A Statistical Model to Identify Factors Affecting Equity Price in the…

Electric Utilities Industry

Brianna L. Palmisano (X03625986)
St. John’s University, The Peter J. Tobin College of Business

Summer 2024

Introduction

This evaluation uses the stock prices of the companies within the Electric Utilities (EU) industry of the Energy sector of the S&P1500 and the indicators within that industry, to examine the linear relationship between these variables.

This includes the earnings per share (EPS), book value per share (BVPS), debt to total assets (DTA) and current ratio (CR) for each company within the EU industry.

Information regarding the stock prices of companies within this industry could be valuable for not only our research within the climate change or sustainability literature, but for investors in energy technology or innovation, or stock traders following the Energy sector of investing as well.

Methodology

The input data here has been collected from the Electric Utilities (EU) industry, within the Energy sector of the S&P1500. The S&P1500 is a stock market index of the top unique companies included in the *S&P LargeCap 500*, *S&P MidCap 400*, and *S&P SmallCap 600* indices (“Cap” refers to market capitalization or equity).

The S&P1500 is inclusive of 11 sectors and 73 industries, with about 1507 publicly traded United States companies. This is secondary data, as these observations have been adjusted for the integrity of the data. Here we have cross-section data, with a total of 23 observations, all taken on the last day of the year, 2021.

The dependent variable here is the price per share (Stock Price), and the independent variables are the EPS, BVPS, DTA, and CR. **Companies *Constellation Energy Corporation* (CEG) and *American Electric Power Company, Inc.* (AEP) have been removed from the data, due to signs of outliers in these observations, as well as missing or ‘NA’ values.**

Methodology, continued…

Listed below are the equations defining the functional specification equation (Eqn. 1), population regression equation (Eqn. 2) and sample regression equation (Eqn. 3).

$$\text{Eqn. 1} \qquad Price = f(\overset{+}{EPS}, \overset{+}{BVPS}, \overset{+}{DTA}, \overset{+}{CR})$$

$$\text{Eqn. 2} \qquad Price = a + ßeps\ EPS + ßbvps\ BVPS + ßdta\ DTA + ßcr\ CR$$

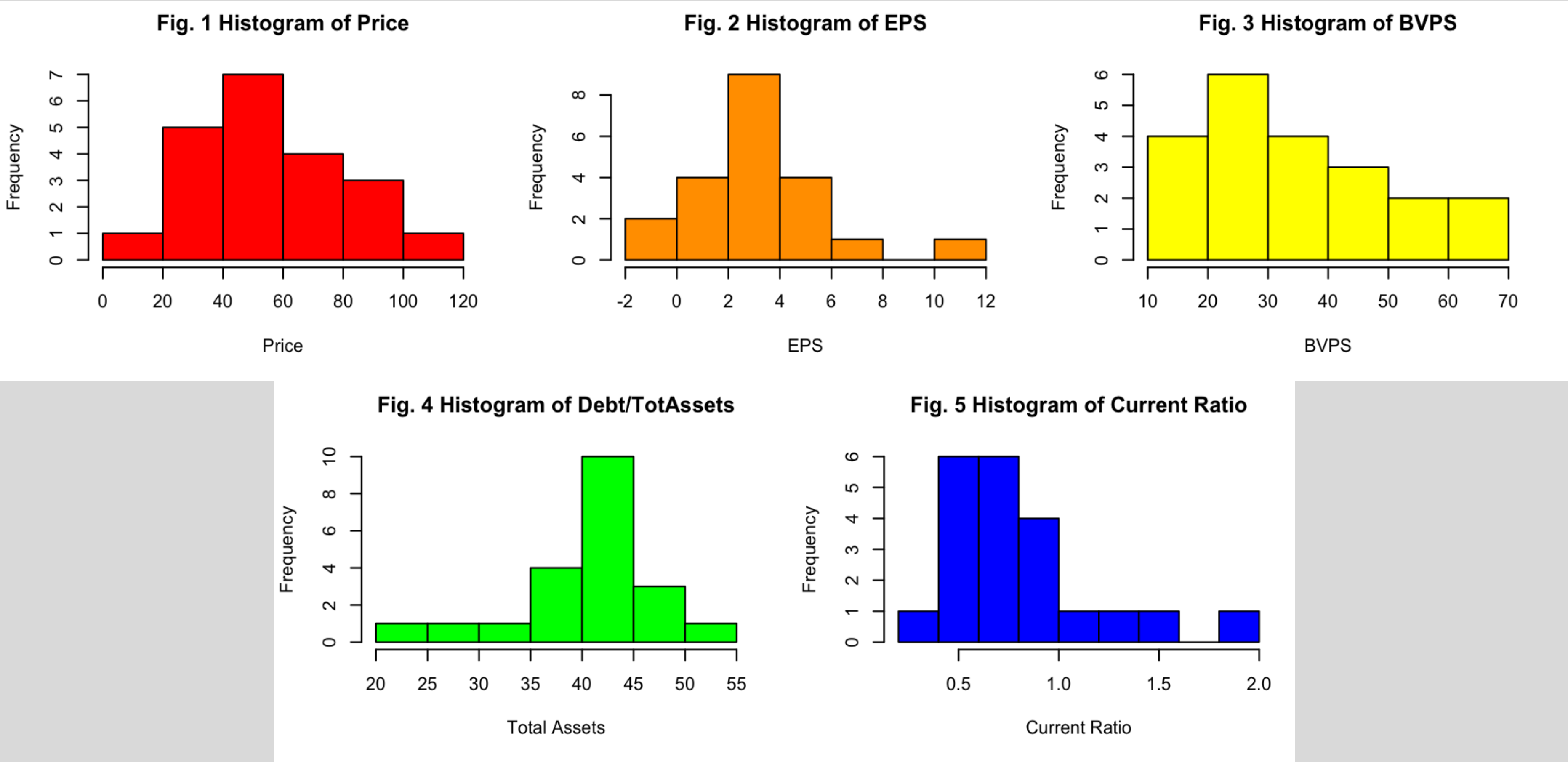
$$\text{Eqn. 3} \qquad Price = a + beps\ EPS + bbvps\ BVPS + bdta\ DTA + bcr\ CR$$

The objective of this model is to evaluate the association or lack thereof that the stock prices of the companies within Electrical Utilities industry have with EPS, BVPS, DTA, and CR.

Graphical techniques including both histograms and scatterplots are used in this analysis. This data has been analyzed using descriptive statistics (for scalable variables), as well as correlation and regression statistical analysis, via both Microsoft Excel and R-Script.

Histograms

Fig. 1 through *Fig. 5* show the histograms of each variable being analyzed. **All histograms are roughly negative and skewed right, excluding DTA.** The histogram of DTA is positive and skewed left. There is less of a recognizable or weak pattern in *Fig. 5*, the graph of CR, having more peaks and irregular symmetry.



Scatterplots

The relationships visible via the scatterplots in *Fig. 6* through *Fig. 9* are the independent variables, each relative to the dependent variable.

The relationships seen in *Fig. 6* and *Fig. 7* are stronger or more positively correlated, non-heteroscedastic, nonlinear and with few outliers. **The stock prices of the companies within the EU industry are most correlated with EPS and BVPS, compared the other variables (bjectively).**

In *Fig. 8* and *Fig. 9,* the data is even more nonlinear, with less of a visible pattern, as well as very congested and clustered. **The DTA data is clustered higher, whereas the clustered CR data is lower.**

The contrasting clustering in *Fig. 8 -9* insists that these companies holding more debt and risk than they are assets**. This also tells us that regardless of debt <u>and/or</u> liability levels relative to assets, the influence that these variables hold on stock prices varies across the companies within this industry.**

Fig. 6 EPS vs. Price

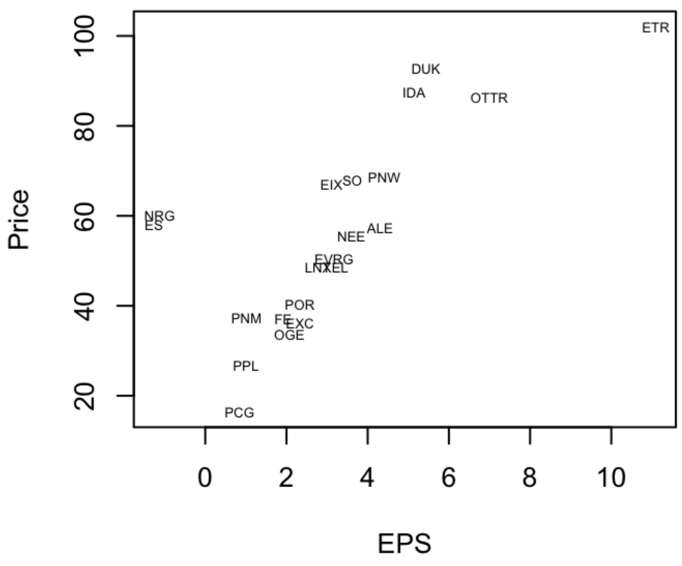


Fig. 7 BVPS vs. Price

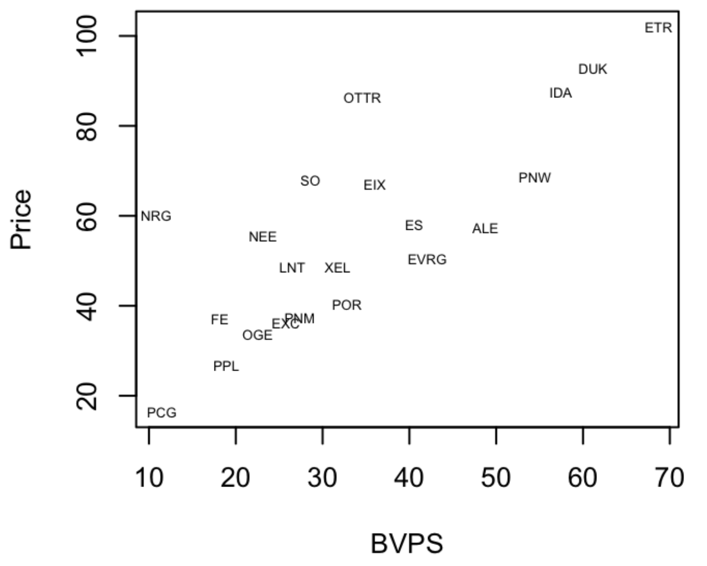


Fig. 8 Debt/Total Assets vs. Price

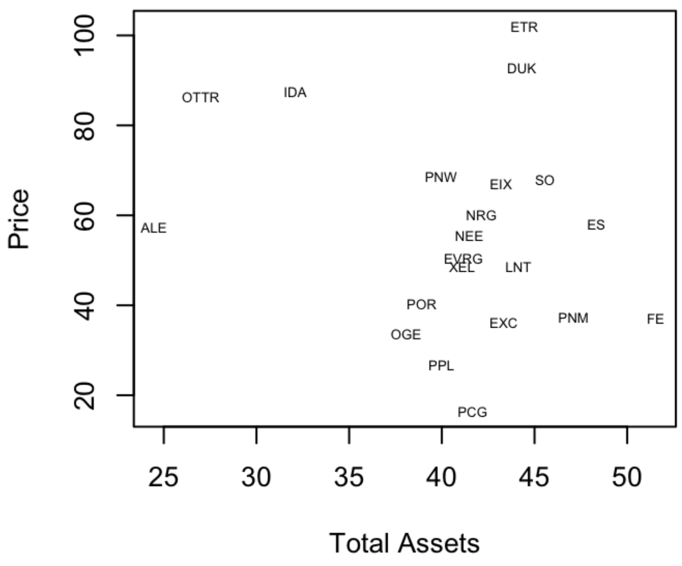


Fig. 9 Current Ratio vs. Price

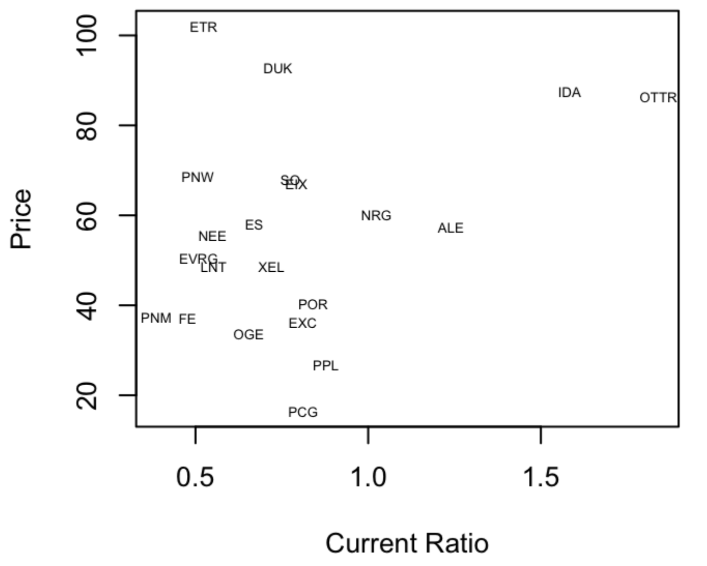


Table 1: *Descriptive Statistics*

In *Table 1*, are the descriptive statistics for all variables that are non-categorical, here, that is all variables. **The standard deviation for each variable is about half (or less than half) of the mean for all variables, except EPS.** The data around EPS is the least disbursed from the mean.

	name	obs	max	min	mean	median	std	skew	kurt
1	price	21	102.040	16.430	56.117	55.400	22.614	0.344	2.420
2	eps	21	11.096	-1.265	3.143	3.112	2.710	0.922	4.919
3	bvps	21	68.700	10.839	34.364	31.744	16.145	0.522	2.438
4	dta	21	51.543	24.467	40.953	41.663	6.449	-1.041	4.125
5	cr	21	1.841	0.386	0.804	0.739	0.364	1.448	4.914

The measure of skewness and kurtosis is asymmetric across all variables in this model, though Price and BVPS are the closest variables to displaying normal distribution, yet still far from. The measure of kurtosis is higher or more peaked than normal distribution in all variables, especially high in EPS, DTA and CR. The higher kurtosis seen in all variables indicates that the data has sharper peaks around the center of the dispersion, with heavier tails. We can also conclude that there is a higher chance of outliers or extreme values.

	price	eps	bvps	dta	cr
price	1.000	0.732	0.785	-0.221	0.341
eps	0.732	1.000	0.723	-0.330	0.210
bvps	0.785	0.723	1.000	-0.235	0.108
dta	-0.221	-0.330	-0.235	1.000	-0.791
cr	0.341	0.210	0.108	-0.791	1.000

Table 2: *Correlation Matrix*

Via the correlation matrix in *Table 2*, it can be observed that the **Price has a strong positive correlation with the EPS (at 0.732) and BVPS (at 0.785).** There is also a high correlation between EPS and BVPS (at 0.723.

It should also be noted that the DTA has a strong correlation with CR (at -0.791), although this correlation is negative. These variables are moving in opposing directions, though there is still a relationship between these variables, one that is greater than all other correlations in this model.

Due to high correlation coefficients, we can assume that there are signs of multicollinearity in this model, or a high correlation in our independent variables than that with our dependent.

Regression Equation

Equation for Sample Regression Line.

			+	+	+	+
Eqn. 4		Price =	f(EPS,	BVPS,	DTA,	CR)
	t-stat	(-2.44)***	(2.16)**	(3.73)***	(2.81)***	(3.44) ***
	p-value	(0.026)*	(0.046)*	(0.002)**	(0.013)*	(0.003)**
	r (corr)		(2.91)	(0.82)	(1.82)	(38.23)

n = 20 r-sq. = 0.812 F = 17.31*** F-Prob = SE =

Confidence Intervals.

*	Significant at the	0%	level of significance (90% Sure, or “are below”) (1.28)
**	Significant at the	5%	level of significance (95% Sure, or “are below”) (1.65)
***	Significant at the	1%	level of significance (99% Sure, or “are below”) (2.33)

Results of an F-test for the entire model.

	Ho: $b_{eps} = b_{bvps} = b_{bdta} = b_{cr} = 0$	(Null Hypothesis)
* 1%	Ha: at least 1 bi not equal to 0 (17.31 > 4.99)	(Alternate Hypothesis)

The above F-test includes the null (Ho) and alternative hypothesis (Ha) for the entire model.

Table 3: *Regression Statistics*

To further interpret the regression statistics in *Table 3*, this F-test includes the null (Ho) and alternative hypothesis (Ha) for the entire model. **The results for the F-Statistic (17.31) indicate that the regression model is statistically significant, as it is above the cutoff (>4.99).**

All variables here were statistically significant. **The null hypothesis should be rejected, while the alternate hypothesis should be accepted.** Rather, the higher value insists that the model explains a significant amount of the variation in the dependent variable with the predictor variables, than without.

The results for the t-statistics show that though EPS is only significant at the 95% confidence interval, all other variables were statistically significant at the 99% confidence level. **This inclines us to reject the null hypothesis at the 99% confidence interval.**

```
Residuals:
     Min       1Q   Median       3Q      Max
-15.7325  -7.2330  -0.7223   4.5938  25.1293

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -86.5527    35.5396  -2.435  0.02695 *
eps           2.9087     1.3486   2.157  0.04657 *
bvps          0.8241     0.2208   3.732  0.00181 **
dta           1.8186     0.6472   2.810  0.01258 *
cr           38.2282    11.1059   3.442  0.00335 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 10.95 on 16 degrees of freedom
Multiple R-squared:  0.8123,    Adjusted R-squared:  0.7653
F-statistic: 17.31 on 4 and 16 DF,  p-value: 1.157e-05
```

Regression Results, continued…

Similarly, the results for our p-value show that all variables were statistically significant and support our t-statistic results in the context of probability. **The BVPS (at 0.002) and CR (at 0.003) were the most significant, having the most effect on the dependent variable, Price.**

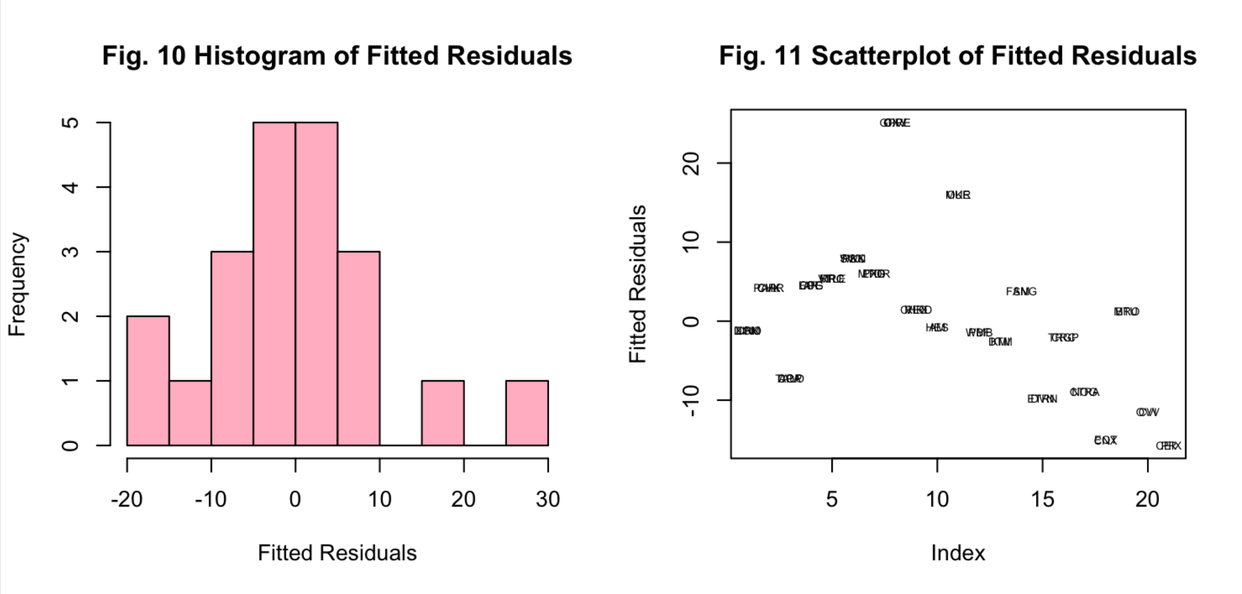
When looking at the results for the coefficient of determination, it can be concluded that that statistical model predicts an association between the dependent and independent variables that is strong. **There is 81.2% of variation in the dependent variable can be explained by variations in the independent variables.**

The standard error is considerably high around the intercept and the variable CR, indicating that there is a lot of room for uncertainty in the analysis of both variables. The high standard error value around the independent variable CR (11.11) tells us that the CR falls much further from the regression line than the other variables, especially BVPS (at 0.22) with the lowest standard error. **The greater standard error value seen in the intercept determines that the dependent variable is being affected by other independent variables, ones not included in this model.**

To discuss the statistically significant regression coefficients, the dependent variable is most impacted by the CR (at 38.23) and the EPS (at 2.91). **That is to say that for every unit increase in the independent variable CR, there is a $38.23 increase in the dependent variable (Price), as well as that there is a $2.91 increase in Price for every unit increase in EPS, holding all other variables constant.**

Residual Analysis

Fig. 10 and *Fig. 11* show the graphed fitted residuals. These graphs give us insight as to the integrity of the assumptions made in this analysis. The histogram is not entirely, but approximately normally distributed, due to a lack of symmetry on either end of the some-what visible bell-curve.



The plotted residuals in *Fig. 11* are fairly negative and non-linear, possibly due to outliers, **insisting that because of there lack of both symmetry and linear pattern in the fitted residual output, our model is possibly lacking statistical significance as a whole.**

Conclusions

For this model, the null hypothesis is rejected, and the alternate hypothesis is accepted. The research presented here can be considered successful, as 81.2% of variation in the dependent variable can be explained by the independent variables**. It can be determined true that the dependent variable is better supported by the independent variables, than if it were not.** This model was not entirely predictive, as the variables here do not influence one-another enough, but it was somewhat predictive when looking at EPS and CR.

This model can be improved on by increasing the number of observations, as there were just enough observations to create a valuable model after removing outliers and the model is still lacking significance. Should there be a need to break down the moments of closer associations, breaking this model up by time periods may add to the findings that this model is capable of. **The highly correlated independent variables or multicollinearity seen in this model can be adjusted by switching those variables out and taking the first difference for those variables to reanalyze the trends in those variables.**

Public Policy Implications

Public policy does not entirely pertain to driving a relationship between these variables but could be influential in the context of regulating the terms of trade, changes in regulations that compromise operations within the Energy industry, as well as climate change efforts (including “green” incentives relative to electricity production or consumption). **This model or an expanded version of this model could be beneficial to investors who require more insight on the variables that influence the stock price of the companies within this industry.**

Appendix I: *Bibliography*

Data Source.

"S&P Composite 1500®." S&P Dow Jones Indices, 2024. https://www.spglobal.com/spdji/en/indices/equity/sp-composite-1500/#data.

Appendix II: *Input Data*

Via R-Script.

	tkr	price	eps	bvps	dta	cr
1	ETR	102.04	11.0960	68.69977	44.45705	0.5238476
2	DUK	92.69	5.4345	61.15435	44.30800	0.7388185
3	IDA	87.41	5.1410	57.44454	32.09545	1.5834920
4	OTTR	86.37	6.9980	34.59573	26.99742	1.8407010
5	PNW	68.61	4.4072	54.46503	39.95434	0.5066100
6	SO	67.98	3.6211	28.61146	45.56775	0.7746343
7	EIX	67.11	3.1117	36.01746	43.18966	0.7921610
8	NRG	60.01	-1.1228	10.83933	42.13457	1.0238950
9	ES	58.14	-1.2651	40.55010	48.33571	0.6698830
10	ALE	57.33	4.3049	48.77778	24.46707	1.2396720

	tkr	price	eps	bvps	dta	cr
11	NEE	55.40	3.5996	23.13255	41.48032	0.5493330
12	EVRG	50.35	3.1727	42.06298	41.16461	0.5093378
13	XEL	48.65	3.2083	31.74388	41.09083	0.7199221
14	LNT	48.51	2.7754	26.50186	44.13316	0.5520833
15	POR	40.41	2.3277	32.80954	38.91604	0.8408273
16	PNM	37.39	1.0168	27.36758	47.10666	0.3863816
17	FE	37.01	1.9199	18.17231	51.54305	0.4767917
18	EXC	36.27	2.3350	25.78078	43.33799	0.8108643
19	OGE	33.59	2.0747	22.52421	38.07610	0.6542572
20	PPL	26.76	1.0011	18.90169	39.98369	0.8778443

Appendix III: *R-Script*

```
# REGRESSION ANALYSIS...
# SECTOR: Energy
# INDUSTRY: Electric Utilities

# IMPORT LIBRARIES...
library(YRmisc)
library(readxl)
# Read SP1500 data file
sp1500 <- read_excel("Downloads/sp1500.xlsx", sheet =
"Tab1")
View(sp1500)
dim(sp1500)
names(sp1500)

# View sector & industry.
data.frame(unique(sp1500$sector))
data.frame(unique(sp1500$industry))

# Choose a sector.
idf<-sp1500[sp1500$industry=="Electric
Utilities",c("tkr","price","eps","bvps","dta","cr")]
dim(idf) #Confirm that the dimensions have decreased.
View(idf)

# Drop CEG & AEP (Outliers/NA Values).
drop_CEG <- "CEG"
idf <- idf[idf$tkr != drop_CEG,]
drop_AEP <- "AEP"
idf <- idf[idf$tkr != drop_AEP,]
View(idf)

# FIGURES 1-5: HISTOGRAMS
par(mfrow=c(3,3)
hist(idf$price, col="red", xlab="Price", ylab="Frequency",
main="Fig. 1 Histogram of Price")
hist(idf$eps, col="orange", xlab="EPS", ylab="Frequency",
main="Fig. 2 Histogram of EPS")
hist(idf$bvps, col="yellow", xlab="BVPS",
ylab="Frequency", main="Fig. 3 Histogram of BVPS")
hist(idf$dta, col="green", xlab="Total Assets",
ylab="Frequency", main="Fig. 4 Histogram of
Debt/TotAssets")
hist(idf$cr, col="blue", xlab="Current Ratio",
ylab="Frequency", main="Fig. 5 Histogram of Current
Ratio")

# FIGURES 6-9: SCATTERPLOTS
par(mfrow=c(2,2))
plot(idf$eps, idf$price, xlab="EPS", ylab="Price",
main="Fig. 6 EPS vs. Price", type="n")
  text(idf$eps, idf$price, as.character(idf$tkr), cex=0.5)
plot(idf$bvps, idf$price, xlab="BVPS", ylab="Price",
main="Fig. 7 BVPS vs. Price", type="n")
  text(idf$bvps, idf$price, as.character(idf$tkr), cex=0.5)
plot(idf$dta, idf$price, xlab="Total Assets", ylab="Price",
main="Fig. 8 Debt/Total Assets vs. Price", type="n")
  text(idf$dta, idf$price, as.character(idf$tkr), cex=0.5)
plot(idf$cr, idf$price, xlab="Current Ratio", ylab="Price",
main="Fig. 9 Current Ratio vs. Price", type="n")
  text(idf$cr, idf$price, as.character(idf$tkr), cex=0.5)

# DESCRIPTIVE STATISTICS
des_stats<-
ds.summ(idf[,c("price","eps","bvps","dta","cr")],3)[,-c(7,8)]
View(des_stats)

# CORRELATION MATRIX
cor_matrix<-
round(cor(na.omit(idf[,c("price","eps","bvps","dta","cr")]))
,3)
View(cor_matrix)

# REGRESSION ANALYSIS
idf1<-na.omit(idf)
fit<-lm(price~eps+bvps+dta+cr,na.action=na.omit, data=idf1)
summary(fit)

# FIGURES 10 & 11: FITTED RESIDUALS
idf$price
fit$fitted.values
fit$residuals
par(mfrow=c(2,2))
hist(fit$residuals,col="pink",xlab="Fitted
Residuals",ylab="Frequency",main="Fig. 10 Histogram of
Fitted Residuals")
plot(fit$residuals,xlab="Index",ylab="Fitted
Residuals",main="Fig. 11 Scatterplot of Fitted
Residuals",type="n")
  text(fit$residuals,as.character(idf$tkr),cex=.5)
```

Thank you!

Questions & Correspondence

Brianna Palmisano
brianna.palmisano21@my.stjohns.edu