Data Translation Challenge: Final Project

By:
Anusha Makam Dileep Kumar
Arpita Debnath
Divya Ramesh
Geetha Sree Kariyavula Narasimha
Niki Gandhi

1. OBJECTIVE

The agenda of this project is to analyze from a CEO's perspective to understand and determine whether having a large amount of NOA increases or decreases the buy-and-hold return for a firm. A simple linear regression method has been used to regress through the NOA with annual buy-and-hold returns for data from 2000 to 2003. Eventually, we will evaluate and recommend a course of action based on the results from our model. Here, we are investigating the association between cash flows and accruals. Hence, the focus is on buy and hold returns rather than specific day returns.

The following formula has been used to calculate NOA in our model:

NOA = (SEQ - CHE +(DLC+DLTT))/SALE_LAG

Where, SEQ is Stockholders Equity (total), CHE is Cash and short-term Investments, DLC is Debt in current Liabilities and DLTT is long-term debts. The sales data is lagged by one year to calculate NOA. The data source for these variables is extracted from Compustat/Fundamentals/Balance Sheet entities. In other words, NOA is a metric which measures the operational performance of a firm.

2. THE MODEL

2.1 DATASET

The datasets used in this model have been downloaded from WRDS, a financial data repository, aggregated by the Wharton School of Business at the University of Pennsylvania.

To begin with, we extracted the data on cash and short term investments, shareholder's equity, total debt accounting for both long-term as well as short-term debts and market securities pertaining to the public companies corresponding to 4 different years between 2000 and 2003 from Compustat.The sales data was then extracted from the same for years 1999-2002, which lagged by an year for bolstering the computation the NOA for the years 2000-2003. The monthly returns data was extracted from CRSP for years 2000-2003.

2.2 STAGES OF MODEL DEVELOPMENT

After the initial process of fetching the required data, the next step included cleaning/preparing the individual datasets and merging them to help calculate the NOA.

The monthly buy-and-hold returns data originally consisted of 373,981 observations. As a part of data cleaning and preparation, we removed the rows with missing return values, special characters and non-numeric values. Initially, the firm's monthly returns appeared as a separate column of data which was then transformed into rows to aid in the calculation of buy-and-hold

returns. On computing the monthly buy-and-hold returns, the monthly returns were merged into annual and thus resulted in a reduced dataset size of 27,411 rows.

For the preparation of NOA data, we substituted zeros for nulls in all individual total debt, shareholder's equity, cash and marketable securities, and sales data and deleted non-numeric data and special characters. We believe that, these numeric fields could be 0 assuming that these firms underwent financial hardships and thus decided to retain them to prevent any kind of bias. At this stage, we removed outliers for four variables: SEQ, CHE, DLC, and DLTT and lagged SALE. The next step was to create a modified year variable in the Sales dataset which was basically fdyear+1 to create a variable to facilitate merging the CHE,SEQ,DLC,DLTT, along with the lagged SALE variables together by sorting and grouping by PERMNO and fdyear, so that one firm's sales do not overlap with another firm's sales for the first year. After this, we obtained 24183 observations in the cleaned NOA dataset.

The next steps involved merging the returns and NOA datasets into one single entity to get it ready for running the simple regression model. We deleted the null values of NOA results from the final merged dataset.We deleted the null values and there were 720 missing values found in the dataset. We are all set for regression.

2.3 VARIABLES USED

We have used the following variables in our model:

PERMNO - PERMNO can be defined as a unique stock level identifier. It is a unique permanent security identification number. It is seen as a share class level identifier as well. PERMNO for most firms is unique and has one class shares. But few companies have more than one class of shares which are traded at different prices. Hence there could be more than one PERMNO.

Year - The data used in this model has been selected from 2000 to 2003 which gives enough room to evaluate and draw conclusions.

CHE - CHE can be defined as Cash and Short Term investments. They are considered to be very liquid assets. These are the assets that can be used during a situation that needs immediate money transfers. Cash and Short Term investments are usually used in liquidity ratios.

DLC - DLC can be defined as Debt in Current Liabilities. They are also known as Short term debts or current liabilities. These are the financial commitments that are ideally expected to be paid off within a year.

DLTT - This can be defined as long-term debts or in financial terms it can be named "Financing debt". These liabilities usually have a maturity date greater than 12 months. Usually, investors

who invest in these liabilities go for long-term debts as it offers certain benefits in terms of interest payments and also considers the time to maturity a liquidity risk.

SEQ - SEQ stands for total stockholder's equity. This is the company's remaining value after liabilities are subtracted from assets.

SALE - The sales data is lagged by one year to calculate NOA.

3. FINDINGS AND FUTURE SCOPE

3.1 RESULTS

The SAS System

The CORR Procedure

2 Variables:	NOA finalret

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
NOA	24183	1.77877	3.31354	43016	-0.92396	36.47059
finalret	23463	0.21027	1.03570	4934	-0.99758	32.00000

Pearson Correlation Coefficients
Prob > |r| under H0: Rho=0
Number of Observations

	NOA	finalret
NOA	1.00000 24183	-0.03784 <.0001 23463
finalret	-0.03784 <.0001 23463	1.00000 23463

The SAS System

The REG Procedure
Model: MODEL1
Dependent Variable: finalret

Number of Observations Read	24183
Number of Observations Used	23463
Number of Observations with Missing Values	720

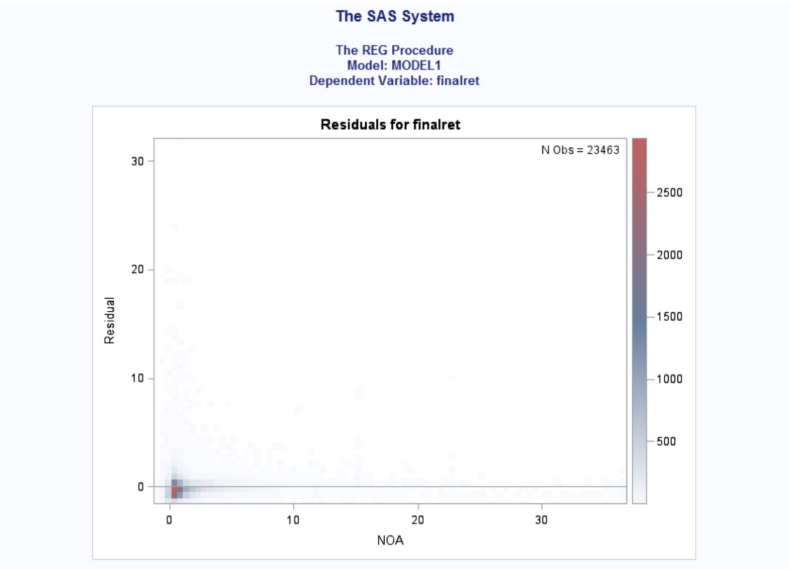
Analysis of Variance

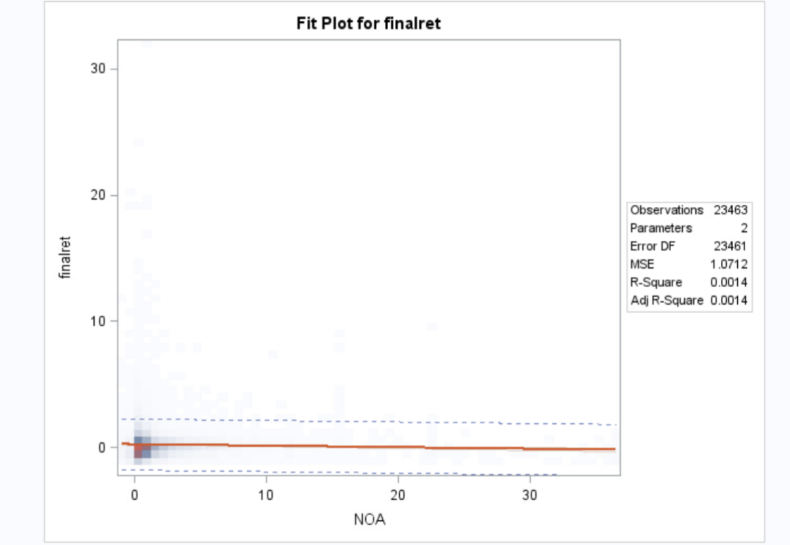
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	36.03480	36.03480	33.64	<.0001
Error	23461	25131	1.07118		
Corrected Total	23462	25167			

Root MSE	1.03498	R-Square	0.0014
Dependent Mean	0.21027	Adj R-Sq	0.0014
Coeff Var	492.20219		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	0.23152	0.00769	30.12	<.0001
NOA	1	-0.01186	0.00205	-5.80	<.0001





We used simple linear regression to capture the effect between NOA and buy and hold returns data of the firm. The R-square value of the model is 0.0014. R-Squared (R^2 or the coefficient of determination) is a statistical measure in a regression model that determines the proportion of variance in the dependent variable that can be explained by the independent variable. R-squared shows how well the data fit the regression model (the goodness of fit). So, we can depict here that there is a negative relationship between these two indicators (NOA and buy and hold returns data) which is explained by the model. From this we can observe that the NOA and buy and hold returns are negatively correlated with a co-effiucient of -0.03784.

The parameter estimate of NOA is -0.01186 in the model with a p-value of <0.0001. In null-hypothesis significance testing, the p-value is the probability of obtaining test results at least as extreme as the result observed, under the assumption that the null hypothesis is

correct. A p-value lesser than 0.0001 means that NOA is statistically significant. So we can conclude NOA has statistical significance in the model for of explaining the returns. The fit plot of NOA and final return also validate the same conclusion.

3.2 AREAS OF IMPROVEMENT

Due to time constraints, we were only able to design a simple linear regression model. Linear regression focuses on linear relationships between dependent and independent variables. It assumes there is a straight-line relationship between them and fails to suit complex datasets properly. Here, we work with real-world data for firms, and the relationship between the variables of the dataset cannot always be linear. Hence, a straight line might not fit the data accurately. In the future, we plan on implementing multiple linear regression models to research and capture the complex effects which might be associated with the data. We plan on using logarithmic data to enhance our current model.

Secondly, we have a limited amount of data. We have only used 2000-2003 data. Three years of data is not sufficient enough to find or establish the relationship between NOA and buy and hold returns of a company. Hence, we plan on performing this process on a larger dataset.

REFERENCES

Wikipedia contributors. (2021, February 16). Net operating assets. Wikipedia. Retrieved September 7, 2022 https://en.wikipedia.org/wiki/Net_operating_assets

Net Operating Assets: Formula and Calculator [Excel Template]. (2022, April 4). Wall Street Prep. Retrieved September 7, 2022
https://www.wallstreetprep.com/knowledge/net-operating-assets/

Cash and Short-Term Investments Definition. (n.d.). YCharts. Retrieved September 7, 2022 https://ycharts.com/glossary/terms/cash_on_hand#:%7E:text=Cash%20and%20short%20term%20investments%20are%20considered%20very,term%20investments%20are%20frequently%20used%20in%20liquidity%20ratios.

Short-Term Debt Definition. (2020, October 31). Investopedia. Retrieved September 7, 2022 https://www.investopedia.com/terms/s/shorttermdebt.asp#:%7E:text=Short-term%20debt%2C%20also%20called%20current%20liabilities%2C%20is%20a,payable%2C%20wages%2C%20lease%20payments%2C%20and%20income%20taxes%20payable.

Johnson, R. (2019, April 25). What Does Total Stockholders Equity Represent? Finance - Zacks. Retrieved September 7,2022
https://finance.zacks.com/total-stockholders-equity-represent-4898.html

APPENDIX I: SAS FOLDER REFERENCE

Please refer to the following files in the SAS folder in the order it has been specified:

I. Datasets extracted from WRDS:
   1. Variables dataset from 2000 to 2003: NOA_variables.sas7bdat
   2. Sales dataset from 1999 to 2002: sale2000.sas7bdat
   3. Returns dataset from 2000 to 2003: returnsdatset.sas7bdat

II. SAS working:

A. Using variables dataset we cleaned the data for each of the variables (CHE, DLC, DLTT, SEQ) by removing special characters and non-numeric values
   1. CHE_cleaned_2000_03.sas – the following output files are created:
      - che2000.sas7bdat
      - che2001.sas7bdat
      - che2002.sas7bdat
      - che2003.sas7bdat
   2. DLC_cleaned_2000_03.sas – the following output files are created:
      - dlc2000.sas7bdat
      - dlc2001.sas7bdat
      - dlc2002.sas7bdat
      - dlc2003.sas7bdat
   3. DLTT_cleaned_2000_03.sas – the following output files are created:
      - dltt2000.sas7bdat
      - dltt2001.sas7bdat
      - dltt2002.sas7bdat
      - dltt2003.sas7bdat
   4. SEQ_cleaned_2000_03.sas – the following output files are created:
      - seq2000.sas7bdat
      - seq2001.sas7bdat
      - seq2002.sas7bdat
      - seq2003.sas7bdat

B. Using sales dataset we removed the outliers and non-numeric values
   1. sales_cleaned_2000_03.sas – sales data is lagged to the years 1999 to 2002 and the following output files are created:
      - sale1999.sas7bdat
      - sale2000.sas7bdat
      - sale2001.sas7bdat
      - sale2002.sas7bdat

C. Merge all the above cleaned data files with all the variables(CHE,DLC,DLTT,SEQ,SALE) and NOA calculation is carried out:
   1. Merge_and_NOA_calc.sas and it generates the file calc_noa.sas7bdat:
   Removes Outliers and cleans NOA data for each year from 2000 to 2003:

NOA_cleaned_2000_03.sas – the following output files are created:

- noa2000.sas7bdat
- noa2001.sas7bdat
- noa2002.sas7bdat
- noa2003.sas7bdat

The above NOA datasets are merged
NOA_Returns_merged.sas and the following file is generated:

- noa_final.sas7bdat

D. Using returns dataset calculates buy and hold returns and transforms vertical returns data (columns) to horizontal (12 months) rows data: projectreturns.sas and generates the following dataset:

- finalret.sas7bdat

E. It merges final returns and NOA final files- Final_model_post_merging_returns_NOA.sas and generates SAS Output_htm#IDX8.mht