ABSTRACT

This study proposes the design of a multi-attribute-decision-support-system that combines the analytical power of two different tools: data envelopment analysis (DEA) and particle swarm optimization (PSO), one of the major algorithms using swarm intelligence. DEA evaluates and measures the relative efficiency of decision making units that use multiple inputs and outputs to provide non-objective measures without making any specific assumptions about data. On the other hand PSO’s main strength lies in exploring the entire search space. This study proposes a modeling technique that jointly uses the two techniques to benefit from the two methodologies.

Keywords: Swarm Intelligence, Particle swarm Optimization, Data Envelopment Analysis

INTRODUCTION

A business organization’s objective is to make better decisions at all levels of the firm to improve performance. Typically organizations are multi-faceted and complex systems that use uncertain information. Therefore, making quality decisions to improve organizational performance is a daunting task. Organizations use decision support systems that apply different business intelligence techniques such as statistical models, scoring models, neural networks, expert systems, neuro-fuzzy systems, case-based systems, or simply rules that have been developed through experience. Managers need a decision-making approach that is robust, competent, effective, efficient, and integrative to handle the multi-dimensional organizational entities. The decision maker deals with multiple players in an organization such as products, customers, competitors, location, geographic structure, scope, internal organization, and cultural dimension (Porter, 1980). Sound decisions include two important concepts: efficiency (return on invested resources) and effectiveness (reaching predetermined goals). However, quite frequently, the decision maker cannot simultaneously handle data from different sources. Hence, we recommend that managers analyze different aspects of data from multiple sources separately and integrate the results of the analysis. This study proposes the design of a multi-attribute-decision-support-system that combines the analytical power of two different tools: data envelopment analysis (DEA) and particle swarm optimization (PSO), one of the major algorithms using swarm intelligence. DEA evaluates and measures the relative efficiency of decision making units that use multiple inputs and outputs to provide non-objective measures without making any specific assumptions about data. On the other hand PSO’s main strength lies in exploring the entire search space. This study proposes a modeling technique that jointly uses the two techniques to benefit from the two methodologies. A major advantage of the DEA approach is that it clearly identifies the important factors contributing to the success of a decision. In addition, I also propose the use of a particle swarm optimization to assess the global minima or maxima that can aid the decision-maker in making decisions regarding the
implications of a decision. One of the important characteristics of population-based search algorithms is their ability to improve the exploration of the search space without falling in the pitfalls of local minima and maxima. The objective of this study is to find a minima solution (ideal loan) nearest to a given loan application using a limited number of iterations. However, not all solutions found using PSO may be true minima so we use the DEA algorithm, a benchmarking technique, to assess the worthiness of the closest optimal solution. The DEA model benchmarks the performance value of the loan application (both original and optimized) against a set of previous loans. Thus, a decision maker can easily analyze and understand any decision using the power of the efficiency frontier algorithm (DEA) and global search algorithm (PSO) to analyze the credit-worthiness of a loan application. The rest of the paper is organized along the following lines. In section II, we provide a review of previous studies on loan evaluation. Section III discusses the model that we use in this study. Section IV discusses the data, methodology, and empirical analysis of our results. Section V summarizes and concludes our study.

LITERATURE REVIEW

Swarm intelligence literature:

Swarm intelligence, based on collective artificial intelligence, is an emerging area in the field of optimization. Many researchers have developed algorithms that model the behavior of different swarm of animals and insects such as ants, termites, bees, birds, fishes, and elephants. In 1990s, researchers introduced two important algorithms – ant colony optimization (Dorigo et. al., 1991) and particle swarm optimization (Kennedy & Eberhart, 1995) based on fish schooling and bird flocking. PSO mimics the movement of birds in a flock sharing information with each other and the way they interact with each other (Acan and Gunay, 2005) defined by topology. The birds in the swarm represent parameter samples called particles. Birds fly around randomly, but keep an eye on others to follow the bird closest to food. Similarly, each particle in the swarm keeps track of its own best solution found so far and shares the information with topological neighbors to fly towards optimal solutions (Brits et. al., 2007). These algorithms have inspired many researchers to create new versions to solve problems in different areas. Researchers have used these models to solve difficult real world problems such as traffic routing, networking, games, industry, robotics, economics, and design of artificial self-organized distributed problem-solving devices. One of the challenges in PSO is to locate global optima without getting trapped in local optima (Hendtlass, 2003). Researchers suggest different topologies to improve the success rate of finding global optima (Brits et. al., 2007). Eberhart et. al. (1996) define gbest and lbest topologies for the original PSO. On the other hand, Kennedy and Mendes (2002) illustrate von Neumann topology, and Suganthan (1999) proposes spatial topology. Similarly, Brits et. al. (2002a) NichePSO, a multi-modal optimization algorithm that employs the Guaranteed Convergence PSO (GCPSO) (van der Bergh and Engelbercht, 2002) to improve local convergence while maintaining the diversity of particles by forming subswarms. Li (2004) introduced the Species-based algorithm (SPSO) to alleviate the shortcomings of the k-means algorithm. Likewise, over the last decades, the intelligent behavior of bee swarm have inspired researchers to develop new algorithms. Abbas (2001) proposed a new swarm intelligence technique, Honey Bee Mating Optimization (HBMO) based on marriage in honey bees. This model simulates the evolution of honey-bees starting with a solitary colony (single queen without a family) to the emergence of an eusocial colony (one or more queens with a family). Abbas applied the model to a fifty propositional satisfiability problem (SAT) with 50 variables and 215 constraints and 3-SAT problem (Abbas 2001a, b). Since then, researchers have proposed many variations of bee swarm intelligence. Jung (2003) proposed an evolution method called queen-bee
evolution simulating the queen-bee role in reproduction process. Lu and Zhou (2008b) developed Bee Collective Pollen Algorithm by simulating the honeybees’ collective pollen as a global convergence searching algorithm. Researchers have used these algorithms to solve the traveling sales person problem using heuristical approach such as the ant colony optimization (Dorigo & Gambardella, 1997), honey bee mating optimization (Marinakis & Marinakis, 2009), and particle swarm optimization (Pang et al., 2004; Lope & Coelho, 2005; Shi et al., 2007).

The basic PSO and its variation has also been used to solve problems such as bin-packing problem (Liu et al., 2008) and flowshop sequencing (Tasgetiren et al., 2007; Tseng & Liao, 2008). Likewise, researchers have also used ant optimization algorithm to develop a hybrid routing protocol such as works by Rajgopalan & Shen, 2006, Baras & Mehta, 2003, and Camara & Loureiro, 2000) among others. Similarly, HBMO has been successfully applied to solve problems such as partitioning and design of embedded systems (Koudil et al., 2007) and cluster analysis (Fathian et al., 2007), just to name a few. These algorithms have found many industrial applications. For example, US manufacturer of America Air Liquids has achieved significant financial savings by using a computer model based on algorithms inspired by the foraging behavior of ants to analyze every permutation of plant scheduling, gas prices, weather and truck movements. Other applications include telecoms data routing and delivery vehicle fleet scheduling (Bogue, 2008). It is only recently, researchers have started using swarm intelligence to financial service industry. Nenortaite & Simutis (2006) use an intelligent decision-making model to calculate one day forward decision for purchase of the stocks. The decision model is based on the application of artificial neural networks and swarm intelligence technology, PSO. Capital One, the giant financial-services company, has replaced its rigid command-and-control management style with a more flexible approach better suited to a fast-growing business. Like a medium-size ant colony whose territory is invaded by a larger competitor, Capital One constantly searches out and targets new market opportunities. To encourage employees to look for opportunities outside their immediate departments, Capital One revamped its employee evaluation system to reward people who actively search for such “food sources.” (Freidman, Catherine, 2003) Martens et al., (2010) illustrate the use of Ant Miner+ to build internal rating systems for credit risk. They use Ant Miner+ to infer a propositional rule set from a given data set, hereby using the principles from Ant Colony Optimization.

As illustrated above, while swarm intelligence has been investigated vastly by the optimization community, not many studies have applied this new technique to financial services industry.

Data envelopment analysis literature:

Recently, many studies have illustrated the use of DEA, a non-parametric methodology to analyze different aspects of business entities. The details of the DEA model are discussed in the next section. In contrast to other methodologies, DEA is one of the methods that have traditionally been used to assess the comparative efficiency of homogenous operating units such as schools, hospitals, utility companies, sales outlets, prisons, and military operations. More recently, it has been applied to banks (Haslem, Scheraga, & Bedingfield, 1999) and mutual funds (Haslem & Scheraga, 2003; Galagedera & Silvapulle, 2002; McMullen & Strong, 1998; Murthi, Choi, & Desai, 1997).

Murthi, Choi, & Desai (1997) examine the market efficiency of the mutual fund industry by different investment objectives. They use a benefit/cost non-parametric analysis where a relationship between return (benefit) and expense ratio, turnover, risk, and loads (cost) is established. They also develop a measure of performance of mutual funds that has a number of advantages over traditional indices. The DEA portfolio efficiency index (DEPI) does not require specification of a benchmark, but incorporates
transaction costs. The most important advantage of DEA method as compared to other measures of fund performance is that DEA identifies the variables leading to inefficiencies and the levels by which they should be changed to restore the fund to its optimum level of efficiency. McMullen and Strong (1998) applied DEA to evaluate the relative performance of 135 US common stock funds using one, three, and five-year annualized returns, standard deviation of returns, sales charge, minimum initial investment, and expense ratio. They illustrate that DEA can assist in selecting mutual funds for an investor with a multifactor utility function. The DEA selects optimum combinations of investment characteristics, even when the desired characteristics are other than the two-factors specified in Capital Market Theory. The DEA enable the user to determine the most desirable alternatives, and pinpoint the inefficiencies in a DEA-inefficient alternative. Sedzro and Sardano (1999) analyzed 58 US equity funds in Canada using DEA with annual return, expense ratio, minimum initial investment and a proxy for risk as factors associated with fund performance. Further, they also find a strong relationship among the efficiency rankings using DEA, Sharpe ratios, and Morningstar data. Galagedera and Silvapulle (2002) use DEA to measure the relative efficiency of 257 Australian mutual funds. The further investigate the sensitivity of DEA efficiency to various input-output variable combinations. They find that more funds are efficient when DEA captures a fund’s long-term growth and income distribution than a shorter time horizon. In general, the overall technical efficiency and the scale efficiency are higher for risk-averse funds with high positive net flow of assets.

Haslem and Scheraga (2003) use DEA to identify efficiencies in the large-cap mutual funds in the 1999 Morningstar 500. They identify the financial variables that differ significantly between efficient and inefficient funds, and determine the nature of the relationships. They use Sharpe index as the DEA output variable. They find that the input/output and profile variables are significantly different between the Morningstar 500 (1999) large-cap mutual funds that are DEA performance-efficient and inefficient. Basso and Funari (2001) propose the use of DEA methodology to evaluate the performance of mutual funds. The proposed DEA performance indexes for mutual funds represent a generalization of various traditional numerical indexes that can take into account several inputs and outputs. They propose two classes of DEA indexes. The first class generalizes the traditional measures of evaluation using different risk indicators and subscription and redemption costs that burden the fund investment. The second class of indexes considers a multiple inputs-outputs structure. Thus, they monitor not only the mean return but also other features such as stochastic dominance and the time lay-out. Morey and Morey (1999) present two basic quadratic programming approaches for identifying those funds that are strictly dominated, regardless of the weightings on different time horizons being considered, relative to their mean returns and risks. They present a novel application of the philosophy of data envelopment analysis that focuses on estimating “radial” contraction/expansion potentials. Furthermore, in contrast to many studies of mutual fund’s performance, their approach endogenously determines a custom-tailored benchmark portfolio to which each mutual fund’s performance is compared. Feroz, Kim, and Raad (2003) illustrate the use of data envelopment analysis to evaluate the financial performance of oil and gas industry. Edirisinghe and Zhang (2007) develop a data envelopment analysis model to evaluate a firm’s financial statements over time in order to determine a relative financial strength indicator that can predict firm’s stock price returns.

in Credit Unions in Ontario, Canada. Ozcan and McCue (1996) use data envelopment analysis for measuring and assessing the financial performance for hospitals. They compute a financial performance index (FPI) as a measure of aggregate financial performance. They show that financial performance index across many financial ratios eases the comparison of an individual hospital with its peers. Halkos and Salamouis (2004) explore the efficiency of Greek banks with the use of a number of suggested financial efficiency ratios for the time period 1997-1999. They show that data envelopment analysis can be used as either an alternative or complement to ratio analysis for the evaluation of an organization's performance. The study finds that the higher the size of total assets the higher the efficiency. Neal (2004) investigates X-efficiency and productivity change in Australian banking between 1995 and 1999 using data envelopment analysis and Malmquist productivity indexes. It differs from earlier studies by examining efficiency by bank type, and finds that regional banks are less efficient than other bank types. The study concludes that diseconomies of scale set in very early, and hence are not a sufficient basis on which to allow mergers between large banks to proceed. Paradi and Schaffnit (2004) evaluate the performance of the commercial branches of a large Canadian bank using data envelopment analysis. Chen, Sun, and Peng (2005) study the efficiency and productivity growth of commercial banks in Taiwan before and after financial holding corporations' establishment. They employ a data envelopment analysis approach to generate efficiency indices as well as Malmquist productivity growth indices for each bank. Howland and Rowse (2006) assess the efficiency of branches of a major Canadian bank by benchmarking them against the DEA model of American bank branch efficiency. Sufian (2007) uses DEA approach to evaluate trends in the efficiency of the Singapore banking sector. The paper uses DEA approach to distinguish between technical, pure technical and scale efficiencies.

Sanjeev (2007) evaluates the efficiency of the public sector banks operating in India for a period of five years (1997-2001) using DEA. The study also investigates if there is any relationship between the efficiency and size of the banks. The results of the study suggest that no conclusive relationship can be established between the efficiency and size of the banks. Lin, Shu, and Hsiao (2007) study the relative efficiency of management in the Taiwanese banking system through DEA. The goal is to estimate the competitiveness of each bank and managerial efficiency is to show the efficiency variation of each bank through Malmquist index. Bergendahl and Lindblom (2008) develop principles for an evaluation of the efficiency of a savings bank using data envelopment analysis as a method to consider the service orientation of savings banks. They determine the number of Swedish savings banks being "service efficient" as well as the average degree of service efficiency in this industry. Hoon and Chunyan (1994) analyzed the productive efficiency of the railway services in 19 Organization for Economic Cooperation and Development (OECD) countries. They report that railway systems with high dependence on public subsidies are less efficient than similar railways with less dependence on subsidies. Cowie and Riddington (1996) evaluate the efficiency of the European railways through the use of a production frontier approach. Yu and Lin (2008) uses a multi-activity network DEA model to simultaneously estimate passenger and freight technical efficiency, service effectiveness, and technical effectiveness for 20 selected railways for the year 2002. Lozano & Gutierrez (2011) illustrate the slacks-based measure of efficiency of 39 Spanish airports using DEA. Liu & Liu (2010) illustrate the use of DEA in evaluating and ranking the research and redevelopement performance of Taiwan’s government-supported research institutes. Saranga & Moser (2010) develop a comprehensive performance measurement framework using the classical and two-stage Value Chain Data Envelopment Analysis model.

As illustrated above, none of the studies illustrate the merger of swarm intelligence and DEA to develop an intelligent decision support system that benefits from the optimal solution finding capability of two diverse techniques. Therefore, the purpose of this study is twofold. Firstly, this study uses DEA to develop a benchmark using actual loan data. The decision support system further uses PSO to find the
global optimal solution beginning with a new loan. The study further validates the PSO solution by benchmarking the optimal loan parameters against the DEA loan base. Thus, the study aims to illustrate the use of PSO systems as a decision-making tool to understand the credit-worthiness of a loan application.

**METHODOLOGY**

This section illustrates the PSO algorithm and the DEA model. The first part illustrates the models in non-technical terms, and the second part presents the mathematical details of the methodology.

**Particle swarm optimization – a swarm intelligence technique:**

Swarm is the collective behaviour of decentralized, self-organized systems, natural or artificial. The concept is employed in work on artificial intelligence. The expression was introduced by Gerardo Beni and Jing Wang in 1989, in the context of cellular robotic systems (Beni, et. al., 1989). SI systems typically consist of a set of simple entities co-existing with each other and their environment. These simple agents follow very simplistic rules. Although, there is no centralized control structure that govern how these individual entities should behave. However, they learn from local interactions leading to the emergence of a complex, globalized behavior. A distinguishing characteristics of SI systems is that, by lacking any hierarchical command and control structure, there is no common-mode failure-point or vulnerability (Bogue, 2008). Some natural examples of SI systems inclue foraging ant colonies, honey bee mating, bird flocking, animal herding, bacteria molding, and fish schooling. Particle Swarm Optimization algorithm is a SI technique that imitates humans (or insects) social behavior. Individuals interact with one another while learning from their own experience, and gradually the population members move into better regions of the problem space (Kennedy & Eberhart, 1995). This technique originates from two separate concepts: the idea of swarm intelligence based off the observation of swarming habits by certain kind of animals (such as birds and fish); and the field of evolutionary optimization. In PSO, each particle represents a parameter sample, and the swarm consists of a population of particles. Particles in the swarm share their information with topological neighbors to move around the search space toward optimal solutions.

**The data envelopment analysis model:**

Data Envelopment Analysis (DEA) (Charnes et al., 1978) model uses linear programming to measure the comparative performance of different organizational units. Further, this generalized optimization technique measures the relative performance of different decision-making entities that have multiple objectives (outputs) and multiple inputs structure. In the DEA terminology, entities/organization units under study are called Decision-Making Units (DMUs). In our study, the DMUs are the loan applications under analysis. DEA measures the efficiency with which a DMU uses the resources available (inputs) to generate a given set of outputs. The DEA methodology assesses the performance of the DMU using the concept of efficiency or productivity, defined as a ratio of total outputs to total inputs. Further, the DEA model estimates relative efficiency, which is with reference to the best performing DMU or DMUs (in case multiple DMUs are most efficient). The DEA allocates an efficiency score of unity or 100 percent to the most efficient unit. The low-performing DMUs’ efficiency can vary between 0 and 100 percent in comparison to the best performance.
ILLUSTRATING SIDE MODEL OF THE DECISION SUPPORT SYSTEM FOR LOAN EVALUATION:

To screen consumer loan applications, loan officers use different methods besides intuitive judgment and experience. Using mathematical techniques, many credit-scoring models have been developed to assist the loan officer in differentiating good loans from bad. Besides these traditional statistical models, many financial institutions use artificial intelligence methods, such as expert systems, artificial neural systems, and fuzzy logic. It is only recently that the finance community has started applying data envelopment analysis, a relatively new technique. This study proposes to assess the creditworthiness of a new loan applicant using a decision support system that applies a combination of two diverse analytical techniques: data envelopment analysis and particle swarm optimization. I propose to use the Data Envelopment Analysis methodology to assess the creditworthiness of an existing set of loans whose outcome (accepted and turned good, accepted and turned bad, and reject) is known. The DEA model benchmarks the given set of loans, and assigns an efficiency score.

Each of the loans is a homogenous unit, and we can apply the DEA methodology to assess comparative performance of these loans. The DEA model is a part of a decision support system that uses a number of variables to determine how good a loan is. A loan application includes information such as the applicant’s age, housing, address time, total income, number of credit cards, number of dependents, job time, other loan obligations, total debt, monthly rent/mortgage payments, number of inquiries for an applicant, and credit rating. The study creates a DEA model that evaluates the relative efficiency of a set of loans that credit unions have already administered, and allocates a score on the scale of 1 to 100. Further, the next step in the design of the decision support system is to use particle swarm optimization algorithm to map a new loan application to verify how close the loan parameters are to an optimal set of values. Once the PSO model comes up with the optimal loan parameters, the loan officer can discern how close the given loan’s variables are to an ideal situation for the given weights on different parameters. To validate the results of PSO, we further benchmark the ideal loan parameters with the given DEA model to check if its 100% efficient, and the PSO algorithm converges to optimal solution.

DATA AND METHODOLOGY

According to Standard & Poor’s industry survey, liquidity, inventory, and profit margin are critical to a retailer’s success. I Therefore, to study the performance of the retail industry (that includes Wal-Mart, Target, Costco, Macys, Sears, J.C. Penney, and BJ Wholesale), we consider seven financial ratios that have been computed on the basis of information contained in the income statement and balance sheet of these firms. The set of ratios that we use to construct the DEA model are: operations (days of sales outstanding/average collection period, inventory turnover, and asset turnover ratios), profitability (operating profit margin, net profit margin, return on equity and return on assets), and financials (quick ratio and total debt/equity ratio). In order to evaluate a firm’s financial performance, a financial analyst usually uses these set of ratios.2 We use the financial statement data available on a quarterly basis from July 2007 to July 2008 from Hoovers Online for this study. Current economic meltdown started in December 2007. Therefore, this time frame allows us to study the financial performance of the retailers before the crisis as well as after onset of economic crisis. Out of these seven ratios, we specify days of

---

sales outstanding and total debt/equity ratio as input, because for a given company the lower these variables are the better the performance of the company. Similarly, higher operating profit margin, net profit margin, return on equity, return on assets, asset turnover, and inventory turnover imply a better-performing company. Thus, we consider these variables as output variables. Finally, the choice of the DEA model is also an important consideration. We should select the appropriate DEA model with options such as input maximizing or output minimizing, multiplier or envelopment, and constant or variable returns to scale. DEA applications that involve inflexible inputs or not fully under control inputs should use output-based formulations. On the contrary, for an application with outputs that are an outcome of managerial goals, input-based DEA formulations are more appropriate. In addition, for an application that emphasizes inputs and outputs, we should use the multiplier version. Similarly, for an application that considers relations among DMUs, envelopment models are more suitable. Furthermore, the characteristics of the application dictate the use of constant or variable returns to scale. If the performance of DMUs depends heavily on the scale of operation, constant returns to scale (CRS) is more applicable, otherwise variable returns to scale is a more appropriate assumption.

In our study, the comparative evaluation among the companies is an important consideration. Therefore, we select the envelopment models for our analysis. In addition, the outputs are an outcome of managerial goals. Therefore, input-based formulation is recommended for our study. The objective of the analysis is to suggest a benchmark for the seven largest retailing firms. Furthermore, to investigate the effect of scale of operations, if any, among the seven companies, we consider both variable returns to scale and constant returns to scale DEA models. Also, the structure of the DEA model (in envelopment form) uses an equation and separate calculation for every input and output. Therefore, all the input and output variables can be used simultaneously and measured in their own units. In this study, we use the Input-Oriented Variables Return to Scale (VRS) to evaluate the efficiency of seven retail companies. Figure 1 illustrates a decision support system using data envelopment analysis. The decision support system uses the DEA methodology to assess the performance of each company. The DEA-based decision support system uses the company attributes – days of sales outstanding (average Collection Period) and total debt/equity ratio as input variables. The system uses the operating profit margin, quick ratio, return on assets, asset turnover, and inventory turnover as output variables to calculate an efficiency score for a firm. This score is a relative value computed by comparing the given firm to a pool of well-performing companies that serve as a benchmark for the company under evaluation. Each firm is evaluated against the existing firms with an identical set of inputs or outputs that is constructed as a combination of performing and non-performing companies. By using the existing good companies as a “role model,” DEA not only helps differentiate well performing (efficient companies from poorly performing (inefficient) firms, but also brings out the reasons why a company may be underperforming.

ILLUSTRATING SIDE MODEL OF THE DECISION SUPPORT SYSTEM FOR LOAN EVALUATION:

To screen consumer loan applications, loan officers use different methods besides intuitive judgment and experience. Using mathematical techniques, many credit-scoring models have been developed to assist the loan officer in differentiating good loans from bad. Besides these traditional statistical models, many financial institutions use artificial intelligence methods, such as expert systems, artificial neural systems, and fuzzy logic. It is only recently that the finance community has started applying data envelopment analysis, a relatively new technique. This study proposes to assess the creditworthiness of a new loan applicant using a decision support system that applies a combination of two diverse analytical techniques: data envelopment analysis and particle swarm optimization. I propose to use the
Data Envelopment Analysis methodology to assess the creditworthiness of an existing set of loans whose outcome (accepted and turned good, accepted and turned bad, and reject) is known. The DEA model benchmarks the given set of loans, and assigns an efficiency score.

Each of the loans is a homogenous unit, and we can apply the DEA methodology to assess comparative performance of these loans. The DEA model is a part of a decision support system that uses a number of variables to determine how good a loan is. A loan application includes information such as the applicant’s age, housing, address time, total income, number of credit cards, number of dependents, job time, other loan obligations, total debt, monthly rent/mortgage payments, number of inquiries for an applicant, and credit rating. The study creates a DEA model that evaluates the relative efficiency of a set of loans that credit unions have already administered, and allocates a score on the scale of 1 to 100. Further, the next step in the design of the decision support system is to use particle swarm optimization algorithm to map a new loan application to verify how close the loan parameters are to an optimal set of values. Once the PSO model comes up with the optimal loan parameters, the loan officer can discern how close the given loan’s variables are to an ideal situation for the given weights on different parameters. To validate the results of PSO, we further benchmark the ideal loan parameters with the given DEA model to check if its 100% efficient, and the PSO algorithm converges to optimal solution.

Data specifications:

The data for this study is a pooled data set of loans made by nine different credit unions with a total of 185. The applicants can be categorized into three major groups: applicants who were accepted, and were good credits (Group 1); applicants who were accepted, but were not good credits (Group 2); and applicants who applied for a loan, but were rejected (Group 3). Table 1 illustrates the characteristics of the data set. Further, the data set also includes information such as the applicant’s age, housing, address time, total income, number of credit cards, number of dependents, job time, co-maker on other loans, total debt, monthly rent/mortgage payments, number of inquiries for an applicant, and credit rating of each applicant. Credit unions in the data set assign loan applicants into four credit groups—excellent (1), good (2), marginal (3), and poor (4). The credit rating is determined on the basis of the number of inquiries. The higher the number of inquiries on an applicant, the lower will be the credit rating. The calculation of credit ratings is consistent across all the credit unions. Thus, based on the information supplied by an applicant, we can calculate the applicant’s total payments, total income, and total debt.

The study considers the following variables:

- Total Debt: Total debt of the applicant at the time of application.
- Number of Loans: Total number of loans outstanding in the applicant’s name.
- Payments: Total monthly payments
- Dependents: No of dependents of the applicants.
- Total Income: Total monthly income from all sources.
- Job time: Time at current employer.

Data envelopment model specifications for loan evaluation:

The study uses factors such as total debt, number of loans, total payments, number of dependants input variables as they should be minimized. Further, we use total income and time spent in employment as output variables as the higher these are the better it is. Hence, these output variables should be maximized. Besides the mathematical and computational requirements of the DEA model, there are many other factors that affect the specifications of the DEA model. These factors relate to the choice of
the DMUs for a given DEA application, selection of inputs and outputs, choice of DMUs for a given DEA application, choice of a particular DEA model (e.g. CRS, VRS, etc.) for a given application, and choice of an appropriate sensitivity analysis procedure (Ramanathan, 2003). Due to DEA’s non parametric nature, there is no clear specification search strategy. However, the results of the analysis depend on the inputs/outputs included in the DEA model. There are two main factors that influence the selection of DMUs – homogeneity and the number of DMUs. To successfully apply the DEA methodology, we should consider homogenous units that perform similar tasks, and accomplish similar objectives. In our study, the loans are homogenous entities as they are from a credit union. Furthermore, the number of DMUs is also an important consideration. The number of DMUs should be reasonable so as to capture high performance units, and sharply identify the relation between inputs and outputs.

In this study, the analysis of a loan emphasizes inputs and outputs. Therefore, I select the multiplier model for my analysis. In addition, factors such as total debt, number of loans, total payments, number of dependants in employment are not very flexible inputs that cannot be immediately controlled. Therefore, output-based formulation is recommended for my study. Furthermore, the quality of the loans does not depend on the scale of operations, thus variable returns to scale is safe assumption. Also, the structure of the DEA model (in multiplier form) uses an equation and separate calculation for every input and output. Therefore, all the input and output variables can be used simultaneously and measured in their own units. Table 2 shows the composition of the data set. There are 23 good loans, 121 loans that were accepted but defaulted and 41 loans that were rejected. Table 3 shows the results of the DEA model to benchmark the loans in the three categories. The DEA model assigns an efficiency score in the range of 1 to 100. Within a tolerance limit of 75%, the DEA model correctly classifies 83% of accepted loans (4 loans less than 75% efficient), 86% bad loans (17 greater than 75% efficient), and 73% reject loans (11 greater than 75% efficient). Thus, DEA model reasonably identifies the good, bad, and reject loans with the probability of type II error very low. However, the cost of type II error, especially for the bad loans is very high. Therefore, we can use a second screening tool to assess the credit-worthiness of a loan. We propose the use of a complementing optimizing tool to examine a loan.

**Particle swarm optimization model for the decision support system:**

For the PSO model, we use the modified PSO algorithm as explained above. We used a fitness function that minimizes the values of three input variables: the ratio of the total payment to the total income (ratio 1), the ratio of debt to the total income (ratio 2), time in employment, number of outstanding loans, and total number of dependants as the factors that discriminate between a good and a bad loan. We used 43% weight for ratio 1, 42% weight for ratio 2, and 5% weight for time in employment, number of dependants, and total number of loans, respectively. We tested the effectiveness of PSO with two loans, one good and one bad. We used 6-dimensional particles with a swarm size of 40. Each dimension represents the loan parameters: total debt, number of loans, total payments, number of dependants, total income, and time spent in employment. In addition, the swarm parameters are: social component (.5), cognitive component (.5), inertia (.5), and clamping factor (.3). The PSO algorithm found the solution in 5 iterations. Table 4 illustrates the example of a bad loan that was 16% efficient as compared to the existing pool of loans using DEA. However, the optimal solution, found using PSO, is 100% efficient as compared to the existing pool of loans using DEA. Thus, PSO successfully finds the optima. To further test PSO, we used a good loan that was 100% efficient using DEA. The PSO simply leaves the parameter unchanged after 5 iterations indicating that the solution is already optimal. Thus, PSO algorithm successfully discerns good and bad loans that are further validated by data envelopment
SUMMARY AND CONCLUSION

This study proposes the modeling and development of a decision support system that uses a combination of data envelopment analysis and particle swarm optimization. Thus, the decision support system derives benefit from both methodologies to recommend a decision. As illustrated in the literature review section, many studies illustrate the use of PSO and DEA. However, no studies illustrate the fusion of DEA and PSO models. This study illustrates the use of two complementing techniques that can aid a decision maker in making decisions regarding the credit-worthiness of a loan where the cost of accepting a bad loan is very high. DEA does not require the manager to attach prescribed weights to each input and output. Moreover, DEA modeling does not require prescription of the functional forms that are needed in statistical regression approaches. DEA uses techniques such as mathematical programming that can handle a large number of variables and constraints. As DEA does not impose a limit on the number of input and output variables to be used in calculating the desired evaluation measures, it’s easier for managers to deal with complex problems and other considerations they are likely to confront. DEA is a methodology based on an interesting application of linear programming allowing a decision maker to use multiple inputs and outputs measured in different units. DEA identifies good units in a given set of DMUs and provides a measure of inefficiency for all others. The DMUs having the most desirable characteristics are rated a score of one (100% efficient), while the DMUs that are inefficient score between zero and one. DEA methodology can identify a bad DMU by comparing its characteristics with a given set of benchmark DMUs having good DMU characteristics.

Similarly, particle swarm optimization does not require restrictive assumptions of the statistical model. PSO model follows the basic principle of swarm intelligence that takes clues from social and cognitive behavior. This model works with costs associated with each parameter. Thus, the decision maker can evaluate the impact of cost on each parameter, something that the DEA model does not associate with the model parameters.

Finally, to illustrate the SIDE model, this study proposed the development of a decision support system to screen consumer loan applications. Loan officers use many different methods besides intuitive judgment and experience. They use mathematical techniques such as credit-scoring models and traditional statistical models. In addition, many financial institutions use artificial intelligence methods such as expert systems, artificial neural systems, and fuzzy logic. This study proposes the development of a decision support system that uses a combination of data envelopment analysis and swarm intelligence. Thus, the decision support system derives benefit from both methodologies to provide a comprehensive review of a loan applicant.

TABLES, FIGURES, & REFERENCES

Tables, figures, references, and full paper available upon request from the authors.