A Statistical Model to Predict the Price of Equity of IBM Company

Pooja Bade

Predictive Analytics and Business Forecasting

Dr. Manuel G. Russon

St. John’s University

May 2024

A Statistical Model to Predict the Price of Equity of IBM Company

I. Introduction

IBM Company used for this research is among the most prominent IT company in the United States. This research tries to develop a predictive model to forecast the price of the IBM Company. In general, this research would be relevant to the IT Analysts, AI Investors, Portfolio Managers and Investment Specialists.

II. Previous Research

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

III. Methodology

The research is based on time series data as the observations are gathered at different point of time from 2000 to 2022. The data is secondary as it has been already collected and organized in Factset. Twenty three observations will be used for developing a statistical model to predict their prices. Histograms, Time series plots and Scatter plots of dependent variable with independent variables will be used as a graphical techniques and analytical methods such as Descriptive Statistics, Correlation and Regression will be used in this research. R will be as a statistical program to conduct a statistical analysis.

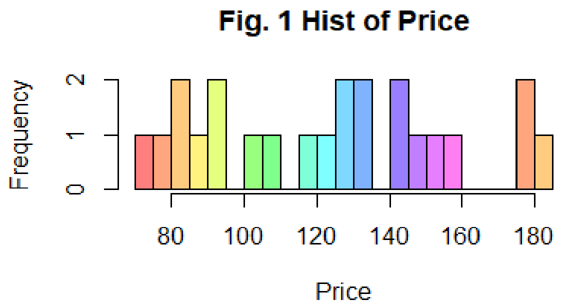
The functional specification, population regression equation and sample regression equation are:

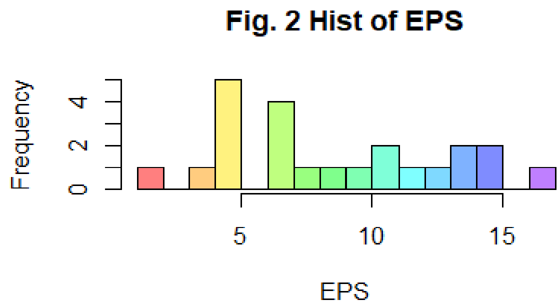
Eqn. 1	$Price = f(\overset{+}{EPS}, \overset{+}{BVPS}, \overset{+}{OBS})$	Functional Specification
Eqn. 2	$Price = \alpha + \beta_{EPS}*EPS + \beta_{BVPS}*BVPS + \beta_{OBS}*OBS$	Population Regression line
Eqn. 3	$Price = a + b_{EPS}*EPS + b_{BVPS}*BVPS + b_{OBS}*OBS$	Sample Regression line

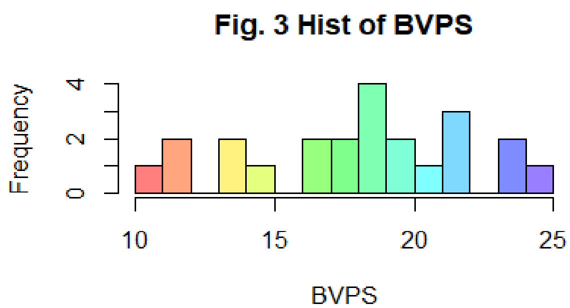
Price is the dependent variable and EPS, BVPS and OBS are independent variables. The above equation shows the mathematical and linear relationship between Price and the independent variables EPS, BVPS and OBS.

IV. Results

Fig. 1 to 4 displays Histograms of all variables







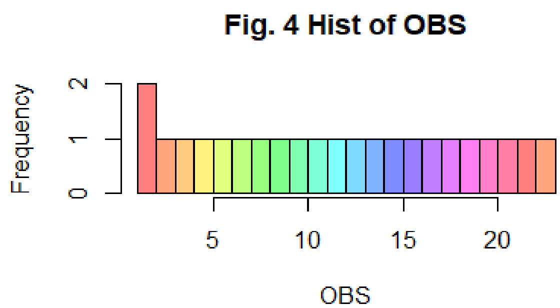


Fig. 1: The histogram of price displays a moderately symmetric distribution, with data spread relatively evenly around the center. There is a slight right skew, suggesting a longer tail to the right compared to the left. Some outliers can be observed on the right side of the distribution, indicated by isolated bars far from the main concentration of bars.

Fig. 2: The histogram of EPS also exhibits a moderately symmetric distribution, with data spread relatively evenly around the center. There is a slight left skew, indicating a longer tail to the left compared to the right. Similarly, some outliers can be present on the left side of the distribution, indicated by isolated bars far from the main concentration of bars.

Fig. 3: In the histogram of BVPS, a distribution that appears moderately symmetric can be observed, with data spread relatively evenly around the center. There is a slight right skew, suggesting a longer tail to the right compared to the left. Some outliers can be visible on the right side of the distribution, indicated by isolated bars far from the main concentration of bars.

Fig. 4: The histogram of OBS displays a moderately symmetric distribution, with data spread relatively evenly around the center. There is a slight left skew, indicating a longer tail to the left compared to the right. Additionally, some outliers are present on the left side of the distribution, indicated by isolated bars far from the main concentration of bars.

Fig. 5 to 8 displays Time series Plot of all the variables

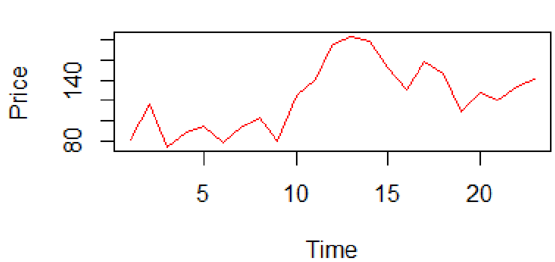


Fig. 5 Ts Plot of Price

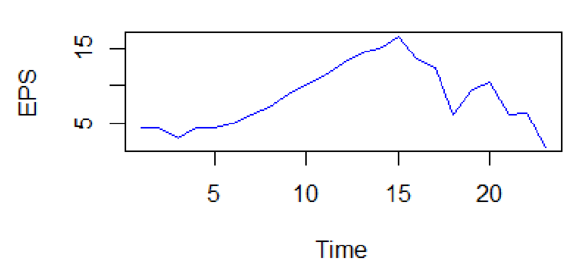


Fig. 6 Ts Plot of EPS

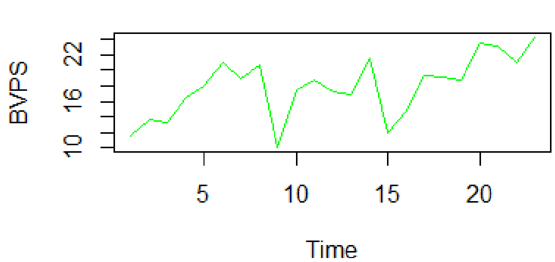


Fig. 7 Ts Plot of BVPS

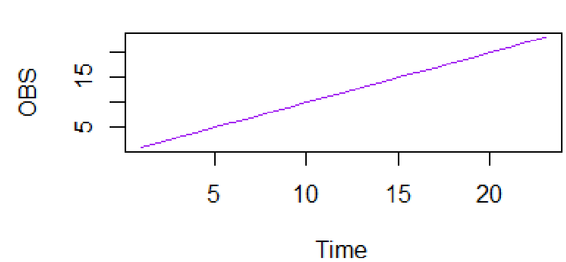


Fig. 8 Ts Plot of OBS

Fig. 5: The time series plot of price is trending upward, however fluctuations can be observed in different time intervals. This trend is volatile as it fluctuates up and down without a consistent pattern.

Fig. 6: The time series plot of EPS is trending upward up to some years and then going downward. Similar to the Price plot, the EPS trend is also volatile, showing significant fluctuations over time.

Fig. 7: The trend in this time series plot of BVPS is cyclical, showing regular up and down movements indicating a pattern that repeats over the given time period.

Fig. 8: The trend in this time series plot of OBS is neither volatile nor cyclical. Instead, it shows a steady upward linear trend, indicating a consistent increase over time.

Fig. 9 to 11 displays Scatter Plot of all the Independent variables vs. Price

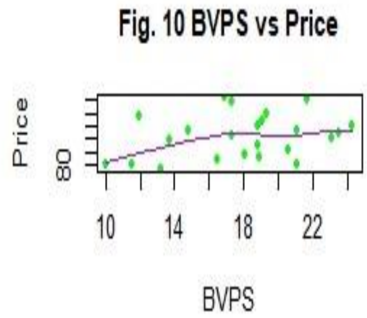


Fig. 9: The relationship between EPS and Price shows a moderate strength, indicating that changes in EPS are somewhat associated with changes in Price. It is a positive relationship, suggesting that as EPS increases, Price tends to increase as well. The pattern is nonlinear, with some curvature observed. There are few outliers, and there's some evidence of heteroscedasticity, indicating unequal variances in Price at different levels of EPS.

Fig. 10: The relationship between BVPS and Price appears moderately strong, indicating that changes in BVPS are associated with changes in Price. It is a positive relationship, implying that as BVPS increases, Price tends to increase as well. The pattern seems somewhat linear, although there might be some curvature present. There are few outliers, and there's some indication of heteroscedasticity, suggesting unequal variances in Price at different levels of BVPS.

Fig. 11: The relationship between OBS and Price seems weak, indicating little association between the two variables. The pattern appears nonlinear, with no apparent linear trend. There are few outliers present, and there's some evidence of heteroscedasticity, suggesting unequal variances in Price at different levels of OBS.

Table 1 displays Descriptive Statistics

<u>Table 1</u> **<u>Descriptive Statistics</u>**

	name	obs	max	min	mean	median	std	skew	kurt
1	price	23	182.96	74.03	123.15	125.03	33.45	0.19	1.99
2	obs	23	23.00	1.00	12.00	12.00	6.78	0.00	1.80
3	eps	23	16.58	1.80	8.47	7.18	4.25	0.30	1.89
4	bvps	23	24.22	10.06	17.88	18.77	3.90	-0.32	2.31

Table 2 displays Correlation Matrix

<u>Table 2</u> **<u>Correlation Matrix</u>**

	price	obs	eps	bvps	cr	dta
price	1.000	0.579	0.706	0.290	-0.070	0.297
obs	0.579	1.000	0.324	0.589	-0.540	0.769
eps	0.706	0.324	1.000	-0.106	0.255	0.064
bvps	0.290	0.589	-0.106	1.000	-0.443	0.303

Variables with correlation coefficients close to 1 or -1 have high correlation with price. From the correlation matrix, Price has a moderate positive correlation with OBS (0.579), a strong positive correlation with EPS (0.706) and a weak positive correlation with BVPS (0.290). So, from the hypothesis that as OBS, EPS, BVPS increase, Price increases, it seems that there's partial agreement with the hypothesis. OBS, EPS and BVPS show positive correlations with Price, supporting the hypothesis.

Regarding multicollinearity, there might be some evidence of multicollinearity, especially between OBS and BVPS, as they have a moderate positive correlation (0.789).

Table 3 displays Regression Results

<u>Table 3</u> **<u>Regression Results</u>**

Eqn. 4	P = 32.16	+ 5.13EPS	+ 1.87BVPS	+ 1.18OBS
t-stat	1.21	4.26 ***	1.22	1.28
p-value	0.24	0.00	0.24	0.22
r (corr)		0.71	0.29	0.58

n = 23 r-sq. = 0.66 F = 12.41 F-Prob= 0.001 SE = 20.92

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

<u>Interpretation of Regression Results:</u>

From Eqn. 4: As EPS increases by $1, price increases by $5.13 on average other things remaining equal. As BVPS increases by $1, price increases by $1.87 on average other things remaining equal. As OBS increases by 1 unit, price increases by $1.18 on average other things remaining equal.

Results of an F-test for the entire model:

Ho: $\beta_{EPS}=\beta_{BVPS}=\beta_{OBS}= 0$

* 5% Ha: at least 1 β_i not equal to 0 (12.41 > 4.99)

As F test at 12.41 is greater than cutoff point 4.99, it is significant, thus Ha is accepted. R-squared (coefficient of determination) is 66%, so 66% variation in price which can be explained by or attributed to variation in EPS, BVPS and OBS. Thus, it's ability to predict is high. Standard Error is 20.92 which is the absolute margin error in the units of Price. As t-stat for EPS i.e 4.26 is greater than 2.34, it is significant at 1%.

Fig. 12 to 14 displays Histogram, Scatter Plot and Time series Plot of Residuals

Fig. 12 Histogram of Residuals

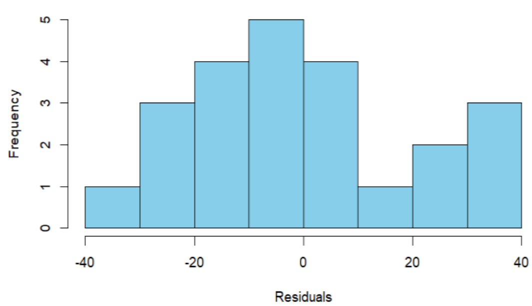


Fig. 14 Time Series Plot of Price

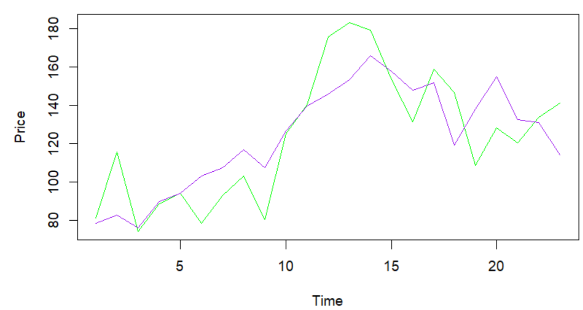
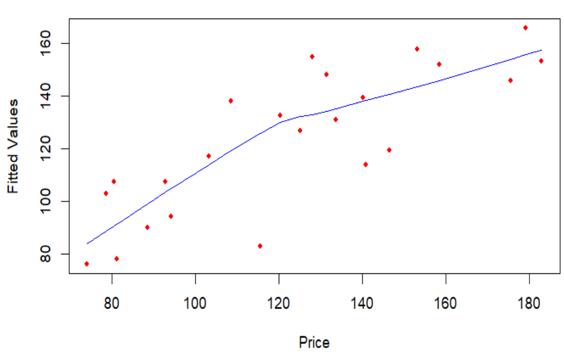


Fig. 13 Price vs Fitted Values



<u>Fig. 12:</u> It shows the histogram of residuals

<u>Fig. 13:</u> It shows the scatter plot of residuals i.e Price vs Fitted values

<u>Fig. 14:</u> It shows the time series plot of residuals i.e Time vs Price

<u>V. Conclusions</u>

The research was highly successful. The explanatory power was high as it's coefficient of determination was 66% and only independent variable EPS was predictive. It seems there is no public implication of this research. This research can be improved in the future by increasing the number of observations and adding more independent variables.

VI. Bibliography

FactSet Universal Screening

VII. Appendix I

Excel

tkr	date	price	eps	bvps	sales	ebit	assets	cso	liabs	ca	cl	cr	dta
IBM	12/31/2000	81.19	4.44	11.56	88396	11634	85381	1762.90	64757	43880	36406	1.21	33.47
IBM	12/31/2001	115.54	4.35	13.70	85866	9295	85918	1723.19	62304	42461	35119	1.21	31.60
IBM	12/31/2002	74.03	3.07	13.23	81186	8943	92170	1722.37	69388	41652	34220	1.22	28.23
IBM	12/31/2003	88.52	4.34	16.44	89131	10513	100169	1694.51	72305	44662	37623	1.19	23.59
IBM	12/31/2004	94.16	4.38	18.08	96293	10013	111003	1645.59	79315	47143	39786	1.18	20.65
IBM	12/31/2005	78.51	4.90	21.03	91134	11148	105748	1573.98	72650	45661	35152	1.30	21.41
IBM	12/31/2006	92.79	6.11	18.92	91424	12218	103234	1506.48	74728	44660	40091	1.11	21.97
IBM	12/31/2007	103.25	7.18	20.55	98786	13834	120431	1385.23	91961	53177	44310	1.20	29.29
IBM	12/31/2008	80.39	8.89	10.06	103630	16675	109524	1339.10	95939	49004	42435	1.15	30.97
IBM	12/31/2009	125.03	10.01	17.34	95758	17173	109022	1305.34	86267	48935	36002	1.36	23.94
IBM	12/31/2010	140.18	11.52	18.77	99870	18656	113452	1227.99	90280	48116	40562	1.19	25.23
IBM	12/31/2011	175.64	13.06	17.31	106916	20786	116433	1163.18	96198	50928	42123	1.21	26.90
IBM	12/31/2012	182.96	14.37	16.88	104507	20989	119213	1117.37	100229	49433	43625	1.13	27.91
IBM	12/31/2013	179.16	14.94	21.62	98367	20191	126223	1054.39	103294	51350	40154	1.28	31.47
IBM	12/31/2014	153.25	16.58	11.98	92793	19105	117271	990.52	105257	47377	39581	1.20	34.72
IBM	12/31/2015	131.45	13.54	14.77	81741	15173	110495	965.73	96071	42504	34269	1.24	36.10
IBM	12/31/2016	158.55	12.43	19.29	79919	12323	117470	945.87	99078	43888	36275	1.21	35.90
IBM	12/31/2017	146.54	6.14	19.08	79139	11678	125356	922.18	107631	47915	37363	1.28	37.35
IBM	12/31/2018	108.57	9.52	18.82	79591	12799	123382	892.48	106452	46846	38227	1.23	37.13
IBM	12/31/2019	128.03	10.56	23.49	77147	10617	152186	887.11	131201	36524	37701	0.97	44.79
IBM	12/31/2020	120.24	6.23	23.07	55179	6657	155971	892.65	135245	38147	39869	0.96	41.73
IBM	12/31/2021	133.66	6.35	21.05	57351	7023	132001	898.07	113005	28442	33619	0.85	41.77
IBM	12/31/2022	140.89	1.80	24.22	60530	7765	127243	906.09	105222	28151	31505	0.89	42.45

VIII. Appendix II

R Script

```
ibrary(readxl)
spMerge <- read_excel("C:/Users/Dell/Downloads/spMerge.xlsx", sheet = "Sheet1")
View(spMerge)

spData<-spMerge
dim(spData)
names(spData)

spMerge <- read_excel("C:/Users/Dell/Downloads/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)) # extract from 7 to 10

#CROSS SECTION ANALYSIS
csdf<-spdf[spdf$year==2022,]
dim(csdf)
names(csdf)
unique(csdf$sector)
#Choose an industry
unique(csdf$industry)

secdf<-csdf[csdf$sector=="Financials",c(3,4,5,13,14)]
secdf
secdf<-csdf[csdf$sector=="Financials",c("tkr","price","eps","bvps","cr","dta")]
secdf
dim(secdf)
secdf

#histograms
#scatterplots

#descriptive statistics
#correlation analysis
#regression

fit<-lm(price ~ eps+bvps+cr+dta,na.action=na.omit,data=secdf)


library("YRmisc")
#TIME SERIES REGRESSION - choose a company tkr
```

```
unique(spdf$tkr)
length(unique(spdf$tkr))

#Enter ticker below  - IBM
names(spdf)
tsdf<-spdf[spdf$tkr=="IBM",c("tkr","price","eps","bvps","cr","dta","year","name")]
tsdf
data.class(tsdf)
is.data.frame(tsdf)
tsdf<-df.sortcol(tsdf,"year",FALSE)
dim(tsdf)
names(tsdf)
tsdf$obs<-1:23 #grab tsdf data and adding vector of 1 to 23 in OBS
names(tsdf)
tsdf[,2:5]<-round(tsdf[,2:5],2)

par(mfrow=c(3,3))
	hist(tsdf$price)
	hist(tsdf$eps)
	hist(tsdf$bvps)
	hist(tsdf$obs)

	# Set up the layout
	par(mfrow = c(2, 2))
	# Define color palette
	colors <- c("#FF5733", "#FFC300", "#C70039", "#900C3F", "#581845")
	# Histogram of Price
	hist(tsdf$price, breaks = 20 , xlab="Price", ylab="Frequency", main="Fig. 1 Hist of Price", col = 
rainbow(20, alpha = 0.5), border = 'black', cex= 0.8, width = 1)
	# Histogram of EPS
	hist(tsdf$eps, breaks = 20, xlab="EPS", ylab="Frequency", main="Fig. 2 Hist of EPS", col = 
rainbow(20, alpha = 0.5), border = 'black', cex= 0.8, width = 1)
	# Histogram of BVPS
	hist(tsdf$bvps, breaks = 20, xlab="BVPS", ylab="Frequency", main="Fig. 3 Hist of BVPS", col = 
rainbow(20, alpha = 0.5), border = 'black', cex= 0.8, width = 1)
	# Histogram of OBS
	hist(tsdf$obs, breaks = 20, xlab="OBS", ylab="Frequency", main="Fig. 4 Hist of OBS", col = 
rainbow(20, alpha = 0.5), border = 'black', cex= 0.8, width = 1)




par(mfrow=c(3,3))
	ts.plot(tsdf$price)
	ts.plot(tsdf$eps)
	ts.plot(tsdf$bvps)
	ts.plot(tsdf$obs)

	# Set up the layout
	par(mfrow=c(2,2))

	# Generate plots with different colors
	ts.plot(tsdf$price, col="red", main=" Fig. 5 Ts Plot of Price", ylab="Price", xlab="Time")
	ts.plot(tsdf$eps, col="blue", main="Fig. 6 Ts Plot of EPS", ylab="EPS", xlab="Time")
	ts.plot(tsdf$bvps, col="green", main="Fig. 7 Ts Plot of BVPS", ylab="BVPS", xlab="Time")
	ts.plot(tsdf$obs, col="purple", main="Fig. 8 Ts Plot of OBS", ylab="OBS", xlab="Time")
```

```
par(mfrow=c(3,3))
        scatter.smooth(tsdf$eps,tsdf$price)
        scatter.smooth(tsdf$bvps,tsdf$price)
        scatter.smooth(tsdf$obs,tsdf$price)

        # Set up the layout
        par(mfrow=c(3,3))

        # Generate scatter plots with different colors and smooth lines
        plot(tsdf$eps, tsdf$price, main="Fig. 9 EPS vs Price", xlab="EPS", ylab="Price", col="red",
pch=20)
        lines(lowess(tsdf$eps, tsdf$price), col="blue")

        plot(tsdf$bvps, tsdf$price, main="Fig. 10 BVPS vs Price", xlab="BVPS", ylab="Price",
col="green", pch=20)
        lines(lowess(tsdf$bvps, tsdf$price), col="purple")

        plot(tsdf$obs, tsdf$price, main="Fig. 11 OBS vs Price", xlab="OBS", ylab="Price", col="orange",
pch=20)
        lines(lowess(tsdf$obs, tsdf$price), col="brown")

names(tsdf)
#analytical methods
ds.summ(tsdf[,c("price","obs","eps","bvps")],2)
round(cor(na.omit(tsdf[,c("price","obs","eps","bvps")])),3)
fit<-lm(price~eps+bvps+obs,na.action=na.omit,data=tsdf)
summary(fit)

# Residual Plots
hist(fit$residuals)
scatter.smooth(tsdf$price,fit$fitted.values)
pl.2ts(tsdf$price,fit$fitted.values,"tsplot")

# Generate a histogram with different colors
hist(fit$residuals, main="Fig. 12 Histogram of Residuals", xlab="Residuals", col="skyblue",
border="black")

# Generate a scatter plot with smooth line
plot(tsdf$price, fit$fitted.values, main="Fig. 13 Price vs Fitted Values", xlab="Price", ylab="Fitted
Values", col="red", pch=20)
lines(lowess(tsdf$price, fit$fitted.values), col="blue")

# Generate a time series plot
plot.ts(tsdf$price, col="green", main="Fig. 14 Time Series Plot of Price", ylab="Price")
lines(fit$fitted.values, col="purple")
```