11th MALAYSIA STATISTICS CONFERENCE 2024
Data and Artificial Intelligence: Empowering the Future
Sasana Kijang, Bank Negara
19th September 2024

Econometric Analysis of Stock Market Movements with Macroeconomic Changes in the US: A Two-Stage Least Squares Approach

Saw Xin Zu[1]

[1] Department of Mathematics and Statistics, Faculty of Applied Sciences and Technology, University Tun Hussein Onn Malaysia, 84600, Pagoh, Johor, Malaysia. xinzusaw@gmail.com

Abstract:

Stock market movements correspond to the fluctuations or changes in stock prices over time. This study aims to use the Two-Stage Least Squares (TSLS) regression technique to explore how the macroeconomic indicators influence the movements of stock market in United States of America. The TSLS approach, a statistical method often employed in instrumental variable analysis, will be applied in this study. The findings indicated that the was an appropriate option for the study's instrumental variable because it met the relevance and exogeneity assumptions. In summary, this study highlights the importance of inflation rate and Treasury yields in shaping the US stock market, suggesting that future research explore additional variables or alternative econometric approach for deeper insights.

Keywords:

Two-Stage Least Squares; Instrumental Variable; Stock Market; Inflation Rates; US Treasury

1. Introduction:

Financial markets, particularly stock markets, are crucial components of international economies, as they reflect the investor sentiment about the future financial prospects of listed companies. According to Agrawalla and Tuteja (2008), investment decisions in the stock market are influenced by its movements, which are affected by various macroeconomic variables [1]. This understanding enables investors, policymakers, and economists to predict future trends and make informed decisions.

Understanding market anomalies and macroeconomic indicators is especially important when analyzing the US stock market. Sathyanarayana and Gargesa (2018) state that inflation, defined as the change in the price level of goods and services in an

economy, impacts stock prices [2]. Factors such as inflation rate, oil price, GDP percent, unemployment rate, trade balance, and US Treasury yields are crucial for economic analysis and policy making. However, analyzing these factors is complex and requires advanced statistical methods to handle endogeneity issues.

The objective of this study is to identify the correlation between the dependent variable and the independent variable, and to determine the appropriate instrumental variables using correlation matrix plot. Besides, the second objective is to resolve the endogeneity issues by using the Two-Stage Least Squares (TSLS) method. TSLS approach is employed to produce unbiased estimates of the effects of the inflation rate and US Treasury yields on stock index prices. In this study, the US Treasury serves as an instrumental variable because it significantly influences the inflation rate but is unlikely to directly affect the stock index price in the US.

The first stage of the TSLS method involves regressing the inflation rate on the US Treasury yields to obtain predicted values of the inflation rate. The second stage involves regressing the stock index price on these predicted values to determine the causal impact of the inflation rate. The analysis reveals significant relationship between the inflation rate, US Treasury yields, and stock market movements in the United States.

The expected outcome is to demonstrate the significant influence of these macroeconomic variables on stock market performance, highlighting their role in shaping market trends. In summary, this study underscores the importance of understanding macroeconomic factors in stock market analysis. Future research could expand on this work by exploring additional variables or employing alternative econometric approaches for deeper insights.

2. Methodology:

In this section, the data and methods used in this study will be discussed in detail. The analysis will be carried out using R software. A Two-Stage Least Squares (TSLS) regression will be built, and a conclusion will be drawn.

2.1 Data Preprocessing

The data that applied in this study is about the movement of stock market with macroeconomic changes in United States of America from 1980 to 2020, which is taken from platform ‘Kaggle’. The data descriptions for each variable are displayed in Table 1 with “index price” as the dependent variable and others hold as independent variables.

Table 1: Data Description

Variables	Description
index price	Stock index price of the US
inflation rate	Inflation rate of the US
oil price	Oil price of the US
gdppercent	GDP percent of the US
unemployment rate	Unemployment rate of the US
manufacturing output	Manufacturing output of the US
trade balance	Trade balance of the US
USTreasury	Investment in US government debt securities

2.2 Multicollinearity Test (VIF)

Multicollinearity occurs when independent variables in a regression model are highly correlated, leading to unreliable and unstable estimates of regression coefficients. The multicollinearity test, which is conducted using the Variance Inflation Factor (VIF), helps identify this issue. VIF values greater than 10 suggest significant multicollinearity, warranting further investigation or remedial measures [3].

2.3 Pearson's Correlation Coefficient

The Pearson's correlation coefficient is a correlation coefficient that measures linear correlation between two sets of data. Pearson's correlation coefficients (r) were calculated to examine the relationships between pairs of variables.

2.4 Instrumental Variable Selection (IV)

The US Treasury is chosen as the instrumental variable due to its relevance and exogeneity. The instrumental variable should be correlated with the endogenous variable (inflation rate) and uncorrelated with the error term in the stock index price equation. These characteristics make the US Treasury a suitable choice.

2.5 Testing for Relevance and Exogeneity

Relevance and exogeneity are key concepts in econometrics, especially for instrumental variables (IV) estimation. Relevance is assessed by regressing the endogenous variable on the instrumental variable to confirm their correlation. The exogeneity test ensures the instrumental variable is not correlated with the error term in the regression equation, thus satisfying the exclusion restriction.

2.6 Ordinary Least Squares (OLS)

Ordinary Least Squares (OLS) is a widely used technique for estimating parameters in linear regression analysis. Assumptions include linearity, independence of errors, homoscedasticity, normality or errors, and absence of multicollinearity among independent variables.

2.7 Two-Stage Least Squares (TSLS)

Two-Stage Least Squares (TSLS) are employed in econometrics to solve the endogeneity problem in regression analysis. The first stage involves predicting the endogenous variables ($\hat{x}_i$) using instrumental variables. The equation for the first stage of TSLS can be written as below [4]:

$$x_i = a_0 + a_1 z_i + \mu_i \quad Eq.2$$

In the second stage of TSLS, the predicted value of $\hat{x}_i$ from the first stage regression is used as the independent variable in a new regression model to estimate the dependent variable, y_i [5]. The equation for the second stage of TSLS can be written as below:

$$y_i = b_0 + b_1 \hat{x}_i + v_i \quad Eq.3$$

TSLS provides unbiased estimates in the presence of endogeneity.

3. Result:

This section provides a detailed discussion of the study's data and analysis. Each step contributes a full analysis of the dataset and its relevance to the study's objectives.

3.1 Multicollinearity Test (VIF)

Table 2 shows the VIF values of all the independent variables to identify multicollinearity. None of the independent variables had VIF value more than 10, so all the variables remained.

Table 2: VIF Value of the Independent Variables

Variables	VIF Value
inflation rate	1.0598
oil price	1.0176
gdppercent	1.0542
unemployment rate	1.1099
manufacturing output	1.1136
trade balance	1.0189
USTreasury	1.0951

3.2 Pearson's Correlation Coefficient

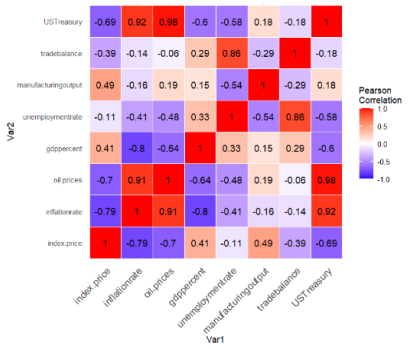

Figure 1: Correlation Matrix Plot

From the figure above, the inflation rate was considered as an endogenous variable due to its strong negative correlation with the stock index price. The US Treasury was identified as an instrumental variable, significantly correlated with the inflation rate but minimally correlated with the stock index price.

3.3 Testing for Relevance and Exogeneity

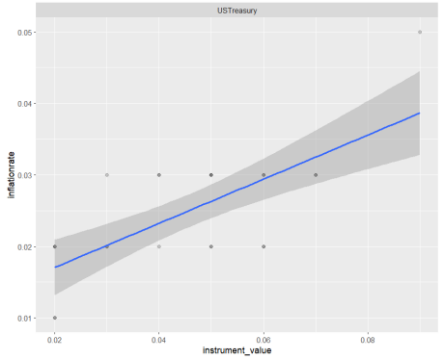

Figure 2: Relationship between US Treasury and Inflation Rate

Relevance refers to the strength of the relationship between the instrumental variable (IV) and the endogenous explanatory variable in a regression model. Based on the figure above, the output visualizes the clear relationships between the instrumental variables (US Treasury) and the endogenous explanatory variable (inflation rate). Both variables show significant coefficients, highlighting their strong correlation.

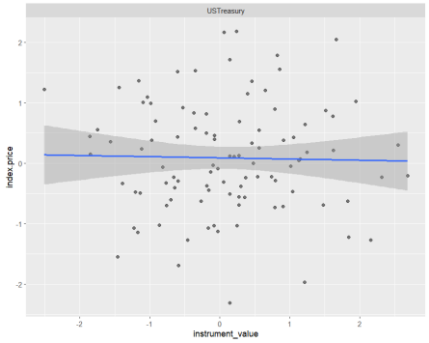

Figure 3: Relationship between US Treasury and Stock Index Price

Exogeneity refers to the instrument being uncorrelated with the error term in the regression model. Based on the figure above, the output visualizes the relationships between instrumental variables (US Treasury) and the dependent variable (stock index price). It indicates that the US Treasury does not directly influence the stock index price. Instead, the US Treasury effect on the stock index price is only through its influence on the inflation rate, and not through any other channel. This ensures that the US Treasury meets the requirement of exogeneity in the regression model.

3.4 Ordinary Least Squares (OLS)

The equation of the OLS is shown below:

Stock index price = 0.08538 – 0.04676 inflation rate

Based on the equation, the stock price index will be explained by the inflation rate. When 1 unit of the inflation rate increases, the stock price index will decrease by 0.04676 units. The p-value of variable inflation rate is less than 0.05 which indicates the variables is statistically significant in the model.

3.5 Two-Stage Least Squares (TSLS) Regression and IV Estimation

```
> first_stage <- lm(inflationrate ~ USTreasury, data = economicdatanew)
> summary(first_stage)

Call:
lm(formula = inflationrate ~ USTreasury, data = economicdatanew)

Residuals:
       Min         1Q     Median         3Q        Max
-0.0093919 -0.0030342  0.0002374  0.0035090  0.0113494

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.010875   0.002918   3.726  0.00105 **
USTreasury  0.308621   0.058959   5.234  2.3e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.005666 on 24 degrees of freedom
Multiple R-squared:  0.5331,    Adjusted R-squared:  0.5136
F-statistic:  27.4 on 1 and 24 DF,  p-value: 2.298e-05
```

Figure 4: Output of First Stage Regression

Inflation rate = 0.010875 + 0.308621 US Treasury + u_i

Based on the output above, the estimated model in the first stage is formed. The coefficient of 0.308621 indicates that a one unit increase in the US Treasury results in an

average increase of 0.3 units in the inflation rate. Both the intercept term and the coefficient for the US Treasury had p-values less than 0.05, suggesting their significance.

```
> second.stage <- ivreg(index.price ~ inflationrate | USTreasury,
+                       data = economicdatanew, x = TRUE)
> summary(second.stage)

Call:
ivreg(formula = index.price ~ inflationrate | USTreasury, data = economicdatanew,
    x = TRUE)

Residuals:
    Min      1Q  Median      3Q     Max
-2889.9 -1391.5  -451.7  1002.9  3976.4

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)      8898       1743   5.105 3.18e-05 ***
inflationrate  -246723      67907  -3.633  0.00132 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2014 on 24 degrees of freedom
Multiple R-Squared: 0.04186,    Adjusted R-squared: 0.001934
Wald test:  13.2 on 1 and 24 DF,  p-value: 0.001323
```

Figure 5: Output of Second Stage Regression

Stock index price = 8898 – 246723 inflation rate_hat + v_i

Based on the output above, the estimated model in the second stage is formed. The coefficient for the inflation rate was estimated to be -246723, indicating that a one unit increase in the inflation rate leads to a decrease of 246723 units in the stock price index, by considering the instrumental variable (US Treasury). Both the intercept term and the coefficient for the inflation rate had p-values less than 0.05, indicating their significance.

4. Discussion and Conclusion:

Overall, there was no missing data in the dataset after data cleaning, and no influential observation remained after eliminating outliers. The study employed the Two-Stage Least Squares approach, identifying the 'US Treasury' as a potential instrumental variable. The US Treasury showed significant correlation with the independent variable (inflation rate) but influencing the dependent variable (stock index price) indirectly. Further investigation into the policy implications of using instrumental variables such as the US Treasury in economic forecasting and decision-making processes could guide effective policy interventions and economic management strategies.

References:

1. Agrawalla, R. K., & Tuteja, S. K. (2008). Share prices and macroeconomic variables in India: An approach to investigate the relationship between stock markets and economic growth. *Journal of Management Research*, 8(3), 136-146.

2. Sathyanarayana, S., & Gargesa, S. (2018). An analytical study of the effect of inflation on stock market returns. *IRA-International Journal of Management & Social Sciences*, 13(2), 48. https://doi.org/10.21013/jmss.v13.n2.p3

3. O'Brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality and Quantity*, 41(5), 673-690. https://doi.org/10.1007/s11135-006-9018-6

4. Chanthawong, A., Dhakal, S., & Jongwanich, J. (2016). Supply and demand of biofuels in the fuel market of Thailand: Two stage least square and three least square approaches. *Energy*, 114, 431-443.

5. Amoah, C., & Jehu-Appiah, J. (2022). Factors driving industrialization in Africa: A panel two-stage least square approach. *Modern Economy*, 13(2), 144-158.