Final Data Analysis of Dow Jones Industrial Average:

2007 - 2017

Josh Cubero

BUS240

Ageno School of Business

Golden Gate University

December 12, 2021

Introduction

Dow Jones Industrial Average

History

The Dow Jones Industrial Average (DJIA) is a list of publicly traded companies' stocks that make up a weighted average index used to indicate the trend of stocks and bonds in the overall market. The index was found in 1896 by Dow Jones & Company, a financial newspaper. The newspaper's founders, Charles Dow & Edward D. Jones, began calculating the industrials average daily, which comprised of 12 stocks. The DJIA became an index of 30 stocks in 1928 and has remained so since, although the companies comprising the DJIA have changed over the years. Additionally, the index's divisors have been adjusted to account for changes stock splits, substitutions, and changes to dividend yields. The DJIA is not intended as an arithmetic mean, but a set of averages that indicate stock market price trends.

Debate

While the DJIA has been in use for over 130 years, the index has received criticism from economists. There's a school of thought that the health US economy is just too complex, too nuanced to be accurately depicted by a set of just 30 equities. Additionally, the DJIA will only perform as well as the underlying assets. Hypothetically speaking, a company such as McDonald's is entangled in litigation, a selloff of the stock could occur bringing down the DJIA. However, at the same time, other fast-food restaurants are experiencing growth. McDonald's performance on the DJIA would indicate a decline in the fast-food industry, when in fact, it was just McDonald's experiencing decline. Despite the complexity of the stock market, the DJIA is likely to remain a leading indicator of US economic health.

Factors Impacting

Furthering the debate surrounding the DJIA, what exactly are the factors that are linked to increases and decreases in the index? There are many factors that are generally accepted to cause the DJIA to increase or decrease. Broadly speaking, these factors could be company news and performance, industry performance, investor sentiment, and economic factors. This analysis will focus on selected, not necessarily exhaustive, economic factors that are believed to explain movement in the DJIA.

Hypothesis

Synopsis

The economy is a major topic of discussion all over the world. US presidential elections of been won and lost on the real or perceived performance of the economy, which many believe to be represented by the DJIA. This analysis seeks to determine exactly which economic factors explain movement in the DJIA with the following hypothesis below.

Ho: All major economic indicators explain the movement of the Dow Jones Industrial Average

Ha: At least one of the major economic indicators does not explain movement in the Dow Jones Industrial Average

Data Collection

Source

The data in this dataset was obtain from the data sharing website https://www.kaggle.com/. Kaggle is a website frequently used by both experienced and aspiring data scientists to code, project ideas, inspiration, and data. Additionally, Kaggle holds competitions and challenges where users can compete and test their skills. This dataset is not, nor intended to be an exhaustive list of economic indicators.

Variables

There are 13 variables in this dataset, with the Dow Jones Industrial Average as the response variable. The other 12 variables are candidates and will be analyzed, with each explanatory variable subject to removal after regression.

Data	Descrption
Consumer Confidence Index	Consumer planned major purchases
Composite Leading Indicator	Provides indications of future business cycle changes
Business Confidence Indicators	Enterprise assesment production orders and inventories
Employment Rate	% Rate of people available to work
Inventory Turnover	% Turnover inventory sold and re-stocked
Purchasing Manager's Index	Indicates health of manufacturing sector
The Advance Report on Durable Goods Manufacturer's Shipments, Inventories and Orders	Measures manufacturing activity levels
Number of Building Permits	Count of requested building permits
Sales	Sum of US retail sales, excluding food services
Payroll	Total US payroll added to work force, non-farm labor
Personal Consumption Expenditures	Total number US consumer spending on goods & services
Federal Funds Rate	% US Federal Reserve interest rate
Dow Jones Industrial Average	Weighted index of 30 stocks that represent health of US economy

Descriptive Statistics

The analysis in this paper will begin with a chart of descriptive statistics. The intent is to analyze the descriptive statistics and gain a better understanding of the data. Table 1 below displays the summary statistics of the data set.

Summary Statistics					
	Mean	Median	Standard Deviation	Minimum	Maximum
ConsConf	99.15	99.00	1.34	96.69	101.30
CompLead	99.70	99.80	1.50	94.59	101.91
BusConf	99.82	99.88	1.04	96.01	101.37
Emp	68.53	68.21	1.70	66.39	72.12
InvToSales	1.32	1.29	0.06	1.24	1.48
PMI	52.35	52.60	5.24	32.40	61.40
MfgOrdDur	214,130.14	221,898	25,346.47	143,769	290,709
BldgPerm	956.59	1,008	282.28	513	1,626
Sales	358,089.67	356,669	36,615.92	292,427	428,747
Payroll	137,105.58	137,094	5,051.64	127,820	147,421
PersCons	11,154.83	11,010.30	1,165.60	9,553.10	13,531.20
Federal Funds Rate	0.83	0.16	1.49	0.07	5.26

Table 1

Correlation Matrix

The next analysis is the correlation matrix, where all variables will be analyzed for correlation. The correlation matrix is important because it reveals the strength of correlation between the variables. The purpose is to identify any variables that could cause multicollinearity during the regression analysis.

	ConConf	*Comp Lead*	*BusConf*	*EmpRate*	*Inventory Turnover*	*PMI*	*MfgOrdDur*	*BldgPerm*	*Sales*	*Payroll*	PersCons	*Funds Rate*	*DJI*
ConConf	1												
Comp Lead	0.4487	1											
BusConf	0.3862	0.6249	1										
EmpRate	0.3225	0.2354	-0.3161	1									
Inventory Turnover	0.3350	-0.5899	-0.5317	0.2590	1								
PMI	0.3780	0.6050	0.9778	-0.2962	-0.5179	1							
MfgOrdDur	0.5605	0.7880	0.3604	0.3830	-0.2011	0.3517	1						
BldgPerm	0.7949	0.5643	0.1147	0.7206	0.2103	0.1127	0.7093	1					
Sales	0.8023	0.3723	0.3643	0.1472	0.3020	0.3562	0.7046	0.6396	1				
Payroll	0.7598	0.2428	0.0329	0.5790	0.5175	0.0375	0.6253	0.7971	0.8400	1			
PersCons	0.7516	0.1510	0.3255	0.0147	0.4460	0.3153	0.5129	0.4819	0.9617	0.7922	1		
Funds Rate	0.1242	0.4298	-0.1258	0.8058	-0.1991	-0.1195	0.2707	0.5428	-0.2081	0.1432	-0.3917	1	
DJI	0.8672	0.4322	0.4185	0.2620	0.2998	0.4074	0.7044	0.7131	0.9680	0.8716	0.9221	-0.0696	1

Table 2

The correlation matrix for the dataset is quite revealing, as there are several variables with strong correlations. These variables could have an adverse effect on the regression analysis by causing

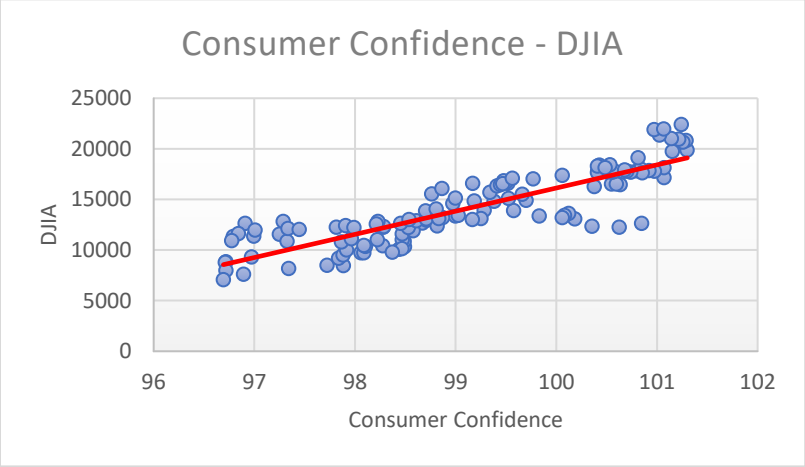
multicollinearity to exist. Multicollinearity causes the analyst to have difficulty determining the exact effect that each independent variable has on the dependent variable. Additionally, multicollinearity could cause volatility in the coefficients depending on the interaction with the other variables. Finally, multicollinearity reduces the precision of the coefficients which will weaken the overall regression model. The possible existence of multicollinearity in the model will be investigated later in this analysis.

Scatter Plots

Next, the analyst completed scatter plots for each independent variable and the dependent variable.

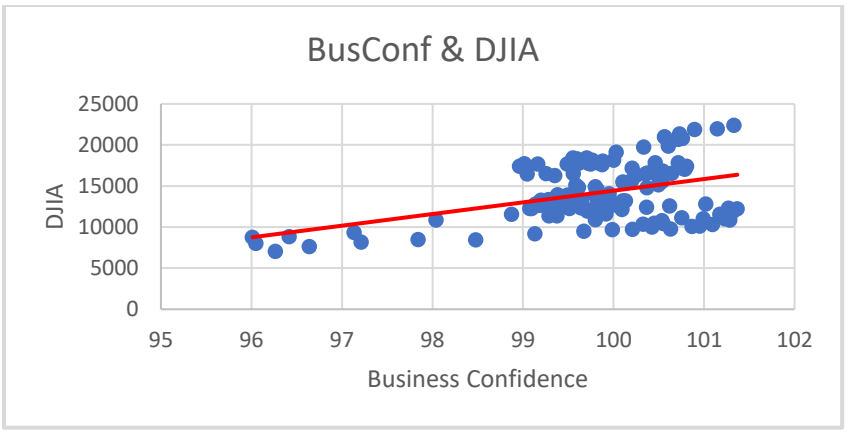
a. Consumer Confidence vs DJIA

The correlation between consumer confidence and DJIA is 0.8672, which would indicate that the two have a strong positive correlation.



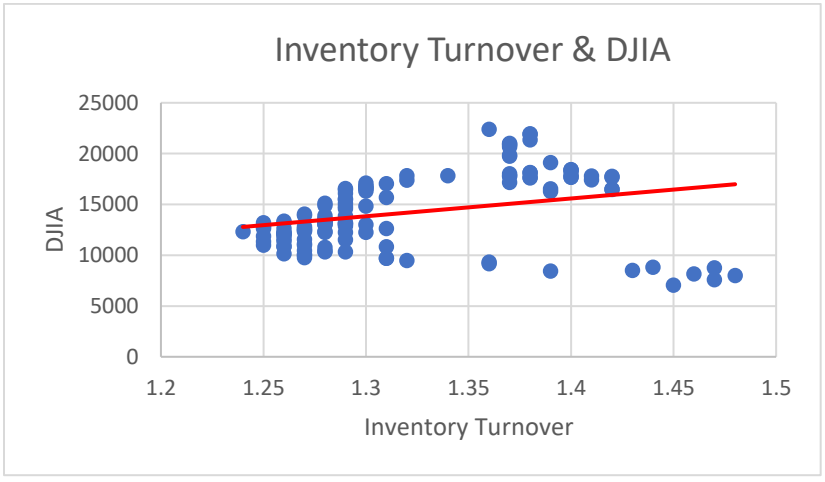
b. Business Confidence vs DJIA

The correlation between business confidence and DJIA is 0.4185, which indicates a weak to moderate correlation between the two variables.

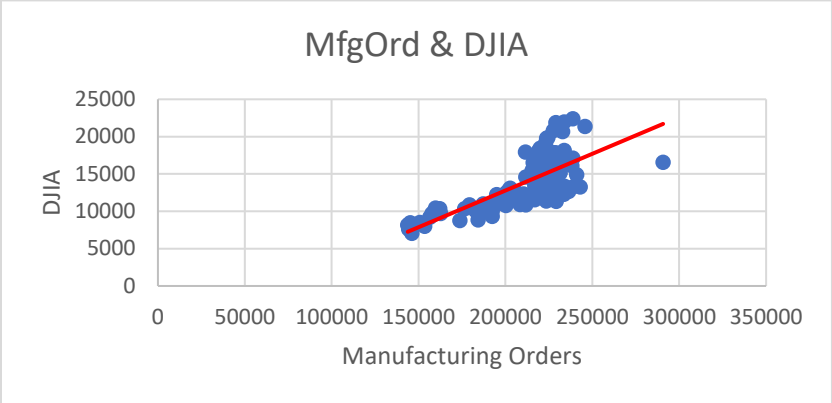


c. Inventory Turnover vs DJIA

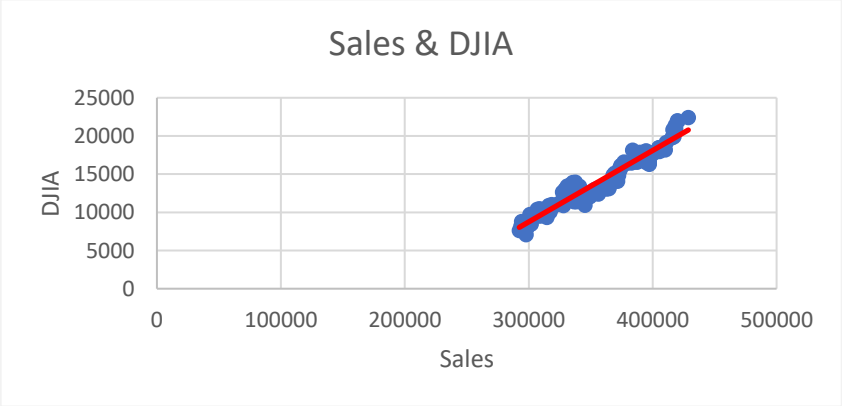
The correlation coefficient between inventory turnover and the DJIA is 0.2998, which indicates a weak positive relationship. A visual analysis of the scatter plot, one can quickly see the weak relationship between inventory turnover and DJIA.



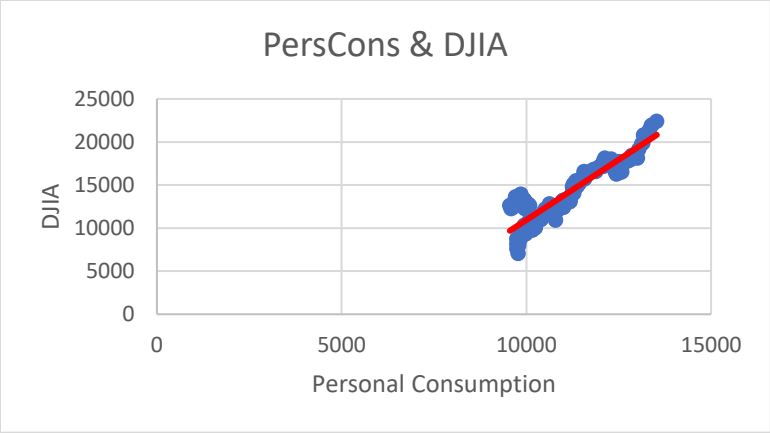
d. The correlation coefficient between Manufacturing Orders and DJIA is 0.7044, which indicates a moderate to strong, positive correlation.



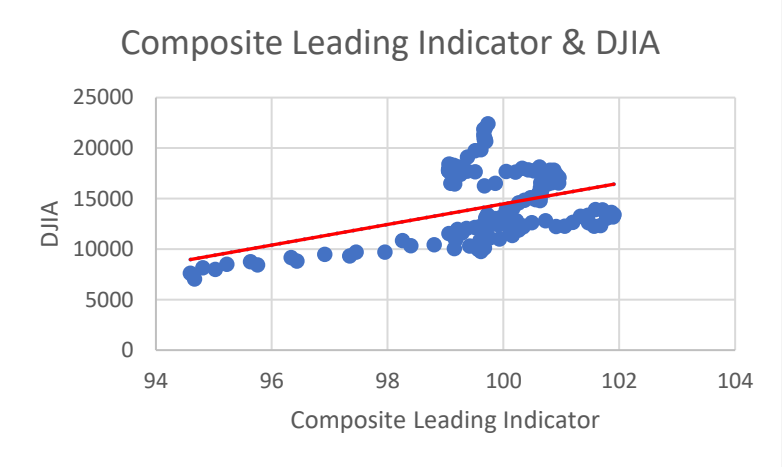
e. The correlation coefficient for Retail Sales and DJIA is 0.968, which indicates a strong positive correlation. Visual review of their scatter plot reveals low residual values near the trendline.



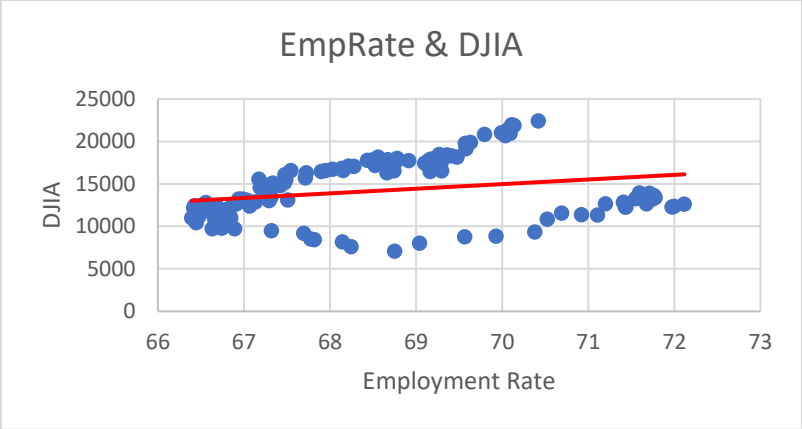
f. The correlation coefficient for Personal Consumption and DJIA is 0.9221. Like retail sales, personal consumption shares a strong, positive correlation with DJIA. Recall from the correlation matrix the correlation coefficient between sales and personal consumption is 0.9617. This analysis may need to include a variance inflation factor for sales and/or personal consumption.



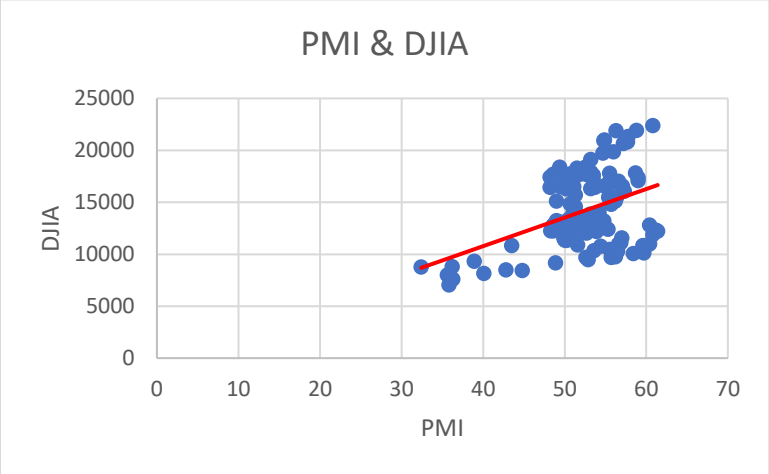
g. The correlation coefficient between composite leading indicator and DJIA is 0.4322, which indicates a weak to moderate, positive correlation. Based on the scatter plot, the correlation appears to be moderate to strong up to CLI of 99, then the correlation disperses to weak to moderate.



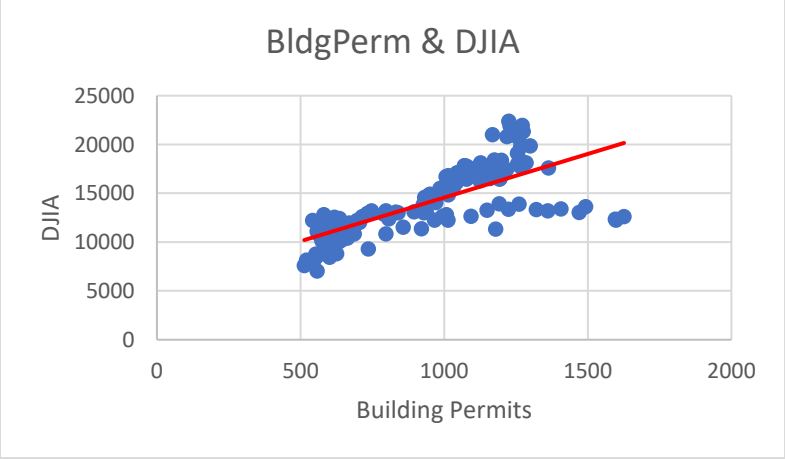
h. The correlation coefficient between employment rate and DJIA is 0.2620, which indicates a weak, positive correlation. This correlation is the second lowest in the dataset and will need to be closely analyzed during the regression phase of the analysis.



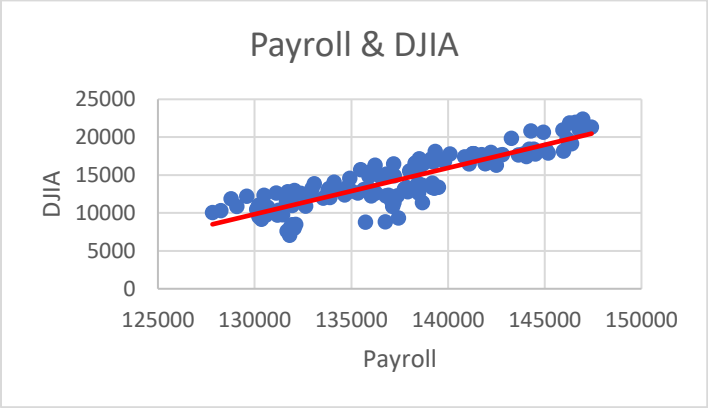
i. The correlation between Purchasing Manager's Index and DJIA is 0.4074, which indicates a weak to moderate correlation. A quick review of their scatter plot shows the weak relationship between PMI and DJIA.



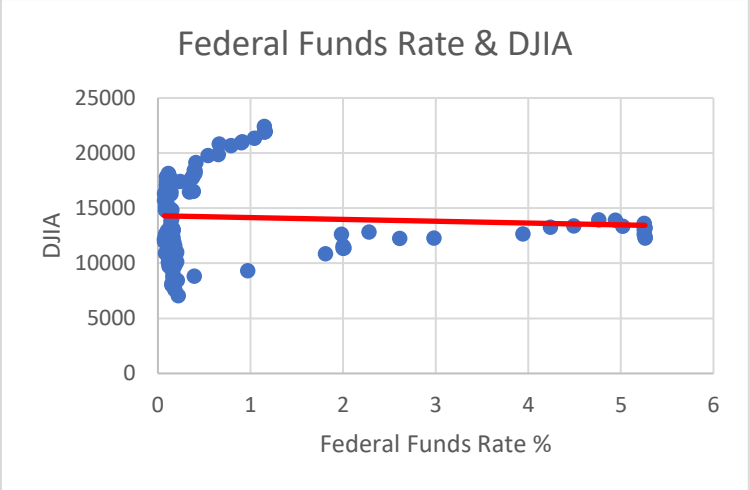
j. The correlation coefficient between Building Permits and DJIA is 0.7131, which indicates a moderate to strong correlation.



k. The correlation coefficient between Payroll and DJIA is 0.8716, which indicates a strong, positive correlation.



l. A visual review of the Federal Funds Rate and DJIA scatter plot quickly reveals the weak correlation. The correlation coefficient between Federal funds rate and DJIA is -0.0696. This supports the scatter plot's depiction of a near zero correlation.



Regression

Regression Analysis

Once the correlation matrix and scatter plots had been completed, the analyst had a suspicion that multicollinearity and insignificant coefficients were present in the model. The analysis would require multiple iterations to eliminate insignificant coefficients and those causing variance inflation.

The first multiple regression iteration confirmed the analyst's suspicions of insignificant variables present in the data set as depicted in the visual below.

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.991661396
R Square	0.983392325
Adjusted R Square	0.981514936
Standard Error	480.8909983
Observations	129

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	13	1574740930	121133917.7	523.8084112	9.50476E-96
Residual	115	26594457.51	231256.1522		
Total	128	1601335387			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-182008.3822	26991.95539	-6.743060276	6.54193E-10	-235474.251	-128542.5134	-235474.251	-128542.5134
Index	99.57460765	30.74590679	3.238629726	0.001570167	38.67288364	160.4763317	38.67288364	160.4763317
Consumer Confidence	111.804499	133.8589481	0.835241129	0.405313856	-153.3443148	376.9533128	-153.3443148	376.9533128
Composite Leading Indicator	140.3159508	131.4493872	1.067452301	0.288003378	-120.0599866	400.6918883	-120.0599866	400.6918883
Business Confidence Indicators	1108.936424	273.7834704	4.05041408	9.31632E-05	566.6240607	1651.248786	566.6240607	1651.248786
Employment Rate	667.9099937	168.2093946	3.97070565	0.000125252	334.7195563	1001.100431	334.7195563	1001.100431
Inventory Turnover	5315.063063	3400.979095	1.562803803	0.120846495	-1421.621975	12051.7481	-1421.621975	12051.7481
Purchasing Manager's Index	-66.22345378	41.2901897	-1.603854433	0.111489118	-148.0113729	15.56446534	-148.0113729	15.56446534
The Advance Report on Durable Goods	-0.004044826	0.005973595	-0.677117563	0.499691205	-0.015877369	0.007787716	-0.015877369	0.007787716
Number of Building Permits	-1.852934341	0.647065832	-2.863594785	0.00497957	-3.13464722	-0.571221462	-3.13464722	-0.571221462
Sales	0.072459361	0.016654417	4.350759413	2.95202E-05	0.039470167	0.105448556	0.039470167	0.105448556
Payroll	0.104373581	0.040359092	2.586123104	0.010954326	0.024429987	0.184317175	0.024429987	0.184317175
Personal Consumption Expenditures	-2.977064564	0.959904861	-3.101416279	0.002423032	-4.878451352	-1.075677775	-4.878451352	-1.075677775
Federal Funds Rate	254.8474889	142.822628	1.784363531	0.077001562	-28.05664972	537.7516276	-28.05664972	537.7516276

Starting with the regression statistics, note the Multiple R is 0.99, which indicates the variables all have strong correlation with the dependent variable. Recall the correlation matrix where there were several coefficients that had weak correlation with the dependent variable DJIA, which is contradictory. Next, moving on to R Square, which is 0.983 and indicates the model is a good fit. Moving to the F statistic, note the model is significant. From there the analyst assessed the P-value column and noted six coefficients with p-values greater than alpha 0.05. Of the six insignificant coefficients, only two, consumer confidence and manufacturing orders, had a moderate correlation with the dependent variable. Thus, the analyst eliminated the six insignificant coefficients, which reduced the influence of weakly correlated variables.

The analyst then turned attention to investigate the possibility that multicollinearity could exist in the model. Recall the correlation matrix revealed strong correlations between several of the independent variables. The first step in the multicollinearity investigation was to conduct another correlation matrix with the now reduced dataset.

	Index	*Business Confidence Indicators*	*Employment Rate*	*BldgPerm*	*Sales*	*Payroll*	*PersCons*	*DJI*
Index	1							
Business Confidence Indicators	0.34798	1						
Employment Rate	-0.17853	-0.316071386	1					
Number of Building Permits	0.33521	0.114719823	0.720562515	1				
Sales	0.910619	0.364339222	0.147235566	0.63962	1			
Payroll	0.668104	0.032894594	0.579002893	0.797107	0.839997	1		
Personal Consumption Expenditures	0.978841	0.32547484	0.014653056	0.481852	0.961734	0.792201	1	
DJI	0.849869	0.41849377	0.262045956	0.713116	0.96801	0.871632	0.9221	1

Note the highlighted cells where a strong correlation exists, which indicates there is likely variance inflation in the model. The analyst then moved on to conduct multiple regression analysis and VIF of personal consumption, payroll, and building permits.

Personal Consumption Regression and VIF

The regression and VIF of personal consumption revealed an astounding VIF of 501. This indicates that personal consumption is indeed causing variance inflation and should be removed from the data model.

Personal Consumption Regression & VIF

Regression Statistics			VIF	501.0031
Multiple R	0.999002			
R Square	0.998004			
Adjusted R Square	0.997906			
Standard Error	53.34026			
Observations	129			

Payroll Regression and VIF

The payroll regression and VIF revealed a Multiple R of 0.9768, which indicates there is a strong correlation between payroll and the other independent variables. The VIF of 21.7588 indicates that moderate variance inflation exists, which eliminates payroll from the dataset.

Payroll Regression & VIF

Regression Statistics		VIF	21.75883
Multiple R	0.976750552		
R Square	0.95404164		
Adjusted R Square	0.951781393		
Standard Error	1109.275923		
Observations	129		

Building Permits Regression and VIF

The building permits regression returned multiple R at 0.946, which indicates a strong positive correlation with the remaining independent variables. Building permits' VIF returned a value of 9.58, which is just shy of being considered strong variance inflation. Recall the baseline correlation matrix, building permits and DJIA have a correlation coefficient 0.7131. Next, building permits' high VIF eliminates this coefficient from the dataset.

Bldg Permits Regression & VIF			
Regression Statistics		VIF	9.581264
Multiple R	0.946377		
R Square	0.89563		
Adjusted R Square	0.890497		
Standard Error	93.40909		
Observations	129		

Final Dataset Regression Analysis

The final dataset regression analysis was conducted. Multiple R was 0.984, indicating strong, positive correlation between the dependent variable and the independent variables. Next, R Square was 0.9685, where 96.9% of the variation in the model can be explained by regression. Adjusted R Square came in at 0.9678, which indicates that all coefficients are significant.

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.984161139
R Square	0.968573147
Adjusted R Square	0.967818902
Standard Error	634.5072556
Observations	129

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	3	1551010455	517003485.1	1284.163393	1.0653E-93
Residual	125	50324932.17	402599.4574		
Total	128	1601335387			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-93685.95338	7379.493236	-12.69544539	3.10548E-24	-108290.8864	-79081.0204	-108290.8864	-79081.0204
Business Confidence Indicators	517.4313323	63.05817712	8.205618301	2.38795E-13	392.6313736	642.2312909	392.6313736	642.2312909
Employment Rate	373.0270646	36.30920293	10.27362306	2.52744E-18	301.1666457	444.8874835	301.1666457	444.8874835
Sales	0.085581499	0.001722415	49.68692253	3.54092E-84	0.082172626	0.088990372	0.082172626	0.088990372

The final step in the regression analysis was to calculate measures of fit: Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percent Error (MAPE). The MAE returned a value of 493.25, RMSE of 624.5825, and a MAPE of 3.78%.

Conclusion

In concluding this study, let's consider the original hypothesis in which this study was founded on.

Ho: All major economic indicators explain movement in the Dow Jones Industrial Average.

Ha: At least one of the major economic indicators does not explain movement in the Dow Jones Industrial Average.

The hypothesis states that all major economic indicators in the study explain the movement of the DJIA. Based on the regression analysis conducted, there is sufficient evidence to reject the null hypothesis. This study began with 12 independent variables and finished with three independent variables. Six of the original independent variables were eliminated due to insignificant p-values, while the other three were eliminated due to strong variance inflation factors. The final data model was proven through regression to be both significant and a good fit. Moving forward, the reader would be wise to approach the reasons for movement in the stock market with a healthy dose of skepticism, as this paper has proven that not all economic measures explain the movement of the US stock market.

References

Understanding the Importance of the Dow Jones Industrial Average. (2019, March 22). Quickenloans. https://www.quickenloans.com/blog/dow-jones-industrial-average.

Doane, D. P., & Lori Welte Seward. (2019). *Applied statistics in business and economics*. Mcgraw-Hill/Irwin.