A Statistical Model to Predict Equity Price for Apple

Amritpal Singh

St. John’s University, NY, USA

A Statistical Model to Predict Equity Price for Apple

I. Introduction

Apple is one of the six companies in the Technology Hardware Storage & Peripherals industry of the Information Technology Sector of the S&P500. This research develops a statistical model to predict the equity price of stock for this company. Financial advisors, investors, portfolio managers and relevant departments at Apple would be interested in the conclusions of this research.

II. Previous Research

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

III. Methodology

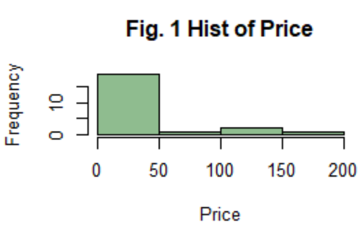
This research will analyze time-series data with 23 observations obtained from FactSet. The research incorporates Graphical Techniques including histograms, time-series plots and scatterplots. It also uses Analytical Methods, namely descriptive statistics, correlation and regression. R is used to create and execute the statistical program.

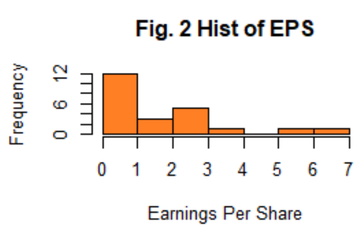
Eqn. 1 price = f($\overset{+}{\text{eps}}$, $\overset{+}{\text{bvps}}$, $\overset{+}{\text{obs}}$)

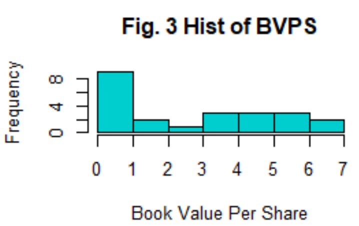
Eqn. 2 price = $\alpha + \beta_{eps}eps + \beta_{bvps}bvps + \beta_{obs}obs + \epsilon$

Eqn. 3 price = $a + b_{eps}eps + b_{bvps}bvps + b_{obs}obs + e$

IV. Results







a. Histograms of price, eps and bvps are displayed below in Figs. 1 to 3. All histograms are skewed to the right. No outliers are present.

Fig. 4 TSPlot of Price

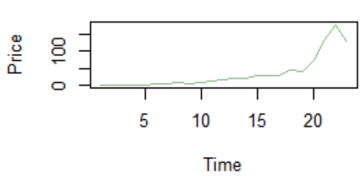


Fig. 5 TSPlot of EPS

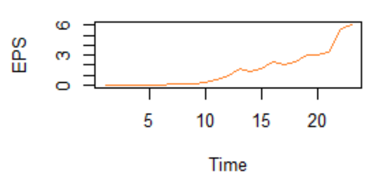
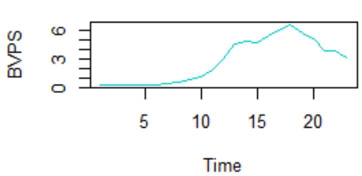


Fig. 6 TSPlot of BVPS



b. Time-Series Plots of all variables are in Figs. 4 to 6. All plots have a positive nonlinear trend.

Fig. 7 EPS vs. Price

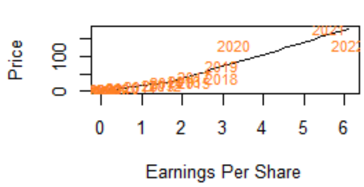
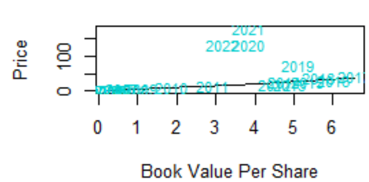


Fig. 8 BVPS vs. Price



c. Scatterplots are in Figs. 7 and 8. All variables have a positive nonlinear correlation and agree with our hypothesis. EPS has a strong correlation while BVPS is weak. Notice that Fig. 8 contains minor outliers in y-space.

d. Table 1 displays the descriptive statistics.

Table 1 Descriptive Statistics

	obs	*mean*	*median*	*std*	*skew*	*kurt*
price	23	33.43	14.45	48.69	1.73	5.22
eps	23	1.51	0.99	1.77	1.18	3.89
bvps	23	2.73	2.94	2.29	0.17	1.44
obs	23	12.00	12.00	6.78	0.00	1.80

e. Table 2 displays the correlation matrix. EPS and Obs have a high correlation with price, while BVPS has moderate correlation. The correlations agree with the initial hypothesis and there is some evidence of multicollinearity between all variables, the strongest being correlations with Obs.

Table 2 Correlation Matrix

	price	*eps*	*bvps*	*obs*
price	1.000	0.921	0.443	0.794
eps	0.921	1.000	0.643	0.897
bvps	0.443	0.643	1.000	0.835
obs	0.794	0.897	0.835	1.000

f. **Table 3 Linear Regression Results**

Eqn. 4	price= (-11.28)	+23.08eps	+(-9.25)bvps	+2.94obs
t-stat	(-1.07)	(4.55)***	(-2.95)***	(1.60)*
p-value	(.30)	(.00)	(.01)	(.13)
r (corr)		.92	.44	.79

n = 23 $r^2 = .899$ F=56.65** F. Prob=.00 SE = 16.68

Significance Legend
* 10% level
** 5% level
*** 1% level

1. Ho: $\beta_{eps}=\beta_{bvps}=\beta_{obs}=0$
   Ha: at least 1 β_ι not equal to 0 (56.65>4.99, Ho is rejected)
2. R-squared- 89.9% variation in equity price can be explained by variation in the independent variables (eps, bvps and obs)
3. Standard Error - There is a $16.68 standard deviation for the residuals around the regression line
4. As eps increases by $1, price increases by $23.08, on average, other things equal
   As bvps increases by $1, price decreases by $9.25, on average, other things equal
   As obs increases by 1 year, price increases by $2.94, on average, other things equal

Fig. 9 Hist of LM Residuals

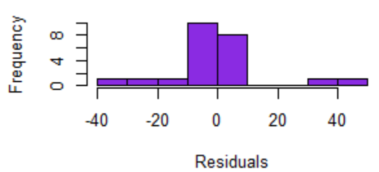


Fig. 10 Predicted vs Actual Price LM

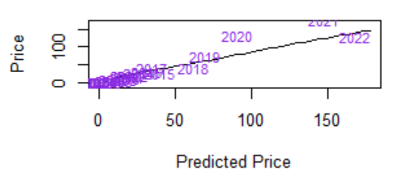
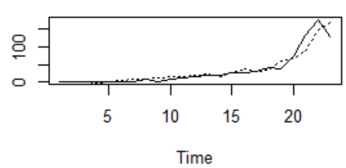


Fig. 11 TSPlot LM



Figs. 9 and 10 showcase the residuals and the predicted vs. actual values of the linear model. The residuals are symmetrical. The smooth scatterplot is positive and linear with no evidence of heteroscedasticity. Fig. 11 displays a timeseries plot of our actual and predicted values.

g. **Table 4 Robust Regression Results**

Eqn. 4	price= (-0.19)	+34.95eps	+(-8.23)bvps	+0.59obs
t-stat	(-0.03)	(8.86)***	(-3.32)***	(0.49)
p-value	(.98)	(.00)	(.01)	(.63)
r (corr)		.92	.44	.79

$r^2 = .564$ SE = 3.09

Significance Legend
* 10% level
** 5% level
*** 1% level

Fig. 12 Hist of Rob Residuals

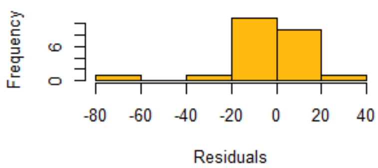


Fig. 13 Predicted vs Actual Price Rob

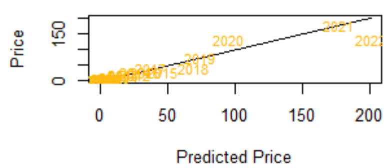
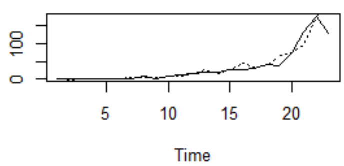


Fig. 14 TSPlot Rob



Compared to linear regression, robust regression had a lower explanatory power of 56.4%. Only eps and bvps were predictive, while all independent variables were for linear regression. The residuals are heavily skewed to the right and the scatterplot is positive and linear.

h. **Figs. 15-17 GAM Regression**

Fig. 15 Gam of EPS

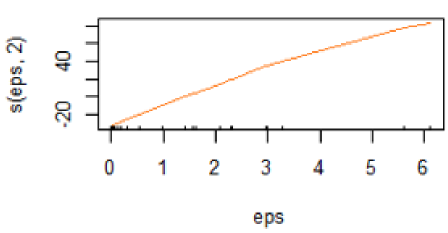


Fig. 16 Gam of BVPS

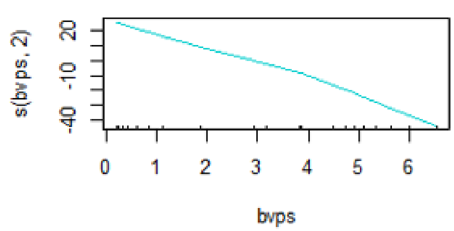
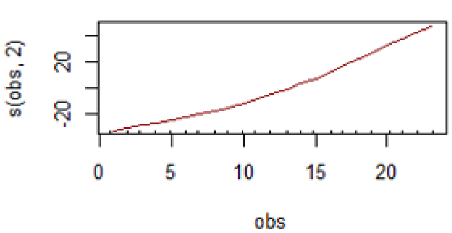


Fig. 17 Gam of Obs



$r^2 = .925$

V. <u>Conclusions</u>

The research was extremely successful. The explanatory power using the linear model was high and all assumptions were not satisfied. All chosen independent variables were significant/predictive (eps, bvps and obs). This research could be further improved by including other independent variables such as current ratio and debt/asset ratio.

VI. <u>Bibliography</u>

FactSet Data

VII. Appendix I

tkr	price	eps	bvps	cr	dta	year	name	obs
AAPL	0.27	0.04	0.21	2.81	4.41	2000	Apple Inc.	1
AAPL	0.39	0	0.2	3.39	5.26	2001	Apple Inc.	2
AAPL	0.26	0	0.2	3.25	5.02	2002	Apple Inc.	3
AAPL	0.38	0	0.21	2.5	4.46	2003	Apple Inc.	4
AAPL	1.15	0.01	0.23	2.63		2004	Apple Inc.	5
AAPL	2.57	0.06	0.32	2.95		2005	Apple Inc.	6
AAPL	3.03	0.08	0.42	2.24		2006	Apple Inc.	7
AAPL	7.07	0.14	0.59	2.36		2007	Apple Inc.	8
AAPL	3.05	0.19	0.85	2.32		2008	Apple Inc.	9
AAPL	7.53	0.32	1.1	2.74		2009	Apple Inc.	10
AAPL	11.52	0.54	1.86	2.01		2010	Apple Inc.	11
AAPL	14.46	0.99	2.94	1.61		2011	Apple Inc.	12
AAPL	19.01	1.58	4.5	1.5		2012	Apple Inc.	13
AAPL	20.04	1.42	4.91	1.68	8.19	2013	Apple Inc.	14
AAPL	27.59	1.61	4.75	1.08	15.22	2014	Apple Inc.	15
AAPL	26.32	2.31	5.35	1.11	22.19	2015	Apple Inc.	16
AAPL	28.95	2.08	6.01	1.35	27.05	2016	Apple Inc.	17
AAPL	42.31	2.3	6.54	1.28	30.82	2017	Apple Inc.	18
AAPL	39.44	2.98	5.63	1.12	31.3	2018	Apple Inc.	19
AAPL	73.41	2.97	5.09	1.54	31.92	2019	Apple Inc.	20
AAPL	132.69	3.28	3.85	1.36	37.75	2020	Apple Inc.	21
AAPL	177.57	5.61	3.84	1.07	38.89	2021	Apple Inc.	22
AAPL	129.93	6.11	3.18	0.88	37.56	2022	Apple Inc.	23

VII. Appendix II

```
#import data
library("YRmisc")
library(readxl)
spMerge <- read_excel("C:/Users/Amritpal/Desktop/BUA 633/Data/spMerge.xlsx",sheet = "Sheet1")
View(spMerge)
spData<-spMerge
dim(spData)
names(spData)
spMerge <- read_excel("C:/Users/Amritpal/Desktop/BUA 633/Data/spMerge.xlsx",sheet = "Sheet2")
View(spMerge)
spInfo<-spMerge
dim(spInfo)
names(spInfo)
names(spData)
names(spInfo)
# note to self merge 2 dfs
spdf<-merge(spData,spInfo,by="tkr")
dim(spdf)
names(spdf)
spdf$date
```

```
spdf$year<-as.numeric(substring(spdf$date,7,10))
#TIME SERIES REGRESSION - choose a company tkr
unique(spdf$tkr)
names(spdf)
tsdf<-spdf[spdf$tkr=="AAPL",c("tkr","price","eps","bvps","cr","dta","year","name")]
tsdf
tsdf<-df.sortcol(tsdf,"year",FALSE)
dim(tsdf)
names(tsdf)
tsdf$obs<-1:23
names(tsdf)
tsdf[,2:6]<-round(tsdf[,2:6],2)\
#Graphical Methods
#Histograms
par(mfrow=c(2,2))
hist(tsdf$price,xlab="Price",ylab="Frequency",main="Fig. 1 Hist of Price", col="darkseagreen")
hist(tsdf$eps,xlab="Earnings Per Share",ylab="Frequency",main="Fig. 2 Hist of EPS", col="chocolate1")
hist(tsdf$bvps,xlab="Book Value Per Share",ylab="Frequency",main="Fig. 3 Hist of BVPS", col="cyan3")
#Time Series Plots
par(mfrow=c(2,2))
ts.plot(tsdf$price,xlab="Time",ylab="Price",main="Fig. 4 TSPlot of Price", col="darkseagreen")
ts.plot(tsdf$eps,xlab="Time",ylab="EPS",main="Fig. 5 TSPlot of EPS", col="chocolate1")
ts.plot(tsdf$bvps,xlab="Time",ylab="BVPS",main="Fig. 6 TSPlot of BVPS", col="cyan3")
#Scatter Plots
par(mfrow=c(2,2))
scatter.smooth(tsdf$eps,tsdf$price,xlab="Earnings Per Share",ylab="Price",main="Fig. 7 EPS vs. Price",type="n")
text(tsdf$eps,tsdf$price,as.character(tsdf$year),cex=.8, col="chocolate1")
scatter.smooth(tsdf$bvps,tsdf$price,xlab="Book Value Per Share",ylab="Price",main="Fig. 8 BVPS vs. Price",type="n")
text(tsdf$bvps,tsdf$price,as.character(tsdf$year),cex=.8, col="cyan3")
#Analytical Methods
#Decriptive Statistics
names(tsdf)
ds.summ(tsdf[,c("price","eps","bvps","obs")],2)[,-c(3,4,7,8)]
#Correlation Matrix
round(cor(na.omit(tsdf[,c("price","eps","bvps","obs")])),3)
#Regression - Linear Model (Parametric)
fit<-lm(price~eps+bvps+obs,na.action=na.omit,data=tsdf)
summary(fit)
# Residual Plots
par(mfrow=c(2,2))
hist(fit$residuals,main = "Fig. 9 Hist of LM Residuals", xlab="Residuals", ylab="Frequency", col="blueviolet")
scatter.smooth(tsdf$price,fit$fitted.values,main="Fig. 10 Predicted vs Actual Price LM", xlab="Predicted Price", ylab="Actual
Price", type="n")
text(fit$fitted.values,tsdf$price,as.character(tsdf$year),cex=.8, col="blueviolet")
pl.2ts(tsdf$price,fit$fitted.values,"Fig. 11 TSPlot LM")
#Regression - robust (linear parametric with outlier mitigation)
library("robust")
fit<-lmRob(price~eps+bvps+obs,na.action=na.omit,data=tsdf)
summary(fit)
par(mfrow=c(2,2))
hist(fit$residuals, main = "Fig. 12 Hist of Rob Residuals", xlab="Residuals", ylab="Frequency", col="darkgoldenrod1")
scatter.smooth(fit$fitted.values,tsdf$price, main="Fig. 13 Predicted vs Actual Price Rob", xlab="Predicted Price", ylab="Actual
Price", type="n")
text(fit$fitted.values,tsdf$price,as.character(tsdf$year),cex=.8, col="darkgoldenrod1")
pl.2ts(tsdf$price,fit$fitted.values,"Fig. 14 TSPlot Rob")
#Regression - General Additive Model (nonlinear nonparametric)
library("gam")
fit<-gam(price~s(eps,2)+s(bvps,2)+s(obs,2),na.action=na.omit,data=tsdf)
plot(fit)
par(mfrow=c(2,2))
plot(fit, main="Fig. 15 Gam of EPS", col="chocolate1") #repeat for all 4 and combine into a final image
cor(tsdf$price, fit$fitted.values)^2
```