Shalini Chhetri, X03742863 (BUA 633)

Statistical Model to Predict Equity Price of Banking Industry

Abstract

This project studies the impact of financial indicators on equity prices of banking industry. Using a robust regression model, this project finds that EPS, BVPS and DTA have a significant impact on the equity prices in the banking sector. The project also suggests ways to further strengthen and improve the results.

I. Introduction

This research will develop an equation to predict the equity price of the banking industry. This research is relevant for investors, bankers, portfolio managers, owners, stakeholders and the government to understand key value drivers and make informed decisions.

II. Previous Research

There has been a lot of previous research but none right on point with this.

III. Methodology

This research is composed of secondary cross-sectional data for banking industry. 104 observations will be studied. Graphical techniques such as histograms and scatterplots and a statistical analysis composed of descriptive statistics for scalable variables, correlation and regression will be used. The research will be developed using R.

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

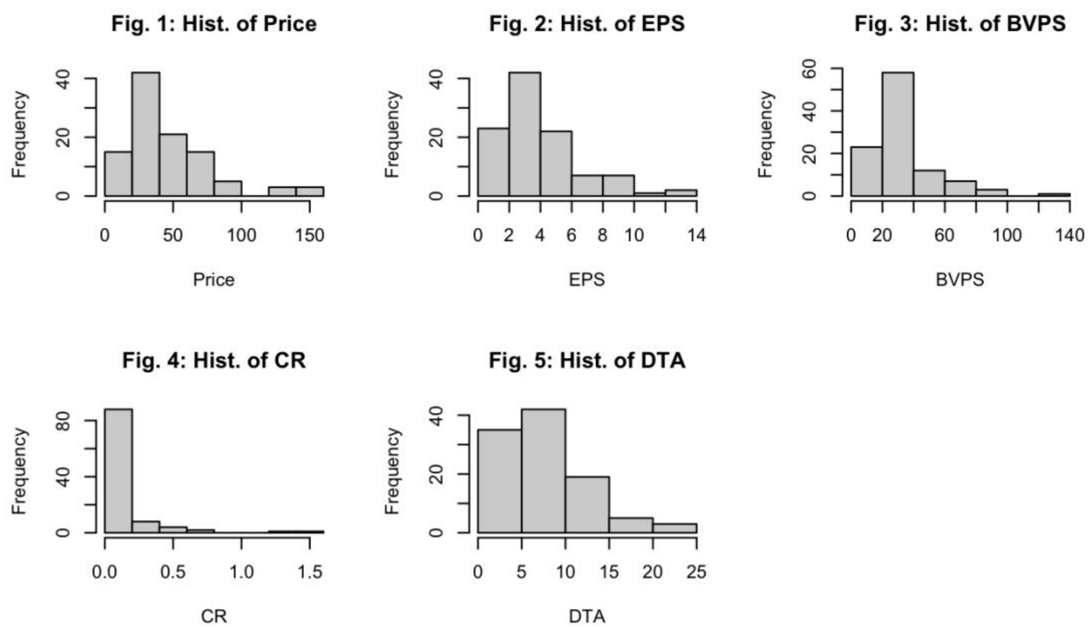
Eqn 2: $\text{Price} = \alpha + \beta_{eps} * EPS + \beta_{bvps} * BVPS + \beta_{cr} * CR + \beta_{dta} * DTA$

Eqn 3: $\text{Price} = a + b_{eps} * EPS + b_{bvps} * BVPS + b_{cr} * CR + b_{dta} * DTA$

An increase in EPS leads to higher profits for shareholders, driving up stock prices. A higher BVPS reflects stronger profitability, also boosting prices. A rising current ratio indicates greater liquidity and financial health, supporting a price increase. However, a higher debt-to-asset ratio suggests greater financial risk, which tends to lower the stock price.

IV. Results

Histograms



All the histograms are positively skewed. The histogram of Current ratio (CR) displays the highest degree of positive skewness. The histogram for Price, BVPS and EPS suggests that there might be some outliers, as seen from the isolated bars on the far-right side.

Scatterplots

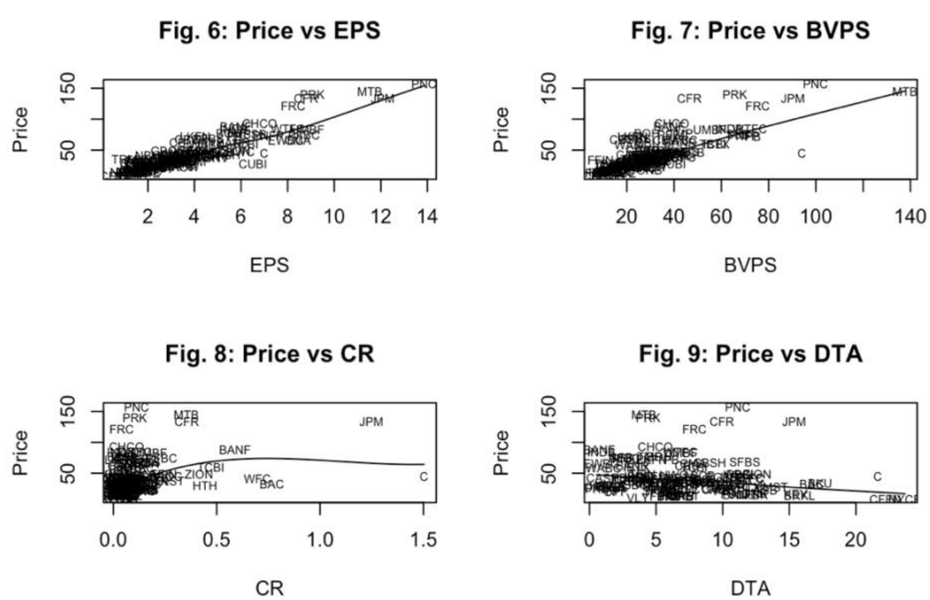


Fig. 6 and 7 shows that the relationship of EPS and BVPS with Price is positive, strong and linear respectively. Fig. 8 shows that the relationship between Price and CR is slightly positive, weak and non-linear whereas, Fig. 9 shows that the relationship between Price and DTA is slightly negative, moderate and linear.

Table 1: Descriptive Statistics

	n	Mean	Median	Std. Dev	Skewness	Kurtosis
Price	104	45.41	36.84	30.8	1.65	5.89
EPS	104	4.01	3.45	2.54	1.38	5.18
BVPS	104	33.42	28.88	21.13	2.04	8.83
CR	104	0.14	0.07	0.22	4.01	22.02
DTA	104	7.37	6.68	4.73	1.03	4.39

Table 1 provides further evidence on the skewness of the variables. All the variables are positively skewed. Price, EPS and DTA are moderately skewed while BVPS and CR are more strongly skewed. All the variables have high kurtosis score, implying that most of the data are clustered around the mean and the tails (higher chance of outliers).

Table 2: Correlation Matrix

	Price	*EPS*	*BVPS*	*CR*	*DTA*
Price	1.000				
EPS	0.899	1.000			
BVPS	0.796	0.840	1.000		
CR	0.274	0.362	0.437	1.000	
DTA	(0.140)	(0.092)	0.000	0.304	1.000

Table 2 shows that both EPS and BVPS have a strong and positive correlation with Price. This is consistent with the hypothesis. Similarly, the correlation between CR and price is weak and positive while that of DTA and price is weak and negative. This is also consistent with the hypothesis. The correlation between EPS and BVPS is also high suggesting that there may be a concern for multicollinearity.

<u>Table 3: Regression Analysis</u>

<u>Sample Regression Equation</u>

Eqn 4: Price = 4.452 + 4.604*(EPS) + 0.623*(BVPS) + 1.825*(CR) – 0.409*(DTA)

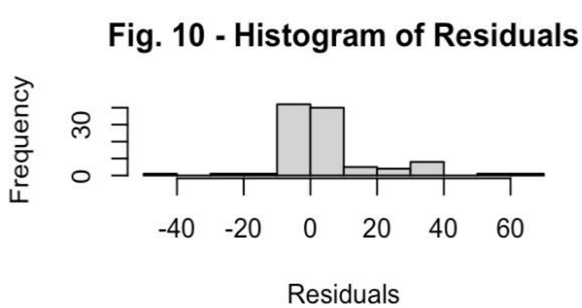
	Intercept	EPS	BVPS	CR	DTA
t- stat	2.632*	8.233***	10.981***	0.366	(2.729)***
p-value	0.020	0.000	0.000	0.715	0.008
r(corr)	-	0.899	0.769	0.274	(0.140)

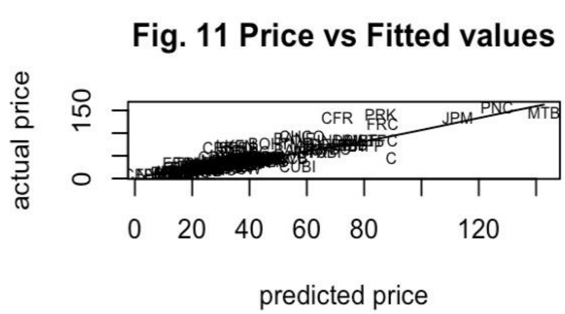
n = 104   r-sq = 0.55   SE = 6.696

* - significant at 10% level
** - significant at 5% level
*** - significant at 1% level

Using a robust linear regression, 55% of the variation in Price is explained by variation in EPS, BVPS, CR and DTA. The results also show that EPS, BVPS and DTA are individually statistically significant predictors of price. On average, ceteris paribus, a $1 increase in EPS leads to a $4.604 increase in price. Similarly, on average, ceteris paribus, a $1 increase in BVPS leads to a $0.623 increase in price. Further, a one unit increase in DTA leads to a $2.729 decrease in price, on average, ceteris paribus.

<u>Residuals</u>





Residuals are centered around 0 as shown in Fig. 10. This implies that the model isn't over predicting or under-predicting. Similarly, the histogram being roughly bell-shaped means that the residuals are normally distributed.

V. Conclusion

The research project was successful. Almost 55% of the variation in prices are explained by the variation in Earnings per share (EPS), Book value per share (BVPS), Current ratio (CR) and Debt to asset ratio (DTA). The hypothesis for this research was also correct since EPS and BVPS have a strong positive correlation with prices while DTA has a strong negative correlation. In general, institutions in the Banking industry should therefore focus on increasing their EPS and BVPS for a higher share price. This research can be improved by adding further explanatory variables and tests for better model selection including confounding variables. More granulated data in terms of time series that explains various economic cycles would help improve the validity of the results.

VI. Bibliography

FactSet Universal Screening

VII. Appendix I

tkr	price	eps	bvps	cr	dta	mv
ABCB	47.14	4.9919	46.0926	0.130371	8.224498	3270.057
ASB	23.09	2.3417	25.40066	0.028537	13.20817	3473.752
AUB	35.14	2.9704	31.75587	0.028072	8.584058	2625.402
AX	38.22	3.9715	27.4846	0.224115	6.377851	2284.713
BAC	33.12	3.1852	30.61233	0.766888	16.67799	264853.3
BANC	15.93	1.8925	16.25869	0.048623	11.23057	940.2182
BANF	88.18	5.7746	38.0476	0.588952	0.738874	2898.967
BANR	63.2	5.6697	42.59318	0.060035	3.225018	2161.062
BHLB	29.9	2.0154	21.50667	0.090709	1.614553	1326.401
BKU	33.97	3.5437	32.19021	0.055736	17.31429	2570.666
BOH	77.56	5.4772	28.54208	0.031928	5.237076	3089.661
BRKL	14.15	1.4188	12.81058	0.065445	15.77847	1095.857
C	45.23	6.9954	94.06055	1.502857	21.641	87609.9
CADE	24.66	2.4594	22.71675	0.052404	9.649049	4498.903
CASH	43.05	5.262	22.41098	0.070057	1.009024	1239.329
CATY	40.79	4.8302	34.01109	0.072743	3.001151	2967.152
CBSH	68.07	3.8544	19.68159	0.144086	9.02448	8526.362
CBU	62.95	3.4598	28.87578	0.032553	7.346263	3382.76
CFFN	8.65	0.622	7.8965	0.01186	22.20963	1201.129
CFG	39.37	4.1021	44.03166	0.106222	7.4784	19381.15
CFR	133.7	8.792	46.48885	0.355636	9.93837	8604.223
CHCO	93.09	6.7984	39.07534	0.064304	4.941426	1376.629
CMA	66.85	8.4361	36.55123	0.156051	7.7758	8755.135
COLB	30.13	3.1989	28.14069	0.020182	5.567674	2369.604
CPF	20.28	2.6817	16.75744	0.037193	1.97025	548.0684
CUBI	28.34	6.5102	39.08009	0.036851	5.381757	917.4706
CVBF	25.75	1.6707	13.93603	0.031674	9.619064	3600.332
DCOM	31.83	3.7256	27.29925	0.027953	10.53092	1227.779
EGBN	44.07	4.3934	39.18476	0.055742	9.940799	1381.458
EWBC	65.9	7.9168	42.45976	0.087533	0.880225	9288.463
FBK	36.14	2.6366	28.35867	0.162999	3.750079	1689.108
FBNC	42.84	4.1188	28.89288	0.038954	2.883234	1529.566
FBP	12.72	1.5892	7.25492	0.079488	5.394556	2324.059
FCF	13.97	1.3653	11.26703	0.042558	6.100674	1304.467
FFBC	24.23	2.3007	21.5128	0.06727	10.0895	2299.211
FFIN	34.4	1.6373	8.93071	0.070245	4.952241	4875.466
FHB	26.04	2.076	17.81521	0.040359	0.566819	3316.541
FHN	24.5	1.5336	13.47606	0.157649	5.666662	13158.97

FITB	32.81	3.3499	33.54181	0.169748	7.255172	22421.89
FNB	13.05	1.2173	15.38546	0.062616	6.00602	4704.135
FRC	121.89	8.235	75.37808	0.040834	7.888619	22336.29
FULT	16.83	1.6724	14.24152	0.050217	10.96588	2820.708
GBCI	49.42	2.7358	25.66674	0.025223	11.24781	5474.638
HAFC	24.75	3.3178	20.91199	0.127097	7.096983	754.5191
HBAN	14.1	1.4503	10.78535	0.046908	6.623074	20347.26
HFWA	30.64	2.3087	22.72766	0.022978	1.324322	1075.669
HMST	27.58	3.4945	30.01258	0.017959	13.63471	516.5839
HOMB	22.79	1.5653	17.33421	0.133053	5.53177	4636.254
HOPE	12.81	1.8118	16.89882	0.034639	6.511005	1530.734
HTH	30.01	1.6019	31.49008	0.4434	8.858272	1941.186
HWC	48.39	5.9781	38.89445	0.06502	6.313159	4158.685
IBOC	45.76	4.78	32.90471	0.264444	3.705412	2843.611
IBTX	60.08	4.6966	57.91075	0.096948	3.538258	2474.736
INDB	84.43	5.6868	63.24765	0.037949	0.902262	3853.49
JPM	134.1	12.0848	90.28791	1.249013	15.36654	393484
KEY	17.42	1.9281	11.73654	0.047536	15.47365	16258.51
LKFN	72.97	4.0376	22.43848	0.030987	4.699605	1849.733
MTB	145.06	11.5313	137.6811	0.353168	4.099385	24556.54
NBHC	42.07	2.1762	29.04135	0.03061	5.177117	1582.19
NBTB	43.42	3.52	27.38249	0.051372	6.90689	1860.887
NFBK	15.73	1.3161	14.78401	0.029195	12.21655	746.2703
NWBI	13.98	1.0502	11.74132	0.021182	6.958368	1775.863
NYCB	8.6	1.2553	12.2149	0.039239	23.67573	5858.469
OFG	27.56	3.4321	21.90786	0.135858	0.543953	1311.343
ONB	17.98	1.4969	29.53542	0.059042	12.37652	2981.767
OZK	40.06	4.5362	37.12846	0.124706	4.091874	4694.108
PACW	22.95	3.4369	29.30001	0.121225	6.741668	2703.881
PB	72.68	5.7259	73.36665	0.038967	6.145365	6636.674
PFBC	74.62	8.7001	39.70593	0.232312	2.625341	1184.77
PFS	21.36	2.3488	21.25476	0.026825	10.20568	1605.614
PNC	157.94	13.8544	99.82045	0.112808	11.13092	63333.94
PNFP	73.4	7.1657	69.35237	0.092129	2.896865	5611.724
PPBI	31.56	2.9794	29.44998	0.087962	6.393883	2998.887
PRK	140.75	9.065	65.74357	0.103943	4.41705	2289.099
RF	21.56	2.2781	15.2896	0.103437	1.827728	20147.64
RNST	37.59	2.9542	38.17511	0.049552	7.029606	2103.277
SBCF	31.19	1.6573	22.44936	0.028204	3.753852	2233.761
SBSI	35.99	3.2563	23.64744	0.117762	7.25616	1135.363
SFBS	68.91	4.6107	23.88145	0.154128	11.65871	3743.641
SFNC	21.58	2.06	25.73355	0.127113	5.1985	2741.667
SNV	37.55	4.9477	27.07229	0.070881	8.111332	5463.023

SSB	76.36	6.5981	67.03595	0.062703	2.410668	5780.8
STBA	34.18	3.4624	30.37608	0.039583	5.352692	1333.011
STEL	29.46	1.4692	26.11984	0.063736	1.796318	1560.054
TBBK	28.38	2.2737	12.46248	0.143514	1.549648	1580.472
TCBI	60.31	6.1753	56.4809	0.471685	7.858082	2942.149
TFC	43.03	4.4282	40.5787	0.221928	12.19602	57093.45
TFIN	48.87	3.9605	35.08712	0.243997	3.97809	1175.499
TMP	77.58	5.8857	42.69622	0.041787	4.531251	1119.246
TRMK	34.91	1.1702	24.47236	0.077355	9.534294	2128.731
TRST	37.59	3.9322	31.53771	0.268444	2.861309	715.1284
UBSI	40.49	2.8022	33.51656	0.061543	8.239959	5455.83
UCBI	33.8	2.5205	24.51689	0.051529	4.465656	3590.329
UMBF	83.52	8.8555	55.19714	0.17916	6.86406	4035.637
USB	43.61	3.692	28.71236	0.155574	10.75051	66765.96
VBTX	28.08	2.712	26.83276	0.100806	11.6707	1517.161
VLY	11.31	1.1392	12.22635	0.07716	3.642401	5727.095
WABC	59.01	4.5354	22.37246	0.113416	1.058052	1588.136
WAFD	33.55	3.3898	30.21975	0.043254	10.32887	2191.826
WBS	47.34	3.7165	44.66591	0.024406	11.08316	8237.519
WFC	41.29	3.1449	41.84903	0.697372	12.12457	158297.8
WSFS	45.34	3.4932	35.79037	0.071087	4.44476	2793.484
WTFC	84.52	8.0165	72.11793	0.103768	6.805792	5138.31
ZION	49.16	5.7895	29.95345	0.414289	12.58138	7308.322

VIII. Appendix II

```
# Load the necessary packages
library("dplyr")
library("ggplot2")
library("tidyverse")
library("readxl")
library("moments")
library("tidyr")
library("broom")
library("YRmisc")
library("robust")
library("gam")
library("quantmod")
library("robustbase")

# Importing the two datasets
#spData - contains financial data of companies
spdata <- read_excel("spData.xlsx")
#spInfo - contains the sector and industry of companies
spinfo <- read_excel("spInfo.xlsx")

# Check the type/class of data and convert to dataframe
#spdata
data.class(spdata)
spdata<-as.data.frame(spdata)
data.class(spdata)
#spinfo
data.class(spinfo)
spinfo<-as.data.frame(spinfo)
data.class(spinfo)

# Merging the two datasets using common variable "tkr"
names(spdata)
names(spinfo)
spdf<-merge(spdata,spinfo,by="tkr")
dim(spdf)
names(spdf)

# Replace mm/dd/yyyy date format with yyyy
spdf$date
spdf$date<-as.numeric(substring(spdf$date,7,10)) # modify date to only keep year
names(spdf)
names(spdf)[2]<-"year"  # rename "date" to "year"
names(spdf)
```

```
#Filter the data for the latest year for cross-section analysis
csdf<-spdf[spdf$year==2022,]
csdf$mv<-csdf$cso*csdf$price ## multiply common stock outstanding with price to obtain
market value
names(csdf)

## Filter for Banking industry as the industry to analyze and create a new data
myIndustry<-"Banks";  myIndustry
idf<-csdf[csdf$industry == myIndustry,
        c("tkr","price","eps","bvps","cr","dta","mv")]
dim(idf)

## Histogram of the variables
par(mfrow=c(2,3))
hist(idf$price,main="Fig. 1 - Hist. of Price", xlab = "price")
hist(idf$eps,main="Fig. 2 - Hist. of EPS", xlab = "Earnings per share")
hist(idf$bvps,main="Fig. 3 - Hist. of BVPS", xlab = "Book value per share")
hist(idf$cr,main="Fig. 4 - Hist. of CR", xlab = "Current ratio")
hist(idf$dta,main="Fig. 5 - Hist. of DTA", xlab = "Debt to asset ratio")

## Generate Scatterplots of each variable against price
par(mfcol=c(2,2))
scatter.smooth(idf$eps,idf$price, main = " Fig. 6 - Price vs EPS",
            xlab = "eps", ylab = "price")
scatter.smooth(idf$bvps,idf$price, main = "Fig. 7 - Price vs BVPS",
            xlab = "bvps", ylab = "price")
scatter.smooth(idf$cr,idf$price, main = "Fig. 8 - Price vs CR",
            xlab = "cr", ylab = "price")
scatter.smooth(idf$dta,idf$price, main = "Fig. 9 - Price vs DTA",
            xlab = "dta", ylab = "price")

## Alternate method to put scatterplots together
# Function of 3 vectors for plot
plotTkr<-function(x,y,z,xxlab,yylab,myTitle){
  scatter.smooth(x,y,type="n",xlab=xxlab,ylab=yylab,main=myTitle)
  text(x,y,z,cex=.6)  }
plotTkr
# Scatterplots
par(mfrow=c(2,2))
plotTkr(idf$eps,idf$price,idf$tkr,"eps","price","Fig. 6 - Price vs EPS")
plotTkr(idf$bvps,idf$price,idf$tkr,"bvps","price","Fig. 7 - Price v BVPS")
plotTkr(idf$cr,idf$price,idf$tkr,"cr","price","Fig. 8 - Price v CR")
plotTkr(idf$dta,idf$price,idf$tkr,"dta","price","Fig. 9 - Price v DTA")

## Summary of data
```

```
names(idf)
idf<-na.omit(idf)
ds.summ(idf[,c("price","eps","bvps","cr","dta")],2)
# Table of summary statistics
summary(idf[,c("price","eps","bvps","cr","dta")])
# Table of correlation
round(cor(idf[,c("price","eps","bvps","cr","dta")]),3)

## Regression Models
#Robust model
fit<-lmRob(price ~ eps+bvps+cr+dta,na.action=na.omit,data=idf)
cor(idf$price,predict(fit,idf))^2
summary(fit)

## Residual Plots
# Histogram
par(mfrow=c(2,2))
hist(fit$residuals, main = "Fig. 10 - Histogram of Residuals", xlab = "Residuals")
#Scatter
plotTkr(fit$fitted.values,idf$price,idf$tkr,"predicted price","actual price","Fig. 11 Price vs Fitted
values")
```