Case Study:
A Linked Linear Regression – Data Envelopment Analysis Model for the Long-Term Prediction of Mutual Fund Performance

Cindy Chao, Greg Higgins
Michael McGee

Department of Computer Science and Engineering
SOUTHERN METHODIST UNIVERSITY
Dallas, Texas

May 14, 1994
## Contents

Abstract ............................................................................................................................. 3
Introduction ...................................................................................................................... 3
What is a Mutual Fund? ................................................................................................. 4
DEA Factors ...................................................................................................................... 4
Model Development and Validation ............................................................................. 6
Examination of Results ................................................................................................... 9
Investment Performance ............................................................................................... 10
Conclusion ...................................................................................................................... 11
References ....................................................................................................................... 12
Professional References ................................................................................................. 13
Appendix ........................................................................................................................ 14
  DEA Formulation ........................................................................................................ 15
  DEA Code ................................................................................................................. 16
  Forcasting Code ....................................................................................................... 19
Case Study:
A Linked Linear Regression – Data Envelopment Analysis Model for the Long-Term Prediction of Mutual Fund Performance

Michael McGee
Greg Higgins
Cindy Chao
Department of Computer Science and Engineering
Southern Methodist University
Dallas, Texas 75275

Data Envelopment Analysis (DEA) can be used to evaluate managerial efficiency of mutual funds. By extending the DEA methodology with linear regression, a new prediction tool for future efficiency is developed. The linked model was tested on a sample of fifty growth mutual funds over a three year time horizon — resulting in accurate predictions of future efficiency with 99% confidence.

The current mutual fund market contains nearly $1.4 trillion dollars; this rapid growth of mutual funds as a long-term investment instrument has led to an increased demand for evaluation and forecasting techniques. The performance of a mutual fund is directly linked to the performance of the management staff. Data Envelopment Analysis (DEA) can be used to evaluate managerial efficiency; however, this is only a current assessment of a fund’s management. DEA is a linear programming methodology that computes relative efficiency scores among a group of homogeneous decision-making units (DMUs). Combined with linear regression, DEA becomes a prediction tool for long-term forecasting. The linked model not only evaluates return on investment but also the quality of the fund’s managerial team on a long-term basis.
What is a Mutual Fund?
A mutual fund is an investment company that pools the money of many institutional and individual investors into one professionally managed investment account. The fund manager creates a portfolio for the shareholders by buying a variety of "true" financial instruments, such as corporate stocks and bonds, and government and municipal securities, in accordance with the fund’s particular investment objectives. Fund managers are professional investors who know the financial market and follow it daily with extensive research capabilities to effectively investigate all investment opportunities. The fund is always monitored, with individual securities constantly being bought, sold, and traded in order to optimize the investment objectives. The beauty of investing in sometimes twenty different securities lies in the protection of the investment through portfolio diversification.

Mutual funds are divided into six major categories according to their investment objectives, which are outlined in the fund’s prospectus, such as goals, level of risk, and the type and frequency of return desired. Since the goal of the linked model is forecasting of long-term performance, growth mutual funds were selected because their objectives emphasize the accumulation of capital gains, and they have the greatest long-term growth potential.

DEA Factors
The inputs and outputs of the envelopment analysis were made up of the main characteristics of a mutual fund. They were selected because they are important factors to investors and are directly linked to the performance of the fund manager. The inputs selected were beta — a measure of risk, expense ratio, and loads & other fees. The outputs used were net asset value (NAV) per share, dividends, and capital gains.

Inputs
Beta is a measure of the fund portfolio’s average volatility, or risk, relative to the S&P 500. The beta value of the market is a constant equal to 1.00. A fund with a beta value greater than 1.00 is more volatile than the market. It is expected to outperform the "average fund" in a bull (rising) market and bring in greater returns, while it is expected to fall more than
the average fund in times of a bear (declining) market. A fund with a beta value less than 1.00 possesses the characteristics of the converse situation. For example, a portfolio with a beta value of 0.75 is expected to be 25% less responsive to the general economic trend than the average fund and return 25% more conservative than the average fund.

Expense ratio is a measure which relates expenses to the average net assets for a period. It is calculated by taking the annual expenses paid by the fund (including management and investment advisory fees, custodial and transfer agency charges, legal fees, and distribution costs) divided by the average shares outstanding for the period. The greater the expense ratio, the larger the percentage of the shareholder’s investment gets paid out directly to the management company. The expense ratio does not include loads or commissions paid for sales or transaction fees.

Loads and other fees represent all other charges made to the shareholder which are not accounted for in the expense ratio. A load is a portion of the offering price of the fund that pays for costs such as sales commissions, brokerage costs, and transaction fees. This value is a percentage of the purchase price taken off either before the money is invested (front-end load) or after investment when the shareholder redeems the shares (back-end load). The amount of the load is established by the managerial company, but is paid to the sales and transaction team. Some funds have a load of 0.00%, meaning that they have the same benefits as load funds, except they are sold directly by the fund company rather than through a sales broker so that no loads or other fees are incurred. With no-load funds, the investor must seek out and contact the investment firm directly. There is a distinct advantage to investing in no-load funds. For example, if we had invested $1000 in a particular portfolio with a 7.50% front-end load, $75 would be taken off the top, and we would be left with only $925 worth of shares in the fund. Had this been a no-load fund, the full $1000 would have been utilized.

**Outputs**

Net asset value per share represents how much each share is worth. It is calculated by taking the total market value of a fund’s shares (securities plus cash plus accrued earnings minus
liabilities) divided by the number of shares outstanding. No profits or losses are realized through NAV until the fund is sold. Interest lies in the amount of increase or decrease in NAV from time of purchase to time of sale.

Dividends are payments declared by a corporation's board of directors and made to shareholders, usually on a quarterly basis. A fund's dividends come directly from the dividends paid out by the individual securities within the portfolio. They are paid out equally among the shareholders on a scheduled time frame.

Capital gains on a mutual fund are the profits made from the sale of a capital asset (such as a security, stock, or bond) within the portfolio. The number used for capital gains is adjusted for losses and stock splits. The sale of an asset occurs when the manager decides that a security needs to be sold for the benefit of the portfolio. Capital gains are distributed to shareholders when they are realized (when a security is sold and a profit is made) and are paid out as capital distributions.

Dividends and capital gains are treated as separate outputs because they are incurred, paid out, and taxed in different ways. Shareholders can reinvest dividends and capital distributions if they elect to do so.

**Model Development and Validation**

The decision-making units (DMUs) in this analysis were the 50 growth mutual funds. A prerequisite of DEA is homogeneous DMUs. By selecting the funds from the same sector (growth funds), this requirement is met.

The DEA model uses the dual formulation of the linear program as described by Charnes, Cooper and Rhodes (1978). The objective function of this form is to minimize the objective value of the current DMU. If a DMU is efficient the optimal value is 1, while an inefficient DMU is driven towards zero. The first step in the development of the linked model is the generation of DEA scores for the mutual funds being studied. The data collected for the mutual funds is used as the inputs and outputs for the DEA code which is documented in the appendix. The DEA code is written using the SAS/OR package allowing the DEA scores to be fed directly into the regression phase of the model.

DEA scores are generated for the fifty mutual funds in the study using
the 1990 data collected from *Morningstar Mutual Funds*. At this point DEA scores are also generated for 1993 and stored for use in validating the model's forecasts. The 1990 DEA scores are then regressed against all of the six DEA factors previously described.

This regression model is unique from previous research in that the goal of the regression is not to create an efficiency score, but to forecast an efficiency score for the future of a particular DMU. Previous work by Thanassoulis (1993) compared the DEA methodology with linear regression as a means of creating efficiency scores. Our model links the two methods into one prediction tool. The full factor model that resulted appeared on the surface to be a good predictor of efficiency \( (R^2 = 0.8239) \). The full factor model included three non-significant factors at the 5% level of significance. Beta (p-value = 0.68), load (p-value = 0.65), and dividend (p-value = 0.50) were all eliminated from the model leaving only expense ratio, net asset value, and capital gains as significant factors. The revised model remained significant with an \( R^2 = 0.8213 \).

As stated before, this model appears to be a good predictor; however, linear regression models must be validated through diagnostic tests. One assumption in regression analysis underpins the entire methodology — errors are a randomly distributed variable following a normal distribution with mean equal to zero, \( N(0,\sigma^2) \).

Regression analysis in SAS is based on a linear relationship between the dependent and independent variables. These assumptions can be verified through the examination of the residuals, or errors, of the regression model. A plot of the residuals versus the predicted values, Figure 1, reveals a distinct pattern indicating a non-linear relationship between the independent and dependent variables. This requires the examination of univariate statistics.
for each of the variables in the regression model. A standard normal quantile plot for each of the variables allows the determination of the distribution of each variable.

From the univariate procedures, some of the variables demonstrated an exponential distribution requiring a log-linear transformation in order to validate the regression model. Net asset value, capital gains, dividends and the actual DEA score follow exponential distributions while beta, load, and expense ratio are all normally distributed. A second data set of the inputs and outputs was created by taking the natural log of each of the exponentially distributed variables. Once normality was confirmed through univariate procedures, a new full factor linear regression model was fitted. The newly formed model resulted in an $R^2 = 0.8163$ and was validated through the residual plot shown in Figure 2. Using the SAS regression report, the regression equation was created using the coefficients of the significant factors: beta, expense ratio, net asset value, and dividend.

The next step is the creation of predicted efficiency scores. The common method for forecast model verification is the prediction of the past. The linked model is verified by using the input and output data for 1993 in the regression equation created from the 1990 data. The predicted values are then compared to the actual 1993 DEA scores generated earlier by the DEA code. The forecast errors are evaluated using a significance test on the mean difference. The test is conducted at the 1% level of significance. $H_0$: mean difference = 0; $H_a$: mean difference $\neq$ 0. The resulting test statistic: $z = -0.5164$ and critical value, $z = 2.57$ leads to the conclusion that the prediction values are not significantly different from the actual DEA scores with 99% confidence.
Examination of Results

In analyzing the results of the sample studied, it is important to remember that a mutual fund is mainly used as a long-term investment, providing stable low-risk growth. A potential shareholder wants to invest in a fund that will be efficient in the future, regardless of its current efficiency. Therefore we focus our examination on the set of funds with efficient scores in either 1990 or 1993. Table 1 shows the 1990 DEA, Predicted 1993, and actual 1993 DEA scores along with the errors for each prediction (Error = 1993 DEA - Predicted 1993).

Of the first four funds listed (all receiving efficient ratings in 1990), the future efficiency of three funds was correctly predicted. Our linked model predicted future behavior of currently efficient funds with 75% accuracy. Although the error for T. Rowe Price New America Fund, PRWAX, is high, we are not concerned with the magnitude of inefficiency, only the state of its efficiency.

Of the last four funds in the table (all receiving efficient ratings in 1993), three funds were predicted accurately. The fifth and sixth funds, Fidelity Destiny Portfolios: Destiny II, FDETX, and Clipper Fund, CFIMX, both received a predicted score of approximately 0.98 for 1993. Based on the previous significance test, these values are not significantly different from 1.00. Thus our model predicts the funds that will be efficient in the future with 75% accuracy.
Investment Performance

Since investors are concerned with the long-term growth and appreciation of their investments, DEA by itself is not an effective tool to aid in the development of investment strategies. However, combined with linear regression, DEA gives us the ability to estimate the long-term efficiency of a fund. To illustrate the effectiveness of the linked model relative to DEA alone, we simulated an investment in the efficient funds in 1990 and the funds predicted to be efficient by our linked model. Since mutual funds allow the purchase of fractional shares, we invested $1000 in each of the mutual funds. The results are given in Table 2.

The percent of appreciation we received in the portfolio of 1990 efficient mutual funds was 46.92%, while the appreciation was significantly higher for the portfolio of funds recommended by the linked model at 60.41%.
Conclusion

The linked model allows us to broaden the range of applications for DEA. The linked model enables the prediction of future efficiency, which is of great benefit to investors because it provides more insight into the long-term viability of a mutual fund. The model also requires little historical data for each DMU unlike other forecasting methods. Another benefit is the relative compactness of the model; very few predictors are required to explain a high degree of variability and achieve accurate predictions. Data envelopment analysis by itself is not an accurate predictor of future efficiency of mutual funds. Only one fund that was efficient in 1990 remained in the efficient set in 1993. Linear regression and DEA combine to give us an accurate prediction for the future behavior of DMUs.

An immediate extension to this research is the definition of an "ideal type" mutual fund. An optimization algorithm will be used to generate values for the regression factors and record the combinations which yield and efficient fund prediction. Once this benchmark is defined a growth mutual fund prospectus can simply be visually inspected to determine managerial efficiency. The next phase of this research would be to compare the accuracy of the linked model with that of other forecasting methods, such as time series decomposition. Another step in the continued development of this area of research is to improve the accuracy of the predictions through further modifications to the regression model using a weighted regression model or the inclusion of adaptive error control.
References


Professional References

Barr, Richard S., Ph.D. Associate Professor, School of Engineering and Applied Sciences. Southern Methodist University, Dallas, Texas.

Dula, Jose, Ph.D. Assistant Professor, School of Engineering and Applied Sciences. Southern Methodist University, Dallas, Texas.

Guerra, Rudy, Ph.D. Assistant Professor, Department of Statistical Science. Southern Methodist University, Dallas, Texas.

Welch, David. Senior Systems Analyst II, Bradfield Computer Center. Southern Methodist University, Dallas, Texas.
Appendix

DEA Formulation

Mathematical formulation of the DEA model used in the Linked Model.

DEA Code

The DEA code is developed using Release 6.0 of the SAS System on an IBM 3090 mainframe. The code presented requires only two input files. One file contains the input and output factor data while the second file contains a listing of the mutual fund ticker symbols. The program is constructed from a series of SAS Macros which fully automate the formulation and solution of the multiple linear programs required in Data Envelopment Analysis.

Forecasting Code

Developed as an extension to the DEA code, the forecasting model uses the same computing platform and software. The program reads the same data files as the DEA code as well as the DEA scores generated by the DEA study. The program generates predicted efficiency scores and all necessary statistical procedures for validating the model.
DEA Formulation

\[ \begin{bmatrix} \text{Inputs} \\ \hline \text{Outputs} \end{bmatrix} \cdot \begin{bmatrix} \lambda \end{bmatrix} + \begin{bmatrix} \text{Slacks} \end{bmatrix} = \begin{bmatrix} \theta \end{bmatrix} \]

Min \( \theta \)

S.T.

\[
\sum A_j^x \lambda_j - A_0^x \Theta \leq 0
\]

\[
\sum A_j^y \lambda_j \geq A_0^y
\]

\( \lambda_j \geq 0 \)

\( \Theta \geq 0 \)

Mathematical formulation of DEA requires that the inputs and outputs for each DMU are formatted in a matrix, \( \mathbf{A} \). The matrix is multiplied by a vector, \( \mathbf{\lambda} \), to construct the linear programming constraints.

Minimize \( \Theta \)

Subject to:

Inputs

Beta: \( x_1^* \text{DMU}_1 + x_2^* \text{DMU}_2 + \ldots + x_{10}^* \text{DMU}_{10} - x_{51}^* (\Theta^* 1) \leq \Theta^* 0 \)

Load: \( x_1^* \text{DMU}_1 + x_2^* \text{DMU}_2 + \ldots + x_{10}^* \text{DMU}_{10} - x_{51}^* (\Theta^* 1) \leq \Theta^* 0 \)

E_Ratio: \( x_1^* \text{DMU}_1 + x_2^* \text{DMU}_2 + \ldots + x_{10}^* \text{DMU}_{10} - x_{51}^* (\Theta^* 1) \leq \Theta^* 0 \)

Outputs

NAV: \( x_1^* \text{DMU}_1 + x_2^* \text{DMU}_2 + \ldots + x_{10}^* \text{DMU}_{10} - x_{51}^* (\Theta^* 0) \geq \Theta^* 1 \)

Cap_Gain: \( x_1^* \text{DMU}_1 + x_2^* \text{DMU}_2 + \ldots + x_{10}^* \text{DMU}_{10} - x_{51}^* (\Theta^* 0) \geq \Theta^* 1 \)

Dividend: \( x_1^* \text{DMU}_1 + x_2^* \text{DMU}_2 + \ldots + x_{10}^* \text{DMU}_{10} - x_{51}^* (\Theta^* 0) \geq \Theta^* 1 \)

All Variables \( \geq 0 \)

where \( \Theta \) is current decision-making unit (DMU)

Linear program solved by SAS for each DMU.
**DEA Code**

**Part 1: Initialization**

This section reads the data files into a SAS dataset. Using the SAS Data Management Language datasets are created for the right hand side vectors, as well as the theta vectors. The two datasets are perfectly matched — each right hand side has exactly one corresponding theta vector. There is one right hand side and one theta vector for each DMU in the study. These vectors are merged with the inputs and outputs to formulate the linear program. The final step in the initialization is the renaming of variables to the proper ticker symbols in order to aid in report interpretation.
Part 2: DEA Macro

This macro accepts one parameter: the index of the DMU to evaluate. The macro merges the input/output dataset with the appropriate theta and right hand side vectors. This dataset now contains the LP formulation for SAS to analyze. The procedure generates detailed printed reports for each of the DMUs.
Part 3: Loop Macro

In order to iterate the execution of a SAS procedure without repetitive statements a macro was written to loop around the calls to the DEA macro. The loop macro accepts one parameter: the number of DMUs in study.

%MACRO LOOP(LIMIT);
%DO INDEX I %TO &LIMIT;
  %DEA(&INDEX);
%END;
%MEND LOOP;

Part 4: Extract Macro

In addition to the written reports generated by the DEA macro, datasets were created in memory containing the results of the LPs. This macro combines the reports for each of the DMUs and extracts the DEA scores into a dataset which is then fed to the forecasting model.

%MACRO EXTRACT;
  MERGE LP OUTPUT DATASETS AND REMOVE UNIMPORTANT VARIABLES /
  DATA RES1;
  SET #SYMBOLS;
  KEEP _VAR_ _VALUE_;
  /* EXTRACT OBJECTIVE FUNCTION VALUES */
  DATA RES2;
  SET RES1;
  IF _VAR_ = "OBJECT" THEN SCORE = _VALUE_;
  ELSE DELETE;
  KEEP SCORE;
  /* GENERATE LISTING OF MUTUAL FUNDS */
  DATA RES3;
  SET TICK;
  ARRAY TICKRI[50] $5;
  DO I = 1 TO 50; /*CHANGE THIS TO 50 IN PRODUCTION RUN */
    NAME = TICKRI[I];
    OUTPUT;
  END;
  KEEP NAME;
  /* MERGE DATASET INTO FINAL FORM */
  DATA RESULTS;
  MERGE RES2 RES3;
  PROC PRINT DATA=RESULTS;
  %MEND EXTRACT;
  %EXECUTABLE CODE
  %LOOP(50);
  %EXTRACT;
LINKED LINEAR REGRESSION — DEA MODEL

Forcasting Code

"* Mutual Fund Data Files */
filename dati 'mfs93 data a'
filename dat2 'mfname data a'
filename dat5 'mfs90 data a'

"* DEA Score Data Files */
filename dat3 'score93 data a'
filename dat4 'score90 data a'

option linesize = 72 nonumber nodate;

"* Read 1993 Mutual Fund Data */
data one;
  infile dati;
  array val(50);
  input name $ val{};
proc transpose; /* Transpose Matrix: from I0*DMU to DMU*I0 */
run;
proc datasets; /* Assign proper names to I0s */
modify data1;
rename col1 = Beta
col2 = Load
col3 = E_Ratio
col4 = NAV
col5 = Cap_Gain
col6 = Dividend
_Name_ = Fund;
data funds; /* Create List of Fund Names */
  infile dat2;
  input Fund $;
data scr93; /* Read 1993 DEA scores from file */
  infile dat3;
  input dea93;
title1 '1993 DEA Analysis';
data comp93;
  set data1;
  merge data1 funds scr93;
  run;

"* Run All possible regression models */
proc reg;
title2 'All Possible Regression Models';
model dea93 = Beta Load E_Ratio NAV Cap_GAIN Dividend/
          selection = Rsquare;
proc reg;
title2 'Full Model';
model dea93 = Beta Load E_Ratio NAV Cap_GAIN Dividend;

"* Forecasting Validation Model */
"* Create predicted values using forecast model compare to actual 1993 values. */
LINKED LINEAR REGRESSION — DEA MODEL

data predict;
  set data;
  Exp_dea=0.177220*E_Ratio + 0.016477*NAV +
0.152344*Cap_GAIN+0.279313;
  if Exp_dea > 1 then Exp_dea = 1;

data forecast;
  merge predict funds scr93 scr90;
  Err = dea93 - exp_dea;
  keep fund dea90 exp_dea dea93 err;

title 'Regression Predictions';
title2'Non-Linear Model';
proc print data=forecast;
  var fund dea90 dea93 exp_dea err;
proc plot;
  plot exp_dea * dea93;
/* added 4/18/94 */
/* Plot 93 Prediction vs 90 Actual */
data forecast;
  merge scr90 forecast;
  keep exp_dea dea90 dea93 fund;

title1 'Predicted 93 vs. Actual 1990';
title2 'Non-Linear Model';
proc plot data=forecast;
  plot exp_dea * dea90;
/* Linear transform to improve predictions */
data linear;
  set comp90;
  Ldea = log(dea90);
  LNAV = log(NAV);
  if Cap_GAIN = 0 then LCG = 0;
  else LCG = log(Cap_GAIN);
  LER = log(E_Ratio);
  if Dividend = 0 then LDIV = 0;
  else LDIV = LOG(Dividend);

/* Transform forecast inputs to linear model */
data trans93;
  set comp93;
  Ldea = log(dea93);
  LNAV = log(NAV);
  if Cap_GAIN = 0 then LCG = 0;
  else LCG = log(Cap_GAIN);
  LER = log(E_Ratio);
  if Dividend = 0 then LDIV = 0;
  else LDIV = LOG(Dividend);

Title 'Linear Forecast Model';
title2 'Model Verification';
proc reg data = linear;
model Ldea = Beta Load E_Ratio LNAV LCG LDIV;
plot residual.*(predicted.);
proc reg data = linear;
model Idea = Beta E_Ratio LNAV LDIV;
/* Linear Regression Equation */
data predict2;
  set trans93;
  PDEA = 0.667459*Beta - 0.525656*E_Ratio + 0.753563*LNAV +
0.229900*LDIV - 2.930810;
  Exp_dea = exp(PDEA);
  if Exp_dea > 1 then Exp_dea = 1;
/* Compute Forecast Errors */
data frcst2;
  merge predict2 funds scr93 scr90;
  Err = dea93 - exp_dea;
  keep fund dea90 exp_dea dea93 err;

title 'Regression Predictions';
title2 'Linear Model';
proc print data=frcst2;
  var fund dea90 dea93 exp_dea err;
title 'Predicted 93 vs. Actual 1990';
title2 'Non-Linear Model';
proc plot data=frcst;
  plot exp_dea * dea90;
/* Linear transform to improve predictions */
data linear;
  set comp90;
  Ldea = log(dea90);
  LNAV = log(NAV);
  if Cap_GAIN = 0 then LCG = 0;
  else LCG = log(Cap_GAIN);
  LER = log(E_Ratio);
  if Dividend = 0 then LDIV = 0;
  else LDIV = LOG(Dividend);
Case Study:
A Linked Linear Regression – Data Envelopment Analysis Model for the Long-Term Prediction of Mutual Fund Performance

Cindy Chao, Greg Higgins
Michael McGee

Department of Computer Science and Engineering
SOUTHERN METHODIST UNIVERSITY
Dallas, Texas

May 14, 1994
Contents

Abstract ............................................................................................................................. 3
Introduction ...................................................................................................................... 3
What is a Mutual Fund? ................................................................................................. 4
DEA Factors ..................................................................................................................... 4
Model Development and Validation ............................................................................. 6
Examination of Results ................................................................................................... 9
Investment Performance ............................................................................................... 10
Conclusion ...................................................................................................................... 11
References ..................................................................................................................... 12
Professional References ............................................................................................... 13
Appendix .......................................................................................................................... 14
  DEA Formulation ........................................................................................................ 15
  DEA Code .................................................................................................................... 16
  Forcasting Code .......................................................................................................... 19
Case Study:
A Linked Linear Regression – Data Envelopment Analysis Model for the Long-Term Prediction of Mutual Fund Performance

Michael McGee
Greg Higgins
Cindy Chao
Department of Computer Science and Engineering
Southern Methodist University
Dallas, Texas 75275

Data Envelopment Analysis (DEA) can be used to evaluate managerial efficiency of mutual funds. By extending the DEA methodology with linear regression, a new prediction tool for future efficiency is developed. The linked model was tested on a sample of fifty growth mutual funds over a three year time horizon — resulting in accurate predictions of future efficiency with 99% confidence.

The current mutual fund market contains nearly $1.4 trillion dollars; this rapid growth of mutual funds as a long-term investment instrument has led to an increased demand for evaluation and forecasting techniques. The performance of a mutual fund is directly linked to the performance of the management staff. Data Envelopment Analysis (DEA) can be used to evaluate managerial efficiency; however, this is only a current assessment of a fund's management. DEA is a linear programming methodology that computes relative efficiency scores among a group of homogeneous decision-making units (DMUs). Combined with linear regression, DEA becomes a prediction tool for long-term forecasting. The linked model not only evaluates return on investment but also the quality of the fund's managerial team on a long-term basis.
What is a Mutual Fund?

A mutual fund is an investment company that pools the money of many institutional and individual investors into one professionally managed investment account. The fund manager creates a portfolio for the shareholders by buying a variety of "true" financial instruments, such as corporate stocks and bonds, and government and municipal securities, in accordance with the fund's particular investment objectives. Fund managers are professional investors who know the financial market and follow it daily with extensive research capabilities to effectively investigate all investment opportunities. The fund is always monitored, with individual securities constantly being bought, sold, and traded in order to optimize the investment objectives. The beauty of investing in sometimes twenty different securities lies in the protection of the investment through portfolio diversification.

Mutual funds are divided into six major categories according to their investment objectives, which are outlined in the fund's prospectus, such as goals, level of risk, and the type and frequency of return desired. Since the goal of the linked model is forecasting of long-term performance, growth mutual funds were selected because their objectives emphasize the accumulation of capital gains, and they have the greatest long-term growth potential.

DEA Factors

The inputs and outputs of the envelopment analysis were made up of the main characteristics of a mutual fund. They were selected because they are important factors to investors and are directly linked to the performance of the fund manager. The inputs selected were beta — a measure of risk, expense ratio, and loads & other fees. The outputs used were net asset value (NAV) per share, dividends, and capital gains.

Inputs

Beta is a measure of the fund portfolio's average volatility, or risk, relative to the S&P 500. The beta value of the market is a constant equal to 1.00. A fund with a beta value greater than 1.00 is more volatile than the market. It is expected to outperform the "average fund" in a bull (rising) market and bring in greater returns, while it is expected to fall more than
the average fund in times of a bear (declining) market. A fund with a beta value less than 1.00 possesses the characteristics of the converse situation. For example, a portfolio with a beta value of 0.75 is expected to be 25% less responsive to the general economic trend than the average fund and return 25% more conservative than the average fund.

Expense ratio is a measure which relates expenses to the average net assets for a period. It is calculated by taking the annual expenses paid by the fund (including management and investment advisory fees, custodial and transfer agency charges, legal fees, and distribution costs) divided by the average shares outstanding for the period. The greater the expense ratio, the larger the percentage of the shareholder’s investment gets paid out directly to the management company. The expense ratio does not include loads or commissions paid for sales or transaction fees.

Loads and other fees represent all other charges made to the shareholder which are not accounted for in the expense ratio. A load is a portion of the offering price of the fund that pays for costs such as sales commissions, brokerage costs, and transaction fees. This value is a percentage of the purchase price taken off either before the money is invested (front-end load) or after investment when the shareholder redeems the shares (back-end load). The amount of the load is established by the managerial company, but is paid to the sales and transaction team. Some funds have a load of 0.00%, meaning that they have the same benefits as load funds, except they are sold directly by the fund company rather than through a sales broker so that no loads or other fees are incurred. With no-load funds, the investor must seek out and contact the investment firm directly. There is a distinct advantage to investing in no-load funds. For example, if we had invested $1000 in a particular portfolio with a 7.50% front-end load, $75 would be taken off the top, and we would be left with only $925 worth of shares in the fund. Had this been a no-load fund, the full $1000 would have been utilized.

Outputs

Net asset value per share represents how much each share is worth. It is calculated by taking the total market value of a fund’s shares (securities plus cash plus accrued earnings minus
liabilities) divided by the number of shares outstanding. No profits or losses are realized through NAV until the fund is sold. Interest lies in the amount of increase or decrease in NAV from time of purchase to time of sale.

Dividends are payments declared by a corporation's board of directors and made to shareholders, usually on a quarterly basis. A fund's dividends come directly from the dividends paid out by the individual securities within the portfolio. They are paid out equally among the shareholders on a scheduled time frame.

Capital gains on a mutual fund are the profits made from the sale of a capital asset (such as a security, stock, or bond) within the portfolio. The number used for capital gains is adjusted for losses and stock splits. The sale of an asset occurs when the manager decides that a security needs to be sold for the benefit of the portfolio. Capital gains are distributed to shareholders when they are realized (when a security is sold and a profit is made) and are paid out as capital distributions.

Dividends and capital gains are treated as separate outputs because they are incurred, paid out, and taxed in different ways. Shareholders can reinvest dividends and capital distributions if they elect to do so.

Model Development and Validation

The decision-making units (DMUs) in this analysis were the 50 growth mutual funds. A prerequisite of DEA is homogeneous DMUs. By selecting the funds from the same sector (growth funds), this requirement is met.

The DEA model uses the dual formulation of the linear program as described by Charnes, Cooper and Rhodes (1978). The objective function of this form is to minimize the objective value of the current DMU. If a DMU is efficient the optimal value is 1, while an inefficient DMU is driven towards zero. The first step in the development of the linked model is the generation of DEA scores for the mutual funds being studied. The data collected for the mutual funds is used as the inputs and outputs for the DEA code which is documented in the appendix. The DEA code is written using the SAS/OR package allowing the DEA scores to be fed directly into the regression phase of the model.

DEA scores are generated for the fifty mutual funds in the study using
the 1990 data collected from Morningstar Mutual Funds. At this point DEA scores are also generated for 1993 and stored for use in validating the model's forecasts. The 1990 DEA scores are then regressed against all of the six DEA factors previously described.

This regression model is unique from previous research in that the goal of the regression is not to create an efficiency score, but to forecast an efficiency score for the future of a particular DMU. Previous work by Thanassoulis (1993) compared the DEA methodology with linear regression as a means of creating efficiency scores. Our model links the two methods into one prediction tool. The full factor model that resulted appeared on the surface to be a good predictor of efficiency ($R^2 = 0.8239$). The full factor model included three non-significant factors at the 5% level of significance. Beta ($p$-value = 0.68), load ($p$-value = 0.65), and dividend ($p$-value = 0.50) were all eliminated from the model leaving only expense ratio, net asset value, and capital gains as significant factors. The revised model remained significant with an $R^2 = 0.8213$.

As stated before, this model appears to be a good predictor; however, linear regression models must be validated through diagnostic tests. One assumption in regression analysis underpins the entire methodology — errors are a randomly distributed variable following a normal distribution with mean equal to zero, $N(0, \sigma^2)$. Regression analysis in SAS is based on a linear relationship between the dependent and independent variables. These assumptions can be verified through the examination of the residuals, or errors, of the regression model. A plot of the residuals versus the predicted values, Figure 1, reveals a distinct pattern indicating a non-linear relationship between the independent and dependent variables. This requires the examination of univariate statistics.
for each of the variables in the regression model. A standard normal quantile plot for each of the variables allows the determination of the distribution of each variable.

From the univariate procedures, some of the variables demonstrated an exponential distribution requiring a log-linear transformation in order to validate the regression model. Net asset value, capital gains, dividends and the actual DEA score follow exponential distributions while beta, load, and expense ratio are all normally distributed. A second data set of the inputs and outputs was created by taking the natural log of each of the exponentially distributed variables. Once normality was confirmed through univariate procedures, a new full factor linear regression model was fitted. The newly formed model resulted in an $R^2 = 0.8163$ and was validated through the residual plot shown in Figure 2. Using the SAS regression report, the regression equation was created using the coefficients of the significant factors: beta, expense ratio, net asset value, and dividend.

The next step is the creation of predicted efficiency scores. The common method for forecast model verification is the prediction of the past. The linked model is verified by using the input and output data for 1993 in the regression equation created from the 1990 data. The predicted values are then compared to the actual 1993 DEA scores generated earlier by the DEA code. The forecast errors are evaluated using a significance test on the mean difference. The test is conducted at the 1% level of significance. $H_0$: mean difference = 0; $H_a$: mean difference ≠ 0. The resulting test statistic: $z = -0.5164$ and critical value, $z = 2.57$ leads to the conclusion that the prediction values are not significantly different from the actual DEA scores with 99% confidence.
Table 1: Efficient Mutual Funds

<table>
<thead>
<tr>
<th>Fund</th>
<th>DEA 1990</th>
<th>Predicted 1993</th>
<th>DEA 1993</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>FTRNX</td>
<td>1.00000</td>
<td>1.00000</td>
<td>0.62320</td>
<td>-0.37680</td>
</tr>
<tr>
<td>PRWAX</td>
<td>1.00000</td>
<td>0.78983</td>
<td>0.29415</td>
<td>-0.49568</td>
</tr>
<tr>
<td>GINLX</td>
<td>1.00000</td>
<td>0.33207</td>
<td>0.23691</td>
<td>-0.09516</td>
</tr>
<tr>
<td>ACRNX</td>
<td>1.00000</td>
<td>1.00000</td>
<td>1.00000</td>
<td>0.00000</td>
</tr>
<tr>
<td>FDETX</td>
<td>0.41944</td>
<td>0.98900</td>
<td>1.00000</td>
<td>0.01100</td>
</tr>
<tr>
<td>CFIMX</td>
<td>0.73415</td>
<td>0.98032</td>
<td>1.00000</td>
<td>0.01968</td>
</tr>
<tr>
<td>LOMCX</td>
<td>0.32965</td>
<td>0.68927</td>
<td>1.00000</td>
<td>0.31073</td>
</tr>
</tbody>
</table>

Examination of Results

In analyzing the results of the sample studied, it is important to remember that a mutual fund is mainly used as a long-term investment, providing stable low-risk growth. A potential shareholder wants to invest in a fund that will be efficient in the future, regardless of its current efficiency. Therefore we focus our examination on the set of funds with efficient scores in either 1990 or 1993. Table 1 shows the 1990 DEA, Predicted 1993, and actual 1993 DEA scores along with the errors for each prediction (Error = 1993 DEA - Predicted 1993).

Of the first four funds listed (all receiving efficient ratings in 1990), the future efficiency of three funds was correctly predicted. Our linked model predicted future behavior of currently efficient funds with 75% accuracy. Although the error for T. Rowe Price New America Fund, PRWAX, is high, we are not concerned with the magnitude of inefficiency, only the state of its efficiency.

Of the last four funds in the table (all receiving efficient ratings in 1993), three funds were predicted accurately. The fifth and sixth funds, Fidelity Destiny Portfolios: Destiny II, FDETX, and Clipper Fund, CFIMX, both received a predicted score of approximately 0.98 for 1993. Based on the previous significance test, these values are not significantly different from 1.00. Thus our model predicts the funds that will be efficient in the future with 75% accuracy.
Since investors are concerned with the long-term growth and appreciation of their investments, DEA by itself is not an effective tool to aid in the development of investment strategies. However, combined with linear regression, DEA gives us the ability to estimate the long-term efficiency of a fund. To illustrate the effectiveness of the linked model relative to DEA alone, we simulated an investment in the efficient funds in 1990 and the funds predicted to be efficient by our linked model. Since mutual funds allow the purchase of fractional shares, we invested $1000 in each of the mutual funds. The results are given in Table 2.

The percent of appreciation we received in the portfolio of 1990 efficient mutual funds was 46.92%, while the appreciation was significantly higher for the portfolio of funds recommended by the linked model at 60.41%.

### Table 2: Investment Simulation Results

<table>
<thead>
<tr>
<th>1990 DEA Efficient</th>
<th>Investment ($)</th>
<th>1990 NAV per share</th>
<th>Number of Shares</th>
<th>1993 NAV per share</th>
<th>Value</th>
<th>Percent Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>FTRNX</td>
<td>$1000</td>
<td>$38.25</td>
<td>26.14</td>
<td>$1000.00</td>
<td>$59.08</td>
<td>$1544.35</td>
</tr>
<tr>
<td>PRWAX</td>
<td>$1000</td>
<td>$14.66</td>
<td>68.21</td>
<td>$1000.00</td>
<td>$28.34</td>
<td>$1933.07</td>
</tr>
<tr>
<td>GINLX</td>
<td>$1000</td>
<td>$66.84</td>
<td>14.96</td>
<td>$1000.00</td>
<td>$16.60</td>
<td>$248.33</td>
</tr>
<tr>
<td>ACRNX</td>
<td>$1000</td>
<td>$32.57</td>
<td>30.70</td>
<td>$1000.00</td>
<td>$70.07</td>
<td>$2151.15</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>$4000.00</td>
<td></td>
<td></td>
<td>$5876.90</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regression Efficient</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACRNX</td>
<td>$1000</td>
<td>$32.57</td>
<td>30.70</td>
<td>$1000.00</td>
<td>$70.07</td>
<td>$2151.15</td>
</tr>
<tr>
<td>FDETX</td>
<td>$1000</td>
<td>$18.97</td>
<td>52.71</td>
<td>$1000.00</td>
<td>$27.15</td>
<td>$1431.08</td>
</tr>
<tr>
<td>CFIMX</td>
<td>$1000</td>
<td>$38.80</td>
<td>25.77</td>
<td>$1000.00</td>
<td>$50.05</td>
<td>$1289.79</td>
</tr>
<tr>
<td>FTRNX</td>
<td>$1000</td>
<td>$38.25</td>
<td>26.14</td>
<td>$1000.00</td>
<td>$59.08</td>
<td>$1544.35</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>$4000.00</td>
<td></td>
<td></td>
<td>$6416.37</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Investment Performance**

Since investors are concerned with the long-term growth and appreciation of their investments, DEA by itself is not an effective tool to aid in the development of investment strategies. However, combined with linear regression, DEA gives us the ability to estimate the long-term efficiency of a fund. To illustrate the effectiveness of the linked model relative to DEA alone, we simulated an investment in the efficient funds in 1990 and the funds predicted to be efficient by our linked model. Since mutual funds allow the purchase of fractional shares, we invested $1000 in each of the mutual funds. The results are given in Table 2.
Conclusion

The linked model allows us to broaden the range of applications for DEA. The linked model enables the prediction of future efficiency, which is of great benefit to investors because it provides more insight into the long-term viability of a mutual fund. The model also requires little historical data for each DMU unlike other forecasting methods. Another benefit is the relative compactness of the model; very few predictors are required to explain a high degree of variability and achieve accurate predictions. Data envelopment analysis by itself is not an accurate predictor of future efficiency of mutual funds. Only one fund that was efficient in 1990 remained in the efficient set in 1993. Linear regression and DEA combine to give us an accurate prediction for the future behavior of DMUs.

An immediate extension to this research is the definition of an "ideal type" mutual fund. An optimization algorithm will be used to generate values for the regression factors and record the combinations which yield and efficient fund prediction. Once this benchmark is defined a growth mutual fund prospectus can simply be visually inspected to determine managerial efficiency. The next phase of this research would be to compare the accuracy of the linked model with that of other forecasting methods, such as time series decomposition. Another step in the continued development of this area of research is to improve the accuracy of the predictions through further modifications to the regression model using a weighted regression model or the inclusion of adaptive error control.
References


Morningstar Mutual Funds (1993), Chicago: Morningstar, Inc.


Professional References

Barr, Richard S., Ph.D. Associate Professor, School of Engineering and Applied Sciences. Southern Methodist University, Dallas, Texas.

Dula, Jose, Ph.D. Assistant Professor, School of Engineering and Applied Sciences. Southern Methodist University, Dallas, Texas.

Guerra, Rudy, Ph.D. Assistant Professor, Department of Statistical Science. Southern Methodist University, Dallas, Texas.

Welch, David. Senior Systems Analyst II, Bradfield Computer Center. Southern Methodist University, Dallas, Texas.
DEA Formulation

*Mathematical formulation of the DEA model used in the Linked Model.*

DEA Code

*The DEA code is developed using Release 6.0 of the SAS System on an IBM 3090 mainframe. The code presented requires only two input files. One file contains the input and output factor data while the second file contains a listing of the mutual fund ticker symbols. The program is constructed from a series of SAS Macros which fully automate the formulation and solution of the multiple linear programs required in Data Envelopment Analysis.*

Forecasting Code

*Developed as an extension to the DEA code, the forecasting model uses the same computing platform and software. The program reads the same data files as the DEA code as well as the DEA scores generated by the DEA study. The program generates predicted efficiency scores and all necessary statistical procedures for validating the model.*
DEA Formulation

\[ A \begin{bmatrix} \lambda \\ \vdots \end{bmatrix} + \begin{bmatrix} \text{Slacks} \end{bmatrix} = \begin{bmatrix} \theta \\ \vdots \end{bmatrix} \]

Min \( \theta \)
S.T.
\[ \sum A_j^x \lambda_j - A_0^x \theta \leq 0 \]
\[ \sum A_j^y \lambda_j \geq A_0^y \]
\[ \lambda_j \geq 0, \quad \theta \geq 0 \]

Mathematical formulation of DEA requires that the inputs and outputs for each DMU are formatted in a matrix, \( A \). The matrix is multiplied by a vector, \( \lambda \), to construct the linear programming constraints.

Minimize \( \theta \)
Subject to:

Inputs
- Beta: \( x_1^{DMU_1} + x_2^{DMU_2} + \ldots + x_{50}^{DMU_{50}} - \bar{x}_1 \theta \leq \theta \cdot 0 \)
- Load: \( x_1^{DMU_1} + x_2^{DMU_2} + \ldots + x_{50}^{DMU_{50}} - \bar{x}_2 \theta \leq \theta \cdot 0 \)
- E_Ratio: \( x_1^{DMU_1} + x_2^{DMU_2} + \ldots + x_{50}^{DMU_{50}} - \bar{x}_3 \theta \leq \theta \cdot 0 \)

Outputs
- NAV: \( x_1^{DMU_1} + x_2^{DMU_2} + \ldots + x_{50}^{DMU_{50}} - \bar{x}_6 \theta \geq \theta \cdot 1 \)
- Cap_Gain: \( x_1^{DMU_1} + x_2^{DMU_2} + \ldots + x_{50}^{DMU_{50}} - \bar{x}_7 \theta \geq \theta \cdot 1 \)
- Dividend: \( x_1^{DMU_1} + x_2^{DMU_2} + \ldots + x_{50}^{DMU_{50}} - \bar{x}_8 \theta \geq \theta \cdot 1 \)

All Variables \( \geq 0 \)

where \( \theta \) is current decision-making unit (DMU)

Linear program solved by SAS for each DMU.
DEA Code
Part 1: Initialization

This section reads the data files into a SAS dataset. Using the SAS Data Management Language datasets are created for the right hand side vectors, as well as the theta vectors. The two datasets are perfectly matched — each right hand side has exactly one corresponding theta vector. There is one right hand side and one theta vector for each DMU in the study. These vectors are merged with the inputs and outputs to formulate the linear program. The final step in the initialization is the renaming of variables to the proper ticker symbols in order to aid in report interpretation.
Part 2: DEA Macro

This macro accepts one parameter: the index of the DMU to evaluate. The macro merges the input/output dataset with the appropriate theta and right hand side vectors. This dataset now contains the LP formulation for SAS to analyze. The procedure generates detailed printed reports for each of the DMUs.
Part 3: Loop Macro

In order to iterate the execution of a SAS procedure without repetitive statements a macro was written to loop around the calls to the DEA macro. The loop macro accepts one parameter: the number of DMUs in study.

Part 4: Extract Macro

In addition to the written reports generated by the DEA macro, datasets were created in memory containing the results of the LPs. This macro combines the reports for each of the DMUs and extracts the DEA scores into a dataset which is then fed to the forecasting model.
LINKED LINEAR REGRESSION — DEA MODEL

Forcasting Code

/* Mutual Fund Data Files */
filename dat1 'm1s93 data a';
filename dat2 'mfname data a';
filename dat5 'mfsgo data a';

/* DEA Score Data Files */
filename dat3 'score93 data a';
filename dat4 'score90 data a';

option linesize = 72 nonumber nodate;

/* Read 1993 Mutual Fund Data */
data one;
  infile dat1;
  array val(50);
  input name $ val{1};
proc transpose; /* Transpose Matrix: from IODMU to DMU*IO */
  run;
proc datasets; /* Assign proper names to lOs */
  modify data 1;
  rename col1 = Beta
           col2 = Load
           col3 = E_Ratio
           col4 = NAV
           col5 = Cap_Gain
           col6 = Dividend
       _Name_ = Fund;
  data funds; /* Create List of Fund Names */
    infile dat2;
    input Fund $;
proc datasets; /* Assign proper names to IOs */
  modify data1;
  rename col1 = Beta
           col2 = Load
           col3 = E_Ratio
           col4 = NAV
           col5 = Cap_Gain
           col6 = Dividend
       _Name_ = Fund;
data scr93; /* Read 1993 DEA scores from file */
  infile dat3;
  input dea93;
title1 '1993 DEA Analysis';
data comp93;
  set data1;
  merge data1 funds scr93;

/* Run All possible regression models */
proc reg;
  title2 'All Possible Regression Models';
  model dea93=Beta Load E_Ratio NAV Cap_Gain Dividend/
             selection=Rsquare;
proc reg;
  title2 'Minimum Spec Model';
  model dea93 = E_Ratio NAV CAP_GAIN;

/* 1990 Data Analysis */
data info90;
  infile dat5;
  array val(50);
  input name $ val{1};
proc transpose; /* Transpose Matrix: from IO*DMU to DMU*IO */
  run;
proc datasets; /* Assign proper names to IOs */
  modify data2;
  rename col1 = Beta
           col2 = Load
           col3 = E_Ratio
           col4 = NAV
           col5 = Cap_Gain
           col6 = Dividend
       _Name_ = Fund;
  data scr90;
    infile dat4;
    input dea90;
  data comp90;
    merge data2 funds scr90;
title1 '1990 DEA Analysis';
proc reg data=comp90;
  title2 'All Possible Regressions';
  model dea90= Beta Load E_Ratio NAV Cap_Gain Dividend/
             selection = rsquare;
proc reg data=comp90;
  title2 'Full Model';
  model dea90= Beta Load E_Ratio NAV Cap_Gain Dividend;
proc reg data=comp90;
  title2 'Minimum Spec';
  model dea90= E_Ratio NAV CAP_GAIN;
  plot residual. '(predicted. E_Ratio NAV Cap_gain);

/* Full Fit Model */
proc reg;
  title2 'Full Model';
  model dea93=Beta Load E_Ratio NAV Cap_Gain Dividend;

/* Forecasting Validation Model */
/* Create predicted values using forecast model compare to actual 1993 values. */
data predict;
set data;
Exp_dea=0.177220*E_Ratio + 0.016477*NAV + 0.152344*Cap_Gain + 0.279313;
if Exp_dea > 1 then Exp_dea = 1;

data forecast;
merge predict funds scr93 scr90;
Err = dea93 - exp_dea;
keep fund dea90 exp_dea dea93 err;

proc print data=forecast;
var fund dea90 dea93 exp_dea err;
proc plot;
plot exp_dea dea93;

/* added 4/18/94 */
/* Plot 93 Prediction vs 90 Actual */

data fcast;
merge scr90 forecast;
keep exp_dea dea90 dea93 fund;

title1 'Predicted 93 vs. Actual 1990';
title2 'Non-Linear Model';
proc plot data=fcast;
plot exp_dea * dea90;

/* Linear transform to improve predictions */
data linear;
set comp90;
Ldea = log(dea90);
LNAV = LOG(NAV);
if Cap_Gain = 0 then LCG = 0;
else LCG = LOG(Cap_Gain);
LER = LOG(E_Ratio);
if Dividend = 0 then LDIV = 0;
else LDIV = LOG(Dividend);

/* Transform forecast inputs to linear model */
data trans93;
set comp93;
Ldea = log(dea93);
LNAV = LOG(NAV);
if Cap_Gain = 0 then LCG = 0;
else LCG = LOG(Cap_Gain);
LER = LOG(E_Ratio);
if Dividend = 0 then LDIV = 0;
else LDIV = LOG(Dividend);

title 'Linear Forecast Model';
title2 'Model Verification';
proc reg data = linear;
model Ldea = Beta Load E_Ratio LNAV LCG LDIV;
plot residual.*(predicted.);
proc reg data = linear;
model Idea = Beta E_Ratio LNAV LDIV;

/* Linear Regression Equation */
data predict2;
set trans93;
PDEA = 0.687459*Beta - 0.525656*E_Ratio + 0.753563*LNavy + 0.229900*LDIV - 2.930810;
Exp_dea = exp(PDEA);
if Exp_dea > 1 then Exp_dea = 1;

/* Compute Forecast Errors */
data frcst2;
merge predict2 funds scr93 scr90;
Err = dea93 - exp_dea;
keep fund dea90 exp_dea dea93 err;

proc print data=frcst2;
var fund dea90 dea93 exp_dea err;