

MAXIMUM PERFORMANCE EFFICIENCY APPROACHES FOR ESTIMATING BEST PRACTICE COSTS

Marvin D. Troutt

**Department of Management & Information Systems
College of Business, Kent State University
Kent, OH 44242-001
Ph: +1 330 672 1145
E-mail: mtroutt@bsa3.kent.edu**

Donald W. Gribbin

**School of Accountancy
Southern Illinois University at Carbondale
Carbondale, IL 62901-4631
E-mail: dgribbin@cba.siu.edu**

Murali S. Shanker

**Department of Management & Information Systems
College of Business, Kent State University
Kent, OH 44242-001
Ph: +1 330 672 1165
E-mail: mshanker@kent.edu**

Aimao Zhang

**School of Information Technology
College of Business Administration
Georgia Southern University
PO Box 8150
Statesboro, GA 30460-8150
Ph: +1 912 681-0755
E-mail: aimao@gasou.edu**

ABSTRACT

Data mining is increasingly being used to gain competitive advantage. In this chapter, we propose a principle of maximum performance efficiency (MPE) as contribution to the data-mining toolkit. This principle seeks to estimate optimal or boundary behavior, in contrast to techniques like regression analysis, which predict average behavior. This MPE principle is explained in the context of activity-based costing situation. Specifically, we consider the Activity-Based Costing situation in which multiple activities generate a common cost pool. Individual cost drivers are assigned to the respective activities but allocation of the cost pool to the individual activities is regarded as impractical or expensive. Our study focuses on published data from a set of property tax collection offices, called rates departments, for the London metropolitan area. We define what may be called benchmark or most efficient average costs per unit of driver. The MPE principle is then used to estimate the best practice cost rates. A validation approach for this estimation method is developed in terms of what we call normal-like-or-better performance effectiveness. Extensions to time series data on a single unit, and marginal cost oriented basic cost models are also briefly described. In conclusion, we discuss potential data-mining applications and considerations.

INTRODUCTION

In recent years, companies have started to realize the potential of using data mining techniques as a form of competitive advantage. For example, in the finance industry, in the decade from 1980 to 1990, the number of credit cards issued doubled to about 260 million. But, in the next ten years, however, there was not another doubling of this number. Given that there are now about 280 million people in the United States, it is widely believed that the credit-card market is saturated (Berson, Smith, & Thearling, 2000). In such situations, any gains by one company leads to a loss for another — a zero-sum game. To gain competitive advantage, credit card companies are now resorting to data-mining techniques to retain and identify good customers at minimal cost.

The cell phone industry is also expected to go the way of the credit card market. Soon, the cellular industry will be saturated; everybody who needs cells phone will have one. Companies who are able to predict and understand customer needs better, will probably be the ones who will survive. The cellular industry, like the credit card industry, is likewise resorting to data-mining techniques to identify traits for retaining good customers.

Research in data mining has so far focused on either developing new techniques or on identifying applications. Being a multidisciplinary field, data mining techniques have originated from areas of artificial intelligence, database theory, visualization, mathematics, operations research, and statistics, among others. Many of the well-known statistical techniques like nearest neighbor, clustering, and regression analysis are now part of the data-mining toolkit.

In this chapter, we present a new technique based on the principal of *maximum performance efficiency* (MPE). While techniques like linear regression analysis are used to predict average behavior, MPE seeks to predict boundary or optimal behavior. In many cases, such models are actually more desirable. For example, in a saturated credit card or cellular market, a company may seek to predict characteristics of its best customers. In essence, choosing to concentrate on customers who are low risk/costs to maximize profit. Such models, usually called ceiling/floor models, can also be used as part of data mining techniques for benchmarking. For example, a company may be interested in comparing the quality of its products over different product lines. The MPE criterion seeks to identify the characteristics of the best performing unit, thus allowing the company to implement these measures in other units to improve their quality, and hence the competitive advantage of the company across product lines.

In this chapter we propose the MPE principle and show how it can be used to estimate the best practice costs in an activity-based costing situation. The rest of the chapter is organized as follows. In the next section, we present our initial motivation for developing the MPE principle. As an example, we consider an activity-based costing situation where multiple activities generate a common cost pool. Our objective is to estimate the best practice cost rates. The following section then distinguishes between basic cost models and the models used to estimate the parameters of the basic cost models. The maximum performance efficiency (MPE) principle is developed using an aggregate efficiency measure that is the sum or average of performance efficiencies. Then the MPE principle is used to derive the linear programming estimation model. The next section describes the data for the study. The MPE criterion is applied to this data, and the results are compared with a previous analysis to assess face validity for the proposed new method. The following section proposes a model aptness theory based on the gamma distribution and a technique called vertical density representation. The fitted performance efficiency scores are required to satisfy a benchmark for validity of the goal assumption called normal-like-or-better performance effectiveness. This is followed by limitations, and

some extensions to other basic cost models. The next section discusses potential data-mining applications and considerations, which is followed by the Conclusions section.

Remarks on notation: In summation notations using \sum the lower limit is omitted when it is clear from the context. Omission of upper limits indicates summation over all values of the index.

MOTIVATION

Managerial accountants are often called upon to provide measures of performance efficiency. Such a task is essentially trivial if there is a single input measure and a single output measure. One could simply divide the output measure by the input measure and use the resulting performance efficiency measure to either: (1) compare the unit's performance over time, or (2) compare the unit's performance with other comparable units.

While there are numerous purposes for performance efficiency measures (i.e., performance evaluation, managerial compensation, benchmarking, and cost control to name a few), the specific purpose of this research is to explain and illustrate how a proposed estimation principle can be used with activity-based costing for determining performance efficiency measures for the purpose of cost control.

Activity-based costing is one approach that could be used to develop a performance measure for such purposes. Activity-based costing consists of disaggregating costs into specific cost pools that can be linked causally with their respective activities. Cooper & Kaplan (1992) illustrate how such an approach can be used with purchase order data. The approach consists of dividing the monthly cost of processing purchase orders by the number of purchase orders processed per month. This purchase order cost can then be used as a benchmark for comparison purposes. This approach is very simple and easy to use if there is only one cost pool and one activity with a single cost driver. But there are many realistic scenarios in which there are multiple cost drivers for a single cost pool. For example, faculty salaries for an academic unit (department or college) appear to be driven by at least two cost drivers. Both student credit hours generated and the level of academic degrees offered (bachelors, masters, and doctoral) appears to drive the amount of faculty salaries for an academic unit. The dilemma is that faculty salaries are in a single cost pool and there is no simple and objective method for disaggregating faculty salaries. The task of attributing the faculty salary pool to separate activities and cost drivers is essentially impossible. There is no easy way for determining how much of the faculty salary pool is due to: (1) the generation of student credit hours, and (2) the level of academic degrees offered by academic unit.

The methodology explained and illustrated in this chapter allows for the inclusion of multiple cost drivers in determining performance efficiency measures. The allocation of the combined cost pool to individual activities might be regarded as not objectively possible, impractical, expensive, or of insufficient additional value for the costing system. We first consider the problem of estimating average costs per unit of cost driver in such situations when cross-sectional data are available for a set of what we call *comparable* business units. We also consider application of the same techniques to basic cost models having marginal cost assessment capability, and briefly discuss the setting in which a single business unit is observed over several time periods

We define benchmark or best practice average costs per unit of cost driver as the average cost rates associated with the most efficient unit(s). A principle of *maximum performance efficiency* (MPE) is proposed and estimation criteria based on efficiency measures are derived from this principle. This is a generalization of the maximum decisional efficiency (MDE) principle introduced in Troutt (1995) and also discussed in Troutt (1997) and Troutt et al. (1997). The efficiency measure used may be

considered analogous to the cost of unused capacity concept proposed by Cooper & Kaplan (1992). These models also provide a different view on the lack of proportionality of costs to driver levels documented in Noreen & Soderstrom (1994).

A basic assumption underlying the estimation principle employed here is that all business units under comparison seek to maximize their efficiencies in performing their services. The data we study are from public service entities that are presumed to have this goal on behalf of the public interest.

However, this assumption needs verification as a kind of model aptness or validation issue similar to the requirement of normally distributed errors in OLS regression. As an estimation model aptness test we propose what may be called a *normal-like-or-better* performance effectiveness measure.

The estimation models proposed here are linear programming (LP) models. Use of such models in cost accounting is not new. See for example, Demski (1967), Onsi (1970), Kaplan & Thompson (1971), and Itami & Kaplan (1980). These previous works have generally utilized LP models assuming that data are known. That is, they assume that technological coefficients and resource levels are given. Then the dual optimal solution (especially the shadow prices) of these fully specified LP models has been employed for (i) overhead allocation (Kaplan & Thompson, 1971), (ii) transfer pricing (Onsi, 1970), and (iii) reallocation of costs to multiple products (Itami & Kaplan, 1980). However, the use of LP models enters in a different way in this chapter. Namely, the estimation models are themselves LP problems in which the decision variables are the unknown best practice cost rates. The next section distinguishes between basic cost model, and the models used to estimate the parameters of the basic cost models. The MPE principle is also presented here.

BASIC COST MODELS AND ESTIMATION MODELS

There are two uses of the word model in this chapter. By the basic cost model we mean the assumed relationship between the cost driver of an activity and its contribution to the total cost pool. The simplest case is that of proportional variation. Let y_r be the amount of the r -th cost driver. Under the basic cost model the contribution to the total cost pool will be $a_r y_r$, where a_r may be called the average cost per unit of y_r or simply the cost rate. This chapter focuses on this basic cost model. Some other models such as for estimating marginal costs are discussed briefly below. The other use of the word model refers to the method used to solve for estimates of the basic cost model parameters — the a_r^* . The estimation models used here are LP models. As noted above, this usage of LP does not appear to be similar to previous uses in the accounting literature, but is a consequence of the estimation principle used.

Suppose there are $r=1 \dots R_r$ activities with associated cost driver quantities of y_r , respectively. Then as in Cooper & Kaplan (1992) the cost of *resources consumed* is given by $\sum a_r y_r$. If x is the actual total cost pool associated with these activities, the cost of resources *supplied*, then there may be a difference, $s \geq 0$, such that

$$\sum a_r y_r + s = x. \quad (2.1)$$

Cooper & Kaplan (1992) call s the *cost of unused capacity*. In their analysis, the x value was a budget figure and the y_r could vary due to decreased demands. We start from the same construct but regard variations of the y_r and x as due to more or less efficiency. By defining what may be called best practice a_r values, we therefore associate the s value with possible inefficiency of a business unit, a cost of inefficiency.

Thus suppose we have $j = 1 \dots N$ comparable business units, achieving y_{rj} units of driver r , respectively, and with associated cost pools, x_j . In addition to having the same activities and cost drivers, we further require that comparable units be similar in the sense that the practices, policies, technologies, employee competence levels and managerial actions of any one should be transferable, in principle, to any other. Define a_r^* as the vector of cost rates associated with the most efficient unit or units under comparison. Then in the equation

$$\sum_{r=1}^R a_r^* y_{rj} + s_j = x_j \quad \text{for all } j, \quad (2.2)$$

the cost of unused capacity, s_j , may be interpreted as an inefficiency suffered by a failure to achieve larger y_{rj} values, smaller x_j value, or some combination of these. The ratio $v_j = \sum a_r^* y_{rj} / x_j$, is an efficiency measure for the j -th unit when the a_r^* are the true benchmark cost rates and $\sum a_r^* y_{rj} \leq x_j$ holds for all units. A technique for estimating parameters in efficiency ratios of the above form was proposed in Troutt (1995). In that paper the primary data were specific decisions. Here a more general form of that approach called *maximum performance efficiency* (MPE) is proposed and applied to estimate the benchmark a_r^* values. Assume that each unit $j = 1 \dots N$, seeks to achieve maximum (1.0) efficiency. Then the whole set of units may be regarded as attempting to maximize the sum of these efficiency ratios, namely, $\sum \sum a_r^* y_{rj} / x_j$. The maximum performance efficiency estimation principle proposes estimates of the a_r^* as those which render the total, or equivalently the average, of these efficiencies a maximum.

Maximum Performance Efficiency (MPE) Estimation Principle: *In a performance model depending on an unknown parameter vector, select as the estimate of the parameter vector that value for which the total performance is greatest.*

Use of the word performance is stressed in the MPE name to emphasize the more general utility of the approach than was indicated by the earlier MDE term. Managerial performance, such as in the units under study here, involves many kinds, levels and horizons of decisions. Define the data elements Y_{rj} by $Y_{rj} = y_{rj} / x_j$. Then the MPE estimation criterion for the benchmark a_r^* values is given by

$$\text{MPE:} \quad \max \sum_{j=1}^N \sum_{r=1}^R a_r Y_{rj} \quad (2.3)$$

$$\text{s.t.} \quad \sum_{r=1}^R a_r Y_{rj} \leq 1 \quad \text{for all } j \quad (2.4)$$

$$a_r \geq 0 \quad \text{for all } r \quad (2.5)$$

Problem MPE is a linear programming problem. Its unknown variables are the best practice cost rates, the a_r^* values. Solution of this model provides values for the a_r^* as well as the unit efficiencies, v_j . A model adequacy or validation approach for this estimation procedure is proposed in a later section. As a linear programming problem, the MPE model is readily solved by a wide variety of commonly available software products. MPE may be seen here to be an estimation criterion analogous to the ordinary least squares errors criterion in regression. The cost model, $\sum a_r Y_{rj}$ is analogous to the regression model.

APPLICATION TO THE DATA

In this section we describe the data and apply the MPE estimation criterion. In the initial application of the MPE criterion one cost rate was estimated to be zero. In this section we concentrate on the case in which all estimated cost rates must be positive and show how the MPE criterion is modified for that case.

Table 1 About Here

The data for this study are due to Dyson & Thanassoulis (1988), (DT) and are reproduced here in Table 1, first six columns. These data were collected for a set of property tax collection offices, called rates departments, in the London Boroughs and Metropolitan Districts. A more complete description of the data is given in Thanassoulis, Dyson, & Foster (1987), (TDF). Total annual costs, measured in units of £100,000, for these offices (units), were collected along with activity driver levels for four activities. The first three activities, collection of non-council hereditaments, rate rebates generated, and summonses issued and distress warrants obtained, were measured in units of 10,000, 1,000, and 1,000, respectively. The fourth, net present value of non-council rates collected, was measured in units of £10,000. This last one was included as a cost driver (called an output by DT, DTF) to reflect the additional administrative effort exerted to ensure the timely payment of large revenue producing cases. TDF briefly discuss the possible disaggregation of the cost pools. They indicate that this would have been possible to some extent but decided against this for several reasons. First, these costs represented the cost of real resources used and available for management deployment. Next they felt that the increased number of variables that would result might have tended to decrease the discrimination power of the data envelopment analysis (DEA) method they were studying. Next, and importantly, it was felt that the disaggregated data were less reliable. This concern is well founded, particularly in the present context. Datar & Gupta (1994) have shown that disaggregation can actually increase errors. The method used in DT and TDF was a modification of data envelopment analysis (DEA), another efficiency estimation technique. It is worthwhile to briefly discuss DEA in regard to the above and other relevant points. An introduction to DEA is contained in Charnes et al. (1994). There are several DEA models but we limit our coverage to that used in DT. If we use the a_r notation for output weights in the DEA model M1 of DT (page 564) we obtain the model(s):

$$MI(j_o) \quad \max \quad h_o = \sum_r a_r Y_{rj_o} \quad (3.1)$$

$$\text{s.t.} \quad \sum_{r=1}^R a_r Y_{rj} \leq 1, \text{ for all } j \quad (3.2)$$

$$a_r \geq 0, \text{ for all } r \quad (3.3)$$

Unlike the one-pass solution of model MPE, Model $MI(j_o)$ is solved for each unit, (j_o) , in turn. The solutions a_{rj_o} therefore depend on which unit is being featured in the objective function. Typically some of the a_{rj_o} values will be zero for various units. In the present cost rates context the following interpretation can be given to the $MI(j_o)$ DEA model. If unit j_o were allowed to choose the activities and drivers to become applicable for the entire group of units then the $MI(j_o)$ model solution obtains these in such a way as to give that unit the most favorable efficiency score. The principal difficulty

here is that no consensus is being achieved on the most efficient cost rates. With the $MI(j_o)$ DEA model each unit can select cost rates that make it look most favorable. DT call this phenomenon weights flexibility. Their work was motivated, in part, to modify the $MI(j_o)$ DEA technique to limit this flexibility. DT gives a more extensive discussion of this DEA limitation.

Except in unusual circumstances, activity cost rates, a_r , can only be positive. In this section we consider this requirement as a preemptive priority. For the solution values, a_r^* , to be strictly positive, it is necessary that they be basic variables in an optimal solution. This may occur gratuitously. Otherwise one or more may be nonbasic, and therefore have value zero. The standard LP solution report provides reduced costs. For a variable with optimal value of zero the reduced cost may be interpreted as follows. It indicates the smallest amount by which the objective function coefficient for the variable must be increased in order that the variable becomes positive in an optimal solution. The last column of Table 1 gives the efficiency scores obtained by model MPE with preemptive positive costs. Descriptive statistics have been added as supplemental information. The MPE model was solved using SAS/IML (1995), which includes a linear programming call function. The initial solution assigned a zero optimal value only to a_1^* . (The full solution for this case was $a_1^*=0$, $a_2^*=0.0882$, $a_3^*=0.2671$, $a_4^*=0.0664$). Thus it was deemed necessary to implement the preemptive positive weights modification. The reduced cost for variable a_1 was given as -1.220440. The objective coefficient was 55.52898. Therefore the modified procedure required increasing the coefficient of a_1 to 56.747. The resulting estimates were as follows:

$$a_1^* = 0.2618, a_2^* = 0.0494, a_3^* = 0.1398, a_4^* = 0.1280 \quad (3.4)$$

The corresponding efficiency scores of the units are shown in the last column of Table 1. Table 2 gives descriptive statistics for the Y_r data.

Table 2 About Here

While not true for the example presented here, it is possible that more than one cost rate is initially estimated as zero, or the reduced cost is also zero when strict positivity is required. Then, it is necessary to proceed as follows. Suppose an auxiliary variable m , and n_r new constraints $a_r \geq m$ are joined to the MPE model. If the optimal value of m is positive, then so it must be also for all the cost rates. Let λ be a nonnegative parameter chosen by the analyst and consider the modified objective function given by

$$\max \sum_j \sum_r a_r y_{rj} + \lambda m \quad (3.5)$$

When the value of λ is zero the objective function is the same as the original one with $m^* = 0$. We have the following theorem whose proof is given the Appendix.

Theorem 1: Let $z^*(\lambda)$, $a_r^*(\lambda)$, and $m^*(\lambda)$ be the solution of the MPE (λ) model:

$$\text{MPE:} \quad \max z(\lambda) = \sum_j \sum_r a_r y_{rj} + \lambda m \quad (3.6)$$

$$\text{s.t. } \sum a_r y_{rj} \leq 1, \text{ for all } j \quad (3.7)$$

$$a_r \geq m, \text{ for all } r \quad (3.8)$$

$$a_r \geq 0, \text{ for all } r, m \text{ unrestricted} \quad (3.9)$$

Then (1): $z^*(\lambda)$ is monotone nondecreasing in λ and $z^*(\lambda) \rightarrow \infty$ as $\lambda \rightarrow 0$; and (2): $\sum a_r^*(\lambda) Y_{rj}$ is monotone nonincreasing in λ .

We propose the solution for positive weights to be that corresponding to the greatest lower bound of λ values for which $m^*(\lambda) > 0$. This may be estimated by trial and error.

We develop a rationale for such positive weights procedures by writing the MPE model objective function in the form $\sum a_r (\sum Y_{rj})$. If maximization of this objective function yields $a_1^* = 0$, then evidently the coefficient, $\sum Y_{1j}$ is too small in some sense relative to the other coefficients. It may be noted that the $Y_{rj} = y_{rj}/x_j$ data are in the nature of reciprocal costs. When the sum of these is too small the implication is that their own reciprocals, the x_j/y_{rj} , are on average too large. This is suggestive of inefficiency with respect to the r -th activity. In such a case, assignment of a zero optimal cost estimate would mask this kind of inefficiency. By using the reduced cost adjustment, or adding the λm term to the objective function, a compensation is made for coefficients that are apparently too small in that sense.

Some comparisons with the results of Dyson & Thanassoulis (1988) can now be discussed. As part of their study, a regression through the origin was obtained. The coefficients of that regression model can be interpreted as average cost rates for these activities. The results were as follows:

$$\bar{a}_1 = 0.5042, \bar{a}_2 = 0.0785, \bar{a}_3 = 0.1765, \bar{a}_4 = 0.1940 \quad (3.10)$$

It will be noted that these average rates are uniformly higher than the presently estimated rates in (3.4), giving a measure of face validity. That is, it is necessary that the cost rates of the most efficient units be lower than the average cost rates for all the units. Also the four departments rated as most efficient by the present method are the same as those indicated by the Dyson & Thanassoulis (1988) approach. It may be observed that the preliminary regression step also gives information on the question of positivity of the cost rates. The positivity of \bar{a}_1 gives further evidence on that for a_1^* . Here \bar{a}_1 is positive and significant. (The significance level was not specified in DT). Since the units are assumed to be comparable it appears unlikely that one or a few could perform activity one with no cost while the typical unit does incur cost for the activity. If a particular unit could produce an activity with zero cost ($a_r^* = 0$) while the average unit does incur cost for the activity, then it must have a radically superior process not actually comparable with the others. Similarly, this regression model also validates activity four as influential on costs. The next section discusses model aptness.

ESTIMATION CRITERION QUALITY ISSUES

A basic assumption underlying the MPE estimation principle's applicability is that the sample of units under analysis does, in fact, have the goal of achieving maximum (1.0) efficiency. This is a model aptness issue that parallels the requirement of $N(0, \sigma^2)$ residuals in OLS regression theory. In the present MPE case the corresponding issue is to specify a measured characteristic of the v_j that indicates consistency with a goal or target of unity (1.0) efficiency. In this section we propose what may be called the normal-like-or-better effectiveness criterion for these fitted efficiency scores.

As a model for appropriate concentration on a target we begin with an interpretation of the multivariate normal distribution, $N(\mu, \Sigma)$, on \mathfrak{R}^n . If a distribution of attempts has the $N(\mu, \Sigma)$ or even higher concentration of density at the mode μ , then we propose this as evidence that μ is indeed a plausible target of the attempts. This is exemplified by considering a distribution model for the results of throwing darts at a bull's-eye target. Common experience suggests that a bivariate normal density represents such data reasonably well. Steeper or flatter densities would still be indicative of effective attempts, but densities whose modes do not coincide with the target would cause doubts about whether the attempts have been effective or whether another target better explains the data. We call this normal-like-or-better (NLOB) performance effectiveness. It is next necessary to obtain the analog of this criterion for the efficiency performance data Y_{ij} relevant to the present context.

If x is distributed as $N(\mu, \Sigma)$ on \mathfrak{R}^n then it well known that the quadratic form, $w(x) = (x - \mu)' \Sigma^{-1} (x - \mu)$ is $\text{gamma}(\alpha, \beta)$, where $\alpha = n/2$ and $\beta = 2$. This distribution is also called the Chi-square distribution with n degrees of freedom (see Law & Kelton, 1982). We may note that for this case $w(x)$ is in the nature of a squared distance from the target set $\{\mu\}$. It is useful to derive this result by a different technique. Vertical density representation (VDR), is a technique for representing a multivariate density by way of a univariate density called the ordinate or vertical density, and uniform distributions over the equidensity contours of the original multivariate density. VDR was introduced in Troutt (1993). See also Troutt (1991), Kotz & Troutt (1996), Kotz et al. (1997), and Troutt & Pang (1996). The version of VDR needed for the present purpose can be derived as follows. Let $w(x)$ be a continuous convex function on \mathfrak{R}^n with range $[0, \infty)$; and let $g(w)$ be a density on $[0, \infty)$. Suppose that for each value of $u \geq 0$, x is uniformly distributed on the set $\{x: w(x) = u\}$. Consider the process of sampling a value of u according to the $g(w)$ density and then sampling a vector, x , according to the uniform distribution on the set $\{x: w(x) = u\}$. Next let $f(x)$ be the density of the resulting x variates on \mathfrak{R}^n . Finally let $A(u)$ be the volume (Lebesgue measure) of the set $\{x : w(x) \leq u\}$. Then we have the following VDR theorem which relates $g(w)$ and $f(x)$ in \mathfrak{R}^n . The proof is given in the Appendix.

Theorem 2: If $A(u)$ is differentiable on $[0, \infty)$ with $A'(u)$ strictly positive, then x is distributed according to the density $f(x)$ where

$$f(x) = \phi(w(x)) \text{ and } g(w) = \phi(w) / A'(w).$$

Theorem 2 can be applied to derive a very general density class for performance related to squared distance type error measures. The set $\{x: (x - \mu)' \Sigma^{-1} (x - \mu) \leq u\}$ has volume, $A(u)$, given by $A(u) = \alpha_n |\Sigma|^{1/2} u^{n/2}$ where $\alpha_n = \pi^{n/2} / \Gamma(n/2)$, (Fleming, 1977), so that $A'(u) = n/2 \alpha_n |\Sigma|^{1/2} u^{n/2 - 1}$. The $\text{gamma}(\alpha, \beta)$ density is given by

$$g(u) = (\Gamma(\alpha) \beta^\alpha)^{-1} u^{\alpha-1} \exp\{-u^2/\beta\}.$$

Therefore Theorem 2 implies that if $w(x) = (x - \mu)' \Sigma^{-1} (x - \mu)$ and $g(u) = \text{gamma}(\alpha, \beta)$, then the corresponding $f(x)$, which we now rename as $\psi(x) = \psi(x; n, \alpha, \beta)$ is given by

$$\psi(x) = \Gamma(n/2) (\pi^{n/2} \Gamma(\alpha) \beta^\alpha)^{-1} \left[(x - \mu)' \Sigma^{-1} (x - \mu) \right]^{\alpha - n/2} \exp\{-1/\beta (x - \mu)' \Sigma^{-1} (x - \mu)\} \quad (4.1)$$

For this density class we have the following observations:

- (i) If $\alpha = n/2$ and $\beta=2$ then $\psi(x)$ is the multivariate normal density, $N(\mu, \Sigma)$.
- (ii) If $\alpha = n/2$ and $\beta \neq 2$ then $\psi(x)$ is steeper or flatter than $N(\mu, \Sigma)$ according to whether $\beta < 2$ or $\beta > 2$, respectively. We call these densities the normal-like densities.
- (iii) If $\alpha < n/2$ then $\psi(x)$ is unbounded at its mode, μ , but may be more or less steep according to the value of β . We call this class the better-than-normal-like density class.
- (iv) If $\alpha > n/2$ then $\psi(x)$ has zero density at the target, μ , and low values throughout neighborhoods of μ . This suggests that attempts at the target are not effective. The data may have arisen in pursuit of a different target or simply not be effective for any target.

For densities in category (iii) the unbounded mode concentrates more probability near the target and suggests a higher level of expertise than that evidenced by the finite-at-mode $N(\mu, \Sigma)$ class. It seems reasonable to refer to α in this context as the expertise, mode, or target effectiveness parameter; while β is a scale or precision parameter. Thus if $\alpha \leq n/2$ we call $\psi(x)$ the normal-like-or-better performance density. To summarize, if attempts at a target set in \mathcal{R}^n have a basic squared distance error measure and this measure is distributed with the gamma(α, β) density with $\alpha \leq n/2$ then the performance with respect to this target set is normal-like-or-better (NLOB).

We extend this target effectiveness criterion to the present context as follows. The target set is $\{Y \in \mathcal{R}^4 : \sum_r Y_r = 1, Y_r \geq 0 \text{ for all } r\}$. If $\sum_r Y_{rj} = v_j$ then the distance of Y_{rj} from the target set is $(1-v_j)/\|a\|^{-1}$. Since $0 \leq v \leq 1$ we employ the transformation $w = (-\ln v)^2 = (\ln v)^2$. This transformation has the properties that $w \cong (1-v)^2$ near $v=1$ and $w \in [0, \infty)$. Therefore $w/\|a\|^2 = (\ln v)^2/\|a\|^2$ is an approximate squared distance measure near the target set. Since the $\|a\|^2$ term is a scale factor it can be absorbed into the β parameter of gamma(α, β). We therefore consider the NLOB effectiveness criterion to hold if w has the gamma(α, β) density with $\alpha \leq n/2=2$. That is, such performance is analogous to that of unbiased normal-like-or-better distributed attempts at a target in \mathcal{R}^n .

There is one additional consideration before applying this effectiveness criterion to the present data. In the LP estimation model MPE at least one efficiency, v_j , must be unity (and hence $w_j = 0$). This is because at least one constraint (2.6) must be active in an optimal solution of the MPE model. We therefore consider the model for the w_j to be

$$p \delta(0) + (1-p) \text{ gamma}(\alpha, \beta), \quad (4.2)$$

where p is the frequency of zero values (here $p = 3/62 = 0.048$ from Table 1), and $\delta(0)$ is the degenerate density concentrated at $w = 0$. We call this the *gamma-plus-zero* density, gamma(α, β)+0. For this data we regard the NLOB criterion to hold if it holds for the gamma density in (4.2). When the gamma(α, β) density is fitted to the strictly positive w values then NLOB requires that $\alpha \leq 2$. For the data of $w_j = (\ln v_j)^2$ based on Table 1, column 7, the parameter value estimates obtained by the Method of Moments (see, for example, Bickell & Doksum, 1977) are $\alpha = 1.07$ and $\beta = 0.32$. This method was chosen because the BESTFIT™ software experienced difficulty in convergence using its default Maximum Likelihood Estimation procedure. The Method of Moments estimates parameters by setting theoretical moments equal to sample moments. For the gamma(α, β) density, $\mu = \alpha\beta$, and $\sigma^2 = \alpha\beta^2$. If \bar{w} and s^2 are the sample mean and variance of the positive w_j values, then the α and β estimates are given by

$$\hat{\alpha} = \bar{w}^2/s^2 \quad \text{and} \quad \hat{\beta} = s^2/\bar{w}.$$

Tests of fit of the w_j data to the gamma ($\alpha = 1.07$, $\beta = 0.32$) density were carried out using the software BestFit™ (1995). All three tests provided in BestFit™, the Chi-square, Kolmogorov-Smirnov, and the Anderson-Darling indicated acceptance of the gamma model with confidence levels greater than 0.95. In addition, for each of these tests, the gamma model was judged best fitting (rank one) among the densities in the library of BestFit™. We therefore conclude that the NLOB condition is met. Use of the NLOB criterion in this way may be regarded as somewhat stringent in that the zero data are only used to define the target and are not used to assess NLOB target effectiveness.

The NLOB criterion is important in establishing whether the estimated cost model is a plausible goal of the units being studied. The MPE model will produce estimates for any arbitrary set of Y_{rj} data. However, if the resulting v_j data were, for example, uniformly distributed on $[0,1]$, there would be little confidence in the estimated model.

LIMITATIONS

Limitations may be discussed for both the new estimation technique itself and for its application to the present context and data. In order to more fully parallel existing OLS theory for model aptness testing, attention should be given to potential outliers, independence of the v_j , and constancy of the distribution of the v_j from trial to trial (analogous to homoscedasticity in OLS theory, see, for example, Madansky, 1988, and Neter et al., 1985). Theory developments for these issues are not yet available for the MPE model.

Hypothesis tests and confidence intervals for the estimates do not appear to be readily derivable from the proposed approach. However information on their variances can be obtained by simulation using additional specific assumptions. As an illustration 100 data sets of 62 observations each were simulated as follows. A value of v_j was generated using the density model (4.2) and the estimates of p , α , and β . Then a vector, Y_{rj} , was generated according to the uniform distribution on the convex polytope $\{Y: \sum_r a_r^* Y_r = v_j, Y_r \geq 0\}$ where a_r^* is given by (3.4). Then the MPE model, (2.3)—(2.5)

was solved for each data set and descriptive statistics for the estimates were obtained. Additional details on the simulation steps are given in the Appendix. The results are shown in Table 3.

Table 3 About Here

The proposed NLOB criterion is a strong standard for performance effectiveness. It requires that squared distance performance with respect to the target set be as good or better than that of unbiased multivariate normal-like performance with respect of a point target in \mathcal{R}^n . A still weaker class of target effectiveness densities might be developed in further research by inclusion of a vector parameter corresponding to possible bias in the multivariate normal-like model.

With regard to limitations of the methodology for the application setting and data used here we discuss first the cost of unused capacity connection again. A cost of unused capacity in the Cooper and Kaplan sense, which can be denoted as s_j^{ck} , might co-exist along with a cost of inefficiency, s_j^I as used in the present chapter; so that $s_j = s_j^{ck} + s_j^I$. The effect of such unused capacities, as distinct from costs of inefficiencies, on the present results would be to understate the true efficiencies. The approach taken with the MPE model is worst-case in the sense that when the s_j^{ck} are identifiable the appropriate data adjustment would be $x_j' = x_j - s_j^{ck}$ and the average performance efficiency would be expected to be

larger. TDF also discuss what we have called comparability of these units. A concern was noted relative to activity four whose monetary driver level might have been affected by the prosperity of the community being served. That is, offices with above average community prosperity and corresponding activity four levels might be considered as being unfairly compared to the others. Other things being equal, units with an inappropriately inflated value of a driver level would be expected to exert heavy downward pressure on the corresponding estimate in model MPE. We believe this kind of incomparability should ideally be removed by some kind of normalization process such as division by a socio-economic index. For the sake of concentrating on essential features of the present technique and maintaining comparability with the results of DT we leave this level of detailed analysis beyond the scope of the chapter.

In the use of the NLOB criterion for this data the α parameter was compared to $n/2$ when $n=4$ was chosen. This assumes that the Y_r data are truly four-dimensional. The discussion of the data in TDF suggested to us that units were free to emphasize or vary all four drivers with the possible exception of the fourth one. If this driver is regarded as not available for improvement by the units, then the data should be considered as three-dimensional. In this case the intended α would be compared with 1.5. Since $\hat{\alpha} = 1.07$ the NLOB criterion is still met by the data under this assumption.

EXTENSIONS TO OTHER BASIC COST MODELS

This section discusses extensions to the three cases: (i) time-series data, (ii) marginal cost oriented basic cost models, and (iii) single driver-single cost pool data.

Time-Series Data

Suppose the data Y_{rt} are given over time periods indexed by t for a single business unit. Then the MPE model with index j replaced by t can be applied. First, it would be necessary to adjust all the x_t cost pool figures, and resulting Y_{rt} data to reflect current dollars using a cost index. This assumes that the estimated a_r^* cost rates are in terms of current dollars. Next, these rates would be interpretable as follows. The estimated a_r^* in the current dollar time series case may be interpreted to be the cost rate vector achieved by the unit during its most efficient observation period or periods. The resulting v_t suggest periods of more or less efficiency, and would be a useful source for self-study aimed at productivity and process improvements.

The comparability issue for the units under comparison should be easier to accept in this case.

However process or technology changes during the data time span could be problematical. A more complete discussion of limitations for this case is left for specific future applications.

In addition to the NLOB effectiveness test, additional considerations can be brought to bear with respect to an improvement over time dimension. Effectiveness in this respect would be supported by establishing a significant fit of the v_t data to a monotone increasing function of time, for example, the reciprocal of a learning curve. Over longer periods of times, learning curve patterns for the estimated gamma parameters could serve similarly. That is, decreasing α indicates improving target effectiveness, while decreasing β would indicate improving precision.

Marginal Cost Oriented Basic Cost Models

Both the time series and cross-sectional versions of the MPE model can be adapted to nonlinear basic cost models with marginal cost features. For example, consider the original cross-sectional case, but using the basic cost model

$$\sum_r a_r y_{rj} + \sum_r b_r y_{rj}^2 + s_j = x_j \quad (6.1)$$

Again s is interpreted as a possible inefficiency due to experiencing higher than benchmark a_r and/or b_r values. The cost of service provided by activity r is $a_r * y_r + b_r * y_r^2$ for efficient units. By differentiation, the marginal cost by this model for activity r becomes $a_r^* + 2b_r * y_r$ at the observed driver level. Defining new data elements $Y_{rj}^{(2)} = y_{rj}^2 / x_j$, the modified MPE model becomes

$$\text{MMPE:} \quad \max \sum \sum a_r Y_{rj} + b_r Y_{rj}^{(2)} \quad (6.2)$$

$$\text{s.t.} \quad \sum a_r Y_{rj} + b_r Y_{rj}^{(2)} \leq 1 \text{ for all } j \quad (6.3)$$

$$\sum a_r Y_r + b_r Y_{rj}^{(2)} \geq 0 \text{ for all } j \quad (6.4)$$

$$a_r \geq 0, \text{ for all } j, \text{ and } b_r \text{ unrestricted} \quad (6.5)$$

The constraints (6.4) ensure that cost contributions can only be nonnegative even if some b_r is negative. In that case the marginal rate is decreasing; while if these coefficients are positive the corresponding marginal rates are increasing. Here a quadratic basic cost model was used. More generally, other models with different marginal cost structures could be employed (e.g. Cobb-Douglas as in Noreen & Soderstrom, 1994).

Implications for the Single Driver - Single Cost Pool Case

The MPE model for this case simplifies to $\max \sum a Y_j$, s.t. $a Y_j \leq 1$, for all j , and $a \geq 0$. The solution of this model is clearly $a^* = \min Y_j^{-1} = \min x_j / y_j$. The NLOB criterion requires $\alpha^* \leq 1/2$ in this case. If this condition fails to hold then this minimum value may be unreasonably low, perhaps due to an outlier. Deletion of one or a few such tentative outliers would be well supported if the remaining data do, in fact, pass the NLOB test. Otherwise no credible a_r estimate is forthcoming from the present method. It should be noted that the simulation method could also be employed for this case, provided the NLOP criterion is met.

DATA MINING APPLICATIONS AND CONSIDERATIONS

Benchmark estimation models, such as that considered here, may also be called frontier regression models. The general application of these in data mining has been discussed in Troutt et al. (2001). They are formed to explain boundary, frontier or optimal behavior rather than average behavior as, for example, in ordinary regression models. Such a model may also be called a ceiling model if it lays above all the observations or a floor model in the opposite case. The cost estimation model of this chapter is a floor model since it predicts the best, in this case, lowest, cost units.

The model considered here is a cross-sectional one. Although data mining is ordinarily thought of from the perspective of mining data from within a single organization, benchmarking type studies must often involve comparisons of data across organizations. Benchmarking partnerships have been formed for this purpose as discussed in Troutt et al. (2000). Such benchmarking oriented data mining might be extended in a number of directions. Potential applications include comparisons on quality and other costs, processing and set-up times and employee turnover. More generally, benchmarking comparisons could extend to virtually any measure of common interest across firms or other entities such as universities, states and municipalities. In the example of this chapter, a simple cost model was used to explain best practice performance. More generally, best practice performance may depend on

other explanatory variables or categories of interest to the firm. Discovery of such models, variables and categories might be regarded as the essence of data mining. With techniques discussed here, the difference is the prediction of frontier rather than average performance. For example, interest often centers on best instances such as customers most responsive to mailings, or safest drivers, etc.

However, cross-sectional applications of benchmark performance models do not necessarily depend on the multiple firm situations. Mining across all a firm's customers can be of interest. Consider a planned mail solicitation of a sales firm. For mailings of a given type, it is desirable to predict the set of most responsive customers so that these can be targeted. Similarly, a charitable organization may be interested in discovering how to characterize its best or worst supporters according to a model.

As noted above, under the topic of times series data, such frontier models can be used within a single organization where the benchmarking becomes across time periods. Models of this type might be used to mine for explanatory variables or conditions that account for best or worst performance periods. Some of the same subjects as noted above for cross-sectional studies may be worthwhile targets. For example, it would be of interest for quality assurance to determine the correlates of best and worst defect production rates.

Ordinary regression is one of the most important tools for data mining. Frontier models, such as considered here, may be desirable alternatives in connection with data mining applications. This is especially the case when it is desired to characterize and model the best and/or worst cases in the data. Such data are typically of the managed kind. In general, such managed data or data from purposeful or goal directed behavior will be amenable to frontier modeling.

CONCLUSIONS

This chapter proposes a method for estimating what may be called benchmark, or best practice unit and marginal cost rates. These rates provide plausible operational goals for the management of the units being compared. This method also provides efficiency measures and suggests which organizational units or time periods are more or less efficient, as well as an estimate of the degree of such inefficiency. Efficient units or time periods provide benchmarks for imitation by other units or can be studied for continuous improvement possibilities. So far as the authors can determine, the proposed methodology is the first technique with the capability to suggest plausible benchmark cost rates.

A principle of maximum performance efficiency (MPE) was proposed as a generalization of the maximum decisional efficiency estimation principle in Troutt (1995). This principle is more broadly applicable than the MDE principle. Also a gamma distribution-based validation criterion was proposed for the new MPE principle. The earlier MDE principle appealed to the maximum likelihood estimation principle for model aptness validation, but required relatively inflexible density models for the fitted efficiency scores.

The estimation models derived here reduce to straightforward linear programming models and are therefore widely accessible. A case was made that an optimal cost rate estimate of zero for some activity may be indicative of generally poor efficiency across the units with respect to one or more activities. Modifications based on reduced costs and augmented objective functions were proposed to compensate in that case.

In these models the unknown cost rate parameters resemble the coefficients of linear regression models. However, the maximum performance efficiency estimation principle is employed rather than a criterion such as ordinary least squares. This principle assumes that for each organizational unit and time period the unit intends to minimize these costs. Model adequacy with respect to this assumption

was judged by a test of normal-like-or better performance effectiveness for the estimated efficiency scores.

These results are also consistent with Noreen & Soderstrom (1994) who found that costs were not generally proportional to activity levels in a cross-sectional study of hospital accounts. The results of this chapter suggest that one source of such nonproportionality, in addition to the possibly nonlinear form of the basic cost model, is what we call performance inefficiency.

The proposed estimation criterion was applied to a published data set previously analyzed by a modified data envelopment analysis method. The resulting estimates were compared with the average costs obtained by the previous method. The estimated benchmark cost rates were uniformly and strictly lower than their average rate counterparts consistent with their definitions and providing a strong measure of face validity.

Benchmarking estimation models, such as discussed here, provide a new tool for data mining when the emphasis is on modeling the best performers.

REFERENCES

- Berson, A., Smith, S., & Thearling, K (2000). *Building Data Mining Applications for CRM*. McGraw Hill, New York.
- BestFit (1995). *User's guide*. Palisade Corporation, Newfield, NY.
- Bickell, P.J., & Doksum, K.A. (1977). *Mathematical statistics: Basic ideas and selected topics*. Holden Day, Inc., San Francisco.
- Charnes, A., Cooper, W.W., Lewin, A., & Seiford, L.M. (1994). *Data envelopment analysis: Theory, methodology, and applications*. Kluwer Academic Publishers, Boston.
- Cooper, R., & Kaplan, R.S. (1992). Activity-based systems: Measuring the costs of resource usage. *Accounting Horizons*, September, 1-13.
- Datar, S., & Gupta, M. (1994). Aggregation, specification and measurement errors in product costing. *The Accounting Review*, 69(4), 567-591.
- Devroye, L. (1986). *Non-uniform random variate generation*. Springer-Verlag, New York.
- Demski, J. (1967). An accounting system structured on a linear programming Model. *The Accounting Review*, 42(4), 701-712.
- Dyson, R.G., & Thanassoulis, E. (1988). Reducing weight flexibility in data envelopment analysis. *Journal of the Operational Research Society*, 39(6), 563-576.
- Fleming, W. (1997). *Functions of several variables*, 2nd ed., Springer-Verlag, New York.
- Itami, H., & Kaplan, R. (1980). An activity analysis approach to unit costing with multiple interactive products. *Management Science*, 26(8), 826-839.
- Kaplan, R., & Thompson, G. (1971). Overhead allocation via mathematical programming models. *The Accounting Review*, 46(2), 352-364.
- Kotz, S., Fang, K.T., & Liang, J.T. (1997). On multivariate vertical density representation and its application to random number generation. *Statistics*, 30, 163-180.
- Kotz, S., & Troutt, M.D. (1996). On vertical density representation and ordering of distributions. *Statistics*, 28, 241-247.
- Law, A.M., & Kelton, W.D. (1982) *Simulation modeling and analysis*. Mcgraw-Hill Book Company, New York.
- Madansky, A. (1988). *Prescriptions for working statisticians*. Springer Valley, New York.
- Neter, J., Wasserman, W., & Kutner, M.H. (1985). *Applied linear statistical models*, 2nd ed., Richard E. Irwin, Inc., Illinois.
- Noreen, E., & Soderstrom, N. (1994). Are overhead costs strictly proportional to activity? *Journal of Accounting and Economics*, 17, 255-278.
- Onsi, M. (1970). Transfer pricing systems based on opportunity cost. *The Accounting Review*, 40(3), 535-543.
- SAS/IML Software (1995). *Usage and reference*. Version 6, First Edition, SAS Institute, Inc., N. C.
- Schmeiser, B.W., & Lal, R. (1980). Squeeze methods for generating gamma variates. *Journal of the American Statistical Association*, 75, 679-682.
- Thannassoulis, E., Dyson, R.G., & Foster, M.J. (1987). Relative efficiency assessments using data envelopment analysis: An application to data on rates departments. *Journal of the Operational Research Society*, 38(5), 397-411.
- Troutt, M.D., Gribbin, D.W., Shanker, M., & Zhang, A. (2000). Cost efficiency benchmarking for operational units with multiple cost drivers. *Decision Sciences*, 31(4), 813-832.

- Troutt, M.D., Hu, M., Shanker, M. & Acar, W. (2001). Frontier versus ordinary regression models for data mining. In Parag C. Pendharkar (ed.), *Managing Data Mining Technologies in Organizations: Techniques and Applications*., Idea Group Publishing Co.
- Troutt, M.D. (1991). A theorem on the density of the density ordinate and alternative derivation of the Box-Muller method. *Statistics*, 22, 436-466.
- Troutt, M.D. (1993). Vertical density representation and a further remark on the Box-Muller method. *Statistics*, 24, 81-83.
- Troutt, M.D. (1995). A maximum decisional efficiency estimation principle, *Management Science*, 41(1), 76-82.
- Troutt, M.D. & Pang, W.K. (1996). A further VDR-type density representation based on the Box-Muller method. *Statistics*. 28, 1-8.
- Troutt, M.D. (1997). Derivation of the maximin efficiency ratio model from the maximum decisional efficiency principle. *Annals of Operations Research*, 73, 323-338.
- Troutt, M.D., Zhang, A., Tadisina, S.K., & Rai, A. (1997). Total factor efficiency/productivity ratio fitting as an alternative to regression and canonical correlation models for performance data. *Annals of Operations Research*, 74, 289-304.

APPENDIX

Proof of Theorem 1:

(i) Consider the potentially feasible solution $m^0 = k = a_r^0$ for all r . Then, $\sum a_r^0 Y_{rj} = k \sum_r (Y_{rj}) \leq 1$ if $k \leq (\sum Y_{rj})^{-1}$ for all j .

Thus the solution $m^0 = k = a_r^0$ is feasible for problem MPE (λ) for all $\lambda > 0$, for $k^0 = \min (\sum Y_{rj})^{-1}$ which is positive due to the positivity of the data. It follows that.

$$z(\lambda^*) \geq \sum_j \sum_r a_r^0 Y_{rj} + \lambda m^0 = k^0 \sum \sum Y_{rj} + \lambda k^0 \geq k^0 \lambda, \text{ for all } \lambda > 0.$$

(ii) Here we note that the $a_r^*(\lambda)$ for the MMPE model while feasible, are not necessarily optimal for the original MPE model. Hence use of these values in the MPE objective function will generally give a lower objective function value.

Proof of Theorem 2: This is a modification of a proof for a version of the theorem given in Troutt (1993). By the assumption that x is uniformly distributed on $\{x: w(x) = u\}$, $f(x)$ must be constant on these contours; so that $f(x) = \phi(w(x))$ for some function, $\phi(\cdot)$. Consider the probability $P(u \leq w(x) \leq u + \epsilon)$ for a small positive number, ϵ . On the one hand this probability is $\epsilon g(u)$ to a first order approximation. On the other hand it is also given by

$$\begin{aligned} \int \dots \int_{\{x: u \leq w(x) \leq u + \epsilon\}} f(x) \Pi dx_i &\equiv \phi(u) \int \dots \int_{\{w: u \leq w \leq u + \epsilon\}} \Pi dx_i \\ &\equiv \phi(u) \{ A(u + \epsilon) - A(u) \} \end{aligned}$$

Therefore

$$\epsilon g(u) \equiv \phi(u) \{ A(u + \epsilon) - A(u) \}$$

Division by ϵ and passage to the limit as $\epsilon \rightarrow 0$ yields the result.

Further Details on the Simulation Experiment

To simulate observations within each data set, a uniform random number was used to choose between the degenerate and continuous portions in the density model

$$p \delta(0) + (1-p) \text{gamma}(\alpha, \beta)$$

where $p = 0.048$, $\alpha = 1.07$, and $\beta = 0.32$. With probability p , $\delta(0)$ was chosen and $w=0$ was returned. With probability $1-p$, the gamma (α, β) density was chosen and a value, w , was returned using the procedure of Schmeiser and Lal (1980) in the IMSL routine RNGAM. The returned w was converted to an efficiency score, v , according to $v = \exp\{-w^{0.5}\}$. For each v , a vector Y was generated on the convex polytope with extreme points $e_1=(v/a_1^*, 0, 0, 0)$, $e_2=(0, v/a_2^*, 0, 0)$, $e_3=(0, 0, v/a_3^*, 0)$ and $e_4=(0, 0, 0, v/a_4^*)$ using the method given in Devroye (1986).

**Table 1: British Rates Departments Data Based on Dyson and Thanassoulis (1988).
Efficiency Rating Based on Model MDE-2 With Preemptive Positive Weights Modifications**

Rates Department	<i>Total Costs</i>	<i>Non-cnl Heredita- ments</i>	<i>Rate rebts grtd</i>	<i>Summons & d'ress wrnts</i>	<i>NPV of non-cnl rates</i>	<i>Efficiency Rating</i>
Lewisham	9.13	7.53	34.11	21.96	3.84	0.7881
Brent	13.60	8.30	23.27	35.97	8.63	0.6920
Stockport	5.76	10.91	13.39	11.53	4.93	1.0000
Bradford	11.24	16.62	36.82	27.55	9.52	1.0000
Leeds	15.57	22.81	95.78	23.61	12.27	1.0000
City of London	5.65	1.78	0.16	1.31	39.01	0.9641
Liverpool	21.60	15.11	70.96	54.22	10.81	0.7577
Walsall	8.57	7.92	48.69	14.03	5.92	0.8361
Rotherham	6.01	7.07	36.30	5.45	2.94	0.7926
Wakefield	8.02	8.86	43.61	13.77	4.27	0.8631
Lambeth	9.93	9.00	36.85	20.66	8.15	0.8122
Sunderland	7.90	8.28	45.22	6.19	5.33	0.7492
Solihull	5.15	6.76	18.70	10.62	3.54	0.8958
Redbridge	6.42	8.98	13.60	12.32	3.75	0.8110
Calderdale	5.94	7.69	25.91	8.24	2.48	0.7994
Haringey	8.68	7.23	16.97	17.58	6.27	0.6864
Barking & Dagenham	4.86	3.36	23.67	4.30	2.48	0.6076
Newcastle-upon-Tyne	10.33	8.56	30.54	17.77	8.01	0.6985
Manchester	21.97	12.23	92.02	29.53	14.76	0.6230
Wolverhampton	9.70	7.67	41.16	13.27	4.50	0.6649
Trafford	6.34	8.17	16.61	8.26	5.05	0.7466
Tameside	7.70	7.88	15.75	14.50	3.03	0.6808
St Helens	5.99	5.67	27.55	5.24	3.41	0.5188
Sutton	5.20	6.92	12.61	4.30	3.04	0.6556
Rochdale	6.36	7.35	23.51	5.74	4.21	0.8471
Barnsley	8.87	6.46	38.10	9.65	3.09	0.5974
Kirklees	10.71	13.64	23.86	14.63	4.63	0.6876
Oldham	6.49	7.68	17.97	8.27	2.76	0.6766
Sheffield	15.32	15.34	55.42	16.36	12.53	0.6905
Havering	7.00	8.37	14.92	9.88	4.33	0.6915
Dudley	10.50	9.61	37.91	13.49	5.04	0.6563
Sefton	10.88	10.65	36.96	14.25	4.84	0.6617
Bexley	8.52	8.97	24.67	11.84	3.75	0.6669
Gateshead	7.61	6.11	31.73	7.66	2.87	0.6031
Wigan	10.91	9.78	42.73	12.17	4.66	0.6363

Kensington & Chelsea	9.72	7.71	5.90	14.60	9.25	0.5646
Coventry	12.63	11.08	41.59	16.42	5.65	0.6289
Sandwell	11.51	9.07	28.49	16.28	5.96	0.5898
Bury	6.22	6.63	14.67	7.70	3.08	0.6294
South Tyneside	5.29	3.96	20.42	1.96	1.84	0.4808
Salford	8.78	6.56	31.72	8.60	4.83	0.5783
Hackney	13.50	4.77	26.47	20.88	4.17	0.4434
Camden	12.60	6.68	30.28	9.09	19.45	0.5478
Hillingdon	8.10	8.10	9.71	8.53	7.50	0.5821
Tower Hamlets	9.67	6.00	19.46	10.71	8.03	0.5187
Barnet	12.37	11.25	28.50	12.53	6.74	0.5604
Bolton	9.50	8.67	23.54	8.99	3.66	0.5411
Ealing	11.47	10.30	15.58	13.74	6.46	0.5388
Bromley	11.78	12.22	14.33	10.10	5.02	0.5039
Wandsworth	12.57	10.43	18.31	16.39	3.92	0.5098
Birmingham	50.26	32.33	150.00	45.10	19.58	0.3565
Enfield	12.70	9.50	22.39	14.90	5.80	0.5030
Southwark	13.30	7.53	21.99	14.66	8.32	0.4608
Knowsley	5.60	3.73	12.21	5.39	2.84	0.4786
Islington	11.75	5.20	13.28	13.62	7.10	0.4079
North Tyneside	8.47	6.15	19.45	6.51	3.30	0.4587
Kingston-upon-Thames	8.36	5.96	17.11	4.66	3.08	0.4107
Hounslow	11.07	7.25	16.34	8.69	6.62	0.4274
Richmond-upon-Thames	10.38	7.76	16.44	6.01	3.31	0.3941
Hammersmith & Fulham	11.83	5.35	12.41	12.24	4.57	0.3622
Newham	12.71	6.32	13.63	8.53	5.16	0.3268
Merton	11.19	6.58	10.90	3.52	3.46	0.2839
Mean	10.45	8.81	29.41	13.33	6.41	0.6314
Standard Deviation	6.21	4.50	23.54	9.38	5.57	0.1699

Table 2: Descriptive Statistics for the Derived Data, Y_r

Variable	Y_1	Y_2	Y_3	Y_4
Mean	0.8948	2.8492	1.2622	0.6665
Standard Deviation	0.3013	1.3959	0.5290	0.8320

Table 3: Descriptive Statistics Estimated from 100 Simulated Data Sets

Estimate	a_1^*	a_2^*	a_3^*	a_4^*
Mean	0.2687	0.0510	0.1446	0.1412
Standard Deviation	0.0625	0.0131	0.0314	0.0297