Rating and Evaluating the Combined Financial and Environmental Performance of Companies in the Metals and Mining Sector

J. S. Yeomans[*]

Management Science Area, Schulich School of Business, York University, Toronto, ON M3J 1P3, Canada

ABSTRACT. The basic intent in financial performance evaluation is to appraise current business operations internally and to benchmark them against similar business operations externally in order to identify best-in-class practices. In this paper, a new positive environmental/financial screening approach that can simultaneously include a wide combination of regulatory, technological, operational and event dimensions is created for the analyzing, rating, ranking, benchmarking, and selecting of companies from an industry sector. A method is provided that advances the use of Data Envelopment Analysis (DEA) for rating and ranking diverse groups of companies using a combination of both financial and environmental performance measures. This novel approach proceeds by first stratifying the sector into comparably efficient groups of companies through the construction of a series of efficient DEA frontiers, and then by ranking each company within these groups relative to DEA-based contextual attractiveness measures calculated from the different partitions. The method is illustrated through an application to a group of companies from the Metals and Mining Industry sector.

Keywords: Benchmarking, financial/environmental screening, rating and ranking

1. Introduction

Historically, investors have employed financial screening mechanisms to evaluate and discriminate between different companies in their investment selection processes. More recently, investors have started to combine "social" screens that incorporate nonfinancial, environmental criteria into this investment decision making process. Positive environmental screening strategies have progressed to the point of identifying favourable companies in which to invest based upon their relatively superior environmental and financial performance within a given industry sector. In these methods, various factors and indicators applying to each company are combined to create a metric that is used to contrast the company's performance relative to others within its specific sector. The aim of these *best-in-class* approaches is to permit investors to recognize which companies can meet their eco-efficiency challenges while simultaneously delivering superior financial profitability. Several major financial advisory companies such as Oekom Research, Innovest Strategic Value Investors, and Storebrand (see Labatt & White, 2002) have constructed environmental ratings and rankings of companies that now provide investors, fund managers, and financial analysts with this material information regarding corporate environmental performance.

Single-measure based gap-analysis has often provided the fundamental methodology for the performance evaluation and benchmarking of companies (Zhu, 2003). For example, output-to-input financial ratios such as return on investment (ROI) have been used as indices to characterize and screen a company's financial performance. Since any company's actual performance represents complex phenomena generally requiring more than a single characterizing criterion, any use of single measures simply ignores all interactions, substitutions and tradeoffs among various performance measures. Although gaps can often be easily identified with respect to individual performance measures, it remains a challenging task to satisfactorily combine multiple disparate measures during the final analysis stage (Camp, 1995). Thus, it is a rare occurrence when any "one-measure-at-a-time" gap-analysis can suffice for the purposes of effective performance evaluation (Camp, 1995; Zhu, 2003). Hence, most single-measure approaches prove unsatisfactory discriminants of "best practice" and are insufficient in the establishment of appropriate benchmarks for measuring a company's relative operating efficiency.

Clearly it is difficult to evaluate an organization's performance or to establish industry benchmarks when there are multiple measurements present, and these difficulties become magnified when the relationships between these measurements are complex and involve unknown tradeoffs (Camp, 1995). If the functional forms of the relationships between performance measures are known, then multi-criteria techniques can be used to estimate best-practice or *efficient frontier* levels of performance. However, the functional forms cannot be specified without *a priori* information on the corresponding tradeoffs and, unfortunately, such information is usually not wholly available in practice. Consequently, because business operations and processes often cannot be adequately characterized by existing analytical approaches, new techniques that can effectively process multiple performance measures, while also providing integrated industry benchmarks, must be considered essential.

---

[*] Corresponding author: syeomans@schulich.yorku.ca

In this paper, a new positive environmental screening approach that can simultaneously include a wide combination of regulatory, technological, operational and event dimensions will be created. The specific aim of the study is to construct and demonstrate this positive screening process for the analysis, ranking, and possible selection of companies from a specific industry sector using *data envelopment analysis* (DEA) (Thanassoulis, 2001) as the performance evaluation and benchmarking tool (Zhu, 2003). DEA's empirical orientation and absence of *a priori* assumptions make it an ideal tool for use in a variety of practical situations and the underlying theoretical basis of DEA is consistent with the practice of rating companies relatively within a specific sector by concurrently examining the relative efficiencies of their multiple performance measures. By using DEA, the focus of performance evaluation and benchmarking will shift from that of characterizing environmental performance in terms of single measures to evaluating performance from a mathematically sound, multidimensional system perspective (Zhu, 2003).

The basic intent in a performance evaluation is to appraise current business operations internally and to benchmark them against similar business operations externally in order to identify best-in-class practices. Thus, such best practices can be empirically identified and one can empirically estimate the resulting *efficient frontier* based upon the observations of similar business operations at a specific point in time. DEA has proved to be an effective tool for identifying such empirical efficient frontiers, for evaluating relative performance efficiencies, and for implicitly estimating the tradeoffs inherent within the empirically designed frontier. Fare *et al*. (2004) provide a thorough analysis of both the advantages and limitations of DEA, together with appropriate mechanisms for circumventing any identified shortcomings.

The DEA method to be described will be used to rank and rate companies according to environmental and financial performance relative to other companies within an industry sector. For investment performance purposes, an empirical frontier approach will be created that recursively partitions companies into sub-groups of relatively superior and inferior performers that can be considered comparable to sectoral rating systems currently employed by financial rating companies. Since companies in the highest-rated partitions can be considered as industry sector benchmarks, a DEA-based scoring system that compares the relative financial and environmental performance of non-benchmarked companies to these benchmarks will be introduced. It will subsequently be demonstrated how companies within each partition can be rank-ordered by constructing a novel contextual attractiveness measure relative to lower-rated companies. By combining all of these procedures into a single method, a new positive environmental screening process for identifying favourable com- panies in which to invest can be created. This entire set of procedures can be implemented using readily available spreadsheet-based software facilitated and linked together by a combination of relatively straightforward VBA programming (Albright, 2001). Consequently, practitioners can readily implement this entire screening method on most currently available personal computers (Ragsdale, 2001).

DEA proves to be an ideal approach for multicriteria systems evaluation when the relationships between the measures are complex and involve unknown tradeoffs. DEA holds advantages over many other methods in that it provides an objective multi-criteria decision-making tool that does not require variables to have the same scale or conversion weights applied to them, while permitting the simultaneous combination of both quantitative and qualitative measures (Thanassoulis, 2001). Thus, DEA permits an objective comparison of variables without requiring any of the *a priori* weightings or conversion factors that have been commonly employed in existing financial rating mechanisms. Consequently, relationships between an organization's environmental strategy and its role in performance ranking can be explicitly brought into the analysis using this approach. For an analytical comparison of the potentially numerous non-cost, non-numeric, and subjective financial/environmental variables of companies during the screening and benchmarking phases, these characteristics prove essential.

2. The Metals and Mining Sector

The specific sector selected for this study consisted of 22 major international companies from the metals and mining industry, with the majority of these companies possessing market captilizations in excess of $100 million (with several exceeding $1 billion). A complete listing of the companies considered appears in Table 1. With significant global operations, these companies provide a full spectrum of the worldwide metals and mining industry and can therefore be considered as representative of the sector with respect to environmental and financial performance. The data examined contained information from the period ending at the first quarter of 2003 with all empirical data furnished by Innovest Strategic Value Investors (henceforth referred to as "Innovest").

Numerous factors, or scores, were measured on each of these companies representing both financial and environmental performance indicators. The factors were broadly partitioned into the 3 major evaluation groupings or *categories* of (i) risk factors, (ii) environmental management capacity, and (iii) opportunity factors, and each of these categories was further subdivided into *sub-categories*. The sub-categories included within the risk factors category consisted of: historic liabilities, operating risk, sustainability risk indicators, and industry specific risk. The environmental management capacity category contained the sub-categories: environmental strategy, corporate governance, environmental management systems, audit, environmental training & development, certification, and products/materials. The opportunity factors category incorporated the sub-categories: strategic competence, environmental opportunity, and performance.

Innovest measured each factor with an ordinal score ranging from 0 (worst in class) to 10 (best in class) and created weighted data by proportionally weighting each ordinal score by the factor's relative contribution (as proprietarily

determined, *a priori*, by Innovest) to the overall company performance. These weighted data were subsequently concatenated into subcategory, category, and overall scores in order to create a relative measure of each company's standing within the metals and mining sector. The overall scores were used for rank ordering the companies within the sector and for assigning one of seven performance category, efficiency group ratings from *AAA* (best) to *CCC* (worst) to each company (see Table 1). Johnson Matthey Plc. and Alcan Inc. were the two highest rated companies, with both receiving the *AAA* rating.

3. Envelopment Modelling with DEA

DEA provides a very different analytical approach to evaluating the metals and mining data by "enveloping" the companies based upon their relative performance scores. In the practice of DEA, the term *decision-making unit* (DMU) has become the accepted term used for designating the specific business operations or processes being studied (Thanassoulis 2001). In this study, the DMUs are comprised of the 22 companies within the metals and mining sector. Charnes *et al.* (1978) originally developed DEA to measure the relative efficiencies amongst a given set of comparable DMUs. The basic concept behind DEA is to measure the performance of each DMU against a projected point on an empirically derived *efficient frontier*. Any DMU whose performance places it directly onto the efficient frontier is considered relatively *efficient*, or non-dominated, in comparison to the other DMUs in the sense that no other DMU's performance is strictly better in either its inputs or its outputs. Through DEA performance evaluation, the efficient frontier representing the best-practice can be identified, and it can be shown how any inefficient DMU can be improved (i.e. moved onto the efficient frontier) with suggested directions for improvement. DEA has been previously used to evaluate the technical efficiency of operations in a wide variety of practical applications including such diverse settings as schools (Charnes *et al.*, 1981), highway maintenance (Cook *et al.*, 1990), robotics (Cook *et al.*, 1992), banking (Barr & Siems, 1997), electrical utilities (Fare *et al.*, 1985), textiles (Zhu, 1996), steel (Ray *et al.*, 1998), physician practice (Chilingerian, 1995), and logistics (Kleinsorge *et al.*, 1989). DEA has also been used to provide insights into a variety of environmental issues (see, for example, Fare *et al.,* 2004; Haynes *et al.*, 1997; Sarkis & Weinrach, 2001; Sarkis & Talluri, 2004; Tyteca, 1996, 1997).

A mathematically based description of the actual enveloping process is as follows (Thanassoulis 2001; Zhu 2003). Each DMU possesses a set of inputs and outputs representing its multiple measures of performance. For the metals and mining data, these inputs and outputs consist of the environmental and financial measures observed for each company (or their aggregations into either sub-category or category data). For the general situation, DEA considers some particular set of *n* DMUs or observations. Each of these observations, DMU_j, $j = 1,..., n$, would consist of *m* inputs x_{ij}, $i = 1,..., m$, that produce *s* outputs y_{rj}, $r = 1,..., s$. In DEA it has become common practice to designate any one of

the *n* observations under specific evaluation as DMU_0 with x_{i0} and y_{r0} representing its *i*th input and *r*th output measure. Then if λ_j, $j = 1,..., n$, represent non-negative scalars, the relative efficiency measure, θ, of each DMU can be determined for each of the *n* observations by iteratively solving the optimization model:

$$\theta^* = min \; \theta \tag{1}$$

subject to:
$$\sum_{j=1}^{n} \lambda_j x_{ij} \leq \theta \, x_{i0} \qquad i = 1,..., m$$

$$\sum_{j=1}^{n} \lambda_j y_{rj} \geq y_{r0} \qquad r = 1,..., s$$

$$\sum_{j=1}^{n} \lambda_j = 1$$

$$\lambda_j \geq 0, j = 1,..., n$$

The decision variables in the enveloping model correspond to "weighting factors" applied to the inputs and outputs. However, unlike subjective scoring approaches, these weighting factors are not determined *a priori* and DEA aims to determine those factors which portray the DMU in its most favourable light. This method determines which units are relatively efficient and defines the relative inefficiency for units dominated by those on the efficient frontier. In summary, DEA determines the relative efficiency of the DMUs by (i) creating an *m* + *s* dimensional surface of the efficient DMUs (where *m* represents the number of inputs and *s* represents the number of outputs), (ii) assigning an efficiency score of θ = 1 to any DMU on the efficient frontier, (iii) determining the distance from the frontier for all inefficient DMUs, and (iv) calculating the value of θ for inefficient DMUs as its proportional, multi-dimensional distance from the nearest point on the efficient frontier.

Enveloping approaches have been broadly divided between two important modelling approaches depending upon the assumed "returns to scale" of the underlying data. Under the assumption of *Constant Returns to Scale* (CRS), if input levels of the input-output correspondences, ($\bar{x}$, $\bar{y}$), of each DMU are scaled up or down by some factor α >0, then another feasible input-output correspondence, ($\alpha \bar{x}$, $\alpha \bar{y}$), is obtained in which the output levels are scaled by the same factor as the input levels. Such an assumption provides for linear interpolation within the observed data. The major implication of the CRS assumption is that the scale of DMU operation does not impact its productivity measurements. In many "real life" contexts, the assumption of CRS is not always appropriate since CRS may not be possible to sustain when arbitrary measurement scales or scale-free variables such as indices have been employed. Consequently, for cases where economies/diseconomies of scale might exist, the more general *Variable Returns to Scale* (VRS) assumption is made which simultaneously permits constant, increasing, and de-

creasing returns to scale at different scale sizes. The DEA model shown above assumes VRS, but the straightforward removal of the condition, $\sum_{j=1}^{n}\lambda_j = 1$, would transform it into a corresponding CRS model. It should be noted that applying the VRS assumption to data is never inappropriate, since the corresponding model will correctly reflect linear interpolation whenever it is applicable to the data. However, VRS models can explicitly account for a lack of linear interpolation within the underlying data while CRS models cannot. Consequently, when VRS and CRS models produce different results, there is a clear indication that linear interpolation does not apply to the data.

This "basic" enveloping model will form the analytical cornerstone for evaluating the performance of the companies in the metals and mining sector. The model will provide the platform for the more complicated procedures described below that will be used to stratify the companies based upon their relative performance, to determine where the companies rank within each of the strata, and to subsequently create a multi-criteria benchmark measure of each company's performance within the sector.

4. Context Dependence and Benchmarking with DEA

The overall aim in evaluating the metals and mining sector is to determine the best-in-class companies by comparing each company to the other companies within the sector. That is, the goal is to contrast each company's performance relative to the performance of every other company using the DEA procedure. Because the relative performance of DMUs is contrasted only to the identified best-practice frontier when using the enveloping model, the actual measures of relative *inefficiency* can change only when the best-practice frontier is altered (that is, when one or more of the efficient DMUs is removed). Specifically, DEA can determine only that a company is efficient or that it is not efficient. Adding or deleting any subset of inefficient companies alters neither the best-practice envelopment frontier nor the calculated efficiency scores. Therefore, the envelopment process of DEA does not provide any direct means to rank the companies based upon their relative "degrees" of either efficiency or inefficiency. Due to this deficiency, combining either bootstrapping or jackknifing into DEA for rating and ranking purposes would not prove beneficial. Hence, it would prove significantly more beneficial to construct an enveloping mechanism that can be used to rank order the DMUs based upon their comparative degrees of attractiveness.

Consumer choice researchers have described how products can appear relatively attractive when compared to a background of less attractive alternatives, but relatively unattractive when contrasted against more attractive alternatives (Simonson & Tversky, 1992; Tversky & Simonson, 1993). This type of contextual attractiveness concept can be incorporated into the DEA evaluation process for the metals and mining sector in order to permit a rating and ranking of the vari-

ous companies under consideration. To obtain the relative attarcttiveness of the companies, the original enveloping methodology must be modified to define relative performance with respect to certain particular best-practice, evaluation *contexts*. These contexts partition the mining companies into groups or sets of like-performing companies that can be thought of as corresponding to the explicit class rating partitions used by Innovest (although, for this situation, the actual number of different groups cannot be known ahead of time).

A recursive, DEA-based enveloping procedure can be implemented that constructs these evaluation contexts by stratifying the DMUs (i.e. the companies) into several levels of relative "best-practice" frontiers. The procedure removes all of the DMUs on the original best-practice frontier and uses the enveloping methodology to form a second-tier best-practice frontier from all remaining DMUs. That is, once the first efficient frontier is calculated, all of the associated efficient companies are removed from further consideration and a new efficient frontier based only upon the remaining, initially inefficient companies is calculated. The DMUs on the second-tier efficient frontier are subsequently removed, permitting a third-tier frontier to be constructed, then a fourth-tier frontier, and so on, until no DMUs remain. In effect, this recursive approach partitions the entire industry sector into multiple groups of comparably-efficient companies based upon their performance measures and these partitions will supply the evaluation contexts for constructing the relative attractiveness scores used to rank those companies placed into the higher levels.

The technical details underlying the stratification procedure can be stated in the following way (Zhu, 2003). Define $\mathbf{J}^l$ = {DMU$_j$, $j = 1,\dots, n$} to be the set of all n DMUs and interactively define $\mathbf{J}^{l+1} = \mathbf{J}^l$ - $\mathbf{E}^l$ where $\mathbf{E}^l$ = {DMU$_k$ $\in$ $\mathbf{J}^l$ | θ^* (l, k) = 1}, and θ^* (l, k) is the optimal efficiency score for DMU$_k$ determined by the model:

$$\theta^*(l, k) = \min_{\lambda_j, \theta(l, k)} \theta(l, k) \qquad (2)$$

subject to:
$$\sum_{j\in F(\mathbf{J}^l)} \lambda_j x_{ij} \leq \theta(l, k)x_{ik}$$

$$\sum_{j\in F(\mathbf{J}^l)} \lambda_j y_{rj} \geq y_{rk}$$

$$\sum_{j\in F(\mathbf{J}^l)} \lambda_j = 1$$

$$\lambda_j \geq 0, \ j \in F(\mathbf{J}^l)$$

In this model, the notation j $\in$ $F(\mathbf{J}^l)$ implies that DMU$_j$ $\in$ $\mathbf{J}^l$ and $F(\bullet)$ represents the correspondence from a DMU set to the corresponding subscript index set.

When $l = 1$, the model becomes the original envelopment model for all n DMUs and $\mathbf{E}^l$ consists of all of the DMUs on the overall best-practice efficient frontier. These DMUs define the first-level best-practice frontier (or the *AAA*-type compa-

nies). When $l = 2$, the model provides the second-level best-practice frontier after the exclusion of the first-level efficient DMUs. This iterative exclusion proceeds in identifying all subsequent lower level best-practice frontiers in which $\mathbf{E}^l$ would correspond to the lth-level best-practice frontier. The iterative algorithm for identifying these best-practice frontiers can be summarized in the following way:

Step 1: Set $l = 1$. Evaluate the entire set of DMUs, $\mathbf{J}^l$, to obtain the set, $\mathbf{E}^l$, of first-level frontier DMUs.

Step 2: Exclude the frontier DMUs from future DEA runs and set $\mathbf{J}^{l+1} = \mathbf{J}^l - \mathbf{E}^l$.

Step 3: If $\mathbf{J}^{l+1} = \varnothing$, then stop. Otherwise evaluate the remaining subset of "inefficient" DMUs, $\mathbf{J}^{l+1}$, to obtain the new best-practice frontier $\mathbf{E}^{l+1}$.

Step 4: Let $l = l + 1$ and go to step 2.

Stopping Rule: The algorithm stops when $\mathbf{J}^{l+1} = \varnothing$.

The recursive procedure stratifies the original set of DMUs into L levels, or groupings, of company efficiencies, with the specific value for L algorithmically determined only *a posteriori* by the "empty-set" stopping rule. When applied to sectoral data, this contextual stratification effectively partitions the sector into distinct groupings of comparably-performing companies. While stratifying the sector into comparable performers, the procedure does not actually rank the standing of any of the companies within these groups. If only a small number of companies are considered in a sector or if a large number of groups each containing only a very small number of companies is produced (i.e. the value of L is very large, but each grouping contains very few elements), then this partitioning process may prove sufficient for actually screening between companies. In general, the stratification will not be restrictive enough to permit sufficient degrees of discrimination between companies. If this is the case, then the subsequent rank-ordering mechanism for the companies within each grouping proves essential.

In Innovest's analysis of the sector, the 22 companies were rank ordered based upon their calculated overall scores. Such rank orderings are not possible with DEA models, since when DMU_x and DMU_y are members of the same best-practice frontier, they both possess efficiency scores of one. Using the original enveloping methodology of DEA, each DMU is evaluated against the set of best-practice DMUs which only permits performance comparisons of inefficient DMU scores with respect to this frontier. Without modification, DEA cannot distinguish which of DMU_x or DMU_y within the same best-practice group provides the better option. In applying the consumer choice analogy directly to DEA, it would be of interest to compare the relative attractiveness of a particular DMU to others when the evaluation context can be constructed from the perspective of alternative options or DMUs. For instance, the relative attractiveness of DMU_x compared to DMU_y would depend on the presence or absence of some third, less-attractive option (say) DMU_z. Since the stratification partitions the set of projects into L efficiency frontiers, it now becomes possible to determine context-dependent attractiveness scores for each grouping of projects at the higher levels as "viewed" from the lower contextual levels. Hence, an evaluation score can be calculated for each project that is based upon how attractive it appears in comparison to projects classified into the less desirable contextual levels.

Once the stratification process has partitioned the set of DMUs into the different efficiency frontier levels characterized by $\mathbf{E}^l$, $l = 1, \dots, L$, it becomes possible to calculate such context-dependent attractiveness measures for the DMUs in the higher levels. These measures can then be used to rank order the DMUs within any specific stratification level. This attractiveness calculation proceeds as follows. Define $H_q^*(d)$ to be the d-degree, $d = 1, \dots, L - l_0$, contextual attractiveness of $DMU_q = (x_q, y_q)$ from some specific level $\mathbf{E}^{l_0}$, $l_0 \in \{1, \dots, L - 1\}$. $H_q^*(d)$ can be calculated by solving the following model:

$$H_q^*(d) = \min H_q(d) \qquad d = 1, \dots, L - l_0 \qquad (3)$$

subject to:
$$\sum_{j \in F(\mathbf{E}^{l_0+d})} \lambda_j x_j \le H_q(d) x_q$$

$$\sum_{j \in F(\mathbf{E}^{l_0+d})} \lambda_j y_j \ge y_q$$

$$\sum_{j \in F(\mathbf{E}^{l_0+d})} \lambda_j = 1$$

$$\lambda_j \ge 0, \ j \in (\mathbf{E}^{l_0+d})$$

DMU_q is viewed as a more attractive option than another DMU if it possesses a larger value for its contextual attractiveness measure, $H_q^*(d)$. Hence, given that the set of metals and mining companies has been stratified into different levels of peer-efficient groupings, these attractiveness measures would subsequently enable decision-makers to select the best option, or to rank all of the various options, within any particular level.

Benchmarking is a method that is widely used to identify and adopt best practices as a means for improving performance and increase productivity. *Benchmarking* is the process of defining valid measures of performance comparison among peer DMUs, using them to determine the relative positions of the peer DMUs, and ultimately in establishing standards of excellence. The stratification process presented above could be thought of as establishing various different sets or performance levels of peer benchmark DMUs to which other non-benchmarked DMUs could be contrasted. Benchmarks are particularly valuable in practical situations in which no objective or "engineered" standards are available to define efficient or effective performance. Furthermore, it can prove very difficult to evaluate an organization's performance when there are multiple inputs and outputs to a system and it is critical to provide benchmarks for situations where multiple measurements exist.

In practice, benchmarking models that can simultane-

ously cope with multiple performance measures and provide an integrated benchmarking measure are essential. In this sense, envelopment modelling can be regarded as a type of benchmarking tool since the identified frontier can be regarded as an empirical standard of excellence. Once a best-practice frontier has been established, it is possible to compare a set of "new" DMUs (those not included in the benchmark set) to the benchmark frontier. If DMU^{new} represents a new (or non-benchmarked) DMU with inputs x_i^{new}, $i = 1,…,$ m, and outputs y_r^{new}, $r = 1,…, s$, then the DEA procedure can be modified to determine a benchmark score, δ, that measures its performance relative to an established set, $\mathbf{E}^*$, of benchmark DMUs. The benchmark score is the solution to the problem:

$$\min \ \delta \quad\quad\quad\quad (4)$$
$$\text{subject to:} \ \sum_{j\in\mathrm{E}^*} \lambda_j x_{ij} \leq \delta \, x_i^{new}$$
$$\sum_{j\in\mathrm{E}^*} \lambda_j y_{rj} \geq y_r^{new}$$
$$\sum_{j\in\mathrm{E}^*} \lambda_j = 1$$
$$\lambda_j \geq 0 \ , \ j \in \mathbf{E}^*$$

The actual benchmark score obtained by this model establishes how any non-benchmarked (or "new") DMU performs in comparison to the benchmark set, with three possible outcomes for the interpretation of δ. If $\delta < 1$, then this indicates that the performance of DMU^{new} is dominated by the members in the benchmark set. A value of $\delta = 1$ indicates that DMU^{new} attains the same performance level as the members in the benchmark set. And if $\delta > 1$, then such a score indicates that either input savings or output surpluses exist within the performance of DMU^{new} in comparison to the benchmark set. These benchmark scores will be subsequently used to evaluate the quality of the contextual stratifications produced by DEA in relation to the ranking system of Innovest. It will also be described how such a benchmarking scoring system can be combined with the stratification to produce a positive screening mechanism for evaluating companies within an industry sector.

5. A DEA-Based Positive Environmental/Financial Screening Procedure

Positive environmental screening strategies are used to identify favourable companies in which to invest based upon their *relatively* superior environmental and financial performance within any given industry sector or group. It is possible to combine all of the above DEA procedures to create such a positive screening process for the analysis, ranking, and selection of best-in-class companies. The initial step in this process would be to use representative data from the various companies within the industry/sector to create several sets of "like-

performing" companies. In effect, this would directly require the development of an industry rating scheme along the lines of the DEA-derived, contextually-based stratification procedure (or could also involve some other, non-DEA rating, akin to the *AAA* to *CCC* rating categories of Innovest, at this stage). The companies within each of these categories/strata would then be considered indicative of the general level of industry performance for companies rated within the identified categories.

If these companies corresponded to representative companies for certain standards of industry performance, then they would naturally establish the industry benchmarks for their respective levels. These benchmarks could then be used to screen, or classify, any new company by allowing the benchmarked companies within each rating category to serve as its contextual evaluation background in the DEA-based, benchmarking procedure. To screen the new company, the different benchmark scores, δ, at each context (or rating) level would be calculated and, via a process of elimination, the industry classification rating of the new company would be assigned to the highest non-dominating contextual level. Such a process would allocate an industry rating to the new company according to the industry benchmarked companies to which it most closely resembles. Since the DEA-based procedure both adopts a best-in-class philosophy and includes environmental criteria measurements, this new multi-level benchmarking approach would provide an effective, positive, environmental screening strategy.

6. The Application of DEA to the Metals and Mining Sector

All of the various DEA approaches were applied to the metals and mining sector data provided by Innovest. In the initial portion of this empirical testing, the recursive partitioning procedure was used to stratify the companies into levels of groupings of "best-practice" frontiers which would be analogous to the rating classification approach of Innovest. However, unlike the predetermined number of classification groups used by Innovest, it could not be known *a priori* exactly how many partitions the DEA method would produce. Since the enveloping model could be operated using either (i) the sub-category data, or (ii) the category data, and since the procedure could simultaneously operate under both the (a) VRS and (b) CRS assumptions, the algorithm was initially run using all 4 possible combinations.

Table 1 provides the resulting partitions created by DEA under the VRS assumption, together with the company classifications (the efficiency group letter ratings) produced in the earlier Innovest analysis. The category-scale data has been stratified into 7 best-practice frontiers, while the sub-category-scale data groups the companies into 3 tiers of best-practice. Arbitrarily mapping the *AAA* to *CCC* efficiency group letter rating scheme to the 7 contextual category partitions produces similar, but not identical, company ratings to those of Innovest. Other than for Rio Tinto, the DEA classification placed all 21 other companies within one rating class

of those determined by Innovest. In comparison to Innovest's *AAA* ratings for Johnson Matthey and Alcan, the DEA stratification adds Rio Tinto, Alcoa, and Falconbridge to the top rated group, while downgrading Alcan to *AA*. Rio Tinto provides the most interesting discrepancy between the two rating approaches, since, although it achieved the highest sector score in the "Environmental Management Capacity" category, it received only a 4th tier, *BBB* company rating by Innovest.

A similar type of ratings scheme was performed on the sub-category partitionings. Mapping three broader-ranging efficiency group letter ratings (*AAA* to *A*, *BBB* to *B*, *CCC* to *C*) to the 3 stratification levels of the sub-category data, once again produced relatively analogous rating schemes between DEA and Innovest. It is interesting to observe, however, that with an enveloping that uses more parameters, the likelihood that a company can attain relative dominance over the other companies with respect to certain measures increases. Hence, although somewhat similar, not all of the previous category classification ratings fell within the ranges produced at the sub-category level. This clearly indicates that it can prove extremely beneficial to view financial and environmental data such as this through multiple analytical angles and lenses.

Table 2 parallels the stratification analysis from Table 1, but under the CRS assumption. Interestingly enough, although DEA again stratified the category data into 7 contextual levels and the sub-category data into 3 levels, there are significant differences between the specific best-practice company memberships within these levels in comparison to the VRS partition. Because VRS and CRS produce identical results only when the CRS assumption is valid, this finding indicates that the specific metals and mining data analyzed does not exhibit the linear interpolation property requisite for CRS.

A comparison between the category data partitioning and the Innovest ratings demonstrates an even greater degree of correspondence between the groupings than experienced under VRS, with all companies falling within one rating category under either method. The *AAA* rated companies in the category data are Alcan, Alcoa, and Johnson Matthey. The most "significant" change from the VRS analysis is that Rio Tinto has fallen from a *AAA* rating to an *A* rating under the CRS assumption. As with the category analysis, the sub-category groupings also exhibit a closer relative congruence to the Innovest ratings, but not all of the category classification ratings fall within the ranges produced at this sub-category level. The most significant finding during this best-practice stratification phase is that the assumption of CRS in the metals and mining data cannot be supported by this analysis. Consequently, the VRS assumption will be applied exclusively in all subsequent analysis. Furthermore, since the Innovest ratings closely parallel those in the CRS findings, this observation provides strong indicative evidence of an unforeseen shortcoming within the Innovest rating system warranting further investigation by the company; namely that Innovest has been implicitly or explicitly applying an incorrect linear interpolation assumption in their rating system. Since Innovest is a major international financial rating org-

anization, the identification of inappropriate evaluation methods is extremely significant and needs to be addressed and corrected for in their own internal system.

While the DEA partitioning algorithm stratifies the companies into distinct levels of best-practice frontiers, it does not provide the relative "scoring" mechanism needed to rank order the companies within these groupings. Such a ranking system would be particularly beneficial within the sub-category partitioning in which very few levels were created, with all but Xstrata being placed into one of two categories. While the best-practice enveloping does not provide this rank ordering, it becomes possible to evaluate the relative standing of each member of a given tier after the partitioning by having a lower tier provide the ranking context. The results of this contextual ranking are illustrated in Table 3 for both the category and sub-category data, and are shown alongside the Innovest company rankings for comparison. While the number of levels produced by the category and sub-category DEA partitionings is different, it is quite striking to observe the extremely similar contextually-based company rankings produced under either approach. Moreover, it can be seen how the rankings produced by DEA and Innovest are relatively similar with the exception of certain conspicuous differences. Notably, the DEA contextual rankings place Rio Tinto, Inco, Homestake, and Barrick much higher within the sector than does Innovest, while placing Alcan, Placer Dome, and Noranda at much lower positions in the relative "pecking" order. All procedures position Johnson Matthey as the top-ranked company within the sector, while simultaneously placing Xstrata at the very bottom.

Clearly, however, for various analytical reasons the approaches of Innovest and DEA produce different top-ranked or best-of-class companies in their performance evaluation of the metals and mining sector. Innovest rates Johnson Matthey and Alcan as the *AAA* companies within the sector, while placing Alcoa, Falconbridge, Noranda and Pechiney into their *AA* category. Conversely, DEA places Rio Tinto, Alcoa, Falconbridge, and Johnson Matthey into the *AAA* group, and categorizes Lonmin and Pechiney as the AA companies. If these *AAA* and *AA* companies are to be viewed as the financial and environmental leaders within this sector, then it would prove instructive to juxtapose, or benchmark, the performance of the lower-ranked companies against these "established" standards of excellence.

As described earlier, once the best-practice group of companies has been established as the industry benchmark, it becomes possible to compare any non-benchmarked companies to these companies by computing their relative performance scores, δ. Scores of $\delta < 1$ signify that a company's performance is strictly inferior to the benchmarked companies, while scores of $\delta > 1$ indicate that certain performance measures of the non-benchmarked company are superior relative to certain members of the benchmark set. Obviously the performance of the best-of-class companies should demonstrate clear dominance over those not included as benchmarks.

Tables 4 and 5 provide the relative benchmark scores

Table 1. DEA Stratification of Metals and Mining Sector under the Variable Returns to Scale (VRS) Assumption

COMPANY	Innovest's Efficiency Group Letter Rating	Using Category Data		Using Sub-Category Data	
		Contextual Level Produced by VRS DEA	Conversion of Contextual Level to an Efficiency Group Letter Rating	Contextual Level Produced by VRS DEA	Conversion of Contextual Level to an Efficiency Group Letter Rating
Rio Tinto	BBB	1	AAA	1	AAA to A
Alcan Inc.	AAA	2	AA	1	AAA to A
Alcoa Inc.	AA	1	AAA	1	AAA to A
Cameco Corporation	B	5	BB	1	AAA to A
Freeport McMoRan Copper and Gold Inc.	B	6	B	1	AAA to A
Inco Ltd.	BBB	3	A	2	BBB to B
Phelps Dodge Corporation	CCC	5	BB	2	BBB to B
Newmont Mining Corporation	CCC	7	CCC	2	BBB to B
Homestake Mining Company	BBB	3	A	1	AAA to A
Barrick Gold Corporation	BB	3	A	1	AAA to A
Falconbridge Ltd.	AA	1	AAA	1	AAA to A
Teck Corporation	BB	5	BB	2	BBB to B
Placer Dome Inc.	A	4	BBB	1	AAA to A
Noranda Inc.	AA	3	A	1	AAA to A
Lonmin Plc.	A	2	AA	1	AAA to A
Billiton Plc.	A	3	A	1	AAA to A
Anglo American Plc.	BBB	3	A	2	BBB to B
Johnson Matthey Plc.	AAA	1	AAA	1	AAA to A
BHP Ltd.	BB	5	BB	2	BBB to B
Mitsui Mining & Smelting Co., Ltd.	BBB	4	BBB	2	BBB to B
Pechiney S.A.	AA	2	AA	1	AAA to A
Xstrata	CCC	7	CCC	3	CCC to C

Table 2. DEA Stratification of Metals and Mining Sector under the Constant Returns to Scale (CRS) Assumption

COMPANY	Innovest's Efficiency Group Letter Rating	Using Category Data		Using Sub-Category Data	
		Contextual Level Produced by CRS DEA	Conversion of Contextual Level to an Efficiency Group Letter Rating	Contextual Level Produced by CRS DEA	Conversion of Contextual Level to an Efficiency Group Letter Rating
Rio Tinto	BBB	5	A	1	AAA to A
Alcan Inc.	AAA	2	AAA	1	AAA to A
Alcoa Inc.	AA	1	AAA	1	AAA to A
Cameco Corporation	B	13	CCC	2	BBB to B
Freeport McMoRan Copper and Gold Inc.	B	11	B	2	BBB to B
Inco Ltd.	BBB	6	A	2	BBB to B
Phelps Dodge Corporation	CCC	10	BB	2	BBB to B
Newmont Mining Corporation	CCC	13	CCC	2	BBB to B
Homestake Mining Company	BBB	8	BBB	1	AAA to A
Barrick Gold Corporation	BB	11	B	2	BBB to B
Falconbridge Ltd.	AA	3	AA	1	AAA to A
Teck Corporation	BB	9	BB	2	BBB to B
Placer Dome Inc.	A	7	BBB	1	AAA to A
Noranda Inc.	AA	4	AA	1	AAA to A
Lonmin Plc.	A	4	AA	1	AAA to A
Billiton Plc.	A	5	A	1	AAA to A
Anglo American Plc.	BBB	7	BBB	2	BBB to B
Johnson Matthey Plc.	AAA	1	AAA	1	AAA to A
BHP Ltd.	BB	12	B	2	BBB to B
Mitsui Mining & Smelting Co., Ltd.	BBB	6	A	2	BBB to B
Pechiney S.A.	AA	3	AA	1	AAA to A
Xstrata	CCC	14	CCC	3	CCC to C

Table 3. Company Within-Sector Rankings Using Contextual DEA

Company	Innovest Ranking	DEA Ranking Based on Category Data	DEA Ranking Based on Sub-Category Data
Rio Tinto	10	4	4
Alcan Inc.	2	6	5
Alcoa Inc.	3	2	2
Cameco Corporation	19	18	17
Freeport McMoRan Copper and Gold Inc.	18	20	19
Inco Ltd.	14	10	9
Phelps Dodge Corporation	20	19	20
Newmont Mining Corporation	21	21	21
Homestake Mining Company	12	8	8
Barrick Gold Corporation	16	11	11
Falconbridge Ltd.	4	3	3
Teck Corporation	15	16	15
Placer Dome Inc.	8	14	13
Noranda Inc.	6	13	14
Lonmin Plc.	9	7	7
Billiton Plc.	7	12	12
Anglo American Plc.	11	9	10
Johnson Matthey Plc.	1	1	1
BHP Ltd.	17	17	18
Mitsui Mining & Smelting Co., Ltd.	13	15	16
Pechiney S.A.	5	5	6
Xstrata	22	22	22

Table 4. Relative Benchmark Scores for Non-Benchmarked Companies Using the Category Data

Company	Innovest Rating	Innovest Benchmark Values Using AAA Companies	Innovest Benchmark Values Using AAA & AA Companies	DEA Level or Rating	DEA Benchmark Values Using Level 1 Companies	DEA Bench-mark Values Using Levels 1 & 2 Companies
Rio Tinto	BBB	1.188	1.082	1	benchmark	benchmark
Alcan Inc.	AAA	benchmark	benchmark	2	0.985	0.981
Alcoa Inc.	AA	1.057	benchmark	1	benchmark	benchmark
Cameco Corporation	B	0.851	0.775	5	0.770	0.723
Freeport McMoRan Copper and Gold Inc.	B	0.783	0.735	6	0.735	0.686
Inco Ltd.	BBB	0.784	0.741	3	0.741	0.693
Phelps Dodge Corporation	CCC	0.637	0.601	5	0.601	0.543
Newmont Mining Corporation	CCC	0.639	0.610	7	0.610	0.552
Homestake Mining Company	BBB	0.943	0.880	3	0.880	0.852
Barrick Gold Corporation	BB	0.797	0.797	3	0.797	0.754
Falconbridge Ltd.	AA	1.098	benchmark	1	benchmark	benchmark
Teck Corporation	BB	0.796	0.776	5	0.776	0.732
Placer Dome Inc.	A	0.910	0.855	4	0.855	0.823
Noranda Inc.	AA	0.945	benchmark	3	0.894	0.869
Lonmin Plc.	A	0.825	0.825	2	0.825	benchmark
Billiton Plc.	A	0.949	0.880	3	0.880	0.852
Anglo American Plc.	BBB	0.821	0.803	3	0.803	0.762
Johnson Matthey Plc.	AAA	benchmark	benchmark	1	benchmark	benchmark
BHP Ltd.	BB	0.835	0.761	5	0.761	0.715
Mitsui Mining & Smelting Co., Ltd.	BBB	0.780	0.743	4	0.743	0.695
Pechiney S.A.	AA	0.865	benchmark	2	0.850	benchmark
Xstrata	CCC	0.597	0.597	7	0.597	0.531

Table 5. Relative Benchmark Scores for Non-Benchmarked Companies Using the Sub-Category Data

Company	Innovest Rating	Innovest Benchmark Values Using AAA Companies	Innovest Benchmark Values Using AAA & AA Companies	DEA Level or Rating	DEA Benchmark Values Using Level 1 Companies
Rio Tinto	BBB	1.903	1.894	1	benchmark
Alcan Inc.	AAA	benchmark	benchmark	1	benchmark
Alcoa Inc.	AA	1.896	benchmark	1	benchmark
Cameco Corporation	B	1.585	1.508	1	benchmark
Freeport McMoRan Copper and Gold Inc.	B	2.655	1.828	1	benchmark
Inco Ltd.	BBB	0.972	0.965	2	0.969
Phelps Dodge Corporation	CCC	1.000	1.000	2	0.987
Newmont Mining Corporation	CCC	0.936	0.838	2	0.825
Homestake Mining Company	BBB	4.090	3.870	1	benchmark
Barrick Gold Corporation	BB	1.238	1.119	1	benchmark
Falconbridge Ltd.	AA	2.498	benchmark	1	benchmark
Teck Corporation	BB	1.215	1.077	2	0.969
Placer Dome Inc.	A	1.537	1.424	1	benchmark
Noranda Inc.	AA	1.453	benchmark	1	benchmark
Lonmin Plc.	A	1.453	1.143	1	benchmark
Billiton Plc.	A	2.655	1.828	1	benchmark
Anglo American Plc.	BBB	1.306	0.915	2	0.932
Johnson Matthey Plc.	AAA	benchmark	benchmark	1	benchmark
BHP Ltd.	BB	1.277	0.966	2	0.966
Mitsui Mining & Smelting Co., Ltd.	BBB	1.153	0.966	2	0.953
Pechiney S.A.	AA	1.215	benchmark	1	benchmark
Xstrata	CCC	0.871	0.804	3	0.793

calculated for the non-benchmarked companies using the category and sub-category data, respectively. Table 4 demonstrates that both the *AAA* and *AAA/AA* companies identified through DEA stratification dominate all of the performance measures of the non-benchmarked companies. Similarly, Table 5 indicates that the top level companies from the sub-category DEA partitioning also dominate all non-benchmarked companies (note: this includes *A* rated companies). Thus, these results provide a solid analytical basis and sound objective support to the DEA stratifying procedure.

Interestingly enough, however, when evaluated using the category data, Table 4 indicates that the *AAA* benchmarked companies of Innovest do not dominate the performance measures of Rio Tinto and Alcoa, and that Innovest's *AAA/AA* benchmarked companies do not dominate Rio Tinto. Furthermore, when using the sub-category data, Innovest's *AAA* benchmarked companies do not dominate the performance of Rio Tinto, Alcoa, Cameco, Freeport McMoRan, Phelps Dodge, Homestake, Barrick, Falconbridge, Teck, Placer Dome, Noranda, Lonmin, Billiton, Anglo American, BHP, Mitsui, and Pechiney (see Table 5). Similarly, the performance of the *AAA/AA* benchmark cannot be shown to demonstrably dominate that of Rio Tinto, Cameco, Freeport McMoRan, Phelps Dodge, Homestake, Barrick, Teck, Placer Dome, Lonmin, and Billiton. While there may prove to be a sound explanation behind these observations, this non-dominance by their best-of-class benchmark companies should highlight some degrees of concern for Innovest and its root causes would need an

appropriately thorough investigation.

7. Conclusions

In this paper, a DEA-inspired positive screening approach for analyzing, ranking, benchmarking, and selecting companies has been investigated. Since the process, in general, is adaptable and invariant to data type, enveloping permits the comparison of non-numeric "environmental" variables that would be "incomparable" using many other methods. In contrast to procedures requiring the subjective assignment of weightings to variables, DEA provides an objective multi-criteria decision-making approach that does not require variables to possess the same scale or have "artificial" conversion weights applied to them. For analytically comparing the non-cost, non-numeric, and subjective financial/environmental variables of companies, this empirical orientation and absence of *a priori* assumptions render DEA an ideal method for practical implementation. Furthermore, for positive, best-inclass purposes, the underlying enveloping basis behind DEA proves consistent with the practice of rating companies relatively within their specific sectors.

Several benefits of the DEA approach were clearly demonstrated during an empirical investigation of the metals and mining sector data provided by Innovest Strategic Value Investors. Firstly, it was shown how DEA could be recursively applied to stratify these companies into similarly efficient groupings in a manner that would be analogous to a sectoral

rating system. Secondly, in order to rank order the companies within these partitions, a novel scoring evaluation based upon relative contextual attractiveness to lower rated companies was introduced. Thirdly, using the highest-rated companies as sector benchmarks, a new benchmark scoring system for comparing non-benchmarked companies was provided. It was demonstrated how all of the non-benchmarked companies in the sector were dominated by these newly established benchmarks. And fourthly, a new process combining all of the DEA concepts was outlined that could serve as the foundation for a positive environmental screening strategy that could identify favourable companies in which to invest within any given industry sector.

Clearly, based upon this empirical investigation, DEA has provided a useful multi-criteria approach for the benchmarking and relative rating of the financial/environmental performance of companies, and its broader applicability to additional applications warrants further study.

References

Albright, S.C. (2001). *VBA for Modelers: Developing Decision Support Systems with Microsoft Excel*, Duxbury, Pacific Grove, CA.

Barr, R.S. and Siems, T.F. (1997). Bank failure prediction using DEA to measure management quality, in R.S. Barr, R.V. Helgason and J.L. Kennington (Eds.), A*dvances in Metaheuristics, Optimization, and Stochastic Modeling Techniques*, Kluwer Academic Publishers, Boston, MA.

Camp, R.C. (1995). *Business Process Benchmarking, Finding and Implementing Best Practices*, ASQC Quality Press, Milwaukee, WI.

Charnes, A., Cooper, W.W. and Rhodes, E. (1978). Measuring the efficiency of decision making units. *Eur. J. Oper. Res.*, 2(6), 429-444.

Charnes, A., Cooper, W.W. and Rhodes, E. (1981). Evaluating program and managerial efficiency: An application of data envelopment analysis to program follow through. *Manage. Sci.*, 27(6), 668-697.

Chilingerian, J.A. (1995). Evaluating physician efficiency in hospitals: A multivariate analysis of best practice. *Eur. J. Oper. Res.*, 80, 548-574.

Cook, W., Johnston, D.A. and McCutcheon, D. (1992). Implementation of robotics: Identifying efficient implementers, *OMEGA*, 20, 227-239.

Cook, W., Roll, Y. and Kazakov, A. (1990). A DEA model for measuring the relative efficiency of highway maintenance patrols, *INFOR*, 28(2), 13-20.

Fare, R., Grosskopf, S. and Hernandez-Sancho, F. (2004). Environmental performance: An index number approach. *Resour. Energy Econ.* (forthcoming).

Fare, R., Grosskopf, S., Logan, J. and Lovell, C.A.K. (1985). Measuring efficiency in production with an application to electric utilities, in R. Fare, S. Grosskopf and C.A.K. Lovell (Eds.), *The Measurement of Efficiency of Production*, Kluwer Academic Publishers, Boston, MA.

Haynes, K.E., Ratick, S. and Cummings-Sexton, J. (1997). Pollution prevention frontiers: A data envelopment simulation, in G.J. Knaup and T.J. Kim, (Eds.), *Environmental Program Evaluation: A Primer*, University of Illinois Press, Urbana, IL.

Kleinsorge, I.K., Schary, P.B. and Tanner, R.D. (1989). Evaluating Logistics Decisions. *Int. J. Physical Distrib. Mater. Manage.*, 19, 200-219.

Labatt, S. and White, R. (2002). *Environmental Finance: A Guide to Environmental Risk Assessment and Financial Products*, John Wiley & Sons, Hoboken, NJ.

Ragsdale, C. (2001). *Spreadsheet Modeling & Decision Analysis*, South-Western College Publishing, Cincinnati, OH.

Ray, S., Seiford, L.M. and Zhu, J. (1998). Market entity behaviour of chinese state-owned enterprises. *OMEGA*, 2, 263-278.

Sarkis, J. and Weinrach, J. (2001). Using data envelopment analysis to evaluate environmentally conscious waste treatment technology. *J. Cleaner Prod.*, 9(5), 417-427.

Sarkis, J. and Talluri, S. (2004). Ecoefficiency measurement using data envelopment analysis: research and practitioner issues. *J. Environ. Assess. Policy Manage.*, 6(1) 91-123.

Simonson, I. and Tversky, A. (1992). Choice in context: Tradeoff contrast and extremeness aversion. *J. Mark. Res.*, 29, 281-295.

Thanassoulis, E. (2001). *Introduction to the Theory and Application of Data Envelopment Analysis*, Kluwer Academic Publishers, Norwell, MA.

Tversky, A. and Simonson, I. (1993). Context-dependent preferences. *Manage. Sci.*, 39, 1179-1189.

Tyteca, D. (1996). On the measurement of the environmental performance of firms: A literature review and a productivity efficiency perspective. *J. Environ. Manage.*, 46, 281-308.

Tyteca, D. (1997). Linear programming models for the measurement of environmental performance of firms: concepts and empirical results. *J. Productivity Anal.* 8, 183-197.

Zhu, J. (1996). DEA/AR analysis of the 1988-1989 performance of the Nanjing textile cooperation. *Ann. Oper. Res.*, 66, 311-335.

Zhu, J. (2003). *Quantitative Models for Performance Evaluation and Benchmarking*, Kluwer Academic Publishers, Norwell, MA.