A SURVEY ON INVESTMENT PERFORMANCE APPRAISAL WITH SPECIAL REFERENCE TO DATA ENVELOPMENT ANALYSIS

Don U.A. Galagedera*
Department of Econometrics and Business Statistics
Monash University
PO Box 197 Caulfield East
Victoria 3145 Australia

Abstract

This paper offers a review of investment performance appraisal methods. The review starts with an exhaustive coverage of various methods ranging from early measures of risk-adjusted return to more recent methods including the rating given by ASSIRT, a financial services organisation that assess managed investment in Australia. We then extend the discussion to performance evaluations based on the concept of production frontier estimation. Primarily, there are two competing theories of frontier estimation, known as stochastic frontier estimation and data envelopment analysis (DEA). DEA enables the inclusion of many factors in the analysis in addition to the usual return and risk measures and therefore is widely used in empirical studies in finance. The DEA methodology and its application in the finance sector are discussed in detail.

Key words: investment performance appraisal, data envelopment analysis

Introduction

Given today's volatile global investment climate, the increasing number of private investors and managed funds, and the growing financial services industry, investment performance appraisal is of paramount importance. Investors, of course, have always been eager to assess the performance of their managed portfolios. In early days, performance was evaluated by comparing the total return of a managed portfolio with that of a randomly chosen unmanaged portfolio (Modigliani and Modigliani, 1997). Later, the concept of an unmanaged 'market' or a capitalisation-weighted portfolio comprising the entire market was introduced so that managed

* Correspondence to: Tissa.Galagedera@buseco.monash.edu.au

ortfolio performance could be evaluated and compared against the market portfolio as a benchmark.

It is well-known that the return earned by a portfolio alone is not an accurate measure of its performance. Further, it is well-established that higher expected returns are associated with higher levels of risk. The downside to this is the possibility of considerable return losses due to market uncertainty. In short, there is a trade-off between risk and return. Investors are generally risk aversive. Therefore, for any risk associated with their investment, investors expect compensation or a risk premium. Consequently, several basic performance appraisal methods emerged in the late 1960s. With the rapid growth and globalisation of finance sectors, the financial services industry responded with new relative performance measures that have now become very popular and are widely used by private and institutional investors. However, there is no consensus in the literature as to what a suitable measure of risk is, and consequently, as to what is a suitable measure for evaluating risk-adjusted performance.

The main shortcoming in the common measures of risk-adjusted return is their inability to incorporate the costs incurred in generating the returns. In the late 1990s, several studies attempted to measure managed portfolio performance by considering the return adjusted for both risk and cost, using a non-parametric methodology of production frontier estimation commonly known as data envelopment analysis (DEA). In this paper, we briefly review the literature on risk-adjusted investment performance measures and production frontier estimation with special reference to DEA and its application to the finance sector.

This paper is organised as follows. First, the basic investment performance measures and the new development of comprehensive performance measures are reviewed. This is followed by a brief description of the types of production efficiency measures at the individual production unit level. A discussion on the relative merits of the two main production frontier estimation methods leads to an outline of the DEA methodology and its application in the finance sector. The final section concludes the paper.

Investment performance evaluation

(a) Early development

The measures used to evaluate asset or fund performance are based on some variations of risk-adjusted returns.

Sharpe index

Sharpe (1966) suggested that the historical performance of a portfolio may be calculated as the excess return earned for bearing risk per unit of total risk. Symbolically, the Sharpe index, S_p, is written as:

$$S_p = \frac{\overline{R}_p - \overline{R}_f}{\sigma_p}. \tag{1}$$

where $\overline{R}_p$ is the mean portfolio return, $\overline{R}_f$ is the mean risk-free asset return and σ_p is the standard deviation of portfolio returns. A higher value for S_p indicates that the portfolio delivers a higher performance for its level of total risk measured by σ_p. There is no benchmark for comparison of performance measures obtained from the Sharpe index. They can mainly be used to compare the performance of several portfolios.

Treynor index

Treynor (1965) considered only the non-diversifiable market risk of an investment. The non-diversifiable market risk, β_p, is defined as:

$$\beta_p = \frac{\sigma_p}{\sigma_m} r_{pm} \tag{2}$$

where r_{pm} is the correlation coefficient between the portfolio return and the market return and σ_m is the standard deviation of market returns. Treynor developed the following relative measure of portfolio performance:

$$T_p = \frac{\overline{R}_p - \overline{R}_f}{\beta_p}. \tag{3}$$

Since this measure does not include diversifiable risk, it can be regarded as a general performance measure and used regardless of the extent of diversification of the portfolio being evaluated.

Now, from (1) and (3) we obtain

$$T_p = S_p\left(\frac{\sigma_m}{r_{pm}}\right). \tag{4}$$

Thus, if the fund under evaluation is perfectly diversified ($r_{pm} = 1.0$), the Treynor index is equal to the Sharpe index times a constant and the portfolio ranking based on these two indices will therefore be identical. If the fund under investigation is not perfectly diversified ($r_{pm} < 1.0$) the performance ranking based on the Sharpe and Treynor indices might be different.

The choice between using the Sharpe or Treynor index depends on the nature of the portfolio being evaluated. If the entire portfolio is considered, the total risk of the investment will be the same as that of the risk borne by the investor. Hence the Sharpe index may be used here. On the other hand, if the evaluation is only on a component of the portfolio, the risk to the investor will only be the non-diversifiable systematic risk. Hence the Treynor measure will be more appropriate.

Jensen's alpha

Jensen (1969) considered an empirical version of the one-period security market line[1] given by:

$$R_{pt} = R_{ft} + \beta_p \left(R_{mt} - R_{ft}\right) + e_{pt} \tag{5}$$

where,

R_{pt} = realised portfolio return during time period t,

R_{ft} = risk-free asset return during time period t,

R_{mt} = realised market return during time period t, and

e_{pt} = error term that reflects portfolio return unrelated to market return.

Jensen introduced an additional term, α_p, to the above model to represent a constant periodic return (positive or negative) that an investor is able to earn in addition to the return of an unmanaged portfolio with identical market risk. Thus, rearranging the terms in (5) together with α_p gives:

$$R_{pt} - R_{ft} = \alpha_p + \beta_p \left(R_{mt} - R_{ft}\right) + e_{pt}\,. \tag{6}$$

Jensen suggested using regression procedures to estimate α_p and β_p and interpreted the estimated alpha based on its sign; if $\alpha_p > 0$ $(\alpha_p < 0)$ and is significant, then the portfolio has outperformed (under-performed) a possible buy-hold strategy, predicted by the market.

Portfolio managers can of course manipulate the alpha through leverage. Therefore, while the Jensen index is a basic risk-adjusted performance measure based on non-diversifiable risk, as measured by the beta, it cannot be used for ranking portfolios.

---

[1] The security market line (SML) expresses the return an individual investor can expect in terms of a risk-free rate and the relative risk of a security or portfolio. The SML with respect to security *i* can be written as $E(R_i) = R_f + \beta_i \left\{E\left(R_m\right) - R_f\right\}$ where, $\beta_i = \dfrac{\sigma_i r_{im}}{\sigma_m} = \dfrac{\mathrm{cov}\left(R_i, R_m\right)}{\sigma_m^2}$ and r_{im} = the correlation between security return, R_i and market portfolio return. The β_i can be interpreted as the amount of non-diversifiable risk inherent in the security relative to the risk of the market portfolio.

From (3) and the average of (6) over the time period, it can be seen that Jensen's alpha is related to the Treynor index as follows:

$$T_p = \frac{\alpha_p}{\beta_p} + \left(\overline{R}_m - \overline{R}_f\right). \tag{7}$$

Since $\left(\overline{R}_m - \overline{R}_f\right)$ is a constant, the Treynor index is simply a transformation of Jensen's alpha divided by the portfolio systematic risk. A criticism of the Treynor and Jensen measures is that their derivations are based on an explicit functional relationship between risk and return only.

(b) Recent developments

An average investor unfamiliar with regression analysis and modern finance theory finds Treynor and Jensen's alpha indices difficult to interpret. Therefore, with more and more private investors showing interest in investing in financial assets, there is a pressing need for performance appraisal methods that an average investor can easily understand. A measure developed without sophisticated theory is the Modigliani and Modigliani (1997) measure, which is described in the next section. Meanwhile, the number of managed funds and the number of institutions managing these funds has grown rapidly. The financial services industry has responded to the needs of investors by establishing companies to do research and rate managed- funds based on many factors in addition to the usual return versus risk. There is no doubt that the basic performance appraisal measures outlined in earlier provide valuable information on management effectiveness. However, factors such as asset class representation, portfolio correlations, expenses and turnover are also very relevant variables that should be taken into account in managed fund performance appraisal. Incorporating these variables will undoubtedly improve the performance measures of funds and their rating. Morningstar Incorporated in the United States and ASSIRT Pty Ltd in Australia are two well-known establishments that provide ratings of a very large number of managed funds in their respective countries. Institutional and private investors heavily rely on these ratings for their investment choices.

Modigliani and Modigliani measure

Modigliani and Modigliani (1997) developed a risk-adjusted performance measure equating the total risk of a managed portfolio with that of the market by creating a hypothetical portfolio comprising a risk-free asset and the managed portfolio. The idea is to adjust the managed portfolio risk to the level of risk of the market portfolio and then measure the returns of the risk-matched portfolio. The Modigliani and Modigliani measure, M_p^2, is calculated as:

$$M_p^2 = \frac{\overline{R}_p - \overline{R}_f}{\sigma_p}\sigma_m + \overline{R}_f \,. \tag{8}$$

Since $\overline{R}_f$ is common to all portfolios a simpler measure of risk-adjusted performance, $M_p^2(adjusted)$ is given as:

$$M_p^2(adjusted) = \frac{\overline{R}_p - \overline{R}_f}{\sigma_p}\sigma_m \,. \tag{9}$$

M_p^2 and $M_p^2(adjusted)$ rank portfolios identically.

From (1) and (9) we obtain,

$$M_p^2(adjusted) = S_p\sigma_m \tag{10}$$

suggesting that the Modigliani and Modigliani measure and the Sharpe index rank portfolios identically. Further, M_p^2 is expressed in percentages similar to portfolio returns and therefore, it is thought to be easily understood by an ordinary investor.

Morningstar rating

Morningstar Incorporated produces a number of managed fund performance measures that take risk and return into account. In some of their measures such as the Morningstar-Sharpe ratio and the Morningstar alpha, each fund receives a numeric rating independent of the performance of other funds. Others such as the category[2] risk-adjusted rating, the three-year risk-adjusted rating and the three-year star rating are relative measures. As with any other risk-adjusted performance measure, Morningstar also calculates the return (Morningstar return– MSRET) and the risk (Morningstar risk– MSRISK) of the funds and defines their risk-adjusted rating as the difference between MSRET and MSRISK. See Sharpe (1998) for details.

ASSIRT rating

ASSIRT is a financial services corporation that assesses managed investments in Australia. To establish a rating, ASSIRT considers a weighted combination of manager capability, past performance and fund issues such as objectives, features, risk issues and strategy information. The weights assigned to manager capability, past performance and fund issues are 55 per cent, 25 per cent and 20 per cent respectively. Each fund is assessed and scored against more than 400 criteria and between one to five 'stars' then awarded. The number of stars measures ASSIRT's assessment of the overall quality of a managed fund and the likelihood that the fund is achieving its investment objectives. See Table 1 for ASSIRT fund rating definitions.

Table 1. ASSIRT fund rating definitions

Rating	Definition
☆☆☆☆☆	An excellent fund with very strong management, a comprehensive investment strategy and strong past performance.
☆☆☆☆	A very good fund with strong management, a sound investment strategy and solid past performance.
☆☆☆	A competently managed fund, but with either an unimpressive or limited performance track record. Potential to improve exists.
☆☆	A fund with a weak investment management capability or strategy, and/or a poor or very limited performance track record.
☆	A poor quality fund with major weaknesses and/or issues affecting the fund's management and performance.

Source: http://funds.comsec.com.au

2 Morningstar categorises managed funds according to the type of securities included in them. The four

Alternative methods of performance evaluation

In this section, the assessment of the performance of individual production units based on the concept of a production frontier is discussed. The concept of a production frontier somewhat reflects desired achievement levels for production units within an industry. So the aim of the individual production unit would be to optimise its efforts to achieve such a level defined by the production frontier. This idea is consistent with the economic theory of optimising behaviour and therefore is a good reason to introduce production frontiers in empirical studies of this nature. While production units typically want to reach the production frontier, in reality, they may fall short due to reasons within, and beyond, their control. This notion of shortfall introduces the concept of inefficiency of production which can be measured.

Efficiency measures

The term 'productive efficiency' is commonly used to describe the level of performance of a production unit in terms of its utilisation of input resources in generating outputs. Koopmans (1951) defined technical efficiency as a feasible input/output vector where it is technologically impossible to increase any output without simultaneously reducing another output. This analogy holds for a reduction in any input or both a reduction in any input and an increase in any output. Farrell (1957) demonstrated that a production unit's 'overall efficiency' is composed of two separate efficiency measures called 'technical efficiency' and 'allocative efficiency'. Farrell measured technical inefficiency as the maximum equi-proportional reduction in all inputs consistent with equivalent production of observed output. A Farrell-efficient unit however, may not be Koopmans-efficient since even after Farrell efficiency is achieved, there may exist additional slack in individual inputs. The efficiency measures are described below for the single-output two-input production function.

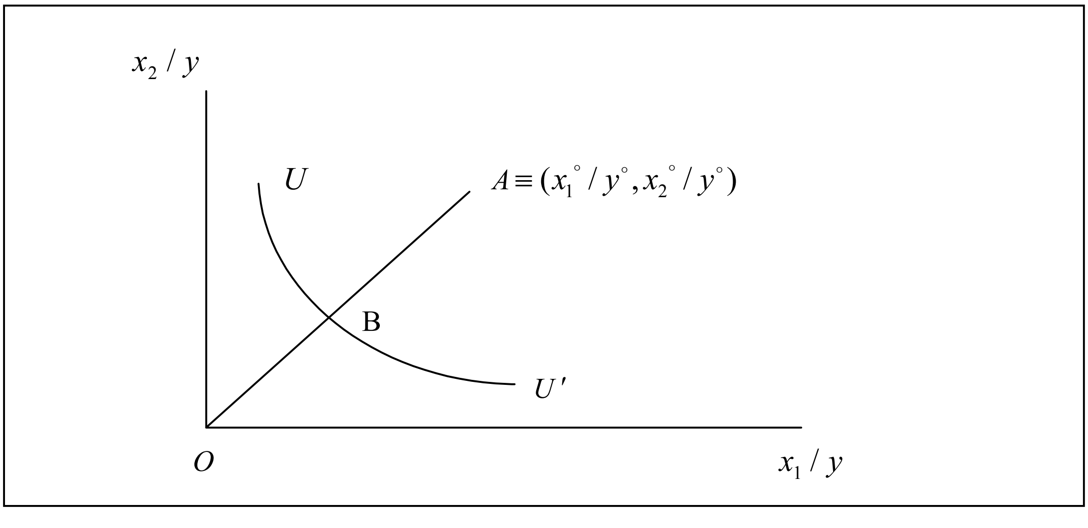
---

categories that they use are: domestic equities, foreign equities, municipal bonds and taxable bonds.

Let x_1 and x_2 denote the two inputs, y the output and $y = f(x_1, x_2)$ the production function. The production function shows the maximum output possible for a given set of inputs, assuming that the firm is technically efficient. Then, assuming that the frontier is characterised by constant returns-to-scale[3] (CRS), it may be written as $1 = f(x_1 / y, x_2 / y)$ a unit (output) isoquant. The unit isoquant may be considered as characterising frontier technology. This is graphically presented as the curve UU' in Figure 2. By definition of the production frontier, any observed point, say $A \equiv (x_1^\circ / y^\circ, x_2^\circ / y^\circ)$, corresponding to a production unit must lie either on or above the unit isoquant. Farrell defined the technical efficiency of production unit *A* as *OB/OA*.

The technical efficiency of production unit *A* may be interpreted as the ratio of inputs needed to produce y° to the inputs actually used to produce y° with input maintained at the same levels of x_1° and x_2°. Therefore, technical efficiency will lie between 0 and 1 inclusive. Any point along the line *OA* will have the same input mix as well. Technical inefficiency results when more output could be produced given the same level of input.

Figure 2. Efficiency frontier



[3] Returns-to-scale refers to how output responds when all input factors are varied in the same proportion.

Allocative efficiency is based on cost considerations, namely, input prices. The type of efficiency measured depends on the data availability and appropriate behavioural assumptions (Yin, 1999). When only quantities are available, technical efficiency can be calculated. When both quantities and prices are available, economic efficiency can be calculated and decomposed into technical and allocative components. We do not consider input prices in the models described in this paper and therefore allocative efficiency measurement is not discussed.

Production frontier estimation

There are two main production frontier estimation methods: parametric and non-parametric. Each method has its own inherent advantages and disadvantages when used in the estimation of production frontiers and individual production unit efficiency. When choosing between these two techniques, usually there is a trade-off between the structure and flexibility.

(a) Parametric methods

Parametric methods are used to estimate the frontier with an explicit functional form given. These types of frontier estimation methods fall under either econometric techniques or stochastic frontier estimation (SFE) methods. The SFE method largely depends on the industry under study as well as data availability. The characteristics of industry and sample data impose restrictions on model specification,[4] which in turn affect the structure and flexibility of the model. An advantage of using the SFE method is that it can handle stochastic noise. However, the requirement of *a priori* (explicit) specification of the production function and assumption of distributions for the error term without regard to the theory are considered as shortcomings in stochastic frontier methods.

[4] Estimating an overly flexible functional form may lose statistical efficiency. On the other hand, the more structure imposed on the model the better the estimates will be.

(b) Non-parametric methods

The methods of estimating the frontier without using an explicit functional form fall under the non-parametric category. One such method is DEA. DEA uses mathematical programming techniques and derives the deterministic frontier instead of estimating it.

Being a non-parametric technique, DEA does not impose any structural form[5], thereby avoiding the danger of misspecification of the frontier. Non-parametric approaches, of course, use less information than parametric approaches and hence the results might be less precise. For DEA to be successful, the data should be assumed to be free from statistical noise. Otherwise, when applying DEA to estimate the technical efficiency at production unit level, inefficiency may include statistical noise as well. In DEA, the production frontier is derived based on sample data and therefore its results could be sensitive to outliers.

A desirable property of the DEA approach is its ability to handle multiple outputs quite easily. Virtually all parametric approaches have been limited to the single output case. This is because the extension of parametric methods for frontier estimation to the multiple output case raises additional theoretical and computational problems (Banker, Conrad and Strauss, 1986).

Variable selection in DEA however, presents problems. The inclusion of many input-output variables is not a viable option in DEA. As the number of variables in the DEA model increases, more and more production units will become efficient. Further, if many variables are used, some of them may be highly correlated and therefore, redundant. On the other hand, when some variables are removed from the DEA model, the production unit efficiency decreases or at most, remains unchanged. There is no standard structured approach to variable selection in DEA. Several methods for variable selection in DEA have been proposed in the literature. For

[5] Recall that the Treynor and Jensen measures are also based on an explicit functional relationship between risk and return.

example, Adler and Golany (2001) suggested using principal component analysis to select a number of variables that are representative of the available data set. Norman and Stoker (1991) proposed a step-wise approach in which they start with a few input-output variables and subsequently add variables to the initial set. Selection of new variables depends on the strength of their correlation with the DEA efficiencies computed using the initial variable set. This is continued until a reasonable set of input-output variables is included. See also Cinca, Molinero and Garcia (2002) for a review of variable selection methods in DEA and a two-stage methodology for variable selection.

(c) Comparison of SFE and DEA performance

The relative superiority of SFE and DEA methods is not just a theoretical issue but also an empirical issue (Gong and Sickles, 1992). Thus, studies comparing the results of the application of SFE and DEA to the same data set emerged. Some of these studies (Banker, Conrad and Strauss, 1986; Bjurek, Hjalmarsson and Forsund, 1990; Whiteman, 1999; Ruggiero and Vitaliano, 1999) contrasted the frontier estimates obtained by the two methods using real-world data. Others used simulated data sets (Banker, Charnes, Cooper and Maindiratta, 1988; Gong and Sickles, 1992; Banker, Gadh and Gorr, 1993; Read and Thanassoulis, 1996). There are advantages associated with working with simulated data, as simulation experiments allow controlling the structure of the underlying technology and the stochastic environment.

The overall findings of these studies are that the efficiency estimates depend, to a large extent, on the choice of the functional form to approximate the underlying production technology and on the measurement methodology employed. The inconsistency of the results with the different techniques makes it imperative that more research is performed to determine the appropriate use of the two measurement methodologies (Craycraft, 1999).

Data envelopment analysis

(a) Methodology

Speaking broadly, the DEA technique defines an efficiency measure of a production unit by its position relative to the frontier of the best performance established mathematically by the ratio of the weighted sum of outputs to the weighted sum of inputs; see, for example, Norman and Stoker (1991) for a detailed description of the DEA technique. The estimated frontier of the best performance is also referred to as efficient frontier, or envelopment surface. The frontier of the best performance characterises the efficiency of production units and identifies inefficiencies based on known levels of attainment. Thus, a production unit attains one hundred per cent efficiency only when it is not found to be inefficient in using the inputs to generate the output when compared with other relevant production units.

In order to motivate the discussion, we begin with the original formulation of the DEA model introduced by Charnes, Cooper and Rhodes (1978), denoted CCR hereafter.

Let us first define the following measures:

$S = \{1,\ldots,s\}$ is the set of outputs considered in the analysis

$M = \{1,\ldots m\}$ is the set of inputs considered in the analysis

y_{rj} = known positive output level of production unit j, $r \in S$

x_{ij} = known positive input level of production unit j, $i \in M$

n = total number of production units evaluated

The CCR model for determining the relative efficiency of a designated production unit 'k' is given as:

$$\text{Max} \left\{ \frac{\sum_{r=1}^{s} u_r y_{rk}}{\sum_{i=1}^{m} v_i x_{ik}} \right\} \tag{11}$$

$$\text{subject to} \qquad \frac{\sum_{r=1}^{s} u_r y_{rj}}{\sum_{i=1}^{m} v_i x_{ij}} \leq 1, \qquad j = 1,2,\ldots,n, \tag{12}$$

$$u_r,\ v_i \geq 0, \qquad r = 1,2,\ldots,s, \text{ and } i = 1,2,\ldots,m. \tag{13}$$

The above formulation assumes constant returns-to-scale (CRS) and the production frontier is a piecewise linear envelopment surface. The variables in the model are the input and output weights u_r and v_i respectively. The objective function (11) is the ratio of the weighted sum of outputs to the weighted sum of inputs of production unit '*k*'. The optimal values of the variables u_r and v_i are determined as a solution to the problem of maximising the efficiency measure of production unit '*k*', subject to the constraint that the efficiency measures of all production units be less than, or equal to, one. The model (11-13) has an infinite number of optimal solutions, since if $\left\{ u_r^*, v_i^* \right\}$ is an optimal solution, then $\left\{ \alpha u_r^*, \alpha v_i^* \right\}$ will also be an optimal solution. One way of avoiding this is to impose the constraint $\sum_{i=1}^{m} v_i x_{ij} = 1$ that results in the following optimisation model:

$$\text{Max} \sum_{r=1}^{s} u'_r y_{rk} \tag{14}$$

$$\text{subject to} \qquad \sum_{r=1}^{s} u'_r y_{rj} - \sum_{i=1}^{m} v'_i x_{ij} \leq 0, \qquad j = 1,2,\ldots,n, \tag{15}$$

$$\sum_{i=1}^{m} v'_i x_{ij} = 1, \tag{16}$$

$$u'_r,\ v'_i \geq 0, \qquad r = 1,2,\ldots,s, \text{ and } i = 1,2,\ldots,m. \tag{17}$$

There is an associated linear programme to the model given in (14-17) called 'the dual'. The optimal solution to one model reveals the optimal solution to the other. Hence, the dual problem, which always has a fewer number of constraints, is the preferred form to handle. The dual of the model given in (14-17) is:

$$\text{Min } \theta \tag{18}$$

subject to

$$\sum_{j=1}^{n} \lambda_j y_{rj} \geq y_{rk}, \qquad r = 1,2,\ldots,s, \tag{19}$$

$$\theta x_{ik} \geq \sum_{j=1}^{n} \lambda_j x_{ij}, \qquad i = 1,2,\ldots,m, \tag{20}$$

$$\lambda_j \geq 0, \qquad j = 1,2,\ldots,n. \tag{21}$$

The variables in the model (18-21) are unrestricted θ and λ_j which is non-negative for all j. The variable θ, as evident in constraint (20), is the proportional reduction in all inputs of the production unit '*k*' required to achieve efficiency. Hence, θ will be the Farrell (technical) efficiency. The constraints in the model ensure that the relative efficiency of unit '*k*' never exceeds 1. The sufficient condition for the efficiency of unit '*k*' is that the optimum value of θ is 1. Otherwise, it is labelled as inefficient compared to the other units in the sample.

The orientation of the model given in (18-21) is an input reduction approach since it provides information on how much proportional reduction of inputs is necessary (while maintaining production levels of output) for an inefficient unit to become DEA-efficient.

Thus far we have discussed CRS models. A measure of efficiency obtained from the solution to model (18-21) therefore, consists of technical as well as scale efficiencies. The variable returns-to-scale (VRS) version of the model (18-21) was proposed by Banker, Charnes and Cooper (1984), hereafter called the BCC model. The BCC model is (18-21) together with the additional constraint,

$$\sum_{j=1}^{n} \lambda_j = 1 , \tag{22}$$

that captures returns-to-scale characteristics. The BCC model measures technical efficiency only. Hence, the efficiency estimates obtained in the BCC model may be considered as "pure" technical efficiency estimates.

A DEA run will produce a relative efficiency score, θ, and a set of λ_j, $j = 1,2,...,n$, values for each production unit. In the DEA literature, the units evaluated are referred to as decision-making units (DMUs). The set of λ_j values of each unit defines a point on the envelopment surface[6] made up of a convex combination of the efficient units. Therefore, for an inefficient unit, the point so defined by the λ_j values becomes a role model that in turn establishes precedence for it to become efficient. The set of efficient production units $\{j: \lambda_j > 0\}$ is called the peer group of the designated unit, '*k*'.

The constraint given in (22) is referred to as the convexity constraint and accounts for VRS. When the convexity constraint is removed the resulting model represents the CRS situation. The relative efficiency score obtained for a designated unit under CRS is a measure of the overall technical efficiency of the unit and is always at least as much as the corresponding value obtained under VRS. The relative efficiency score obtained under VRS is a measure of pure technical efficiency. The difference in overall and pure technical efficiencies is attributed to scale efficiency. A measure of scale efficiency is simply the ratio of overall and pure technical efficiencies.

[6] Efficient units determine a piecewise linear envelopment surface. The entire mean variance frontier also may be generated by linear combinations of any frontier portfolios (Cass and Stiglitz, 1970).

The efficiency of certain production units obtained as the solution to model (18-22) sometimes can be misleading due to what is known as input slack. Input slack results when the section of the linear piecewise frontier used in the measurement of efficiency of a certain unit lies parallel to the axis of measure. Input slack can be obtained from the solution to the model (18-22) by substituting the optimal values of θ (θ^*) and λ_j (λ_j^*) in (20). For the designated unit '*k*', the slack of input '*i*' will be

$$\theta^* x_{ik} - \sum_{j=1}^{n} \lambda_j^* x_{ij} \, . \tag{23}$$

The value of (23) will be either zero or positive. Most studies ignore this and simply solve the BCC model for θ which is the Farrell technical efficiency.

(b) Application

The seminal paper of Charnes, Cooper and Rhodes (1978) introduced the CRS model to measure technical efficiency only. Their model was initially applied to the public sector (Bessent and Bessent, 1980), non-profit institutions (Charnes and Cooper, 1980), and the education sector (Charnes, Cooper and Rhodes, 1981). Later Banker, Charnes and Cooper (1984) extended the CCR model to accommodate the VRS assumption that enables measurement of scale efficiency. This led to the rapid expansion of the application of DEA to a number of areas, including hospitals (Conrad and Strauss, 1983; Nunamaker, 1983), electric utilities (Fare, Grosskopf and Logan, 1983), courts (Levin, Morey and Cook, 1982), agriculture (Fare, Grabowski and Grosskopf, 1985) and marketing (Charnes, Cooper, Learner and Phillips, 1985), to name a few. In the 1990s, DEA became very popular due to significant advances in model development and computational efficiency. See Seiford (1996) for an evolution map that illustrates the growth of DEA in theory and application from 1978 to 1995.

The DEA approach can be problematic when some of the inputs and/or outputs of the decision-making unit (DMU) are stochastic. In situations where the input or output variables of the

DMUs are assumed to be random variables, a number of studies have resorted to analytical approaches where a random component is added to the efficient frontier (Olesen and Petersen, 1995; Retzlaff-Roberts and Morey, 1993; Sengupta, 1987). Premachandra, Powell and Shi (1998) used a DEA-based numerical approach to investigate the relative performance of New Zealand managed portfolios under a stochastic environment. Now, DEA application is becoming more sophisticated and is used as a versatile and effective tool in empirical analysis.

(c) Application in finance

Banking

A growing number of studies on bank branches can be found in the literature in many different countries. See Berger and Humphrey (1997) for a survey of 130 studies that apply frontier efficiency analysis to financial institutions in twenty-one countries. The reason for the rapid growth of such studies was mainly due to intensified competition among major banking players at the local level and their having to operate under different regulatory regimes in foreign markets.

Efficiency measurement techniques generally separate bank branches that perform better, relative to a benchmark, from the others. Since DEA is a relative efficiency measurement technique, the use of DEA to measure bank branch efficiency is now becoming increasingly popular. See, for example, Parkan (1987) for an assessment of the branches of a Canadian chartered bank, Oral and Yolalan (1990) of a Turkish bank, Giokas (1991) of the Greece Commercial Bank, Al-Faraj, Alidi and Bu-Bshait (1993) of a Saudi Arabian bank and Athanassopoulas (1998) of a commercial bank in the United Kingdom.

Insurance

DEA has also been applied to the financial services sector, in particular the insurance industry, although there are only a limited number of studies. See, for example, Berger and Humphrey

(1997) for a survey of eight studies in the US, France and Italy and Worthington and Hurley (2000) for a study of a sample of Australian general insurers.

In general, most DEA applications in the banking and insurance sectors concentrated on US financial institutions. The Berger and Humphrey (1997) survey reported that of the 116 single country studies, US financial institutions accounted for 66 of these.

Securities

Powers and McMullen (2000) applied the DEA technique with weight restrictions to distinguish between strong performers and others in a set of financial securities. Weight restrictions are generally imposed to avoid production units achieving efficiency while having undesirable input-output levels (Thompson, Langemeier, Lee, Lee and Thrall, 1990; Wong and Beasley, 1990). They argued that security selection could be thought of as a multi-criteria decision-making problem since security selection is usually based on an examination of several attributes. Considering 1-, 3-, 5-, and 10-year average returns and earnings per share as output variables, and price to earnings ratio, beta risk and 3-year standard deviation of returns as input variables, Powers and McMullen estimated the DEA-efficiency of 185 of the largest market cap securities in the US. They highlighted that DEA is able to (i) provide a single composite score for each security, (ii) inform the decision-maker as to which securities are consistently the best when several attributes are considered and (iii) provide information as to how much improvement is needed for each security to become efficient with respect to given inputs and outputs.

Managed Funds

Investment performance measures such as the Sharpe, Treynor and Jensen indices can be used to evaluate the risk-return performance of managed funds based on the risk-adjusted return or its variations. Murthi, Choi and Desai (1997), were the first to apply DEA to mutual fund appraisal. Motivated by this application and the results, they argued that the superiority of DEA over the above three indices comes from the fact that DEA can accommodate important variables such as transaction costs, while the indices do not make use of such information. Another drawback of the Treynor and Jensen indices is the requirement of a benchmark[7] for performance comparisons.

Transaction costs include loads and/or other fees that financial institutions charge investors for their expertise and for conducting financial transactions on their behalf. Murthi, Choi and Desai (1997) modified the idea of the Sharpe index by incorporating transaction costs. Their index, denoted by *I*, is expressed as:

$$I = \frac{R}{\sum_{i=1}^{n} w_i X_i + v\sigma} \tag{24}$$

where, R is the excess return, σ is the standard deviation of returns, n is the number of components of the total transaction costs and X_i is the transaction costs associated with the cost component ***i***. w_i and v are the weights associated with variables X_i and σ. The index *I* is interpreted as the excess return after controlling for the level of risk of the investment and the expenses incurred through transactions.

The weights $w_i : i = 1,...,n$ and v can be determined by employing a parametric approach and specifying a functional form for the association between the output variable R and input variables $X_i : i = 1,...,n$, and σ. However, acknowledging the criticism of Varian (1990) for using parametric specifications here, Murthi, Choi and Desai (1997) employed DEA to appraise 731 mutual funds using the actual return as the output variable and four input variables: expense ratio (accounts for management fees, marketing expenses and other operational expenses), load (a charge at the time of investment and/or withdrawal also referred to as sales charge), turnover

[7] Grinblatt and Titman (1993) introduced a measure that does not require the use of a benchmark. However, they failed to account for transaction costs.

(captures the trading activity of the fund manager proxied by min{monthly purchases, sales}/average net asset value) and the standard deviation of returns.

Murthi, Choi and Desai (1997) found strong evidence that mutual funds are approximately mean-variance efficient and that efficiency is not related to transaction costs. However, their study assumed a CRS frontier and therefore was unable to examine the issue of scale effects on the mutual funds.

McMullen and Strong (1998), on the other hand, analysed 135 common stock mutual funds using DEA. Their choice of the input-output variable set differed slightly from that of Murthi, Choi and Desai (1997). McMullen and Strong postulated that an investor's choice of a mutual fund would be typically a function of recent performance, long-term performance, the associated risks of these returns and transaction costs. In particular, they considered 1-, 3- and 5-year annualised returns as output variables and sales charge, expense ratio, minimum initial investment and standard deviation of return measured over three years as the input variables.

Apart from the choice of the input-output variable set, the McMullen and Strong (1998) study differed from Murthi, Choi and Desai (1997) in two other aspects. These are: (i) relaxing the CRS assumption and (ii) imposing weight restrictions on the input-output variables. McMullen and Strong demonstrated that DEA results could assist investors to decide which funds to buy or not to buy, by providing them with reasons.

Sedzro and Sardano (1999) analysed 58 US equity funds in Canada using DEA. Their study differs from McMullen and Strong (1998) in two aspects: (i) the use of another proxy (Vos ratio[8]) for risk, different from the usual standard deviation of returns and (ii) comparison of the DEA results with three other performance measures – the Morningstar rating, the Sharpe index

[8] Vos ratio is an ordinal classification of funds based on the numerical evaluation of five variables that capture several dispersion measures.

and the Vos ratio (Vos, 1997). Sedzro and Sardano (1999) treat annual return as the output variable and expenditure ratio, minimum initial investment and inverse of Vos risk measure[9] as the input variables.

Sedzro and Sardano (1999) reported that DEA yields results similar to those of the Sharpe, Vos and Morningstar measures, and through critical examination of the DEA results emphasised the advantage of using DEA over the other measures. In particular they highlighted the possibility of identifying the causes for the under-performance of inefficient funds.

Morey and Morey (1999) addressed the issues of integrating fund performance over different time horizons and identification of dominant funds. They suggested a method of eliminating subjectivity in the selection of weights in the integration of fund performance over different time horizons by adopting a DEA-based approach. Premachandra, Powell and Shi (1998) on the other hand, proposed a spreadsheet-based stochastic DEA model for ranking a set of portfolios created by mixing three alternative investments, namely, securities in the New Zealand stock exchange, the NZSE40 index and a risk-free asset.

Conclusions

In this paper, we reviewed the literature on investment performance appraisal methods. Early measures were based on some form of risk-adjusted return and did not include the costs incurred in generating the return. Later, companies developed comprehensive performance evaluation methods that have now become very popular among institutional and private investors. Some methods award 'stars' to managed funds based on their risk-adjusted returns after due consideration given for costs such as sales charges. A feature of the Morningstar company-sponsored measures is that they publish ratings separately for different groups of funds. The groups are categorised on the basis of fund characteristics.

[9] The more risky the fund the higher the inverse of Vos risk measure.

The choice of the variable set in DEA models is generally more of an empirical issue. Apart from freedom of choice in the input-output variable set the analyst will have to select the appropriate DEA model from a wide range available in the literature. In short, model selection is a major issue in DEA. Therefore, it will be beneficial to the analyst to be aware of the dangers of model misspecification in DEA. Misspecification in DEA models can result due to the omission of relevant variables, inclusion of irrelevant variables and incorrect assumption on returns-to-scale. A few studies have addressed some of these issues under some specific circumstances. All of them investigated a variety of model misspecifications under different production processes using simulation studies. DEA performance can also be sensitive to the choice of the sample size and the number of, and the association among, the variables used in the model.

References

Adler, N. and Golany, B. (2001). Evaluation of deregulated airline networks using data envelopment analysis combined with principal component analysis with an application to Western Europe, *European Journal of Operational Research*, 132, 260-273.

Al-Faraj, T.N., Alidi, A.S. and Bu-Bshait, K.A. (1993). Evaluation of bank branches by means of data envelopment analysis, *International Journal of Operations and Production Management*, 13, 45-52.

Athanassopoulas, A. (1998). Nonparametric frontier models for assessing the market and cost efficiency of large-scale bank branch networks, *Journal of Money, Credit and Banking*, 30, 300-313.

Banker, R.D., Charnes, A. and Cooper, W.W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis, *Management Science*, 30, 1078-1092.

Banker, R. D., Charnes, A., Cooper, W.W. and Maindiratta, A. (1988). A comparison of data envelopment analysis and translog estimates of production frontiers using simulated observations from known technology, in A. Dogramaci and R. Fare (eds.), *Applications of Modern Production Theory: Efficiency and Productivity*, Kluwer Academic Publishers: Boston, 33-54.

Banker, R.D., Conrad, R.F. and Strauss, R.P. (1986). A comparative application of data envelopment analysis and translog methods: an illustrative study of hospital production, *Management Science*, 32, 30-43.

Banker, R.D., Gadh, V.M. and Gorr, W.L. (1993). A Monte Carlo comparison of two production frontier estimation methods: corrected ordinary least squares and data envelopment analysis, *European Journal of Operational Research*, 67, 332-343.

Berger, A.N. and Humphrey, D.B. (1997). Efficiency of financial institutions: international survey and directions for future research, *European Journal of Operational Research*, 98, 175-212.

Bessent, A.M. and Bessent, E.W. (1980). *Determining the comparative efficiency of schools through data envelopment analysis,* Education Administration Quarterly*, 16, 57-75.*

Bjurek, H., Hjalmarsson, L. and Forsund, F.R. (1990). Deterministic parametric and nonparametric estimation of efficiency in service production, *Journal of Econometrics*, 46, 213-227.

Cass, D. and Stiglitz, J. (1970). The structure of investor preferences and asset returns, and separability in portfolio allocation: a contribution to the pure theory of mutual funds, *Journal of Economic Theory*, 2, 122-160.

Charnes, A. and Cooper, W.W. (1980). *Auditing and accounting for program efficiency and management efficiency in not-for-profit entities,* Accounting, Organizations and Society*, 5, 87-107.*

Charnes, A., Cooper, W.W., Learner, D.B. and Phillips, F.Y. (1985). Management science and marketing management, *Journal of Marketing*, 49, 93-105.

Charnes, A., Cooper, W.W. and Rhodes, E. (1978). Measuring the efficiency of decision making units, *European Journal of Operational Research*, 2, 429-444.

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

Cinca, C.S., Molinero, C.M. and Garcia, F.C. (2002). On model selection in data envelopment analysis: A multivariate statistical approach, Discussion Paper No. M02-7, Department of Management, University of Southampton, United Kingdom.

Conrad, R.F. and Strauss, R.P. (1983). A multiple-output multiple-input model of the hospital industry in North Carolina, *Applied Economics*, 341-352.

Craycraft, C. (1999). A review of statistical techniques in measuring efficiency, *Journal of Public Budgeting, Accounting & Financial Management*, 11, 19-27.

Fare, R., Grabowski, R. and Grosskopf, S. (1985). Technical efficiency of Philippines agriculture, *Applied Economics*, 17, 205-214.

Fare, R., Grosskopf, S. and Logan, J. (1983). The relative efficiency of Illinois Electric Utilities, *Resources and Energy*, 5, 349-367.

Farrell, M.J. (1957). The measurement of productive efficiency, *Journal of the Royal Statistical Society*, 120, 253-290.

Giokas, D. (1991). Bank branch operating efficiency: a comparative application of DEA and the loglinear model, *Omega*, 19, 549-557.

Gong, B. and Sickles, R.C. (1992). Finite sample evidence on the performance of stochastic frontiers and data envelopment analysis using panel data, *Journal of Econometrics*, 51, 259-285.

Grinblatt, M. and Titman, S. (1993). Performance measurement without benchmarks: an examination of mutual fund returns, *Journal of Business*, 66, 47-68.

Jensen, M.C. (1969). Risk, the pricing of capital assets and the evaluation of investment performance, *Journal of Business*, 42, 167-247.

Koopmans, T.C. (1951). An analysis of production as an efficient combination of activities, in T.C. Koopmans (ed.), *Activity Analysis of Production and Allocation*, Wiley: New York, 33-97.

Levin, A.Y., Morey, R.C. and Cook, T.J. (1982). Evaluating the administrative efficiency of courts, *Omega*, 10, 401-411.

McMullen, P.R. and Strong, R.A. (1998). Selection of mutual funds using data envelopment analysis, *Journal of Business and Economic Studies*, 4, 1-12.

Modigliani, F. and Modigliani, L. (1997). Risk-adjusted performance, *Journal of Portfolio Management*, 23, 45-54.

Morey, M.R. and Morey, R.C. (1999). Mutual fund performance appraisals: a multi-horizon perspective with endogenous benchmarking, *Omega*, 27, 241-258.

Murthi, B.P.S., Choi, Y.K. and Desai, P. (1997). Efficiency of mutual funds and portfolio performance measurement: a non-parametric approach, *European Journal of Operational Research*, 98, 408-418.

Norman, M. and Stoker, B. (1991). *Data Envelopment Analysis: The Assessment of Performance*, Wiley: New York.

Nunamaker, T.R. (1983). Measuring routine nursing service efficiency: a comparison of cost per patient day and data envelopment analysis models, *Health Science Research*, 18, 183-205.

Olesen, O.B. and Petersen, N.C. (1995). Chance constrained efficiency evaluation, *Management Science*, 41, 442-457.

Oral, M. and Yolalan, R. (1990). An empirical study on measuring operating efficiency and profitability of bank branches, *European Journal of Operational Research*, 46, 282-294.

Parkan, C. (1987). Measuring the efficiency of service operations: an application to bank branches, *Engineering Costs and Production Economics*, 12, 237-242.

Powers, J. and McMullen, P.R. (2000). Using data envelopment analysis to select efficient large market cap securities, *Journal of Business and Management*, 7, 31-42.

Premachandra, I.M., Powell, J.G. and Shi, J. (1998). Measuring the relative efficiency of fund management strategies in New Zealand using a spreadsheet based stochastic data envelopment analysis model, *Omega*, 26, 319-331.

Read, L.E. and Thanassoulis, E. (1996). A comparison of data envelopment analysis (DEA) and stochastic frontiers (SF) under variation of fit, Warwick Business School Research Paper 234, Warwick Business School Research Bureau.

Retzlaff-Roberts, D.L. and Morey, R.C. (1993). A goal programming method of stochastic allocative data envelopment analysis, *European Journal of Operational Research*, 71, 379-397.

Ruggiero, J. and Vitaliano, D.F. (1999). Assessing the efficiency of public schools using data envelopment analysis and frontier regression, *Contemporary Economic Policy*, 17, 321-331.

Sedzro, K and Sardano, D. (1999). Mutual fund performance evaluation using data envelopment analysis, Working Paper, School of Business, University of Quebec at Montreal, Canada.

Seiford, L.M. (1996). Data envelopment analysis: the evolution of the state of the art (1978-1995), *Journal of Productivity Analysis*, 7, 99-137.

Sengupta, J.K. (1987). Data envelopment analysis for efficiency measurement in the stochastic case, *Computers and Operations Research*, 14, 117-129.

Sharpe, W.F. (1966). Mutual fund performance, *Journal of Business*, 39, 119-138.

Sharpe, W.F. (1998). Morningstar's risk-adjusted ratings, *Financial Analysts Journal*, 54, 21-33.

Thompson, R.G., Langemeier, L.N., Lee, C., Lee, E. and Thrall, R.M. (1990). The role of multiplier bounds in efficiency analysis with application to Kansas farming, *Journal of Econometrics*, 46, 93-108.

Treynor, J.L. (1965). How to rate management of investment funds, *Harvard Business Review*, 44, 131-136.

Varian, H. (1990). Goodness-of-fit in optimising models, *Journal of Econometrics*, 46, 125-140.

Vos, W. (1997). Measuring mutual fund performance, *Canadian Investment Review*, 10, 33-37.

Whiteman, J.L. (1999). The potential benefits of Hilmer and related reforms: electricity supply, *Australian Economic Review*, 32, 17-30.

Wong, Y.B. and Beasley, J.E. (1990). Restricting weight flexibility in data envelopment analysis, *Journal of the Operational Research Society*, 41, 829-835.

Worthington, A.C. and Hurley, E.V. (2000). Technical, allocative and cost efficiency in the Australian general insurance industry, Discussion Paper No. 74 in Economics, Finance and International Competitiveness, School of Economics and Finance, Queensland University of Technology, Australia.

Yin, R. (1999). Production efficiency and cost competitiveness of pulp producers in the Pacific Rim, *Forest Products Journal*, 49, 43-49.