

Stocks Fundamental Analysis

Vanessa Hu

Project 2 - Web Scraping & Linear Regression
October 9, 2020

Stock Valuation

Over value? Under value?



EPS Equity per Share

Balance Sheet

Income Statement

Cash Flow

Web Scraped

Yahoo Finance

Stock Market

NASDAQ

Sectors

Technology, Healthcare
Finance, Consumer Services

Time

2019 Annual Report

Rows, Columns

1206, 76

THE DATA

Correlation Heat Map

Null < 5%

1	-0.33	-0.47	-0.023	0.13	0.15	-0.39	-0.33	0.15	-0.32	0.092	0.029	0.014	0.038	0.038	-0.21	0.048	0.2	0.88	0.32	0.32	0.36	0.32	0.32
-0.33	1	0.94	0.84	0.78	0.69	0.94	0.94	0.69	0.89	-0.54	0.83	0.71	0.51	0.51	-0.36	0.76	0.61	-0.44	0.54	0.54	0.52	0.54	0.53
-0.47	0.94	1	0.62	0.53	0.44	0.83	0.81	0.45	0.81	-0.52	0.67	0.62	0.5	0.5	-0.16	0.65	0.48	-0.54	0.4	0.4	0.37	0.4	0.4
-0.023	0.84	0.62	1	0.98	0.91	0.89	0.92	0.91	0.8	-0.45	0.87	0.67	0.4	0.4	-0.6	0.76	0.67	-0.17	0.63	0.63	0.63	0.63	0.62
0.13	0.78	0.53	0.98	1	0.94	0.81	0.85	0.94	0.73	-0.45	0.88	0.67	0.42	0.42	-0.61	0.77	0.71	0.00094	0.69	0.69	0.7	0.69	0.69
0.15	0.69	0.44	0.91	0.94	1	0.72	0.75	1	0.67	-0.46	0.83	0.65	0.4	0.4	-0.5	0.72	0.68	0.05	0.68	0.68	0.68	0.68	0.68
-0.39	0.94	0.83	0.89	0.81	0.72	1	1	0.72	0.89	-0.56	0.83	0.62	0.48	0.49	-0.34	0.75	0.61	-0.52	0.52	0.52	0.49	0.52	0.51
-0.33	0.94	0.81	0.92	0.85	0.75	1	1	0.75	0.89	-0.56	0.85	0.64	0.49	0.49	-0.39	0.77	0.63	-0.46	0.55	0.55	0.53	0.55	0.55
0.15	0.69	0.45	0.91	0.94	1	0.72	0.75	1	0.67	-0.46	0.83	0.65	0.4	0.4	-0.5	0.72	0.68	0.05	0.68	0.68	0.68	0.68	0.68
-0.32	0.89	0.81	0.8	0.73	0.67	0.89	0.89	0.67	1	-0.51	0.79	0.75	0.49	0.49	-0.32	0.72	0.54	-0.44	0.46	0.46	0.46	0.47	0.46
0.092	-0.54	-0.52	-0.45	-0.45	-0.46	-0.56	-0.56	-0.46	-0.51	1	-0.75	-0.44	-0.89	-0.89	-0.34	-0.84	-0.85	0.13	-0.81	-0.81	-0.78	-0.81	-0.8
0.029	0.83	0.67	0.87	0.88	0.83	0.83	0.85	0.83	0.79	-0.75	1	0.73	0.72	0.72	-0.33	0.97	0.9	-0.076	0.87	0.87	0.86	0.87	0.87
0.014	0.71	0.62	0.67	0.67	0.65	0.62	0.64	0.65	0.75	-0.44	0.73	1	0.47	0.47	-0.39	0.69	0.57	0.021	0.58	0.58	0.58	0.59	0.58
0.038	0.51	0.5	0.4	0.42	0.4	0.48	0.49	0.4	0.49	-0.89	0.72	0.47	1	1	0.3	0.82	0.85	0.042	0.82	0.82	0.81	0.82	0.82
0.038	0.51	0.5	0.4	0.42	0.4	0.49	0.49	0.4	0.49	-0.89	0.72	0.47	1	1	0.29	0.82	0.85	0.042	0.82	0.82	0.81	0.82	0.82
-0.21	-0.36	-0.16	-0.6	-0.61	-0.5	-0.34	-0.39	-0.5	-0.32	-0.34	-0.33	-0.39	0.3	0.29	1	-0.16	-0.06	-0.1	-0.085	-0.085	-0.12	-0.086	-0.086
0.048	0.76	0.65	0.76	0.77	0.72	0.75	0.77	0.72	0.72	-0.84	0.97	0.69	0.82	0.82	-0.16	1	0.95	-0.041	0.92	0.92	0.91	0.92	0.92
0.2	0.61	0.48	0.67	0.71	0.68	0.61	0.63	0.68	0.54	-0.85	0.9	0.57	0.85	0.85	-0.06	0.95	1	0.13	0.99	0.99	0.98	0.99	0.99
0.88	-0.44	-0.54	-0.17	0.00094	0.05	-0.52	-0.46	0.05	-0.44	0.13	-0.076	0.021	0.042	0.042	-0.1	-0.041	0.13	1	0.27	0.27	0.31	0.28	0.28
0.32	0.54	0.4	0.63	0.69	0.68	0.52	0.55	0.68	0.46	-0.81	0.87	0.58	0.82	0.82	-0.085	0.92	0.99	0.27	1	1	0.99	1	1
0.32	0.54	0.4	0.63	0.69	0.68	0.52	0.55	0.68	0.46	-0.81	0.87	0.58	0.82	0.82	-0.085	0.92	0.99	0.27	1	1	0.99	1	1
0.36	0.52	0.37	0.63	0.7	0.68	0.49	0.53	0.68	0.46	-0.78	0.86	0.58	0.81	0.81	-0.12	0.91	0.98	0.31	0.99	0.99	1	0.99	1
0.32	0.54	0.4	0.63	0.69	0.68	0.52	0.55	0.68	0.47	-0.81	0.87	0.59	0.82	0.82	-0.086	0.92	0.99	0.28	1	1	0.99	1	1
0.32	0.53	0.4	0.62	0.69	0.68	0.51	0.55	0.68	0.46	-0.8	0.87	0.58	0.82	0.82	-0.086	0.92	0.99	0.28	1	1	1	1	1

Features Engineering

"Diluted Equity" = Diluted EPS * Diluted Average Share

TARGET

'ROA', 'Total Liabilities Net Minority Interest',

'Total Equity Gross Minority Interest', 'Common Stock Equity',

'Tangible Book Value', 'Invested Capital', 'End Cash Position',

'Financing Cash Flow', 'Operating Cash Flow', 'Total Revenue',

'Investing Cash Flow', 'Tax Effect of Unusual Items', 'Pretax Income'

FEATURES

Train

Validation

Test

0.9899

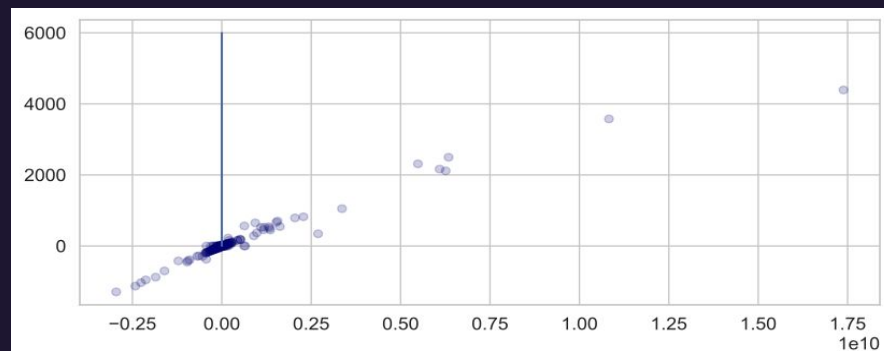
0.9819

534384176.33

—— R-squared ——

MAE

Lasso Regression Model



Why?

Residuals:

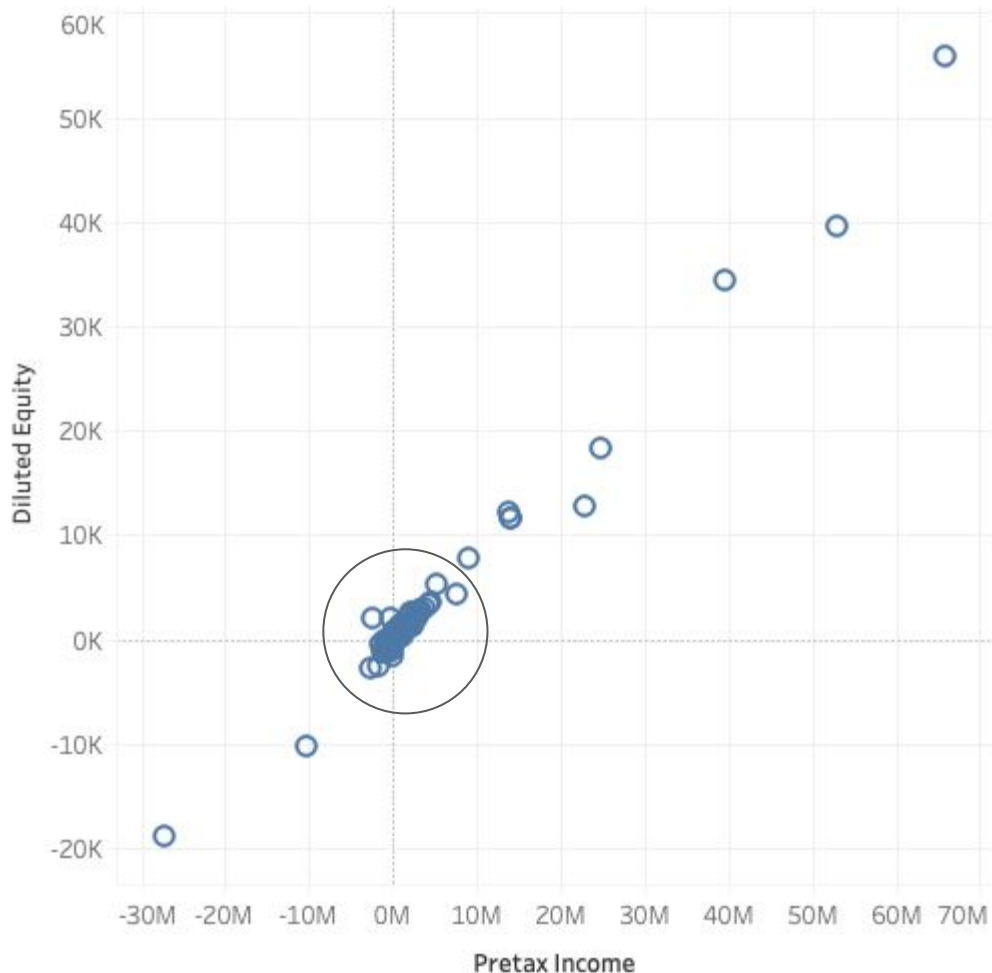
Min	1Q	Median	3Q	Max
-3709.8523	-6.958	2.4008	15.4998	2778.448

Coefficients:

	Estimate	Std. Error	t value	p value
_intercept	122.183520	9.150959	13.3520	0.000000
x1	0.000000	9.153609	0.0000	1.000000
x2	0.000000	54.071287	0.0000	1.000000
x3	0.000000	187.965176	0.0000	1.000000
x4	0.000000	148.544682	0.0000	1.000000
x5	0.000000	24.026073	0.0000	1.000000
x6	0.000000	102.745873	0.0000	1.000000
x7	0.000000	47.157220	0.0000	1.000000
x8	-55.212619	63.394978	-0.8709	0.384093
x9	148.888806	76.289080	1.9516	0.051382
x10	0.000000	19.299316	0.0000	1.000000
x11	0.000000	39.837822	0.0000	1.000000
x12	-0.000000	51.329737	-0.0000	1.000000
x13	2131.718488	41.220481	51.7150	0.000000

R-squared: 0.98992, Adjusted R-squared: 0.98973

F-statistic: 5152.68 on 13 features



Outlier Issue +

Only 285 positive EPS

Tech Giants

GOOG

MSFT

YNDX

OUTLIERS

Foreign ADR

JFU IQ

YJ YY

-10 M < Pretax Income < 10 M

Simple Linear Regression

Train

Validation

Test

0.9445

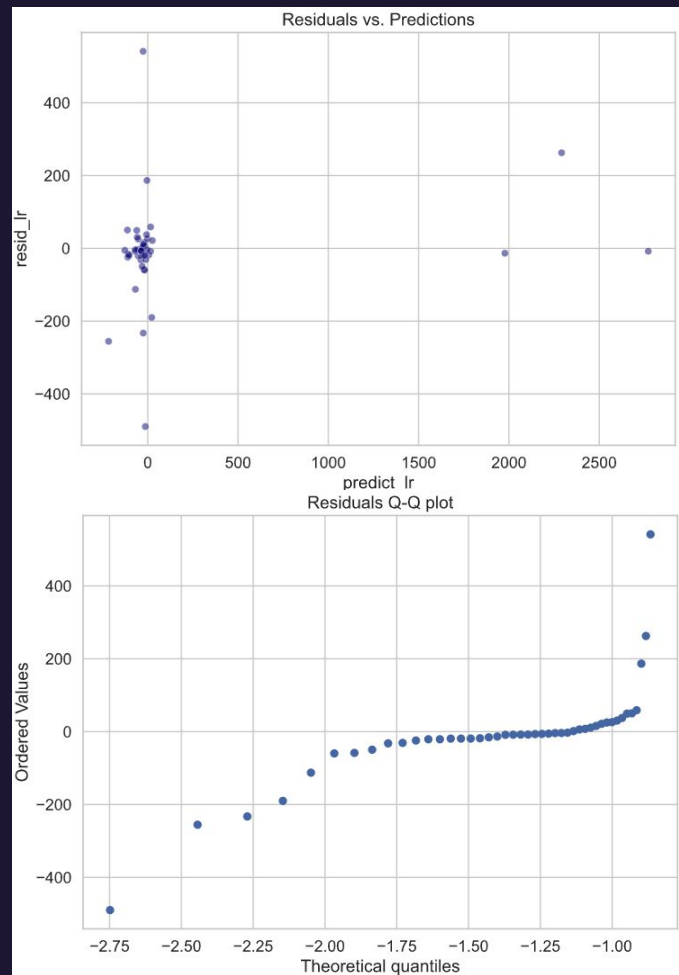
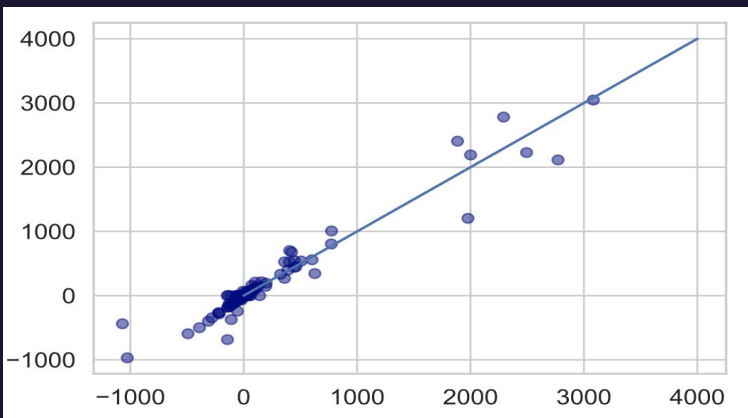
0.6988

0.9342

R-squared

45.60

MAE



Future Steps

Large Dataset

Generalized Nature

Matured Company

S&P 500

Questions?