FIN 503 Group Project: Predict Stock Market Crash

Group members: Xiaolei Chang, Wanying Mao

Introduction

This report outlines how to use regression methods to find effective variables and methodology for predicting the U.S. stock market from 1990 to 2022. The study aims to design a model to predict the bear market according to the significant variables correlated to S&P500 monthly return.

Description of Data

The preparation of eight variables which belong different categories from the FRED economic data website (Federal Reserve Economic Data, n.d.) utilized in this project included:

- The 3-month treasury bill secondary market rate (TB3MS)
- Unemployment rate (UNRATE)
- Gross domestic product (GDP)
- Industrial production of total index (INDPRO)
- Sticky price consumer price index less food and energy (CORESTICKM159SFRBATL)
- Share prices for all shares/broad: total for United States (SPASTT01USM661N)
- Labor force participation rate for Illinois (LBSSA17)
- Equity market volatility tracker for housing and land management (EMVHOUSELANDMGMT)

The study chooses above variables after a careful consideration based on previous literature reviews and economic studies.

- According to "How Do Interest Rates Affect the Stock Market?" (Hall, 2022), interest rates and the stock market have a negative relationship which indicates that when interest rates rise, share prices fall and therefore raise the volatility level.
- According to "The Relationship Between Unemployment and Stock Market Returns Explained" (Groette, 2023), the stock market goes up when unemployment is high. Moreover, according to a case study related to unemployment rate and its effect on S&P500 (BetterTrader.co, n.d.), the release of the unemployment rate causes a lot of volatility in the markets.
- According to "U.S. GDP and S&P500-An Inquiry into the Nature and Causes of the Econometric Relation between GDP and Stock Market in the U.S" (Jordi Mora Ros, 2018-2019), there is a positive historical relation between current U.S. GDP growth and past S&P 500 growth.
- According to the article "Another Look at The Stock Market & The Business Cycle?" (EconomicsLOL, 2011) , there is an uncertain relationship between industrial production and S&P500 for its positive in certain period and sometimes negative, and the cross correlations for the S&P and industrial production suggest that the equity market anticipates changes in

industrial production at roughly six to eight months in advance.

- According to "Inflation's Impact on the Stock Market" (Hackett, 2023), correlation exists between CPI and stock market performance.
- As for share prices for all US stocks, it has some relations with S&P500 return in fact because S&P500 is a collection of some representative large-cap U.S. equities from all the stocks in US market.
- According to the fiscal year report of the state of economy in Chicago, Illinois, the workforce participation rate will affect production in the society directly and make contribution to the stock market performance eventually.
- According to the report "What Do Stock Indices Reveal About Housing?" (Shapiro, 2021) there is some evidence that housing price will reflect the investor's confidence about the stock market.

In terms of keeping the stability and availability of chosen variables, the data processing uses the percent change of raw data. In addition, since the GDP only has quarterly data as a minimum period, the quarter GDP is evenly divided into three parts and transferred into monthly data to keep consistency of statistical calibers with other variables.

Establishment of Model

In order to build a valid model, the study first used the following regression methods to test the significance of each variable:

1. **Simple Linear Regression**

   First, R studio runs a regression of the S&P500 monthly price return with individual independent variables in a sequence of the above independent variables. The results conclude that the unemployment rate, industrial production, share prices, and Illinois workforce participation are significantly related to the S&P500 at a confidence level of 0.1%. Among them, the share price is the most significant variable, which shows the p-value is <2e-16. However, GDP has a significant relation with the S&P500 at a confidence level of 1%, and the interest rate has a significant relation with the S&P500 at a confidence level of 5%. In contrast, the linear relation between the S&P500 monthly return and the other two variables (CPI and housing market indicators) is not significant.

2. **Multiple Linear Regression**

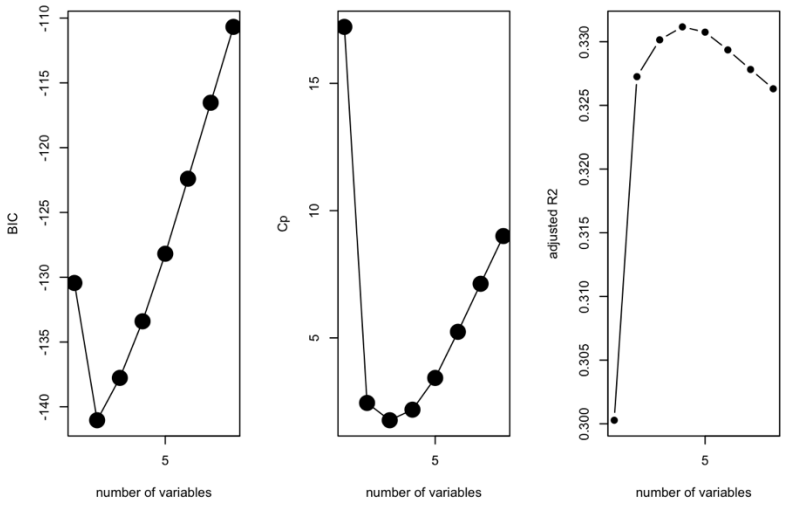
   The next regression is about the S&P500 with all independent variables to achieve a better comparison. According to the result, the linear relation between the S&P500 monthly return and all the eight variables is significant as a whole. In terms of individual variables in this multiple linear regression model, only two variables are significant. For example, share prices are significant at a confidence level of 0.1%. Workforce participation in Illinois is only significant at a confidence level of 10%.

According to the above regression results, the multiple linear regression model can give a better explanation. Its R-squared is 0.3399, meaning 33.99% of the variation in the S&P500 monthly return can be explained by the independent variables in the model. At the same time, this multiple linear

regression model can explain more proportions than the eight single linear regression models above. So, the next best model selection is based on multiple linear regression and divided into three steps.

Step1: Using BIC, Cp, Adjusted R-squared

First, the study makes plot graphs of BIC, Cp and adjusted R-squared in order to have a preliminary intuition of the number of variables that the optimal model may contain. As the graphs show, when there are two independent variables in the model, BIC is the smallest. When there are three independent variables in the model, Cp is the smallest. When there are four variables, adjusted R2 is the largest. So, the further analysis is needed to find the optimal model.



(Graphical outputs of best variables chosen in BIC, Cp, Adjusted R-squared)

Step2: Using stepAIC

The study passed stepAIC screening, and the results left three highly significant variables: "Unemployment rate (UE)," "Share prices (SP)," and "Labor force participation rate for Illinois (WP)." In summary, this regression model can effectively explain 33.52% of the S&P500 monthly return. Meanwhile, the p-value<2.2e-16 monitored by F statistics also shows that the above variables are meaningful. The regression function is "S&P500 monthly return = -0.00873UE + 0.006357SP + 0.026377WP".

Step3: Using step Cp

By applying Cp, the lowest Cp is with three explanatory variables, which can correspond correctly to the stepAIC's result. The best three explanatory variables are UE, SP, WP, and their lowest Cp is 1.764426. In order to ensure the accuracy of the regression, "ncvTest" is used to test heteroskedasticity. This model has heteroskedasticity because the p-value is 0.0016398, less than 0.05. So, the null hypothesis can be rejected. Then, using robust regression to fix the model, the variable "WP" is removed based on its p-value of 0.1086, which is higher than 0.05 and is insignificant.

Finally, the optimal model is "S&P500 monthly return = -0.00873UE + 0.006357SP". The share prices have a positive relationship with the S&P500 monthly return, and the unemployment rate has a negative relationship with the S&P500 monthly return.

Test Risky Months

After choosing the optimal model, the study tries to create dummy variables to test whether August, September, and October are risky months for stock markets. In the regression, the variables of these three months are run respectively with the selected variables above.

Firstly, they showed the same results, which these three months are insignificant to predict the stock market. All their p-values are 0.306, higher than 0.05. Due to the fact that there is multicollinearity between the three months, it cannot regress together; otherwise, the result will be "NA" in the September and October regressions.

Second, it performs regression comparisons across all months.

- Create a dummy variable with intercept and without December: The results show that the p-value of the intercept term is 0.7007, which is not significant and does not need to exist in the model.
- Create a new dummy variable with December and without intercept. The R-squared of this model is 0.2282, which explains that the impact of August, September, and October on the overall S&P500 monthly returns in this model is not significant.

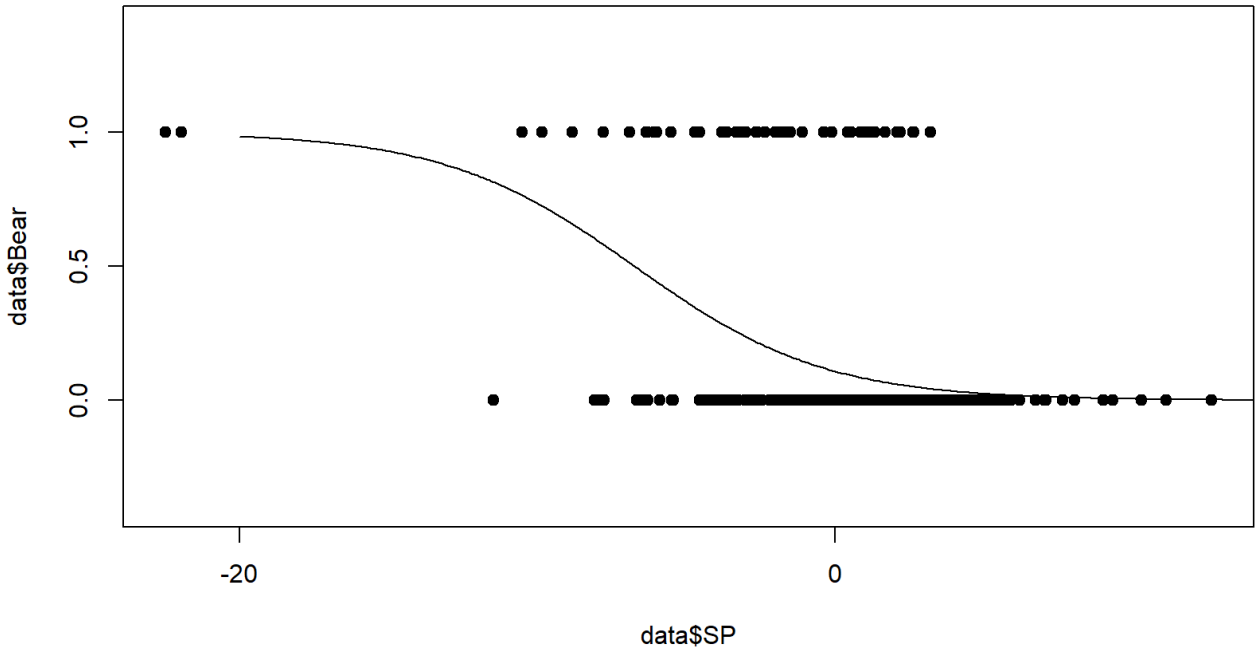
d_monthsmonths8	-0.0048298	0.0061307	-0.788	0.4313
d_monthsmonths9	-0.0067861	0.0061300	-1.107	0.2690
d_monthsmonths10	0.0049095	0.0061919	0.793	0.4283

(Partial dummy variable regression)

None of these coefficients are statistically significant at conventional significance levels ($p < 0.05$). This suggests that, based on this model, there is no strong evidence to conclude that August, September, or October have a significant impact on the SP500.

Predict the Bear Market

For this project, the bear market is defined when the stock market declines more than equal to 4% a month. Under the logit regression, the variable "UE" is removed from the regression (p-value = 0.234). The final prediction function is "P(Bear market) = 1/(1+exp(-2.12009 - 0.3.819SP)". In summary, the probability that there is a bear market has a positive relationship with the share prices volatility. According to the prediction plot, when the percent change of stock prices for all stocks in US market is negative, there is a larger probability that the stock market will shrink. In general, the model works well and can give investors a good picture of prediction of bear market.



(Plot of logit regression)

Conclusion

The best multiple regression model only contains unemployment rate and share prices, and there is no evidence to show the specific months will add risk to stock market. It excludes the influence of timing as the result of regression has shown. The share prices have a positive relationship with the S&P500 monthly return, and the unemployment rate has a negative relationship with the S&P500 monthly return. In addition, the logit model indicates that the more sharply the share prices for all stocks in US decrease, the more probability that the stock market crash is upcoming. In brief, investors in stock market should make full preparations to face the market risk when historical prices for all stocks have a highly volatile trend over a fairly continuous period.

References

BetterTrader.co. (n.d.). *Case study - Unemployment rate and its effect on the S&P 500: it's less intuitive than you think | BetterTrader.co*. Bettertrader.co. https://bettertrader.co/case-studies/unemployment-rate-and-its-effect-on-the-SP-500.html

EconomicsLOL. (2011, October 24). *Another Look at The Stock Market & The Business Cycle? | The Capital Spectator.* https://www.capitalspectator.com/another-look-at-the-stock-market-the-business-cycle/

Federal Reserve Economic Data. (n.d.). *Categories of Economic Data | FRED | St. Louis Fed.* Fred.stlouisfed.org. Retrieved December 12, 2023, from https://fred.stlouisfed.org/categories/

Groette, O. (2023, April 18). *The Relationship Between Unemployment and Stock Market Returns Explained (Backtest and Statistics) – Quantified Strategies. Trading And Investing for Traders And Investors.* Quantified Strategies. https://www.quantifiedstrategies.com/unemployment-rate-and-stock-market-returns/

Hackett, M. (2023, March 15). *Inflation's impact on the stock market.* Nationwide Financial. https://blog.nationwidefinancial.com/markets-economy/capital-market-impact/inflations-impact-on-the-stock-market/

Hall, M. (2022, February 2). *How Do Interest Rates Affect the Stock Market?* Investopedia. https://www.investopedia.com/investing/how-interest-rates-affect-stock-market/

Mora, J. (2018). *U.S. GDP and S&P 500 An Inquiry into the Nature and Causes of the Econometric Relation between GDP and Stock Market in the U.S.*

https://repositori.upf.edu/bitstream/handle/10230/42502/TFGECO1819USGDPMora.pdf?sequence=1&isAllowed=y

Shapiro, M. (2021, February 24). *Council Post: What Do Stock Indices Reveal About Housing?* Forbes. https://www.forbes.com/sites/forbesrealestatecouncil/2021/02/24/what-do-stock-indices-reveal-about-housing/?sh=4c8402d666c4