

What Does Macroeconomics tell us about the Dow?

**SYS302 Final Project
Professor Tony Smith
Alfred Kong**

CONTENTS

Objective

Solutions Approach

- Explanations
- Choosing Variables

Regression Modeling I

- Mixed Regression (Choosing the best adjusted R^2 Value)
- Coefficients Analyses
- Multicollinearity Consideration
- Conclusion

Regression Modeling II

- Alternative Regression Model
- Gauss-Markov Assumptions and Heteroscedasticity Testing
- Multicollinearity Consideration
- Autocorrelation Testing

Final Conclusion with Coefficients Analyses (Economic Explanation)

Objective

In the financial market, many analysts and economists try to predict various financial and economic data and explain relations between them using different mathematical and statistical models. One of the most exciting, challenging, and popular set of data that modern financial analysts and economists try to predict is the Dow Jones Industrial Average. When Charles H. Dow first unveiled his industrial stock average on May 26, 1896, the stock market was not highly regarded. Stocks moved on dubious tips and scurrilous gossip because solid information was hard to come by. Nowadays, however, economists believed that the Industrial Average strongly depends on the well being of the economy and solid information.

In the short run, the stock market is very volatile. The Dow tends to rally and plunge base on many political and financial jitters and rumors. Many of these political and financial instability in the short run is very unpredictable, and therefore forecasting the Stock Market and Dow Index in the short run become very difficult. In this project, I focus on the performance of the stock market measured by the Dow Jones Industrial Average over a 39 year period from 1959 to 1997. This would eliminate many of the short-term instability. Also, by selecting a period over the past 4 decade instead of over the past century, we can reduce the chances of getting “incorrect” data for the model since economic data is collected and measured differently in the first part of the century than those in the modern days. In addition, all data in this project is annualized. For example, Dow Jones Industrial Average in this project is calculated by summing the closing price of each month of the year and divided by 12 months. This calculated average is used to

represents the DJIA for the year. By taking the annual average of all the variables, we can significantly reduce the impact of short-term fluctuations of the economy and the volatility of the index due to short-term jitters and rumors.

The goal of this project is to formulate a regression model that can on one hand provides the highest predictability for the DJIA, and on the other hand enables reasonable economic explanation of the relations between selected economic data and the Dow Jones Industrial Average. At the end, we will analyze the final regression model and explain the impact of individual variables on the Stock Market.

Solutions Approach

Explanation

In this project, we make two assumptions: Stock market is volatile in the short run due to political and economic jitters and rumors. In order to reduce the effect of the short-run volatility on the model, all data are collected in a year to year basis. For economic data that are reported monthly, we simply take the average over the 12-month period and use the average to represent the sample data for the year. For quarterly reported data, we take the average of the four quarters. For example, Dow Jones Industrial Average in this project is calculated by summing the closing price of each month of the year and divided by 12 months. This calculated average is used to represents the DJIA for the year. Similar calculations are made in calculating all the economic variables in this project. By taking the annual average of all the variables, we can significantly reduce the impact of short-term fluctuations of the economic data and the volatility of the index.

Choosing Variables

In order to find a long-run relationship between the Dow Jones Industrial Average and U.S. Macroeconomics, first we must find a list of explanatory macroeconomic variables that has high correlation with the industrial average and best represents the economy.

Here is a list of the candidates for the X-variables in the final regression model:

- Federal Fund Rate
- Consumer Installment Credit
- Unemployment Rate
- Federal Government Surplus (+) or Deficit (-)
- Export
- Corporate Profit
- Bank Loan Prime Rate
- One-Year Treasury Bill Rate (Auction Average)
- Inflation
- Federal Debt Held By Private Investors
- Personal Savings
- M3 Money Supply

Each one of these variables, except all interest rates and unemployment rate (boldfaced below) is quite significantly related to the DJIA as indicated below by the P- value in a single regression. The value on the right-hand column represents the P-value of each individual variable in a single regression with the DJIA.

Lock	Entered	Parameter	Estimate	nDF	SS	"F Ratio"	"Prob>F"
X	X	Intercept	1573.84324	1	0	0.000	1.0000
-	-	Fed Fund Rate	?	1	1499886	0.961	0.3336
-	-	con-credit	?	1	16400100	14.453	0.0006
-	-	unempl. r	?	1	35303.61	0.022	0.8829
-	-	(G-T)	?	1	23699110	25.587	0.0000
-	-	export	?	1	49458124	259.993	0.0000
-	-	corp. profit	?	1	51623886	402.213	0.0000
-	-	fed debt	?	1	51475872	388.267	0.0000
-	-	bank loan	?	1	7383.95	0.005	0.9463
-	-	1 yearT-bill	?	1	1455993	0.932	0.3409
-	-	inflation	?	1	6218745	4.362	0.0441
-	-	fdhbpi	?	1	52718530	543.075	0.0000
-	-	p. saving	?	1	23552705	25.315	0.0000
-	-	M3	?	1	46801581	175.860	0.0000

Although most of the variables are quite significant in single regression with DJIA, developing a regression model simply by introducing all these variables as X-variables would not produce an optimal regression model. The most direct approach to pick the most relevant and best combination of the variables in the model is to run a stepwise regression in JUMPIN. There are two things we have to keep in mind when running the stepwise regression: 1) the objective is to find a combination of variables that yields the highest adjusted R-squared value. 2) In the final regression model, each dependent variable should have a significant positive relation (i.e. P-value < 0.10).

The final regression model should fulfil the Gauss-Markov Assumptions. At this point we should do a normal plot of the residual to see if they are normally distributed with a constant variance (Homoscedasticity). If not, then we should redo a new regression model that follows the Gauss-Markov Assumptions.

Since we're studying a time-based Y-variable (DJIA from year 1959 to 1997) autocorrelation problem might occur. In order to check for autocorrelation, we should perform the Durbin-Watson Test on the final regression model. Also, we will check for inconsistency caused by multicollinearity among the dependent variables.

Finally, we will analyze the effect of each explanatory variable on the DJIA in the final model.

REGRESSION MODELING I

Mixed Regression (Choosing best adjusted R² value)

The first step in formulating a regression model is to run a stepwise regression using all of the X-variables listed above against the DJIA using the mixed direction option. This would give a regression model that gives the highest adjusted R² value (highest predictability). The result of the stepwise regression is shown below.

		Response: DJIA-avg						
		Stepwise Regression Control						
		Prob to Enter 0.250						
		Prob to Leave 0.250						
		Direction						
		2 rows not used due to missing values.						
		Current Estimates						
		SSE	DFE	MSE	RSquare	RSquare Adj	Cp	AIC
		491960.28	28	17570.01	0.9912	0.9887	5.546717	369.3237
Lock	Entered	Parameter	Estimate	nDF	SS	"F Ratio"	"Prob>F"	
X	X	Intercept	1086.12789	1	0	0.000	1.0000	
-	X	Fed Fund Rate	92.7588472	1	82309.06	4.685	0.0391	
-	-	con-credit	?	1	6042.404	0.336	0.5671	
-	-	unempl. r	?	1	326.7433	0.018	0.8944	
-	X	(G-T)	0.00872216	1	411674.3	23.431	0.0000	
-	X	export	3.10512272	1	152149.2	8.660	0.0065	
-	-	corp. profit	?	1	3613.28	0.200	0.6585	
-	X	fed debt	5.34978853	1	197403.3	11.235	0.0023	
-	X	bank loan	-154.09315	1	201720.1	11.481	0.0021	
-	-	1 yearT-bill	?	1	21981.28	1.263	0.2710	
-	-	inflation	?	1	3777.12	0.209	0.6513	
-	X	fdhbpi	-4.0940349	1	114743.5	6.531	0.0163	
-	X	p. saving	2.59838443	1	37281.19	2.122	0.1563	
-	X	M3	-3.0893391	1	181644.5	10.338	0.0033	

The eight X-variables in this model are Federal Fund Rate (Fed Fund Rate), Federal Surplus or Deficit (G-T), Export, Federal Debt, Bank Loan Prime Rate, Federal Debt Held by Private Investors, Personal Savings, and M3 Money Supply.

Coefficients Analysis

We then run a least squared regression using these eight selected variables. The results are as follows:

Response: DJIA-avg					
Summary of Fit					
RSquare					0.987793
RSquare Adj					0.984538
Root Mean Square Error					192.1637
Mean of Response					1700.222
Observations (or Sum Wgts)					39
Parameter Estimates					
Term	Estimate	Std Error	t Ratio	Prob> t	
Intercept	1208.5368	163.3857	7.40	<.0001	
Fed Fund Rate (G-T)	108.42216	61.88525	1.75	0.0900	
export	4.3123613	1.500782	2.87	0.0074	
fed debt	6.59581	2.252209	2.93	0.0064	
bank loan	-191.588	65.25194	-2.94	0.0063	
fdhbpi	-5.332499	2.244065	-2.38	0.0241	
p. saving	5.3824295	2.494501	2.16	0.0391	
M3	-4.118706	1.358573	-3.03	0.0050	
Effect Test					
Source	Nparm	DF	Sum of Squares	F Ratio	Prob>F
Fed Fund Rate (G-T)	1	1	113345.48	3.0695	0.0900
export	1	1	957366.42	25.9260	<.0001
fed debt	1	1	304885.86	8.2565	0.0074
bank loan	1	1	316709.68	8.5767	0.0064
fdhbpi	1	1	318340.86	8.6208	0.0063
p. saving	1	1	208512.94	5.6466	0.0241
M3	1	1	171922.13	4.6557	0.0391
			339389.11	9.1908	0.0050
Durbin-Watson					
Durbin-Watson		Number of Obs.		AutoCorrelation	
1.8189815		39		-0.0250	

The main thing to focus on is the Estimated Coefficients above. Notice that the coefficients for Federal Fund Rate and Bank Loan Prime Rate are in opposite sign, and also notice that Federal Fund Rate turns out to be not too significant with P-Value equals to 0.09. In general, these two coefficients should have the same sign since both represents the interest rates and both moves in the same direction in the economy. What went wrong?

To answer the question, let's consider the possibility of multicollinearity.

Multicollinearity Consideration

Since the coefficient for Federal Fund Rate and prime rate have opposite signs and the Federal Fund Rate turns out to be not too significant, the existence for multicollinearity is highly probable. So let's do a correlation plot with the DJIA and all eight selected X-variables:

Correlations Variable	DJIA-avg	Fed Fund Rate	(G-T)	export	fed debt	bank loan	fdhbp	p. saving	M3
DJIA-avg	1.0000	-0.1453	-0.4414	0.9287	0.9260	0.0340	0.9431	0.5643	0.8671
Fed Fund Rate	-0.1453	1.0000	-0.1326	0.0636	-0.0509	0.9606	-0.0801	0.3386	0.0462
(G-T)	-0.4414	-0.1326	1.0000	-0.6956	-0.7220	-0.2933	-0.6782	-0.9219	-0.7954
export	0.9287	0.0636	-0.6956	1.0000	0.9837	0.2717	0.9806	0.8040	0.9731
fed debt	0.9260	-0.0509	-0.7220	0.9837	1.0000	0.1506	0.9979	0.7838	0.9863
bank loan	0.0340	0.9606	-0.2933	0.2717	0.1506	1.0000	0.1206	0.5092	0.2427
fdhbp	0.9431	-0.0801	-0.6782	0.9806	0.9979	0.1206	1.0000	0.7461	0.9757
p. saving	0.5643	0.3386	-0.9219	0.8040	0.7838	0.5092	0.7461	1.0000	0.8625
M3	0.8671	0.0462	-0.7954	0.9731	0.9863	0.2427	0.9757	0.8625	1.0000

By spotting through the correlations between the X-variables, as predicted, we found that Federal Fund Rate and Bank Loan Prime Rate **does** have a higher correlation with each other (0.9606) than either one of them with the DJIA (-0.1453 and 0.0340). Therefore multicollinearity **does** exist in this model. We can then conclude that multicollinearity *might* cause two general inconsistencies on the coefficients and the X-variables: 1) it tends to lower the significance of the X-variables that are highly correlated with each other. 2) It changes the sign of the coefficients.

Conclusion for Regression Model I

Although this regression model with eight X-variables does yield the highest adjusted R^2 value of (0.984538), according to the multicollinearity consideration, multicollinearity does exist in the model. Furthermore, according to the coefficient analysis, the existence of multicollinearity creates an opposite sign for the coefficient of Federal Fund Rate. Since one of the objectives of this project is to relate independent variables to the DJIA in the final model, and since multicollinearity creates disturbance in this model, we can conclude that this model should be rejected.

REGRESSION MODELING II

Modifying the Regression Model

In the first regression model, we concluded that multicollinearity causes inconsistency in the signs of coefficients (signs of coefficients for fed fund rate and bank prime rate do not agree). Therefore, in modifying the regression model, we should remove either one of fed fund rate and prime loan rate and create another set of variables that yield the highest adjusted R^2 value and at the same time gives the minimal inconsistency by multicollinearity. After many many many many relentless and heart-pouring trial and errors, we come up with this revised regression model:

Response: DJIA-avg
Stepwise Regression Control
Prob to Enter 0.250
Prob to Leave 0.250

Direction
2 rows not used due to missing values.

		Current Estimates						
		SSE	DFE	MSE	RSquare	RSquare Adj	Cp	AIC
		796283.43	32	24883.86	0.9858	0.9840	12.73114	379.1413
Lock	Entered	Parameter	Estimate	nDF	SS	"F Ratio"	"Prob>F"	
X	X	Intercept	875.791702	1	0	0.000	1.0000	
-	-	Fed Fund Rate	?	1	107147.2	4.820	0.0357	
-	-	con-credit	?	1	129822.3	6.039	0.0198	
-	-	unempl. r	?	1	87348.93	3.820	0.0597	
-	X	(G-T)	0.00363948	1	1209447	48.604	0.0000	
-	X	export	4.18367189	1	910804.4	36.602	0.0000	
-	X	corp. profit	5.20603296	1	405981.5	16.315	0.0003	
-	-	fed debt	?	1	56600.28	2.372	0.1337	
-	X	bank loan	-83.61035	1	2410245	96.860	0.0000	
-	-	1 yearT-bill	?	1	1391.772	0.054	0.8173	
-	-	inflation	?	1	21050.04	0.842	0.3660	
-	-	fdhbpi	?	1	44890.94	1.852	0.1834	
-	-	p. saving	?	1	16408.12	0.652	0.4255	
-	-	M3	?	1	1347.738	0.053	0.8202	

In this model, only four variables are selected. They are Federal Surplus or Deficit, Export, Corporate Profit, and Bank Loan Prime Rate. Then we run a least squared regression on these four new-selected X-variables and the results are as follows:

Response: DJIA-avg				
Summary of Fit				
RSquare				0.986577
RSquare Adj				0.984997
Root Mean Square Error				189.2854
Mean of Response				1700.222
Observations (or Sum Wgts)				39

Parameter Estimates				
Term	Estimate	Std Error	t Ratio	Prob> t
Intercept	812.60458	98.20781	8.27	<.0001
(G-T)	0.0046264	0.000555	8.34	<.0001
export	3.9017166	0.824776	4.73	<.0001
corp. profit	6.8093085	1.472223	4.63	<.0001
bank loan	-81.71751	10.05418	-8.13	<.0001

In this revised model, the adjusted R^2 value is still quite high (0.984997), very close to the adjusted R^2 value of the original model (0.9887). Although the adjusted R^2 value is not as high as the original model, all four of the X-variables in this model becomes very significant with P-value < 0.0001 for every single one of them. The result is quite pleasing.

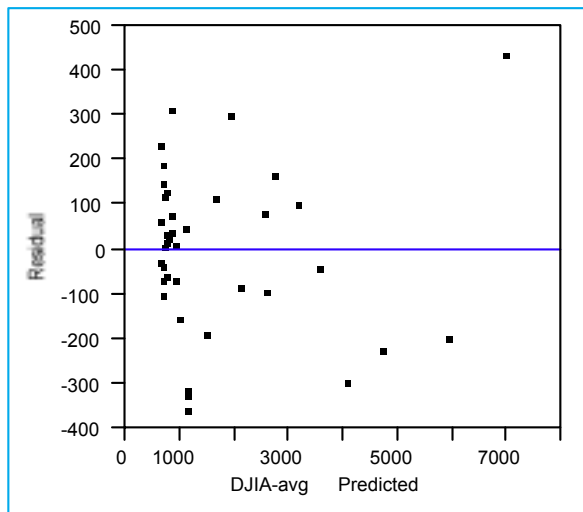
Gauss-Markov Assumptions and Heteroscedasticity Testing

So far the new regression model provides very significant result with very high significance for all X-variables and a very high adjusted R^2 value. However, in order to accept this regression model, there are several tests we must perform on the model. One of them is the Gauss-Markov Assumptions and the heteroscedasticity testing.

Gauss-Markov Assumptions state that

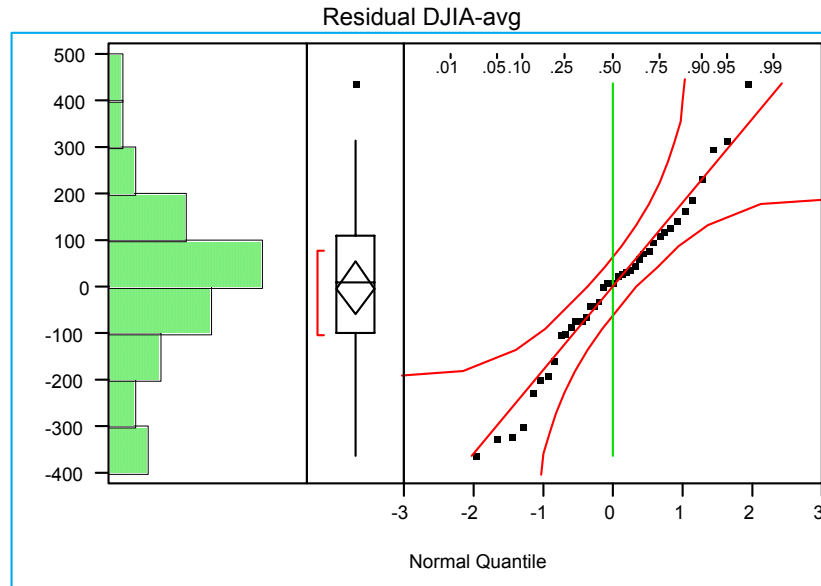
- (i) $\varepsilon_i \sim N(0, \sigma^2)$, $i = 1, \dots, n$
- (ii) $(\varepsilon_1, \dots, \varepsilon_n)$ mutually independent

In order to test for heteroscedasticity, first we perform a residual plot for the regression model and look for any specific trend for the residuals (i.e. does higher DJIA associates with larger variance?). If a trend does not exist, then we can conclude that the variance can assumed to be constant.



According to the residual plot above, we can conclude that no significant trend is presence. We can then stated that this model does not violate the constant-variance (homoscedasticity) assumption in the Gauss-Markov Assumption.

Next, we perform a normal plot for the residuals and see if the residuals are normally distributed.



Moments	
Mean	0.0000
Std Dev	179.0461
Std Error Mean	28.6703
Upper 95% Mean	58.0397
Lower 95% Mean	-58.0397
N	39.0000
Sum Weights	39.0000

According to the normal quantile plot above, the residual samples are normally distributed. Therefore we can conclude that Gauss-Markov Assumption holds in this regression model.

Multicollinearity consideration

In the first regression model, we found out that the presence of multicollinearity might either lower the significance of highly correlated X-variables and the signs of their coefficients. In this model we plot out the correlation table of all the X-variable and the dependent variable:

Variable	Correlations				
	DJIA-avg	(G-T)	export	corp. profit	bank loan
DJIA-avg	1.0000	-0.4414	0.9287	0.9632	0.0340
(G-T)	-0.4414	1.0000	-0.6956	-0.6005	-0.2933
export	0.9287	-0.6956	1.0000	0.9789	0.2717
corp. profit	0.9632	-0.6005	0.9789	1.0000	0.2142
bank loan	0.0340	-0.2933	0.2717	0.2142	1.0000

From the table, we found that Corporate Profit and Export have a very high correlation. This implies that multicollinearity does exist in this model. However, if we take a look at the P-values of all X-variables below, we could find that all of the X-variables are VERY significant. This implies that although multicollinearity does exist in this model, it does not affect the significant level of the X-variables. Therefore, we can conclude that the effect of multicollinearity in this model can be ignored.

Term	Parameter Estimates			
	Estimate	Std Error	t Ratio	Prob> t
Intercept	812.60458	98.20781	8.27	<.0001
(G-T)	0.0046264	0.000555	8.34	<.0001
export	3.9017166	0.824776	4.73	<.0001
corp. profit	6.8093085	1.472223	4.63	<.0001
bank loan	-81.71751	10.05418	-8.13	<.0001

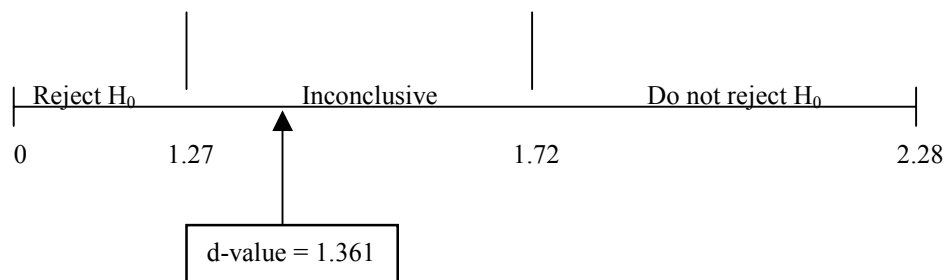
Autocorrelation Testing

In this project, we are observing the effect of macroeconomic variables on the Dow Jones Industrial Average over a 39 year period. This means that the dependent variable (DJIA) is a time-series based variable. Although the technique for performing time-series analysis is beyond the scope of this course, we can perform an autocorrelation testing to see if there is any autocorrelation in this model. The simple way to do an autocorrelation test is to check on the Durbin-Watson value in the JMPIN output on the regression model:

Durbin-Watson		
Durbin-Watson	Number of Obs.	AutoCorrelation
1.3613327	39	0.2407

Durbin-Watson value is 1.361 in this model. According to the Durbin Watson Table, at $\alpha=0.05$ level, with number of observation equals to 39 and $k=4$ (four X-variables), the critical value for the rejection region (d_L) is 1.27. This means that Durbin-Watson value in this model cannot be rejected (i.e. there is no significant indication of autocorrelation). However, the d-value in this model does not exceed the d_U value (1.72). This means that the d-value for the model lies in the inconclusive range, and therefore we cannot confidently conclude that autocorrelation does not exist.

$H_0: \rho = 0$ versus $H_1: \rho \neq 0$,



At this point we are unable to further investigate the possibility of autocorrelation in this model. Since the model provide an overall good fit (Adj. $R^2 = 0.984$) and very high significance of X-variables, and since the Durbin-Watson Test provide inconclusive results, we simply assume that autocorrelation does not exists in this model.

Final Conclusion with Coefficients Analyses (Economic Explanation)

In the final conclusion, we have formulated a very predictive model with adjusted R^2 equals to 0.9845. Also each independent variable is very significant (each with a P-value <0.0001) for predicting the DJIA. Furthermore, this model fulfils the Gauss-Markov Assumptions and passes many tests including the Heteroscedasticity test. However, we can not be fully confident that autocorrelation does not exist in this model. Further investigation on autocorrelation requires time-series studies that are beyond the scope of this course. For now, we will just assume that autocorrelation does not exist. In addition, multicollinearity has almost no effect on this regression model. This is proven by the very very high significance of the explanatory variables. At the end of the project, I attached an overlay plot of the predicted and actual DJIA over the 39-year period.

Given these results, I think that this regression model is a success and now I will spend the rest of the project relating the explanatory variables to the dependent variable.

Coefficients Analyses

So far we performed many statistical analysis on the results of the regression model, now we will focus on the economic explanation of these results. The most interesting thing to analyze is the coefficients of the X-variables. How does each individual explanatory variable influence the dependent variable (DJIA)?

Term	Parameter Estimates			
	Estimate	Std Error	t Ratio	Prob> t
Intercept	812.60458	98.20781	8.27	<.0001
(G-T)	0.0046264	0.000555	8.34	<.0001
export	3.9017166	0.824776	4.73	<.0001
corp. profit	6.8093085	1.472223	4.63	<.0001
bank loan	-81.71751	10.05418	-8.13	<.0001

A) Annual Federal Surplus or Deficit

Economic explanation:

United States adopt many fiscal policies and programs that act as Automatic Stabilizers in the economy, meaning government spending rises or taxes falls automatically without legislative action when GDP falls. These programs include unemployment insurance and the income tax system. This means that government deficit tends to rise in recessions and fall during booms. Since the coefficient for Federal Deficit in the regression model is slightly positive (0.0046264), the DJIA should rise (drop) as federal surplus or deficit decreases (increases). This means that the DJIA rises during boom and drop during the recession. This is very likely to happen, as stock prices are more likely to increase when the economy is doing well. Therefore this economic analysis supports the result in the model.

B) Exports

Economic explanation:

In this model, the coefficient for export is a positive number (3.90), this means that as export increases, DJIA increases. We can look at the increase in export as an increase of foreign demand for domestic goods. All of the 30 companies are multinational corporations, thus, if we assume that increases in export in national level imply increases in export of these 30 companies in the DJIA, then we can interpret that increase in export as increase in sales of these 30 companies in the Industrial Average. With more sales, revenue increases and company valuations should also increase. Therefore, we can conclude that export has a positive influence in DJIA.

C) Corporate Profits

Economic explanation:

From the model, the coefficient of Corporate Profits is 6.81. This means that as corporate profit increase, DJIA increases. This is quite straightforward, if we assume that the profits of the 30 companies in the Dow rises and falls with the national aggregate corporate profits. High profit means higher profitability, and thus drives the stock price higher.

D) Bank Loan Prime Rate

Economic explanation:

Usually Prime Rate is the basic rate that banks charge on loans to their best customers. However, in this model, we simply use Prime Rate to represent the interest rate of the economy since all types of interest rate goes up and down together most of the time. In this model, the coefficient of Prime Rate is -81.7 . This means that as interest rate is higher, DJIA goes lower. Interest Rate in a broad definition is the cost of borrowing. As interest rate goes higher, corporations and the public incur a higher borrowing cost. Therefore, a higher interest rate reduces consumption and investment. Lower consumption from the public and lower investment from corporations might also lower corporate profit and valuation. Therefore, we can conclude that interest rate is negative related to the DJIA.

In fact, in the stock market, investors closely watch the interest rate as an investment parameter. For instance, the stock market recovered over 1500 points in the past three months based on the three consecutive rate cut by the Fed. This further supported that DJIA moves in opposite direction as the interest rate.