ABSTRACT

Forecasting is an important tool used by businesses to plan and evaluate their operations. One of the most commonly used techniques for forecasting is regression analysis. Often forecasts are produced based upon a set of comparable units which could be individuals, groups, departments or companies that perform similar activities such as a set of banks, a group of managers and so on. We apply a methodology that includes a new independent variable, the comparable unit’s DEA relative efficiency, into the regression analysis. The results of applying this methodology to the performance of commercial banks over time will be presented.

(Keywords: Forecasting, Data Envelopment Analysis, Regression)

INTRODUCTION

Quantitative forecasting techniques use historical data to predict future outcomes. Most quantitative forecasting techniques can be categorized into either time series or causal models. Time series forecasting techniques use only the time series data itself to build the models. These time series approaches isolate and measure the impact of trend, seasonal, and cyclical time series components. Causal models use a set of independent (predictor) variables, possibly including the time series components, believed to influence the forecasted dependent variable. One of the most popular causal model approaches is regression analysis. Regression techniques employ the least squares method to establish a statistical relationship between the dependent (forecasted) variable and the set of independent (predictor) variables.

In many forecasting situations analysts must produce forecasts for comparable units. Comparable units could be individuals, groups of individuals, departments or business and operational entities. In this analytical environment each comparable unit should be performing a similar set of tasks. For example, preparing forecasts for a number of corporate divisions will predict the sales results in future periods for similar products based upon prior investments in similar
marketing promotions. When applying regression analysis, the established statistical relationship is an average relationship using one set of weights assigned to the independent variables. However, when regression is applied to a set of comparable units the relative weight of each of the independent variables will most likely vary from comparable unit to comparable unit. For example, if advertising is an independent variable, one comparable unit might emphasize advertising more (or less) than other comparable units. As a result the regression model could provide forecast estimates that are too high or too low.

In this paper we will apply and extend some of our recent work in which we introduced a methodology that incorporates into the regression forecasting analysis a new variable that captures the unique weighting of each comparable unit [3, 4, 5]. This new variable is the relative efficiency of each comparable unit. It is generated by a non-parametric technique called data envelopment analysis (DEA). The DEA efficiency variable is a nonlinear variable that takes into account a set of weighted inputs and outputs. In each of our previous studies the inclusion of this multivariate variable has improved the regression forecasting model. The main objective of this paper is to present the results of a longitudinal study applying this methodology.

In the next section, we provide a brief introduction to DEA. Subsequently, we discuss the methodology and present the results of applying our methodology to a data set of commercial banks. Finally, the conclusions and future extensions are discussed.

**DATA ENVELOPMENT ANALYSIS (DEA)**

One of the major concerns of managers in evaluating the performance of an operation within any type of organization is efficiency. Efficiency measures whether resources are being put to good use. One dimension of the efficiency of an operation of any organization is the manner by which that organization selects and uses resources to produce its products. The more product produced for a given amount of resources the more efficient (i.e., less wasteful) is the operation. In order to evaluate the relative efficiency of comparable components of an organization, Charnes, Cooper and Rhodes [1] proposed an innovative quantitative technique that they named data envelopment analysis (DEA).

DEA utilizes linear programming to produce measures of the relative efficiency of comparable decision-making units that employ multiple inputs and outputs. The decision-making unit (DMU) is the component of the organization being evaluated. For example, a hospital may use the technique to evaluate different care-giving units. DEA takes into account multiple inputs and outputs to produce a single aggregate measure of relative efficiency for each DMU. It requires only that the selected inputs and outputs be quantifiable. The technique can analyze these multiple inputs and outputs in their natural physical units without reducing or transforming them into some common unit of measurement. Finally, DEA evaluates all the DMUs and all their inputs and outputs simultaneously, conservatively identifying the sets of relatively efficient and inefficient DMUs. Thus, the solution of a DEA model provides a manager a summary with comparable DMUs grouped together and ranked by relative efficiency. Since the appearance of the seminal paper in 1978, there have been thousands of theoretical contributions and practical applications in various fields using DEA. DEA has been applied to many diverse areas such as
health care, military operations, criminal courts, university departments, banks, electric utilities, mining operations, and manufacturing productivity [2, 6, 7].

For DEA, efficiency is defined as the ratio of weighted outputs to weighted inputs:

\[
\text{Efficiency} = \frac{\text{weighted sum of outputs}}{\text{weighted sum of inputs}}
\]

The DEA approach identifies the set of weights (all weights must be positive) that maximizes individually each DMU’s efficiency while requiring the corresponding weighted ratios of the other DMUs to be less than or equal to one. A DMU is considered relatively inefficient if its efficiency rating is less than one (E<1). The degree of inefficiency for a DMU is measured relative to a set of more efficient DMUs. However, a DMU identified as being efficient (E=1), does not necessarily imply absolute efficiency. It is only relatively efficient as compared to the other DMUs that are being considered.

The Charnes, Cooper and Rhodes (CCR) DEA model [1] is a linear program that compares the ratio of weighted outputs to weighed inputs, i.e., efficiency, for each comparable unit. The efficiency \( E_k \) of comparable unit \( k \) is obtained by solving the following linear formulation:

\[
\begin{align*}
\text{MAX } E_k &= \sum_{r=1}^{t} u_r Y_{rk} \\
\text{s.t.} & \quad \sum_{i=1}^{m} v_i X_{ik} = 1 \\
& \quad \sum_{r=1}^{t} u_r Y_{rj} - \sum_{i=1}^{m} v_i X_{ij} \leq 0 \quad j=1,\ldots,n \\
& \quad u_r, v_i \geq \varepsilon \quad \forall r, i
\end{align*}
\]

\( u_r, v_i \geq \varepsilon \) \quad \forall r, i

where:

Parameters

\( Y_{rj} \) = amount of the rth output for the jth comparable unit;
\( X_{ij} \) = amount of the ith input for the jth comparable unit;
\( t \) = the number of outputs;
\( m \) = the number of inputs, and;
\( n \) = the number of comparable units;
\( \varepsilon \) = is a small infinitesimal value;

Decision Variables

\( u_r \) = the weight assigned to the rth output, and;
\( v_i \) = the weight assigned to the ith input.
REGRESSION FORECASTING METHODOLOGY

Our regression forecasting methodology is designed for application to a historical data set of multiple inputs and outputs variables from a set of comparable units [3, 4, 5]. Additionally, one output variable (e.g., sales, production, or demand) is selected as the principal dependent variable to be forecast.

Since the data set we studied in this paper has a relatively small number of inputs and outputs we adjust our procedure and eliminate the initial stepwise regression. As a result, the first step is to run a data envelopment analysis for each comparable unit. We use the resultant efficiency scores as surrogate measures of the unique emphasis of the variables and of performance. Using a principal output variable as the regression dependent variable, all the input variables plus the DEA efficiency score as regression independent variables, we run the regression. This regression model with the DEA efficiency variable should be superior to the model without the DEA efficiency variable. Superiority would be demonstrated with a significantly lower standard error of the mean and increased $R^2$.

EXAMPLE

Seiford and Zhu, [8], applied DEA to the 55 U.S. commercial banks that appeared in the Fortune 1000 list in April 1996. The DEA input variables were the number of employees, assets, and stockholder’s equity; and the DEA output variable were revenue and profit. The selection of these variables were “based on Fortune’s original choice of factors for performance characterization”, [8]. We retrieved the same Fortune 1000 list of U.S. commercial banks from 1995 to 2005. Thirteen banks were common over this time period. Ten of these thirteen banks are regional and during this time period almost all were mid cap, profitable, and had a higher Return on Assets than the industry average.

Using the data for these 13 banks we ran ten DEA models for 1995 to 2004. Table 1 below lists some descriptive statistics of the DEA efficiency scores for these years. As shown in Table 1, at least half of the banks consistently, for the ten-year period examined, had efficiency scores greater than 90% and usually at least greater than around 80%, except in 2004 where one bank drop to 62.1%.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Min</td>
<td>85.89</td>
<td>73.13</td>
<td>78.89</td>
<td>77.69</td>
<td>80.40</td>
<td>81.75</td>
<td>86.72</td>
<td>84.47</td>
<td>81.41</td>
<td>62.10</td>
</tr>
<tr>
<td>Max</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
</tr>
<tr>
<td># 100</td>
<td>5</td>
<td>3</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>7</td>
<td>5</td>
<td>5</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Average</td>
<td>96.78</td>
<td>91.44</td>
<td>93.30</td>
<td>92.81</td>
<td>93.00</td>
<td>94.72</td>
<td>96.12</td>
<td>96.07</td>
<td>91.97</td>
<td>90.57</td>
</tr>
<tr>
<td>median</td>
<td>97.29</td>
<td>90.98</td>
<td>96.80</td>
<td>95.53</td>
<td>95.52</td>
<td>100.00</td>
<td>98.14</td>
<td>98.39</td>
<td>91.22</td>
<td>92.48</td>
</tr>
</tbody>
</table>

Table 1. Descriptive statistics of the DEA efficiency scores for each year.
Using the DEA efficiency scores as an input and revenue as our primary output variable, we ran regression models for 1996 to 2005. The basic regression equation used was:

\[
\text{Revenue}(t) = \text{Employees}(t-1) + \text{Assets}(t-1) + \text{Equity}(t-1) + \text{DEA}(t-1)
\]

where \( t = 1996, \ldots, 2005 \). We refer to this model using the DEA variable as w/DEA and conversely, the same regression without the DEA efficiency score variable is refer to as NoDEA.

Tables 2 and 3 summarize the regression models results with \( R^2 \) values and standard errors. In all years the \( R^2 \) values were greater with the w/DEA models than with the NoDEA models. The average improvement was only 1.65%. The standard error values for the w/DEA models, in Table 3 showed significantly more improvement in eight out of the ten years, averaging a 23.8% decrease in the standard errors.

<table>
<thead>
<tr>
<th>YEAR</th>
<th>( R^2 )</th>
<th>( R^2 )</th>
<th>( R^2 )</th>
<th>( R^2 )</th>
<th>( R^2 )</th>
<th>( R^2 )</th>
<th>( R^2 )</th>
<th>( R^2 )</th>
<th>( R^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>NoDEA</td>
<td>97.58</td>
<td>97.57</td>
<td>96.42</td>
<td>91.41</td>
<td>95.59</td>
<td>93.82</td>
<td>98.53</td>
<td>95.70</td>
<td>97.29</td>
</tr>
<tr>
<td>w/DEA</td>
<td>99.09</td>
<td>99.57</td>
<td>96.76</td>
<td>93.89</td>
<td>98.88</td>
<td>95.20</td>
<td>99.55</td>
<td>96.49</td>
<td>97.47</td>
</tr>
<tr>
<td>Difference</td>
<td>1.51</td>
<td>2.00</td>
<td>0.34</td>
<td>2.48</td>
<td>3.29</td>
<td>1.38</td>
<td>1.02</td>
<td>0.79</td>
<td>0.18</td>
</tr>
</tbody>
</table>

Table 2. The regression \( R^2 \) values for each year and for the two models.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>NoDEA</td>
<td>333.47</td>
<td>343.29</td>
<td>478.24</td>
<td>669.27</td>
<td>517.96</td>
<td>575.61</td>
<td>236.38</td>
<td>402.67</td>
<td>349.88</td>
<td>645.26</td>
</tr>
<tr>
<td>w/DEA</td>
<td>216.69</td>
<td>153.24</td>
<td>482.42</td>
<td>598.84</td>
<td>276.53</td>
<td>538.25</td>
<td>139.03</td>
<td>385.65</td>
<td>358.57</td>
<td>374.16</td>
</tr>
<tr>
<td>Difference</td>
<td>-116.78</td>
<td>-190.05</td>
<td>-126.92</td>
<td>-210.73</td>
<td>-241.43</td>
<td>-422.41</td>
<td>-107.35</td>
<td>-216.02</td>
<td>-191.31</td>
<td>-271.10</td>
</tr>
</tbody>
</table>

Table 3. The regression standard errors for each year and for the two models.

**CONCLUSIONS**

In this paper, we applied a new regression forecasting methodology to forecasting comparable units. This approach included in the regression analysis a surrogate measure of the unique weighting of the variables and of performance. This new variable is the relative efficiency of each comparable unit that is generated by DEA. The results of applying this new regression forecasting methodology including a DEA efficiency variable to a data set demonstrated that this provides an enhanced approach to forecasting comparable units. We plan to perform further testing with other data sets from other industries, some with more comparable units and possibly more years of data.

**REFERENCES**


