A Statistical Model to Estimate Equity Price of Microsoft Corporation

By: Laura Adeniyi

I. Introduction

Microsoft Corporation stands as a prominent titan in the software industry, headquartered in Redmond, WA. The purpose of this research is to develop a statistical model that will forecast the equity price of Microsoft. This model will be of high use to portfolio managers, investment bankers and stockholders.

II. Methodology

This research uses time-series data derived from FactSet.com and there are 23 observations. The graphical techniques used are histograms, time series plots, and scatter plots. The analytical techniques used are descriptive statistics for the scalable variables, a correlation matrix and a multivariate regression. Additionally, this research uses R to perform the statistical analysis.

Eqn. 1: Functional specification

Eqn. 1 $$\text{Price} = f(\overset{?}{\text{OBS}}, \overset{+}{\text{EPS}}, \overset{+}{\text{BVPS}})$$

Eqn. 2: Population regression equation

Eqn. 2 $$\text{Price} = \alpha + \beta_{obs} * \text{OBS} + \beta_{eps} * \text{EPS} + \beta_{bvps} * \text{BVPS} + \varepsilon$$

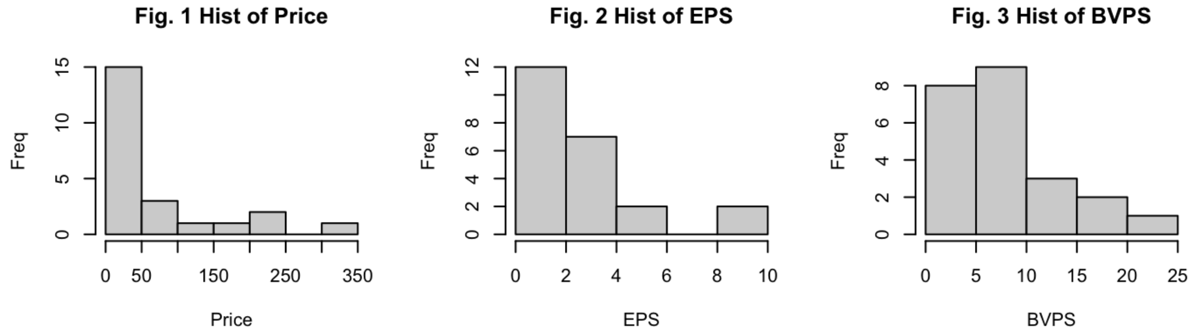
Eqn. 3: Sample regression equation

Eqn. 3 $$\text{Price} = a + b_{obs} * \text{OBS} + b_{eps} * \text{EPS} + b_{bvps} * \text{BVPS} + e$$

Price is hypothesized to be a positive function of earnings per share (EPS) and book value per share (BVPS) whereas the relationship between price and the trend also referred to as observations (OBS) is unclear.

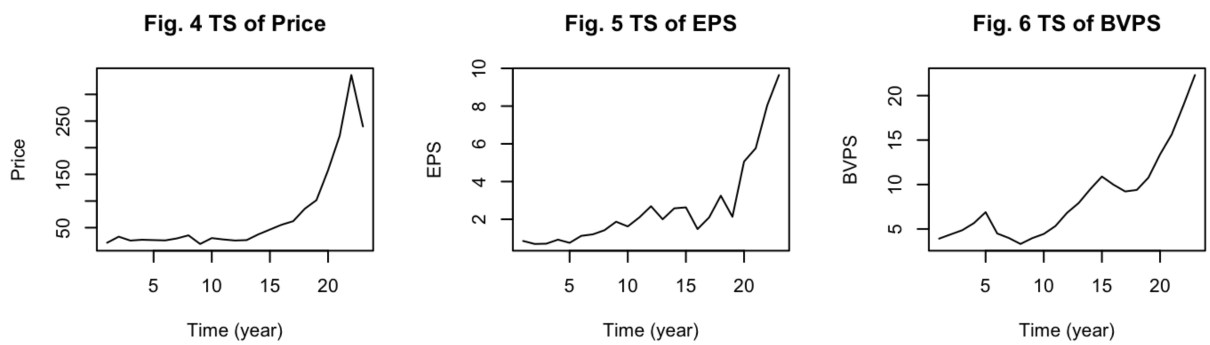
III. Results

Fig. 1 - Fig. 3: Histograms of Price, EPS and BVPS



As seen in Fig. 1 - Fig. 3 all three of the variables – Price, EPS, and BVPS – are skewed to the right.

Fig. 4 - Fig. 6: Time Series Plots of Each Variable



All three of the time series plots shown in Fig. 4 - Fig. 6 demonstrate an upward trend. The time series plot of price (Fig. 4) shows no cyclicality and little to no volatility. The time series plot of earnings per share (Fig. 5) is volatile while the time series plot of book value per share (Fig. 6) is cyclical.

Fig. 7 - Fig. 8: Scatterplots of Each Dependent Variables (BVPS, EPS) with the Independent Variable (Price)

Fig. 7 Price v BVPS

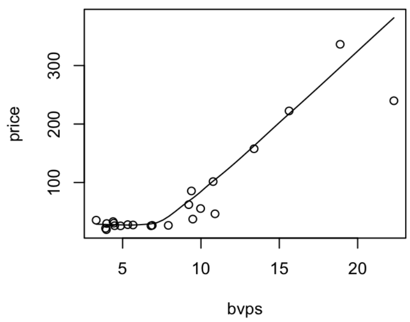


Fig. 8 Price v EPS

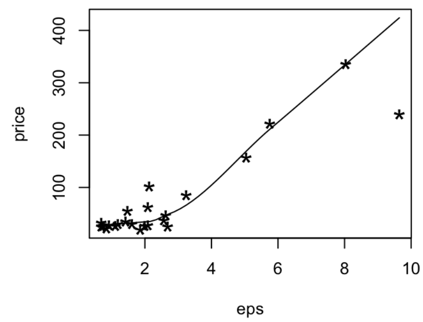


Fig. 7 displays a strong, positive, non-linear relationship between price and book value per share with few outliers and no heteroscedasticity. Fig. 8 displays a strong, positive, non-linear relationship between price and earnings per share with few outliers and no heteroscedasticity.

Table 1: Descriptive Statistics

	Name	***Obs***	***Max***	***Min***	***Mean***	***Median***	***Std***	***Skew***	***Kurt***
1	Price	23	336.32	19.4	73.99	33.12	84.53	1.8	5.67
2	Eps	23	9.65	0.69	2.64	2	2.35	1.68	5.4
3	Bvps	23	22.31	3.32	8.52	6.89	5.06	1.17	3.92

Table 1 shows the descriptive statistics.

Table 2: Correlation Matrix

	Price	**Obs**	**Eps**	**Bvps**
Price	1	0.75	0.92	0.9
Obs	0.75	1	0.8	0.86
Eps	0.92	0.8	1	0.93
Bvps	0.9	0.86	0.93	1

The correlation matrix in Table 2 shows a high correlation between all of the variables. EPS displays the highest correlation of 0.92 and OBS displaying the lowest correlation of 0.75 to the dependent variable – Price. The demonstrated correlations agree with the original hypothesis that price is hypothesized to be a positive function of EPS and BVPS. Note, there is a high correlation of 0.93 between EPS and BVPS, thus, there is evidence of multicollinearity.

Table 3: Regression Results

	Estimate	***Std. Error***	***t value***	***Pr(>|t|)***
(Intercept)	-28.24	17.045	-1.657	0.114
Obs	-1.071	2.059	-0.52	0.609
Eps	22.343	8.334	2.681	0.0148**
Bvps	6.595	4.55	1.449	0.1635

n = 23 r-sq. = 0.863 F = 40.2 SE = 33.56 DW=1.534

* - Significant at the 10% level of significance
** - Significant at the 5% level of significance
*** - Significant at the 1% level of significance

<u>Interpretation</u>

Eqn. 4: Estimated sample regression equation

Price = -28.240-1.071*OBS+22.343*EPS+6.595*BVPS

For every one unit increase in OBS, price decreases by $1.071. For every one unit increase in EPS, price increases by $22.343. For every one unit increase in BVPS, price increases by $6.595.

Note that the Durbin Watson (DW) statistic is < 1.7, thus, there is forecast bias.

1. The F-statistics of 40.2 > 4.99, therefore, the entire equation is statistically significant.
2. As determined by r-squared, 86.3% of the variation in price is explained by the variation in obs, eps, and bvps. The explanatory power is high.
3. For regression coefficients all of the t-statistics are significant at least at the 90% confidence level (t-stat >1.29) except for obs which is not statistically significant.
4. The standard error is 33.56

IV. Conclusion

The research was successful as it demonstrated a high explanatory power. Earnings per share and book value per share proved to be predictive and in agreement with the hypothesis that price is a positive function of earnings per share and book value per share. The trend was not predictive. This research can be improved by increasing the number of observations and including additional independent variables.

V. Appendix I: The Dataset

	tkr	price	eps	bvps	name	year	obs
20636	MSFT	21.6875	0.85	3.9152	Microsoft Corporation	2000	1
20646	MSFT	33.125	0.69	4.39244	Microsoft Corporation	2001	2
20637	MSFT	25.85	0.705	4.86845	Microsoft Corporation	2002	3
20647	MSFT	27.37	0.92	5.66521	Microsoft Corporation	2003	4
20641	MSFT	26.72	0.75	6.88869	Microsoft Corporation	2004	5
20644	MSFT	26.15	1.12	4.49253	Microsoft Corporation	2005	6
20652	MSFT	29.86	1.2	3.98569	Microsoft Corporation	2006	7
20638	MSFT	35.6	1.42	3.31525	Microsoft Corporation	2007	8
20633	MSFT	19.44	1.87	3.96525	Microsoft Corporation	2008	9
20634	MSFT	30.48	1.62	4.44073	Microsoft Corporation	2009	10
20650	MSFT	27.91	2.1	5.32707	Microsoft Corporation	2010	11
20651	MSFT	25.96	2.69	6.81507	Microsoft Corporation	2011	12
20654	MSFT	26.7097	2	7.91827	Microsoft Corporation	2012	13
20649	MSFT	37.41	2.58	9.47935	Microsoft Corporation	2013	14
20653	MSFT	46.45	2.63	10.89744	Microsoft Corporation	2014	15
20639	MSFT	55.48	1.48	9.9767	Microsoft Corporation	2015	16
20632	MSFT	62.14	2.1	9.22093	Microsoft Corporation	2016	17
20643	MSFT	85.54	3.25	9.39206	Microsoft Corporation	2017	18
20635	MSFT	101.57	2.13	10.77478	Microsoft Corporation	2018	19

20648	MSFT	157.7	5.06	13.38872	Microsoft Corporation	2019	20
20645	MSFT	222.42	5.7635	15.62594	Microsoft Corporation	2020	21
20642	MSFT	336.32	8.0535	18.88389	Microsoft Corporation	2021	22
20640	MSFT	239.82	9.647	22.3127	Microsoft Corporation	2022	23

VI. Appendix II: R Script

```
library("YRmisc")
library("readxl")

# 1. import spInfox excel file

spInfox <- read_excel("Desktop/BUA 633/Class 3 June 1/spInfox.xlsx")
View(spInfox)
spInfoSave<-spInfox
spInfox<-as.data.frame(spInfox)
dim(spInfox)
names(spInfox)


# 2. import spDatax  excel file

spDatax <- read_excel("Desktop/BUA 633/Class 3 June 1/spDatax.xlsx")
View(spDatax)
spDataSave<-spDatax
data.class(spDatax)
spDatax<-as.data.frame(spDatax)
dim(spDatax)
names(spDatax)

# 3. Merge 2 dataframes by "tkr" overlap

names(spInfox)
names(spDatax)
spmdf<-merge(spInfox,spDatax,by="tkr")
names(spmdf)
dim(spmdf)

# 4. create new numeric variable year extracted from date

spmdf$date
spmdf$year<-as.numeric(substring(spmdf$date,7,10))
names(spmdf)

# 5. Time Series Data Analysis

# extract time series data for 1 company MSFT (Microsoft)
tsdf<-spmdf[spmdf$tkr=="MSFT",c("tkr","price","eps","bvps","name","year")]
tsdf

#sorting in ascending order
```

```
tsdf<-df.sortcol(tsdf,6,desc = F)
tsdf

#creating obs variable
tsdf$obs<-1:nrow(tsdf)
tsdf
                                              ?    +     +
# function specification   price = f(obs, eps,bvps)
dim(tsdf)
names(tsdf)
tsdf

#6. Graphical Analysis

par(mfrow=c(3,3))  # par - partion
hist(tsdf$price,xlab = "Price",ylab="Freq",main="Fig. 1 Hist of Price")
hist(tsdf$eps,xlab = "EPS",ylab="Freq",main="Fig. 2 Hist of EPS")
hist(tsdf$bvps,xlab = "BVPS",ylab="Freq",main="Fig. 3 Hist of BVPS")

par(mfrow=c(3,3))  # par - partion
ts.plot(tsdf$price,xlab = "Time (year)",ylab="Price",main="Fig. 4 TS of Price")
ts.plot(tsdf$eps,xlab = "Time (year)",ylab="EPS",main="Fig. 5 TS of EPS")
ts.plot(tsdf$bvps,xlab = "Time (year)",ylab="BVPS",main="Fig. 6 TS of BVPS")

par(mfrow=c(2,2))
scatter.smooth(tsdf$bvps,tsdf$price,xlab="bvps",ylab="price",main="Fig. 7 Price v BVPS")
scatter.smooth(tsdf$eps,tsdf$price,xlab="eps",ylab="price",main="Fig. 8 Price v EPS",
    pch="*",cex=2)

#7. Statistical Analysis

# des stats
ds.summ(tsdf[,c("price","eps","bvps")],2)[,-c(7,8)]

# cor
tsdf
round(cor(tsdf[,c("price","obs","eps","bvps")]),2)

#lm
~  - "modeled as a function of obs, eps, bvps."
fit<-lm(price~obs+eps+bvps,data=tsdf,na.action=na.omit)
reg.dw(fit)
   summary(fit)
   names(fit)
```

```
  coefficients(fit)
  names(summary(fit))
  predValues<-predict(fit,tsdf)
  residValues<-residuals(fit)

# 7. Model Validation
  vdf<-data.frame(tsdf,predValues,residValues)
  par(mfrow=c(2,2))
  hist(vdf$residValues)
  plot(vdf$predValues,vdf$price,type="n")
  text(vdf$predValues,vdf$price,vdf$tkr)
  scatter.smooth(vdf$predValues,vdf$price,type="n")
  text(vdf$predValues,vdf$price,vdf$tkr)
  pl.2ts(vdf$price,vdf$predValues,"TSPlot Act v Pred")
```