input

A modified additive DEA model (see Appendix B) was developed to consider simultaneously undesirable outputs and exogenous inputs aside from the controllable inputs and desirable outputs. The modified additive DEA model was used to compute for efficiency scores of DMUs. The efficiency scores were used as a means to identify whether a DMU was relatively efficient or relatively inefficient and from these, identify the rankings of all DMUs. In addition, excess resources were identified from DMUs that contributed to the system's inefficiencies.

Undesirable output. Undesirable outputs were treated similar as a resource input such as the model developed by Korhonen and Luptacik (2004) and further examined by Yang and Pollitt (2009). It was considered as an input in a manner that for a DMU to be efficient, it should be able to produce more output with the least amount of undesirable outcomes.

Exogenous input. The impact of exogenous inputs or environmental factors in efficiency measurement was considered through benchmarking selection. The most representative model within this option was the one proposed by Banker and Morey (1986). In the Banker and Morey (1986) model, the comparison may include units that operate in a similar or more unfavorable environment compared with the assessed unit. However, this assumes a positive impact on the desirable output. According to Hua, Bian, and Liang (2007), such treatment was not as applicable when both inputs and outputs were simultaneously considered in performance assessment of DMUs. Positive impact of the exogenous input cannot be assumed since its impacts to the inputs and undesirable output indicators cannot be always affirmed to be positive. Instead, to be able to consider exogenous inputs, this requires that reference comparisons utilize the same levels of the transformed exogenous inputs as that of the assessed business units.

C. Reallocation Model

The third functionality of the DEA-based performance measurement and reallocation software was a reallocation model (see Appendix C). The excess resources identified during efficiency comparisons of DMUs were allocated to other DMUs needing these excess resources with the goal of maximizing overall system output under the assumption of maintaining current relative operational efficiency. To provide for the reallocation decision functionality, a forecast model in the form of an inverse DEA model is utilized. Inverse to the DEA, the input reallocation problem determines how much an input should be allocated given that outputs are increased and efficiency remains the same (Wei, Zhang, & Zhang, 2000). Another type of the inverse DEA problem was the forecasting problem which determines how much an output should increase given that inputs are increased and efficiency remains the same (Wei et al., 2000). With the forecast model, it can now determine how additional input can affect output changes.

The objective of the reallocation model was to reallocate resources that can increase the overall output of the whole system. The DEA-based performance measurement and reallocation software was developed to incorporate reallocation of excess inputs to enhance output production of the entire system. The process of selection on reallocation depends on the influence of that reallocation to the changes in all output indicators. The software chooses to allocate resources that have the highest increase in percentage output change. The increase in output production was presented through an index that aggregates the percentage increase in output production of each output indicator (e.g., the model will choose to allocate to a DMU if percentage increase in an output or a combination thereof is greater than percentage increase in other output when allocated to another unit).

SOFTWARE DEVELOPMENT

With considerations to the functionalities and mathematical foundations of the tool, the software prototype was developed using Microsoft.NET framework 4.0 with C# as its primary language. The software was designed and built following the architectural design shown in Figure 2. The software has three core modules:

1. Data Reader Module – The module responsible for reading inputted data from Microsoft Excel, data analysis and storing as a .NET data table.
2. Solver Module – The module responsible for estimating missing data from the input data set and execution of DEA analysis.
3. Visualization Module – The module responsible for generating the charts and table outputs from the input data sets and the resulting output.

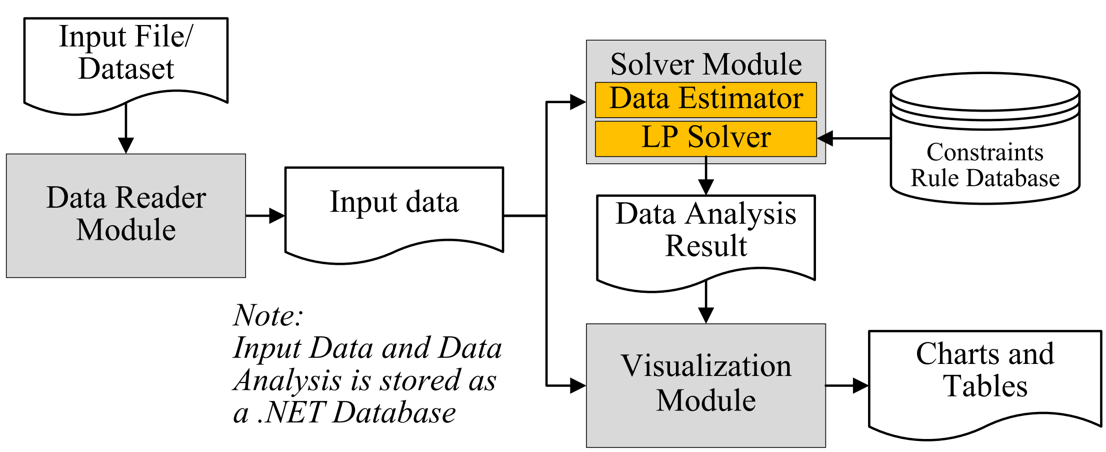

Figure 2. Architectural design.

Users of the software would place the data they wanted to be analyzed in table form on a single Excel spread sheet. The software will then allow the users to specify the excel file where their data is located for DEA analysis. The specified file is then processed by the Data Reader module that reads the data from the excel spread sheet and stores the data in memory as a .NET System.Data.DataTable. The DataTable is then processed by the Solver module to estimate missing data and then performs the DEA analysis. For the Solver module to perform the DEA analysis, the model constraints discussed above are retrieved from the Constraints Rule Database. Finally, the results of the LP Solver is stored, to another DataTable and is processed by the Visualizer module to format the data on screen and generate charts and graphs for the user to easily visualize the results.

A. Data Reader Module

The Data Reader module is responsible for reading the data set from the user (currently only supports MS Excel files) and converts it into a System.Data.DataTable that the Solver module takes as input to perform the DEA analysis. This module makes use of Microsoft Excel COM Interop to read excel files (.xls and .xlsx files) into the software. Microsoft Excel COM Interop makes use of Microsoft Excel's shared libraries to natively read and write excel files. This method provides a fast and direct way of reading Excel files but imposes a restriction that Microsoft Excel has to be installed on the machine where it is going to be used.

Once the data is read from the excel file, it is stored in memory as a System.Data.DataTable. The DataTable class stores the read file as a series of columns and rows. The module further annotates the columns in the DataTable for analysis, by asking the user to identify which fields are considered as Input, Exogenous Input, Desirable Output, and Undesirable Output. The annotated DataTable is then finally passed to the Visualizer module, to display on-screen the data read from the Excel file, and to the Solver module, to perform the DEA analysis.

B. Solver Module

The Solver module is responsible for estimating missing data from the input data set as well as performing the analysis by solving the modified additive DEA model discussed above. The Solver module forwards the DataTable to the Data Estimator sub-module that first checks for missing data from the input DataTable. This is done by checking every cell in the table if they are empty or not. For all empty cells found in the DataTable, the data estimation algorithm described in the section above is run to complete the DataTable.

Once a complete DataTable is obtained, the data is forwarded to the LP Solver sub-module. The LP Solver sub-module makes use of Microsoft Solver Foundation (MSF) to solve the modified additive DEA model using the Simplex method. This sub-module takes as input the DataTable as well as the linear constraints stored in the Constraints Rule Database written in Optimization Modelling Language (OML). Decision variables from the model are then read and stored into a DataTable and passed to the Visualizer Module for display.

C. Visualizer Module

The Visualizer module converts the DataTable results of the DEA analysis into individual GridViews and generates charts for better visualization of the user. This module splits the resulting DataTable into smaller tables and displays them into individual GridViews located in different tabs. The DataTable is divided into two smaller tables which contains its calculated efficiency and reallocation. Finally, these data are then converted into bar charts using Microsoft Charting Controls.

D. Verification of Software Results

Comparison of results from the software developed was conducted to verify if output results were logical and valid. To test the resulting values, software results were verified against reference models ran in General Algebraic Modeling System (GAMS). Verification runs resulted to the same output in the software developed and on GAMS. Comparison of standard error shows that the LP estimates resulted in lower standard error as compared to using regression analysis. Short discussion on the details of the verification can be found on Appendices D and E.

E. Usability Evaluation and Enhancement

The usability of the software interface was evaluated using the Nielsen (1993) heuristics as it was considered to be one of the most dependable usability heuristics. Using Nielsen heuristics for software usability ensures that users can easily use and understand the software. The usability evaluation focused on two things: (a) data entry and (b) presentation of results.

The initial software interface with respect to data entry was found to have usability problems. Labels and dialogue boxes were confusing for the user and the software did not provide feedback or status for the actions done on the software. This means that the terminologies initially used confused the user on what to do and did not give the user an idea on what was happening while extracting the data file. For the presentation of results, heavy user's mental load was an initial problem because a lot of numbers were presented but were not actually significant for interpretation. Graphs and tables were not appropriate for the results presented, which adds to the confusion as to how the different sets of results were related. In addition, coded labels for the different variables were not easily identifiable to a specific variable and were consistently in places that required the user to memorize the labels.

Enhancements were done to ensure users would not have a difficult time to use the software and interpret the results. In the data entry, commonly used terms were used; visibility of the system status were also done as the user extracts the file and while analysis was in progress. Graphs with the corresponding variable names were used instead of coded variables and tabulated results were shown. Graphical representation of the results were also added to make it easy to differentiate values among the input and output variables.

SELECTION OF INPUTS AND OUTPUTS

A. A Case Study on NCR Hospitals

This study considered the analysis of efficiency of hospitals located in the National Capital Region (NCR) of the Philippines. Categorization of hospitals was done in choosing the final set of hospitals to be included in the study following the Department of Health's (2012) hospital classification. Department of Health classified hospitals in the Philippines based on ownership (government or private), scope of services (general or specialized), and functional capacity (level 1, 2 or 3). For the purpose of this study, hospitals that are government-owned, offering general services, and have Level 3 functional capacity were chosen.

Government hospitals are hospitals that are owned by the Philippine government and receive government funding. Meanwhile, general hospitals provide medical and surgical care to the sick and injured, maternity care and shall have as minimum the following clinical services: medicine, pediatrics, obstetrics and gynecology, surgery and anesthesia, emergency services, outpatient, and ancillary services. General hospitals are further classified in three levels of functional capacity. Functional capacity of Level 3 means that a hospital has the necessary equipment and manpower for all the clinical services aforementioned with the presence of teaching with accredited residency training program in major clinical services.

A total of 14 government, general, and Level 3 hospitals were chosen to form the final data set for this study. Some hospitals were not included because of inadequate information in the statistical reports gathered from the Department of Health.

B. Input Selection

Measuring efficiency of hospitals requires appropriate selection of inputs to be considered in the evaluation. Hospital input categories generally fall into three broad sub-categories: capital investment, labor, and other operating expenses (O'Neill, Rauner, Heidenberger, & Kraus, 2008). Capital investment usually refers to beds, equipment, and different facilities that can be directly used for clinical services. The number of fully staffed hospital beds is most often used as a proxy for hospital size and capital investment. O'Niell et al. (2008) made a comparison and taxonomy of hospital efficiency studies and showed that 55 out of 79 research studies included the number of beds as an input category. Some of these research studies were from Ballestero and Maldonado (2004), Chern and Wan (2000), and Grosskopf, Margaritis, and Valdmanis (2004), among others.

Moreover, about two-thirds of hospital operating costs is due to payroll expenses that usually refer to labor costs (Sahin & Ozcan, 2000). Hospital clinical staff consists of physicians, nurses, and other health/medical personnel. Sommersguter-Reichmann (2000) defined number of personnel as a general labor input category. Finally, other than labor costs, non-labor costs are also being incurred by hospitals during operations. Non-labor costs include medical supply, food, drug and pharmaceutical, material costs, among others.

Hosseinzadeh Lofti et al. (2007) purported that influences of external conditions affect the performance of each organization that is regarded as the exogenous inputs that are identifiable but uncontrollable by the organization in consideration. For this study, the population of a city where a hospital is situated was considered as exogenous input. Thus, the corresponding group of inputs that describe the health care services offered in the hospitals under consideration were: (a) authorized bed capacity, (b) total personnel, (c) total expenditure, and (d) population.

C. Output Selection

The Agency on Health Care Research and Quality (AHRQ, 2011) of the United States of America distinguished two types of outputs in healthcare, namely: health services which refer to the products and services that healthcare units provide to constituents (e.g., visits, admissions, drugs, etc.), and health outcomes which refer to resulting output of the services availed (e.g., preventable deaths, functional status, and blood pressure control). Selection of outputs was considered to reflect the general range of hospital activities (Al-Shammari, 1999). In the case of this study, the corresponding group of outputs that describe the health care services offered in the hospitals in assessment were: (a) total patients administered (health services), (b) laboratory services (health services), and (c) net death rate (health outcome).

Number of patients (Al-Shammari, 1999; Katharaki, 2008; Steinmann, Dittrich, Karman, & Zweifel, 2004) and laboratory services (Katharaki, 2008) were selected as criteria for efficiency assessment of DMUs similar to the study conducted by Maria Katharaki on the management of Greek hospitals' gynecological and obstetrics unit. Meanwhile, outputs used in efficiency measures were usually a mix of hospital services such as discharges, visits, and procedures (AHRQ, 2011). In the case of this study, the mix of hospital services was represented through the total patients administered and the laboratory services.

Net death rate, on the other hand, refers to a health outcome measure for quality of hospital system-level performance. Although there were debates on the drawbacks on the use of mortality as a quality measure, it was one of the most widely used (Kroch & Duan, 2008). Mortality rate was considered to be a simple measure as it is easily observable by counting deaths from discharges. In addition, a document review of hospital performance reports indicated net death rate to be a common measure among NCR tertiary hospitals (Katharaki, 2008).

DISCUSSION

The case study uses the DEA-based performance measurement and reallocation software to assess 14 NCR hospital units in the perspective of a managing body. The case study uses seven efficiency factors. Inputs used in analysis were authorized bed capacity, total personnel, and total expenditure. Exogenous input used is the population of the local government unit being served. Outputs considered are total patients served and the laboratory services offered. Undesirable output considered is the net death rate of the hospital unit.

A. Missing Data

Ideally, DEA analysis should only be carried out with all available data on hand and complete. However, even at the best of efforts to gather data, it is a reality that there will be cases of data that will not be available. Performance data of each of the hospitals in this case analysis was gathered from DOH performance reports. A thorough document review of reports was done to gather the most complete information on hospital performance. Unfortunately, there are still instances of missing data as shown in Table 1. It was observed that 83% of efficiency factors data is available; however nine of the 14 hospitals have at least one efficiency factor missing. In traditional DEA software, the efficiency analysis can now only be carried out for the five remaining hospitals with complete data. With just five remaining hospital qualified for analysis, only 36% of total data available will be utilized.

With the DEA-based performance measurement and reallocation software developed, it allows the flexibility of carrying out efficiency analysis despite some cases of missing data. The LP predictive model optimizes the best predictor in the data set to estimate values for missing data. Using the software with the LP predictive model developed, it estimated data for: total personnel, total patients, laboratory services, and net death rate (see Table 2). As such, DEA efficiency analysis can now be performed and the data set with estimated values was used.

B. Insights on Efficiency

The case analysis using the DEA-based performance measurement and reallocation software in Figure 3 determined that two hospitals are below par of its peers. These are East Avenue Medical Center and Tondo Medical Center.

Tondo Medical Center's efficiency is rated at 93.40%, and East Avenue Medical Center's efficiency is rated at 91.84%. The efficiency ratings indicated that they are only performing at a rate to what is considered as efficient based on other hospital peers' input-output performance.

C. Excess Inputs/Resources Identification

Aside from the ratings that indicate relative efficiency; the software was also capable of identifying excess inputs. Excess inputs were

Table 1. ***Initial Data Set for Case Analysis of NCR Hospitals***

DMU	Authorized Bed Capacity	Total Personnel	Total Expenditure	Population	Total Patients	Laboratory Services	Net Death Rate
Amang Rodriguez Memorial Medical Center	300.00			424,150.00	84,005.00	566,159.00	0.0380
Dr. Jose N. Rodriguez Memorial Hospital	200.00	97.00	54,980,664.4500	744,500.00	70,007.00	453,005.00	
East Avenue Medical Center	600.00	1,025.00		345,215.00	155,396.00		0.0332
Gat Andres Bonifacio Memorial Medical Center	150.00			330,434.20	89,842.00	244,754.00	0.0290
Jose R. Reyes Memorial Medical Center	450.00	1,074.00	909,947,310.53	330,434.20	224,044.00	1,154,747.00	0.0530
Justice Jose Abad Santos General Hospital	150.00	454.00	49,112,110.56	330,434.20	62,795.00	70,716.00	
Las Pinas General Hospital & Satellite Trauma Center	150.00	292.00	137,627,154.70	552,573.00	49,456.00	128,625.00	0.0300
Mandaluyong City Medical Center	150.00	547.00		328,699.00	72,672.00	64,849.00	0.2950
Pasay City General Hospital	150.00			196,434.50	54,925.00	173,849.00	0.0200
Philippine General Hospital	1,346.00	3,653.00	2,248,344,940.00	330,434.20	525,741.00	1,341,067.00	0.0453
Diosdado Macapagal Memorial Medical Center	2,000.00			744,500.00	36,384.00	54,392.00	0.0193
Quezon City General Hospital	250.00	424.00	297,597,919.78	345,215.00	85,355.00	277,814.00	0.0130
Quirino Memorial Medical Center	350.00	192.00	269,836,926.00	345,215.00			
Tondo Medical Center	200.00	419.00	249,406,489.13	330,434.20	78,775.00	209,112.00	0.0326

Table 2. ***Final Data Set with Estimates for Case Analysis of NCR Hospitals***

DMU	Authorized Bed Capacity	Total Personnel	Total Expenditure	Population	Total Patients	Laboratory Services	Net Death Rate
Amang Rodriguez Memorial Medical Center	300.00	518.4033	272,796,295.2854	424,150.00	84,005.00	566,159.00	0.0380
Dr. Jose N. Rodriguez Memorial Hospital	200.00	97.00	54,980,664.4500	744,500.00	70,007.00	453,005.00	0.0294
East Avenue Medical Center	600.00	1,025.00	592,073,857.6638	345,215.00	155,396.00	409,626.37	0.0332
Gat Andres Bonifacio Memorial Medical Center	150.00	559.8232	298,900,750.0710	330,434.20	89,842.00	244,754.00	0.0290
Jose R. Reyes Memorial Medical Center	450.00	1,074.00	909,947,310.53	330,434.20	224,044.00	1,154,747.00	0.0530
Justice Jose Abad Santos General Hospital	150.00	454.00	49,112,110.56	330,434.20	62,795.00	70,716.00	0.0294
Las Pinas General Hospital & Satellite Trauma Center	150.00	292.00	137,627,154.70	552,573.00	49,456.00	128,625.00	0.0300
Mandaluyong City Medical Center	150.00	547.00	222,112,418.9643	328,699.00	72,672.00	64,849.00	0.2950
Pasay City General Hospital	150.00	312.0491	142,743,606.3754	196,434.50	54,925.00	173,849.00	0.0200
Philippine General Hospital	1,346.00	3,653.00	2,248,344,940.00	330,434.20	525,741.00	1,341,067.00	0.0453
Diosdado Macapagal Memorial Medical Center	2,000.00	180.4806	598,223,836.5998	744,500.00	36,384.00	54,392.00	0.0193
Quezon City General Hospital	250.00	424.00	297,597,919.78	345,215.00	85,355.00	277,814.00	0.0130
Quirino Memorial Medical Center	350.00	192.00	269,836,926.00	345,215.00	83,343.2800	220,681.3083	0.0310
Tondo Medical Center	200.00	419.00	249,406,489.13	330,434.20	78,775.00	209,112.00	0.0326

identified in comparison to other hospitals, that is, a select hospital with consideration of its efficiency is using more than what it needs to produce its current level of output.

With the help of the software (see Table 3), excess resources have been identified for authorized bed capacity (133.84), personnel (total 103.84), and total expenditure (5,040,625).

The tool was able to identify that East Avenue Medical Center have excess resources in authorized bed capacity by approximately 134 excess beds (see Figure 4). This is indicative that the hospital has a larger hospital size in terms of capital investment (O'Neill et al., 2008) relative to its output performance. Similarly, Tondo Medical Center is identified to have excess in total expenditure amounting to as much as 5 million pesos. Meanwhile, Mandaluyong Medical Center is considered relatively efficient but has an excess of 103.84 personnel. Excess personnel at the efficient level indicate that the hospital can afford to lessen human capital and still expect to achieve the same level of relative efficiency.

Identifying excess resources is useful to indicate which aspects or area a hospital unit

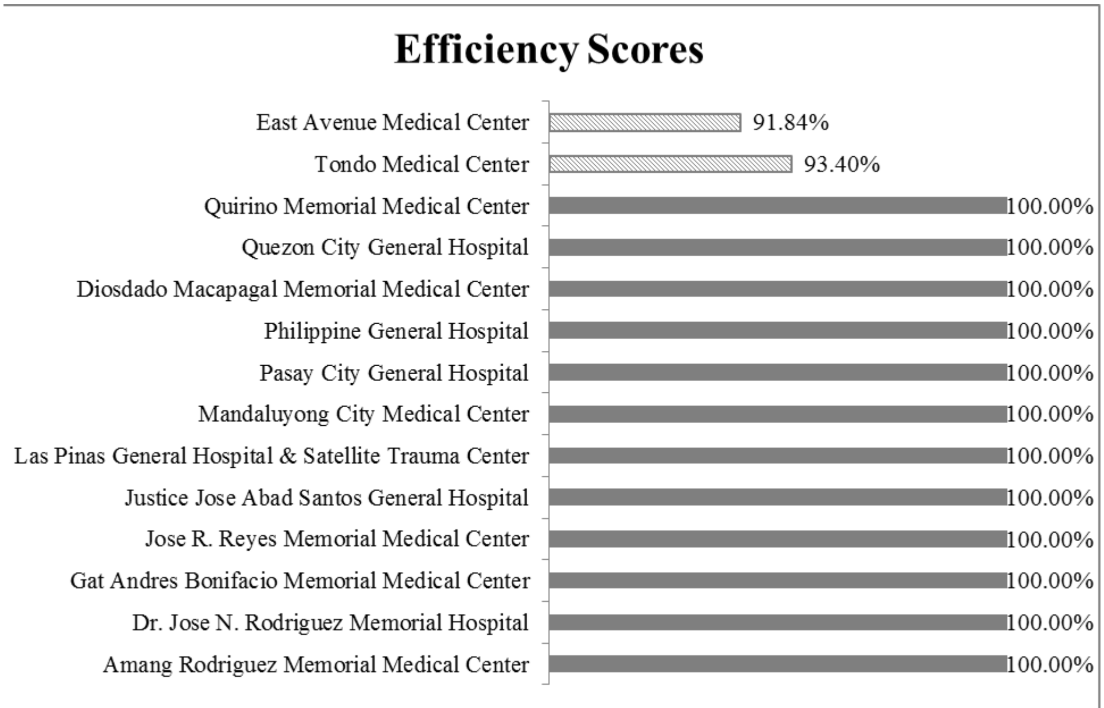

Figure 3. Efficiency results and excess inputs.

is weak at. This gives insights for hospitals on where to put focus to improve operational performance. However, as a managing body there are strategic decisions to be made on how best to utilize limited resources that would maximize the overall system (all hospitals) total output performance. Given that units are already identified with excess resources, a managing body has to make a decision whether other hospitals may need the additional budget or resources to produce more output.

The software developed has an additional capability to provide analysis on reallocation of excess inputs. The objective of the reallocation decision is to enhance the percentage increase in the output indicators as a system (all hospitals considered), namely: total patients administered and laboratory services. The increase in output production is indicative through an index that aggregates the percentage increase in output production of each output category.

Reallocation analysis provides insights in two areas: (1) target operational improvements for inefficient DMUs, and (2) reallocation excess inputs and expected improvements in system output performance.

D. Targets on Operational Improvements

Summary results of the reallocation analysis presented in Table 4 showed that the inefficient units could still increase their level of output given current levels of relative efficiency. For East Avenue Medical Center to retain its current level of relative efficiency, even with 22% less capital investment (bed capacity), it has to focus on improving operations with targets in increasing the number of laboratory services by 51.24%. This is also a cue for East Avenue Medical Center to review their capital expenditures and target to minimize unnecessary capital spending, while focusing on improvements to increase laboratory

Table 3. ***Excess Inputs***

DMU	Efficiency Score	Excess Authorized Bed Capacity	Excess Total Personnel	Excess Total Expenditure
Amang Rodriguez Memorial Medical Center	100.00%	0.00	0.00	0.00
Dr. Jose N. Rodriguez Memorial Hospital	100.00%	0.00	0.00	0.00
East Avenue Medical Center	91.84%	133.84	0.00	0.00
Gat Andres Bonifacio Memorial Medical Center	100.00%	0.00	0.00	0.00
Jose R. Reyes Memorial Medical Center	100.00%	0.00	0.00	0.00
Justice Jose Abad Santos General Hospital	100.00%	0.00	0.00	0.00
Las Pinas General Hospital & Satellite Trauma Center	100.00%	0.00	0.00	0.00
Mandaluyong City Medical Center	100.00%	0.00	103.84	0.00
Pasay City General Hospital	100.00%	0.00	0.00	0.00
Philippine General Hospital	100.00%	0.00	0.00	0.00
Diosdado Macapagal Memorial Medical Center	100.00%	0.00	0.00	0.00
Quezon City General Hospital	100.00%	0.00	0.00	0.00
Quirino Memorial Medical Center	100.00%	0.00	0.00	0.00
Tondo Medical Center	93.40%	0.00	0.00	5,040,625.27

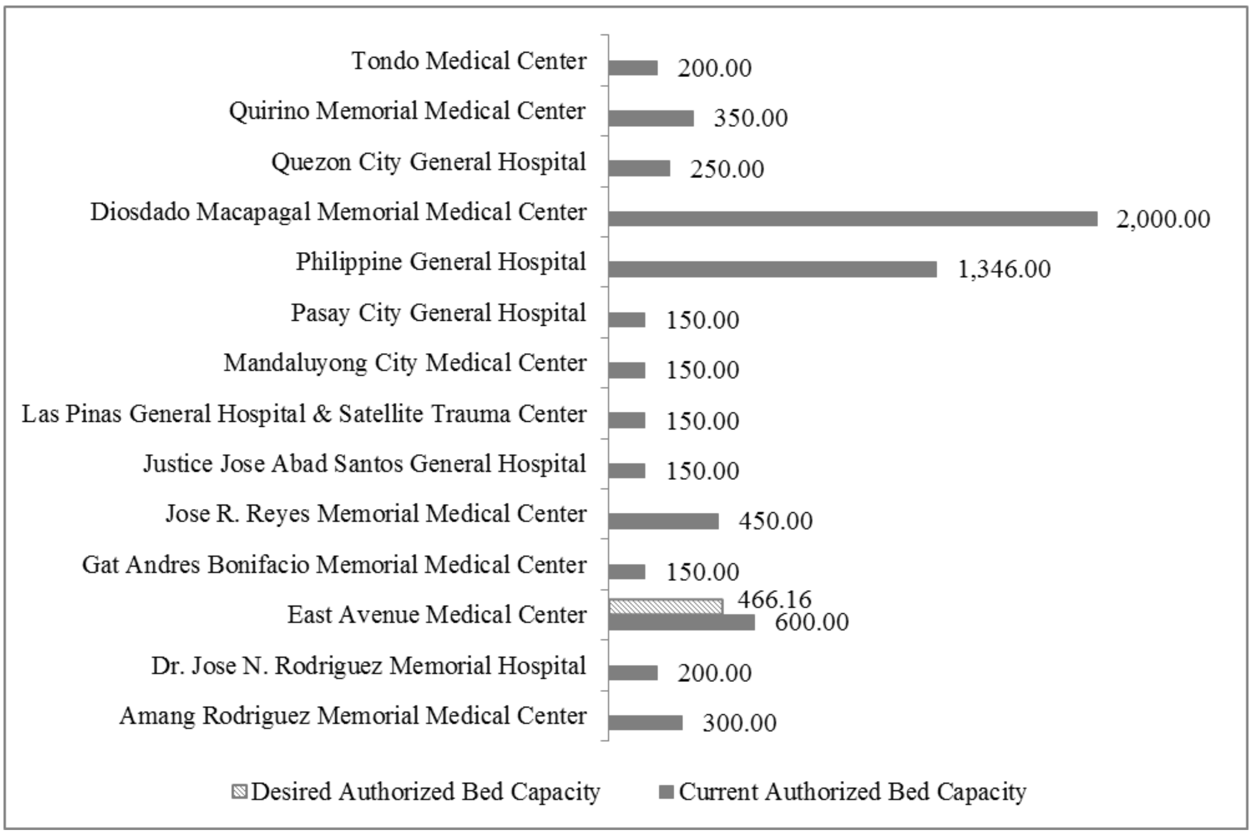

Figure 4. Identification of excess authorized bed capacity.

Table 4. ***Summary of Reallocation of Inputs and Expected Change in Output***

DMU	Reallocated Bed Capacity	Reallocated Total Personnel	Reallocated Total Expenditure	Change in Total Patients	Change in Laboratory Services	Change in Net Death Rate
Amang Rodriguez Memorial Medical Center	-	-	-	-	-	-
Dr. Jose N. Rodriguez Memorial Hospital	-	-	-	-	-	-
East Avenue Medical Center	-	-	-	-	209,786.08	0.01
Gat Andres Bonifacio Memorial Medical Center	-	-	-	-	-	-
Jose R. Reyes Memorial Medical Center	-	-	-	-	-	-
Justice Jose Abad Santos General Hospital	1.07	-	5,040,625.77	-	10,641.23	-
Las Pinas General Hospital & Satellite Trauma Center	51.43	-	-	26,158.79	286,696.31	-
Mandaluyong City Medical Center	81.34	-	-	1,218.82	332,519.51	-
Pasay City General Hospital	-	-	-	-	-	-
Philippine General Hospital	-	-	-	-	-	-
Diosdado Macapagal Memorial Medical Center	-	-	-	33,623.00	398,613.00	0.01
Quezon City General Hospital	-	-	-	-	-	-
Quirino Memorial Medical Center	-	103.84	-	3,500.26	97,182.57	-
Tondo Medical Center	-	-	-	-	112,045.91	-

services. However, East Avenue Medical Center should take caution or precautionary measures in increasing the laboratory services with less capital investment as it will also potentially affect health outcome by potentially increasing net death rate by 19%.

Meanwhile, Tondo Medical Center has excess expenditure of about 2% of their total allocation. While budget for total expenditure can be reduced, it should focus on improving its laboratory services from 53.58% to 93.4% efficiency.

Efficiency results also showed that Diosdado Macapagal Memorial Medical Center was efficient relative to its peers. However, reallocation results as seen in the summary results in Table 4 showed that health services outcome can still be maximized given the current resources of the hospital. Diosdado Macapagal Memorial Medical Center can further study their internal operations to focus on increasing its health services particularly on laboratory services. However, the hospital should take precautionary measures against increasing its health services while maintaining current resources as this will also potentially affect health outcome by potentially increasing net death rate up to 52%.

While there is room for improvement for the two inefficient units, there is also reason to believe that output performance of efficient hospital units can be expected to increase given that excess budgets are reallocated to these hospital units. Reallocation analysis has identified areas for improvement for the inefficient units as well as identified candidates for potential reallocation of resources and targeted increase in output performance.

E. Resource Reallocation Decision

Decision variables on the reallocation of excess inputs have certain economic impact on the hospitals. Economic impact on the hospitals can result to improved services that translate to serving more patients and increasing their satisfaction level. Additional income generated from the increase in patients can be used to buy additional input resources such as equipment to further strengthen the service capabilities. In addition, excess inputs can be translated to realignment of capital investment within the same hospital.

The software in Figure 5 identified hospitals as candidates for reallocation of capital investments (authorized bed capacity); and these are: Justice Jose Abad Santos Medical Center, Las Piñas General Hospital & Satellite Trauma Center, and Mandaluyong Medical Center. By allocating

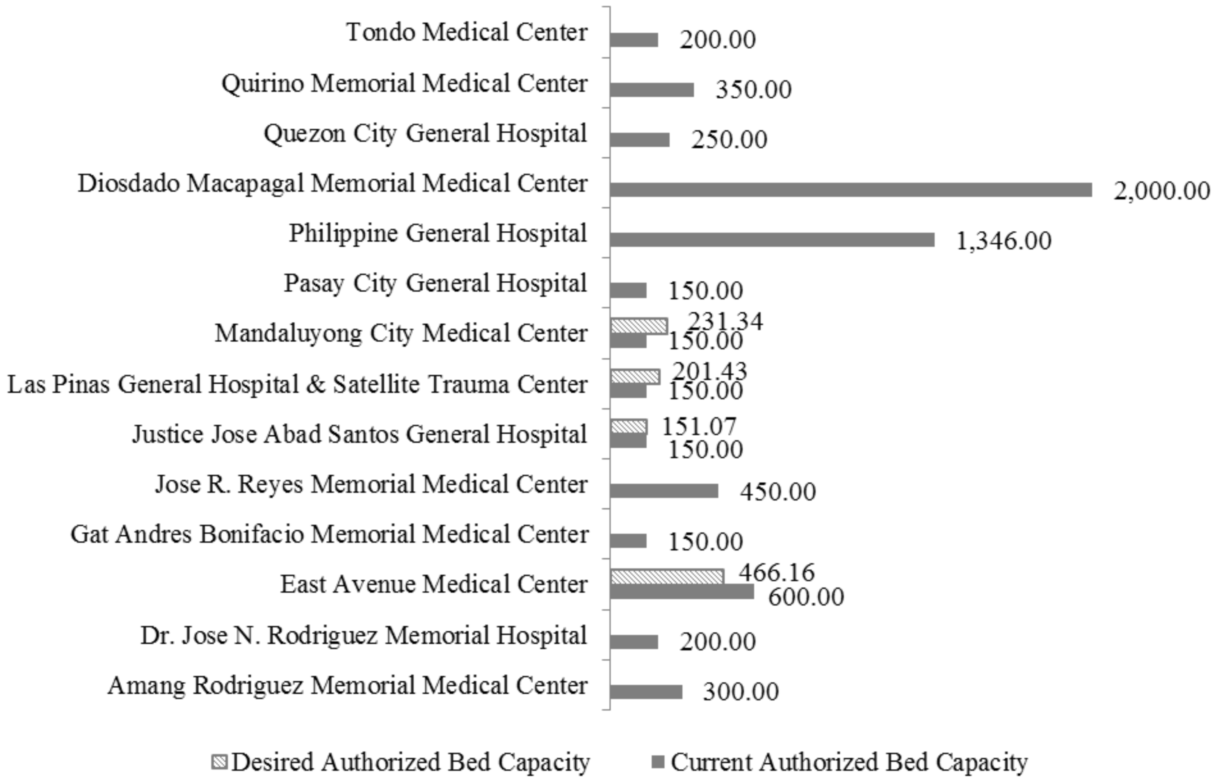

Figure 5. Identification of reallocation of authorized bed capacity.

capital investment to these hospitals, using the same relative efficiency, increase in both health services of total patients administered and laboratory services can be expected. An increase by 34.29% of capital investment in Las Piñas General Hospital & Satellite Trauma Center can potentially increase administered patients by 42% and laboratory services by as much as three times its current performance (see Table 4); while an increase by 54% of capital investment in Mandaluyong Medical Center can potentially increase patients administered by 2% and laboratory services by as much as six times its current output performance. Reallocation for bed capacity should be in line with the approved bed capacity issued with the hospital license.

The reallocation candidate identified by the software as shown in Figure 6 for excess personnel is Quirino Memorial Medical Center. An increase of 54% in personnel can potentially increase health services by 4% and 44% for patients administered and laboratory services respectively. The reallocated excess personnel should have the same position and skill needed to render the services.

For other excess inputs that can be reallocated to hospitals such as medical equipment, there is a need to have a change in accountability of the assigned resource. The additional services generated from the reallocated resources lead to revenues that can be used for additional operational funds in the form of additional incentive to personnel or increase in medical supplies/equipment. In the case of Justice Jose Abad Santos Medical Center, to have an increase in capital investment by 1% and total expenditure

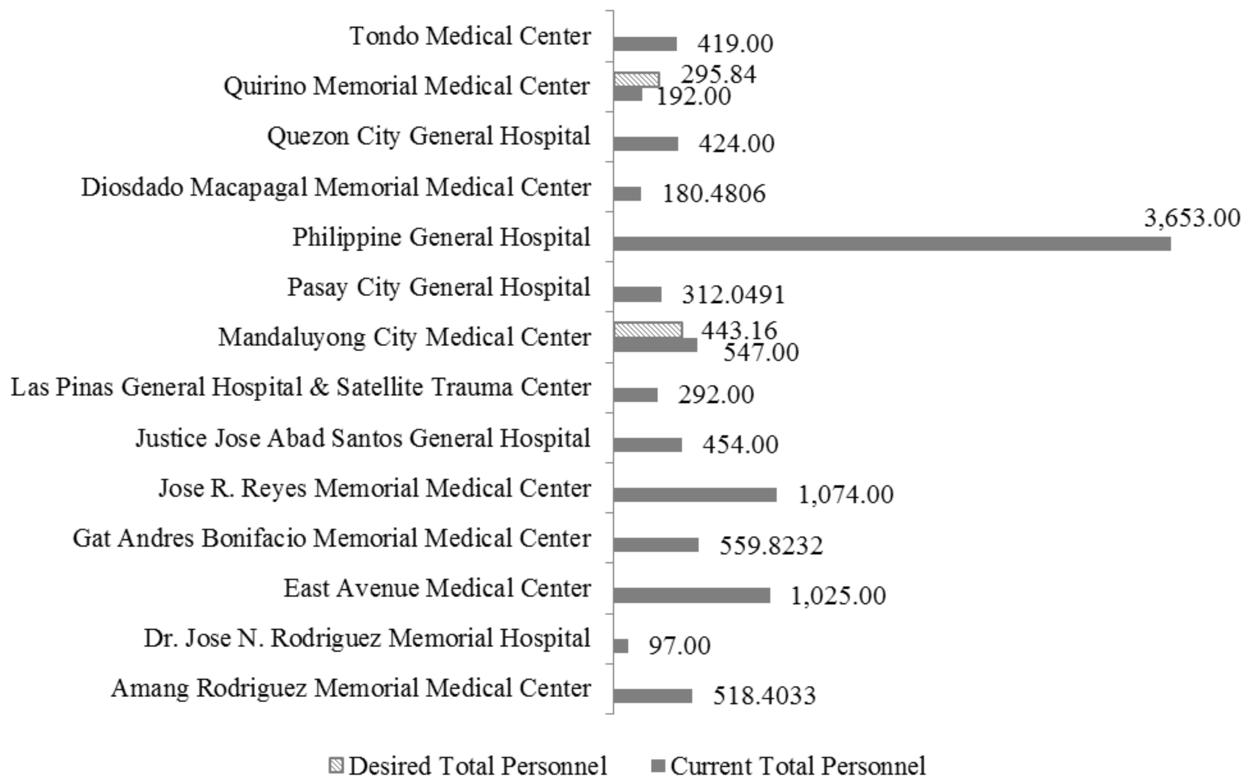

Figure 6. Identification of reallocation of total personnel

by 10%, at its current efficiency level, it can potentially increase laboratory services by 15%. The increase in laboratory services can generate income that can help the hospital to further equip the laboratory with the latest medical equipment.

CONCLUSION

The case analysis was able to showcase functionalities and application of the DEA-based performance measurement and reallocation software. Through the software application system, DEA analysis was carried out despite instances of missing data in some of the hospital units. Also the tool was able to identify that Amang Rodriguez Memorial Medical Center, East Avenue Medical Center, and Tondo Medical Center were performing relatively below its peers. Through the use of the software excess input analysis, targets for output performance and candidates for resource reallocation were identified.

To improve performance of East Avenue Medical Center, it is suggested that capital expenditure is reviewed as this was identified to be in excess. Internal operations of East Avenue Medical Center should be reviewed with focus on increasing health services, particularly in laboratory services. The medical center should also take caution that focusing on increasing health services should not affect service outcome. Similar to East Avenue Medical Center, Tondo Medical Center's internal operations should also be improved to increase health services with particular focus on laboratory services.

If savings from capital expenditure (from East Avenue Medical Center), personnel expenditure (from Mandaluyong Medical Center), and total expenditure (from Tondo Medical Center) are realized these excess resources or budget can be re-allocated to other medical centers. To maximize increase in health services, excess budget for capital investments should be reallocated to Justice Jose Abad Santos Medical Center, Las Piñas General Hospital & Satellite Trauma Center, and Mandaluyong Medical Center. Excess personnel resources on personnel can be maximized with Quirino Memorial Medical Center. Excess budget on total expenditure should be best re-allocated to Mandaluyong Medical Center.

ACKNOWLEDGMENT

I would like to extend my utmost gratitude on some people who have helped on this study in one way or another: (a) to Engr. Cecille Matienzo, Division Chief of the Center for Medical Device Regulation and Radiation, Food and Drug Administration, Department of Health for her valuable inputs in the economic interpretation of the results; (b) to Dr. Nicolas Lutero III, CESO III and Director IV of the Bureau of Health Facilities and Services, Department of Health and his staff for allowing me to gather data from hospital statistical reports and for entertaining my questions about hospital performance indicators and procedures on performance measurement; and lastly (c) to the University Research Coordination Office of De La Salle University for funding this research.

REFERENCES

Agency on Health Care Research and Quality [AHCRQ]. (2011). *Selecting quality and resource use measures: A decision guide for community quality collaboratives - Part III.* Retrieved from http://www.ahrq.gov/qual/perfmeasguide/perfmeaspt3.htm

Al-Shammari, M. (1999). A multi-criteria data envelopment analysis model for measuring the productive efficiency of hospitals. *International Journal of Operations & Production Management*, *19*(9), 879–890.

Avkiran, N. K., & Rowlands, T. (2008). How to better identify the true managerial performance: State of the art using DEA. *Omega*, *36*(2), 317–324.

Ballestero, E., & Maldonado, J. A. (2004). Objective measurement of efficiency: Applying single price model to rank hospital activities. *Computers & Operations Research*, *31*(4), 515–532.

Banker, R. D., & Morey, R. C. (1986). Efficiency analysis for exogenously fixed inputs and outputs. *Operations and Research*, *34*(4), 513–21.

Cantor, V. J. M., Tan, M. L., & Yu, R. S. (2008). *A performance measurement framework considering undesirable outputs, exogenous inputs and missing data using additive DEA model* (Unpublished undergraduate thesis). De La Salle University, Manila Philippines.

Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operations Research*, *2*(6), 429–444.

Chen, Y., Li, Y., Xie, Q., An, Q., & Liang, L. (2014). Data envelopment analysis with missing data: A multiple imputation approach. *International Journal of Information and Decision Sciences*, *6*(4), 315–337.

Chern, J. Y., & Wan, T. T. H. (2000). The impact of the prospective payment system on the technical efficiency of hospitals. *Journal of Medical Systems*, *24*,159–172.

Colledani, M., & Tolio, T. (2009). Performance evaluation of production systems monitored by statistical process control and off-line inspections. *International Journal of Production Economics*, *120*(2), 348–367.

de Lima, E. P., da Costa, S. E. G., & de Faria, A. R. (2009). Taking operations strategy into practice: Developing a process for defining priorities and performance measures. *International Journal of Production Economics*, *122*(1), 403 – 418.

Department of Health. (2012). *Proposed classification of hospitals and other health facilities*. Retrieved from http://www.doh.gov.ph/content/proposed-classification-hospitals-and-other-health-facilities.html

Fare, R., Grosskopf, S., Lovell, C. A., & Pasurka, C. (1989). Multilateral productivity comparisons when some outputs are undesirable: A nonparametric approach. *The Review of Economics and Statistics*, *71*(1), 90–98.

Gerschewski, S., & Xiao, S. S. (2015). Beyond financial indicators: AN assessment of the measurement of performance for international new ventures. *International Business Review*, *24*(4), 615–629.

Grosskopf, S., Margaritis, D., & Valdmanis, V. (2004). Competitive effects on teaching hospitals. *European Journal of Operational Research,154*(2), 515–527.

Hoissenzadeh Lofti, F., Jahanshahloo, G. R., & Esmaeli, M. (2007). Sensitivity analysis of efficient units in the presence of nondiscretionary inputs. *Applied Mathematics and Computation*, *190*(2), 1185–1197.

Hua, Z., Bian, Y., & Liang, L. (2007). Eco-efficiency analysis of paper mills along the Huai River: An extended DEA approach. *Omega*, *35*(5), 578–587.

Kao, C., & Liu, S. T. (2000). Data envelopment analysis with missing data: An application to university libraries in Taiwan. *The Journal of the Operational Research Society*, *51*(8), 897–905.

Katharaki, M. (2008). Approaching the management of hospital units with an operation research technique: The case of 32 Greek obstetric and gynaecology public units. *Health Policy*, *85*(1), 19–31.

Korhonen, P., & Luptacik, M. (2004). Eco-efficiency analysis of power plants: An extension of data envelopment analysis. *European Journal of Operational Research, 154*(2), 437–446.

Kroch, E., & Duan, M. (2008). *Care science risk assessment model: Hospital performance measurement.* Rockville, MD: CareScience.

Macpherson, A. J., Principe, P. P., & Shao, Y. (2013). Controlling for exogenous environmental variables when using data envelopment analysis for regional environmental assessments. *Journal of Environmental Management*, *119*, 220–229.

Nielsen, J. (1993). *Usability engineering*. Boston: Academic Press.

O'Neill, L., Rauner, M. S., Heidenberger, K., & Kraus, M. (2008). A cross-national comparison and taxonomy of hospital efficiency studies using data envelopment analysis. *Socio-Economic Planning Sciences*, *42*(3), 158–189.

Sahin, I., & Ozcan, Y. A. (2000). Public sector hospitals efficiency for provincial markets in Turkey. *Journal of Medical Systems*, *24*(6), 307–320.

Seiford, L. M., & Zhu, J. (2002). Modeling undesirable factors in efficiency evaluation. *European Journal of Operational Research*, *142*(1), 16–20.

Sherman, H. D., & Zhu, J. (2006). *Service productivity management: Improving service performance using data envelopment analysis*. New York: Springer Science + Business Media.

Smith, P., & Street, A. (2004). Measuring the efficiency of public services: The limits of analysis. *Journal of the Royal Statistical Society: Series A*, *168*(2), 401–417.

Sommersguter-Reichmann, M. (2000). The impact of the Austrian hospital financing reform on hospital productivity: Empirical evidence on efficiency and technology changes using a non-parametric input-based Malmquist approach. *Health Care Management Science*, *3*(4), 309–332.

Steinmann, L., Dittrich, G., Karman, A., & Zweifel, P. (2004). Measuring and comparing the (In) efficiency of German and Swiss hospitals. *The European Journal of Health Economics, 5*(3), 216–226.

Tung, A., Baird, K., & Schoch, H. P. (2011). Factors influencing the effectiveness of performance measurement systems. *International Journal of Operations & Production Management*, *31*(12), 1287–1310.

Wei, Q., Zhang, J., & Zhang, X. (2000). An inverse DEA model for inputs/outputs estimate. *European Journal of Operations Research*, *121*(1), 151–163.

Winston, W. (2004). *Operations research: Applications and algorithms* (4th ed.). UK: Thomson Brooks/Cole.

Verbeeten, F. H. M., & Boons, A. N. A. M (2009). Strategic priorities, performance measures and performance: An empirical analysis in Dutch firms. *European Management Journal*, *27*(2), 113–128.

Yang, H., & Pollitt, M. (2009). Incorporating both undesirable outputs and uncontrollable variables into DEA: The performance of Chinese coal-fired power plants. *European Journal of Operational Research*, *197*(3) 1095–1105.

Zha, Y., Song, A., Xu, C., & Yang, H. (2013). Dealing with missing data based on data envelopment analysis and halo effect. *Applied Mathematical Modelling*, *37*(9), 6135–6145.

Zanella, A., Camanho, A. S., & Dias, T. G. (2015). Undesirable outputs and weighting schemes in composite indicators based on data envelopment analysis. *European Journal of Operational Research*, *245*(2), 517–530.

APPENDIX A: *Model Formulation of Linear Programming Predictive Model for Missing Data*

Indices

i = Data set category

Parameters

Y_i = Actual value of parameter to be estimated in data set i
X_i = Predictor parameter for data set i

Variables

A_1 = Positive slope coefficient
A_2 = Negative slope coefficient
B_1 = Positive y intercept coefficient
B_2 = Negative y intercept coefficient
δ_1 = Estimate error for data set i

Linear programming data estimation model

$$Min \sum_{i=1}^{n} \delta_i$$

Subject to:

$$Y_i - [(A_1 - A_2) * X_i + B_1 - B_2] + \delta_i \geq 0, \quad \forall i$$

$$Y_i - [(A_1 - A_2) * X_i + B_1 - B_2] - \delta_i \geq 0, \quad \forall i$$

$$A_1, A_2, B_1, B_2, \delta_i, \quad \forall i$$

Objective function. The objective of the model is to find the best combination of coefficients to minimize the sum of estimate error from what is estimated by the model, and the actual value of the parameter to be estimated (this is denoted by the variable δ).

$$Min \sum_{i=1}^{n} \delta_i \tag{1}$$

Constraints. Estimate errors are obtained and calculated through error constraints. These error constraints limit that the error should be the difference of the predicted and actual value. Two error types are possible, one would be that the predicted value is greater than the actual value (negative error), and the other type would be that the predicted value is less than the actual value (positive error). To differentiate positive and negative errors two constraints are introduced for the two possible error types.

(a) *Negative error constraint.* The inequality constrains that if an estimated value is less than the actual value, the estimate error should be greater than zero (2).

$$Y_i - [(A_1 - A_2) * X_i + B_1 - B_2] + \delta_i \geq 0, \quad \forall i \tag{2}$$

(b) *Positive error constraint.* This constraint limits that if an estimated value is less than the actual value, the estimate error should be less than zero (3).

$$Y_i - [(A_1 - A_2) * X_i + B_1 - B_2] - \delta_i \geq 0, \quad \forall i \tag{3}$$

(c) *Nonnegativity constraint.*

$$A_1, A_2, B_1, B_2, \delta_i, \quad \forall i \tag{4}$$

Utilizing the LP predictive model, a series of steps is done to generate estimate values for missing data. Having a data set with missing data, initially use all data that are complete, that is, those cases that are complete across all inputs and outputs. The LP predictive model is then run for each pair wise combination between and among inputs and outputs. The coefficients, sum error, and standard residuals of each pair wise comparison were recorded. After all runs were done, the results of the sum errors are compared. For every input and output, the linear function with the least sum of errors was noted; these linear functions will be used to estimate missing data.

The LP predictive model then generates a single point estimate of the missing data, however, since the values are just an estimate there is no guarantee that the calculated value will represent the "real" value of the missing data. An interval estimate is deemed better to represent the range of values in which the missing value can be expected. From each of the pair wise combination comparisons, the standard error of the linear function was calculated using the following formula given in (5):

$$S_e = \sqrt{\frac{SSE}{n-2}} \tag{5}$$

It is usually true that approximately 68% of the estimated values of y will be within S_e, and approximately 95% of estimated y will be within $2S_e$ (Winston, 2004). In the library data set used from a DEA study (Hoissenzadeh Loftiet al., 2007), it was found that approximately 83% are within S_e.

APPENDIX B: *Model Formulation for DEA with considerations to exogenous input and undesirable output*

Model indices

j = Decision Making Unit (DMU) (j = 1, 2, 3 … n)
k = DMU in assessment (k = 1, 2, 3 … n)
i = Type of input (i = 1, 2, 3 … m)
r = Type of output (r = 1, 2, 3 … s)

Decision variables

θ_k = Proportion of input use in output production
λ_{jk} = Proportion of input and output benchmarked from DMU for DMU

Model parameters

E_{ij} = Exogenous input of DMU j
X_{ij} = Input i used by DMU j
Y_{rj} = Output r of DMU j
U_{rj} = Undesirable output r of DMU j

Modified DEA model

$$M\,in\ z = \theta_k$$

Subject to:

$$\sum_{j=1}^{n}\left(\lambda_{jk} * X_{ij}\right) \le \theta_k * X_{ik}, \qquad \forall k \forall i \qquad \text{(Input constraint)}$$

$$\sum_{j=1}^{n}\left(\lambda_{jk} * E_{ij}\right) = \theta_k * E_{ik}, \qquad \forall k \forall i \qquad \text{(Exogenous input constraint)}$$

$$\sum_{j=1}^{n}\left(\lambda_{jk} * Y_{rj}\right) \ge Y_{rk}, \qquad \forall k \forall r \qquad \text{(Output constraint)}$$

$$\sum_{j=1}^{n}\left(\lambda_{jk} * U_{rj}\right) \le U_{rk}, \qquad \forall k \forall r \qquad \text{(Undesirable output constraint)}$$

$$\sum_{j=1}^{n} \lambda_{jk} = 1, \qquad \forall k \qquad \text{(Variable returns} - \text{to} - \text{scale constraint)}$$

$$\lambda_{jk} \geq 0, \qquad \forall j \forall k \qquad \text{(Nonnegativity constraint)}$$

$$\theta_k \geq 0, \qquad \forall i \forall k \qquad \text{(Nonnegativity constraint)}$$

The DEA model is similar to that of Charnes et al. (1978), however to accommodate variations in input and output, additional constraints were introduced. Such that of exogenous input and undesirable output:

Exogenous input. The impacts of the exogenous inputs are reflected by the way of reference units selection. The most representative model within this option is the one proposed by Banker and Morey (1986). In the Banker and Morey (1986) model, the reference set may include units that operate in the similar or more unfavorable environment compared with the assessed unit (Banker & Morey, 1986). Furthermore, it assumes a positive impact on the desirable output, such that the effect of the exogenous variable is deducted from the objective function of the primal model. However, according to Hua et al. (2007), such treatment is not as applicable when both inputs and outputs (both good and bad) are simultaneously considered in the objective function. Positive impact of the exogenous cannot be assumed to the objective function since the impacts to the inputs and bad outputs cannot be affirmed to be positive. Instead, to be able to consider exogenous inputs, this requires that reference units utilize the same levels of the transformed exogenous inputs as that of the assessed DMU on average, which is depicted on the equality constraint:

$$\sum_{j=1}^{n} \left(\lambda_{jk} * E_{ij}\right) = \theta_k * E_{ik}, \qquad \forall k \forall i \tag{6}$$

Undesirable output. Undesirable outputs can be treated as an input, such as the model developed by Korhonen and Luptacik (2004) and further examined by Yang and Pollitt (2007). The undesirable outputs are constrained that a DMU being evaluated should only benchmark itself to units that are performing better than itself with the undesirable output concerned. That is, the benchmarked undesirable outputs of DMUs in the reference set should have less undesirable output than that of the DMU in evaluation.

$$\sum_{j=1}^{n} \left(\lambda_{jk} * U_{rj}\right) \leq U_{rk}, \qquad \forall k \forall r \tag{7}$$

The expected output from the model would be the relative efficiency of each DMU (i.e., $\theta_k, \forall k$) and excess inputs from DMU, that is, $(\theta_k * X_{ik}) - \sum_{j=1}^{n}\left(\lambda_{jk} * X_{ij}\right)$.

APPENDIX C: *Model Formulation of the DEA Resource Reallocation Model*

Inverse to the DEA, the input allocation problem determines how much an input should be allocated given that outputs are increased and efficiency remains the same (Wei et al., 2000). Another type of the inverse DEA problem is the forecasting problem; which determines how much an output should increase given that inputs are increased and efficiency remains the same (Wei et al., 2000). With the forecast model, decision makers can now determine how additional input can affect output changes. This model is further enhanced to incorporate re-allocation of excess inputs to enhance output production of the entire system.

Reallocation model indices

j	=	Decision Making Unit (DMU)	(j = 1, 2, 3 … n)
k	=	DMU in assessment	(k = 1, 2, 3 … n)
i	=	Type of input	(i = 1, 2, 3 … m)
r	=	Type of output	(r = 1, 2, 3 … m)

Decision variables

λ_{jk}	=	Proportion of input and output benchmarked from DMU j for DMU k
A_{ik}	=	Increase in input i for DMU k
C_{rk}	=	Change in desirable output r for DMU k
N_{rk}	=	Change in undesirable output r for DMU k

Model parameters

θ_k	=	Proportion of input use in output production of DMU k
E_{ij}	=	Exogenous input of DMU j
X_{ij}	=	Input i used by DMU j
Y_{rj}	=	Output r of DMU j
U_{rj}	=	Undesirable output r of DMU j
S_i	=	Excess input i

Reallocation Model

$$Max\ w = \sum_{k=1}^{n} \sum_{r=i}^{s} \left(\frac{C_{rk}}{Y_{rk}} - \frac{N_{rk}}{U_{rk}} \right)$$

Subject to:

$$\sum_{j=1}^{n}\left(\lambda_{jk} * X_{ij}\right) \leq \theta_k * (X_{ik} + A_{ik}), \qquad \forall k \forall i \qquad \text{(Input constraint)}$$

$$\sum_{j=1}^{n}\left(\lambda_{jk} * E_{ij}\right) = \theta_k * E_{ik}, \qquad \forall k \forall i \qquad \text{(Exogenous input constraint)}$$

$$\sum_{j=1}^{n}\left(\lambda_{jk} * Y_{rj}\right) \geq Y_{rk} + C_{rk}, \qquad \forall k \forall i \qquad \text{(Output constraint)}$$

$$\sum_{j=1}^{n}\left(\lambda_{jk} * U_{rj}\right) \leq U_{rk} + N_{rk}, \qquad \forall k \forall r \qquad \text{(Undesirable output constraint)}$$

$$\sum_{j=1}^{n} \lambda_{jk} = 1, \qquad \forall k \qquad \text{(Variable return} - \text{to} - \text{scale constraint)}$$

$$\sum_{k=1}^{n}\sum_{i=1}^{n} A_{ik} = S_i, \qquad \forall i \qquad \text{(Resource allocation limit constraint)}$$

$$\lambda_{jk} \geq 0, \qquad \forall j \forall k \qquad \text{(Nonnegativity constraint)}$$

$$A_{ik} \geq 0, \qquad \forall i \forall k \qquad \text{(Nonnegativity constraint)}$$

$$C_{rk} \geq 0, \qquad \forall r \forall k \qquad \text{(Nonnegativity constraint)}$$

$$N_{rk} \geq 0, \qquad \forall i \forall k \qquad \text{(Nonnegativity constraint)}$$

APPENDIX D: *Software Development Verification – Data Estimation*

Table D.1 *Data set for data estimation*

DMU	X1	X2	E1	Y1	Y2	U1
1	163,523.00	26.00	49,196.00	5,561.00	105,321.00	63.53
2	338,671.00	30.00	78,533.00	18,106.00	314,682.00	90.47
3	281,655.00	51.00	176,381.00	16,498.00	542,349.00	108.23
4	400,993.00	78.00	189,397.00	90,810.00	847,872.00	228.79
5	363,116.00	69.00	192,235.00	52,279.00	158,704.00	69.87
6	541,658.00	114.00	194,091.00	66,139.00	1,438,746.00	223.87
7	508,141.00	61.00	228,535.00	35,295.00	839,597.00	166.96
8	338,804.00	74.00	238,691.00	33,188.00	540,821.00	142.08
9	511,467.00	84.00	267,385.00	65,391.00	1,562,274.00	192.23
10	393,815.00	68.00	277,402.00	41,197.00	978,117.00	152.81
11	509,682.00	96.00	330,609.00	47,032.00	930,437.00	236.04
12	527,457.00	92.00	332,609.00	56,064.00	1,345,185.00	236.52
13	601,594.00	127.00	356,504.00	69,536.00	1,164,801.00	156.48
14	528,799.00	96.00	365,844.00	37,467.00	1,348,588.00	259.86
15	394,158.00	77.00	389,894.00	57,727.00	1,100,779.00	137.08

For LP predictive model, additional verification was done where the results were compared to regression analysis results. Table D.1 above contains data set used for LP estimation verification. Table D.2 shows a sample comparison of standard error results of predicting target X_1 by predictor data (X_2, E, Y_1, Y_2, Y_3). Comparison of standard error shows that the LP estimates resulted in lower standard error as compared to using regression analysis. Best predictor identified is the same between the two methods that is X_2 is best used to estimate for missing X_1 data.

Table D.2 *Comparison of standard error estimates*

Predictor	Target	Standard Error	
		LP Estimate	Regression
X_2	X_1	56,361.00	65,118.06
E	X_1	79,156.64	246,792.90
Y_1	X_1	93,898.55	101,941.67
Y_2	X_1	59,855.47	82,825.22
Y_3	X_1	62,056.26	69,455.33

To verify if the results are accurate and acceptable, the returned results of the software are compared to that of results from a hard-coded model ran in GAMS. Comparing software results to reference model ran in GAMS; both results are similar and comparable (see Table D.3).

Table D.3 *Comparison of results between GAMS and DEA software*

System	Predictor	Target	Slope	Intercept
GAMS	X_2	X_1	3521.68	154340.0
DEA Software	X_2	X_1	3521.68	154340.9

APPENDIX E: *Software Development Verification – Efficiency Results and Reallocation*

Using the same data set; results are compared using GAMS and using the software developed. For the test data run (see Table E.1), the same DMUs are identified under the efficient frontier; both from GAMs and from the software developed. Resulting efficiency scores index are also the same. Table E.2 presents the sample run efficiency result using the test data.

Table E.1 *Data set for test run*

DMU	X1	X2	E1	Y1	Y2	U1
1	67,126,924.17	371.06	324,356.00	106,879.70	11,355.03	1,186.00
2	45,874,108.45	249.00	231,717.00	96,195.81	10,170.13	1,116.00
3	106,201,374.50	490.88	555,272.00	219,551.56	21,986.98	2,055.00
4	206,524,724.17	496.00	1,445,209.00	409,966.00	9,828.00	4,151.00
5	62,964,402.70	262.21	445,510.00	123,555.00	9,663.52	2,325.00
6	244,920,515.58	313.00	557,297.00	232,901.70	17,054.98	2,652.00
7	16,485,119.10	66.00	60,378.00	17,136.00	1,402.00	232.00
8	102,579,154.32	774.00	630,161.00	261,366.68	26,680.83	1,433.00
9	247,710,131.71	905.00	2,468,417.00	1,008,935.00	58,466.00	10,762.00
10	731,442,498.98	760.05	389,478.00	818,226.29	48,316.02	2,659.00

Table E.2 *Efficiency results of test run*

DMU	DEA Software	GAMS
1	0.7896	0.7896
5	0.7773	0.7773
6	0.9477	0.9477

Likewise, re-allocation decisions are similarly compared both from the output of the DEA software developed and reference model run in GAMS. For the test data run; the same re-allocation decisions were derived from both output. Table E.3 presents the sample run allocation result using the test data.

Table E.3 *Reallocation results of test run*

Allocation	DEA Software		GAMS	
DMU4	X_1	90921182.92	X_1	9.092118E+7
DMU8	X_2	10.18	X_2	10.18

Output	DEA Software		GAMS	
DMU4	Y_1	34411.48	Y_1	34410.53
	Y_2	4727.50	Y_2	4727.37
	U_1	414.14	U_1	414.12