Gettysburg College Headquarters

Volume 2 Article 5

2023

The Effect of Trading Volume on Stock Price

Jackson Dino
Gettysburg College, dinoja01@gettysburg.edu

The Effect of Trading Volume on Stock Price

Abstract

Knowledge of the relationship between trading volume and trading equities enhances investors and public policy maker's knowledge of market structure. In this paper, we examine the effects of trading volume on stock prices using a panel of stock prices from the S&P 500 Index. We develop an ordinary least squares regression model, implementing control variables, fixed effects and an instrumental variable to minimize statistical bias. We find evidence that an increase of trading volume by its mean is associated with a $2.42 increase in average stock price. We also find stronger evidence that an increase of trading volume by its average initiates stock returns of 3.20%. Our results suggest trading volume can be a cause of higher stock prices, allowing for arbitrage opportunities in the market and disputing the efficient market hypothesis.

Keywords

Stock price, trading volume, regression analysis, efficient market hypothesis, trading returns, S&P 500

Cover Page Footnote

The author would like to thank Professors Drew Murphy and James O'Brien for their input on this project. Furthermore, I would like to thank my peers for their feedback and insights.

The Effect of Trading Volume on Stock Price – *Jackson Dino, Gettysburg College*

Introduction

For decades, traders have been intrigued by the question of how trading volume impacts stock prices and returns. Understanding this impact may reveal important details regarding financial market structure and future market event (Bajzik, 2021). Trading volume is the total number of shares of a security traded within a certain timeframe. In this paper, we examine the extent of the causal relationship between trading volume and stock prices. The effect of trading volume on stock price reveals fundamental information significant for our knowledge of how financial markets operate and offers an opportunity to shed light on the efficient market hypothesis. This paper will be of use to investors, traders and public policy makers.

Nearly 56% of Americans are invested in the stock market. (Saad, 2019) As volume and stock price are fundamental characteristics of the financial markets, understanding their relationship bears significant implications for millions of Americans, who have significant sums of money invested in the market or in retirement funds. This relationship also bears significance for traders and investment banks, whose livelihoods and existence depend on their ability to make quality returns in the market. Finally, these results are important for public policy makers to understand. New legislation that raises the capital gains tax rate, for example, would initiate a locked-in effect, where investors hold on to financial assets longer to avoid taxation. (Tatom, 2021) Trading volume would thus decrease. Understanding the resulting impact on stock price, and thus its impact on millions of Americans with savings tied to the equity markets, is imperative.

Prior literature primarily concentrates on addressing the relationship between trading volume and stock returns. Past research is divided on whether there is a positive or negative relationship between these variables, which is corroborated by a literature review conducted by Ruhani et al. (2018). Estimating whether this relationship truly exists, and its magnitude, bears significance for the efficient market hypothesis, which argues that it is impossible to outperform the market as share prices reflect all publicly available information. If a relationship exists, the hypothesis is dispelled, as an opportunity for arbitrage is indicated. We develop a linear framework where stock price is contingent on trade volume and other market factors.

To estimate the relationship between trade volume and stock price, we rely on a regression framework and a dataset of S&P 500-listed companies from 2013 to 2018. Unlike the majority of previous research into related topics, we examine the impact of trading volume on stock price, not stock returns. We contribute to existing literature by utilizing an ordinary least squares model, while other researchers tend to perform a time series analysis or implement a volatility-based GARCH model. Our framework uniquely positions us to control for variation in our model, as we implement control variables that address the impact of the stock's performance the previous day and the magnitude of its daily volatility. Stock and time fixed effects, as well as an instrumental variable, are added to our model, minimizing heterogeneity, while all estimates are made with heteroskedastic-robust standard errors.

We find that an increase in trade volume positively affects stock prices of companies listed in the S&P 500 index. In our headline regression, stock price increases by $5.606e-07 for each additional share traded. When multiplied by the mean trading volume of these stocks, mean stock price increases by $2.42. Through the addition of nonlinearities into our model, we find that stock returns in the index increase by 3.20%. Our findings corroborate certain prior literature and reveal that an arbitrage opportunity exists for traders and investors to exploit, disputing the efficient market hypothesis.

Conceptual Framework

Most prior literature relating to the topic of the relationship between trading volume and stock prices focuses on stock returns. Both positive and negative relationships have been discovered between these two variables. (Ruhani et al., 2018) We will first define the key variable of trade volume before examining past research that finds both positive and negative relationships. Then, we will address the implications of potential findings on the efficient market hypothesis. Finally, we will develop a basic framework that will serve as the foundation of our empirical analysis of trading volume and stock price later in the paper.

Trading volume is characterized as the total number of shares of a given stock that was traded or exchanged hands on a given day. Volume is a key technical indicator for investors, as it reflects the liquidity and efficiency of order execution in a given market. Higher trade volume can reflect a more active market (Twin, 2022) and the flow of new information into the market (Tetlock, 2007). Chordiaa et. al (2000) emphasize that liquidity can affect asset returns. Their analysis of NYSE and AMEX stocks from 1966 to 1995 ultimately yielded a negative relationship between volume and expected stock returns. Chordiaa speculates that increased volume is sourced from lower trading costs, meaning that investors demand a lower return when purchasing securities included in the sample. Chen (2012) finds a negative relationship between volume and returns during bear markets, but a positive relationship during bull markets. Conversely, a positive relationship between volume and returns is found by Karpoff (1988), Gallant (1992) and Shen and Wang (1998). On a related topic, Bessembinder and Seguin (1992) finds that volume and volatility are positively related, emphasizing the need to address stock variance in our econometric model. A clear disparity exists in literature related to volume and stock prices, as findings range from negative to positive. This disparity is corroborated by a literature review completed by Ruhani et al. (2018).

In completing a meta-analysis of the relationship between trade volume and stock returns, Bajzik (2021) argues that determining the relationship between these variables would bear significance in confirming or denying the efficient market hypothesis. Downey (2021) defines the efficient market hypothesis as a posit that share prices reflect all available information, and that outperforming the market through generating consistent outsized returns is therefore impossible. According to this theory, stocks trade at their fair market value on all exchanges. Karpoff (1988) regards volume as a representation of the flow of information into the market, which may present opportunities for traders, as Tetlock (2007) believes that a sentiment-based trading strategy could be possible following his analysis of *Wall Street Journal* columns on stock returns. If a relationship between trading volume and stock prices is discovered, it would indicate that markets are not indeed fully efficient and reflective of all information. This would thus generate an opportunity for arbitrage and stand to dismiss the efficient market hypothesis, despite its empirical backing. (Bajzik, 2021)

Prior literature establishes that the movement of volume relates to the price of equities. Shen and Wang (1998) repeat an adage of technical analysis held by many traders that abnormally large volume is a sign of price changes. Gebkaa and Wohar (2013) find a complex relationship between these variables. Shen (2012) observes market folklore suggesting that prices and volume are positively related, as volume is necessary for prices to move. Given these fundamental market viewpoints and past research, we estimate an equation where stock price depends on trading volume and other financial market characteristics. We will add to existing literature by using a linear model, which contrasts with prior researchers, who primarily utilize time series analysis and GARCH models. This linear model permits us to address heterogeneity by instituting control variables, fixed effects, and an instrument. We will be assuming a linear model consistent with an ordinary least squares (OLS) regression:

1. $stockprice_i = \beta_0 + \beta_1 tradingvolume_1 + X_i\gamma + \varepsilon_{it}$

Where $stockprice_i$ is the measure of daily stock price, $tradingvolume_1$ measures the number of shares bought or sold on the same day and ε_{it} is a random error term. This framework will serve as the basis for the remainder of our analysis.

Data

To estimate the relationship between trading volume and stock price, we rely on a dataset of the Standard & Poor's 500 Index (S&P 500) sourced from the New York Stock Exchange. The S&P 500 Index is a market-capitalization float-weighted index of 500 premier leading publicly traded American companies. Containing a large sample of diverse large cap stocks from varying industries, it is considered the benchmark and gauge of the entire equities market. (Kenton, 2022) The dataset is provided by user Cam Nugent on the online data science platform Kaggle. Nugent acquired the dataset through the Investor's Exchange API (IEX), a program that tracks the stock data of all companies listed on the S&P 500 composite index. This data is available to download on the Kaggle website. Spanning from February 8, 2013 to February 7, 2018, this panel dataset includes detailed information regarding publicly traded corporations over the duration of this period. (Nugent, 2018)

The individual unit of observation in this dataset are the daily stock prices of all companies listed on the S&P 500 index. There are 619,040 observations in this dataset. Other variables listed in this dataset include the date, the open, high, low and close price of each stock, and finally, the volume. For this sake of this analysis, the daily closing price of each stock will be utilized as the regression dependent variable. Utilizing the closing price will permit us to control for the impact of the stock's performance the previous day in our model. In tracking the daily returns of Pacific Basin countries, Gebkaa and Wohar (2013) indicate that accounting for market movement the previous day is important.

Table 1 shows a summary of relevant variables included in the S&P 500 dataset. Variable stock price, the closing daily stock price of each company during the timeframe, has a mean of $83.044 and a standard deviation of 97.39. The standard deviation of this variable exceeding the mean indicates a distribution that includes extreme values. The difference between the minimum stock price ($1.59) and maximum stock price ($2,049) corroborates this notion, suggesting that a small number of corporations with extremely high stock price cause an upward effect in the estimation of the mean and standard deviation of the dataset. These companies will not be dropped from the dataset, however, as many of them have a significantly high enough market capitalization to warrant inclusion given their weighting in the S&P 500. The standard deviation of the trading volume variable (8693609.5 shares) also exceeds its mean (4,321,823.4 shares). Evidently, stocks with high trading volumes skew the distribution. However, they will not be dropped from the dataset due to their market capitalization warranting their inclusion in the Index. The mean number of trading volume, 4,321,823.4 shares, will be utilized in the interpretation of our results. As there is a significant disparity between the magnitude of the mean stock price ($83.044) and mean volume (4,321,823.4), the regression coefficient for trade volume will be extremely small. Thus, we will multiply the mean number of shares traded by this regression coefficient to attempt estimation the true impact of volume increases on price.

Other variables included in Table 1 include the mean and standard deviation of stock variables open, high and low. Similarly, to stock price and volume, the standard deviation of these figures exceeds their means. This is consistent with intuition suggesting extreme values are present in the dataset. Also included in Table 1 are control variables Highlowdiff and Prevday, which we will elaborate on in the Econometric Identification segment of this paper. Binary variable Prevday has a mean exceeding .5, indicating that stocks closed higher than they opened over 50% of the time on average. This figure indicates that the overall stock market was in a bull run during this timeframe.

Econometric Identification

We estimate the relationship between stock price and trading volume using a linear ordinary least squares (OLS) model, as shown in equation 2. Though Bajzik (2021) does not raise questions regarding the viability of OLS in stock analysis, this contradicts prior literature. The main challenge presented by utilizing the OLS framework is controlling for unobserved heteroskedasticity in the stock market. Shen and Wang (1998) note that conditional heteroskedasticity is common for stock returns, particularly in the short run. Given these concerns, we will implement control variables, fixed stock and time effects and an instrumental variable into our model. This will minimize the impact of heterogeneity originating from the behavior of individual firms and variation sourced from time itself. The basic OLS framework that will be utilized for this analysis is presented as follows:

$$2. \quad stockprice_i = \beta_0 + \beta_1 tradingvolume_1 + \varepsilon_{it}$$

Where $stockprice_i$ is the measure of daily stock price, $tradingvolume_1$ measures the number of shares bought or sold on the same day and ε_{it} is a random error term. We will term this framework Model 1. As it lacks sophisticated controls, it can only be considered as a preliminary estimate of the true impact of trading volume on stock price. To address conditional heteroskedasticity contained in the error term, we will implement several control variables in Model 2:

3. $$stockprice_i = \beta_0 + \beta_1 tradingvolume_1 + \beta_2 Highlowdiff_1 + \beta_3 Prevday_1 + \beta_4 month_1 + \varepsilon_{it}$$

This framework introduces three primary control variables. $Prevday_1$ is a lagged binary variable reflecting whether the stock closed higher than it opened the previous trading day. It is coded 0,1, with value 0 representing the stock closing lower than its opening price and 1 if it closes higher than its opening price. Control variable $Highlowdiff_1$ is introduced to control for short-run volatility in stocks. It is calculated as the difference between the High and Low price of a stock on a given day. Thus, the effects of a stock's volatility throughout a trading day will be held constant throughout this analysis. Finally, we follow Gallant et al. (1992) in introducing monthly controls. Variable $month_1$ represents a value coded contingent on the month of the year, holding the impact of the month on stock price constant.

As we are working with a panel dataset, we will introduce fixed effects into our model. The fixed effects will hold constant the impact of individual publicly traded companies and the impacts of time. By implementing these fixed effects, we are controlling for the average differences of both observed and unobserved variation across all companies in the dataset. This greatly minimizes the effect of omitted variable bias in our analysis. We are instituting both fixed effects for the individual stocks and daily time fixed effects. The time fixed effects will eliminate any observed and unobserved heteroskedasticity across all individual days included in this dataset. Thus, Model 3 can be outlined with the following equation:

$$4.\quad stockprice_i = \beta_0 + \beta_1 tradingvolume_1 + X_i\gamma + \alpha_i + \lambda_t + \varepsilon_{it}$$

Where X_i is a set of control variables Highlowdiff, Prevday and month, α_i is represents stock fixed effects and λ_t contains daily time fixed effects. Though we have now controlled for a significant quantity of omitted variable bias, we can maximize the robustness of our final model by introducing an instrumental variable. This variable must be exogenous, only impacting stock price through trade volume. Shen and Wang (1998) included weekday dummy variables in their analysis of trading volume and price limits on the Taiwan Stock Exchange. We will build off this previous research by introducing instrumental variable day of week. The day of the week will not impact the prices of individual stocks except through the quantity of stocks traded on that individual day. A Tuesday trading day will not influence stock price except through the number of shares exchanged on a Tuesday. Our final, headline 2 Stage Least Squares regression can be modeled with the following equations:

$$5.\quad stockprice_i = \beta_0 + \beta_1 trad\widehat{ingvo}lume_1 + X_i\gamma + \alpha_i + \lambda_t + \varepsilon_{it}$$

$$6.\quad trad\widehat{ingvo}lume_1 = \Pi_0 + \Pi_1 Z_{dow} + X_i\theta + \alpha_i + \lambda_t + \upsilon_{it}$$

Where $trad\widehat{ingvo}lume_1$ is an instrumented variable representing trading volume and Z_{dow} is instrument variable day of week, which does not impact stock price except through trade volume. Equation 5 outlines Model 4, the primary regression utilized in this analysis. When combined with our existing stock and time fixed effects, our model appears to be robust and controls for a large quantity of heterogeneity in the data.

A further way to approach this question is through the percentage change of a stock in response to an increase in volume. The percentage change of a financial asset is defined as its return (Hayes, 2021). Considering much of related literature on this topic focuses on the impact of trade volume on stock returns, introducing a nonlinearity will function as a robustness check and a point of comparison with prior research. Furthermore, we add to prior research conducted by Gebkaa and Wohar (2013), who find a complex, nonlinear relationship between volume and stock returns. An updated model (Model 5), modified to include nonlinearities, reveals the percentage change in stock price due to a single trade increase in volume:

$$7.\quad lnstockprice_i = \beta_0 + \beta_1 trad\widehat{ingvo}lume_1 + X_i\gamma + \alpha_i + \lambda_t + \varepsilon_{it}$$

Where $lnstockprice_i$ represents the percentage change in stock price given an increase of shares traded.

This OLS framework represents a strong basis for estimating the impact of volume on stock price. However, it is imperfect. Other sources of bias remain and were unable to be controlled for due to limitations in the available dataset. Information regarding the specific industry of all companies listed in the S&P 500 was not available. Thus, industry-specific effects are unable to be controlled for in this analysis. Theoretically, financial stocks may be subjected to higher volume than agricultural stocks, which could lead to a disproportionate impact on stock prices. This may lead to an overestimation or underestimation of the true effect of trading volume on stock price. Macroeconomic conditions during the period reflected by the data are generally consistent, limiting the application of this analysis beyond the conditions of an expanding U.S. economy.

Results

We begin with a preliminary, first-stage ordinary least squares regression including only independent variable trade volume and dependent variable stock price. This regression, Model 1, is included in Column 1 of Table 2. No further control variables, stock or time fixed effects or instrumental variables are implemented in this model. We find a negative relationship between trade volume and stock price (-1.600e-06), statistically significant at the one percent level (standard error of 2.590e-08). This suggests that, on average, a single trade will decrease stock prices by a small quantity. Multiplying the regression coefficient by 4,321,823.4, the mean trading volume of S&P 500 stocks in Table 1, an average stock price decrease of -$6.91 is found. However, this model lacks adequate control variables, as indicated by a small adjusted R^2 of .02039084.

To build a more sophisticated model, we introduce several control variables. Included in Model 2, depicted in Column 2 of Table 2, are control variables highlowdiff and Prevday. Monthly control variables are also implemented. In this second model, we find another negative relationship between trade volume and stock price (-9.698e-07), statistically significant once more at the one percent level (standard error of 1.720e-08). Multiplying the regression coefficient by the mean trading volume reveals a decrease in stock price of -$4.19. This result suggests that our preliminary Model 1 overestimated the magnitude of stock price decrease.

The coefficients for control variables Highlowdiff (34.921249) and Prevday (2.7617624) are both positive and statistically significant at the one percent level (with standard errors of (.35171411 and .15351793, respectively). Based on these results, we can infer that larger volatility and an equity's performance the previous day are both positive sources of a stock price increase. We further conduct an F-Test of Highlowdiff in Model 2, finding a result of 9858,25, indicating that it is a significant control variable. Coefficients for the monthly variables are all positive and statistically significant at the one percent level. We find an adjusted R^2 of .62175199, indicating that this model is a more effective fit for the data.

In Model 3, presented in Column 3 of Table 2, and given our panel dataset, we implement both stock and time fixed effects. We include these fixed effects to control for any daily variation in specific equities over time. The average effects of each stock (stock effect) and each trading date (time effect) are held constant, eliminating any observed or unobserved conditional heteroskedasticity in individual companies. After completing the fixed effects regression, we once again find a negative relationship between trading volume and stock price (-5.090e-07), statistically significant at the one percent level (standard error of 1.075e-07). We find a decrease of stock price by -$2.20 after multiplying the volume regression coefficient by the mean trading volume. This represents a substantial change in the magnitude of the stock price decline. Model 3 suggests a difference of $1.89 in the decline of stock price relative to Model 2, and a $2.82 difference in the decline of stock price relative to Model 1. Without fixed effects, the magnitude of stock price decline is overestimated.

Examining the other control variables yields notable results. The coefficients for Highlowdiff (7.5999057) and Prevday (1.1828226) are both positive, but of smaller magnitudes relative to Model 2. They remain statistically significant at the one percent level. The robust standard error of Highlowdiff increases from Model 2, suggesting greater variance of volatility values, while the standard error of Prevday decreases. Coefficients for month variables remain positive and statistically significant at the one percent level, with the exception of the month of October (-2.531538), which has a negative coefficient, and November (.08794807), which has no statistical significance. This result indicates that the month of November contributes neither positively nor negatively to a change in stock price. The Adjusted R^2 of the model declines significantly relative to Model 2, however, decreasing to .10410542. The model is a less effective fit for the data at hand.

As a final step to develop a maximally robust result, we introduce an instrumental variable for day of the week. Each weekday (Monday, Tuesday, Wednesday, etc.) is coded with a specific value. Day of the week is an exogenous instrument within our regression framework because the effect of each weekday will not impact individual stock prices except through the quantity of trades on that given day. For example, the trading day of a Monday will not impact specific stock prices except through the number of shares traded because it is Monday. To evaluate the strength of this instrument, we complete a regression with day of week as a control variable. We find the instrumental variable, *dow*, to have an F-statistic of 13.75. As the value of the F-statistic is greater than ten, it is a sufficiently strong instrumental variable for the sake of our analysis. We will proceed with Day of Week as an instrumental variable.

Our final headline 2 Stage Least Squares regression includes regressor trade volume, control variables highlowdiff and Prevday, monthly controls, stock and time fixed effects, and the day of week instrumental variable. Our result in Model 4 differs drastically from our previous Models. Included in Column 4 of Table 2, we find a positive relationship between trade volume and stock price (5.606e-07), significant at the ten percent level (standard error of 3.273e-07). When multiplied by the mean trading volume included in Table 1, this represents an increase of stock price by $2.42. This is a positive change in stock price relative to Model 3 by $4.62. This result significantly alters our interpretation. With the day of week instrument *dow* included in our model, an increase of trading volume by an additional trade increases stock price by $.0000005606. It is important to note the loss of statistical significance with this result, suggesting that this positive value is harder to differentiate from a zero change in stock price due to an increase in trading volume.

Model 4 yields further results of interest in control variables. Variables highlowdiff (7.187223) and Prevday (1.2723824) maintain positive regression coefficients, which are statistically significant at the one percent level (standard errors of .12977547 and .08369697, respectively). Neither the coefficients nor standard errors are substantially different from Model 3, suggesting that the inclusion of the instrumental variable *dow* does not alter the impact of the market's volatility or previous day's performance in comparison to the fixed-effects only regression. Coefficients for monthly variables are all positive and statistically significant at the one percent level with exceptions of the months of August and November. August has a negative regression coefficient (-.42586915) that is not statistically significant at any level. The impact of the month of August on stock price is not substantially different from zero. The coefficient of November remains positive (.61060662) and adds two levels of significance, now significant at the five percent level.

We extend our findings and further test for robustness by introducing nonlinearities to our model. Introducing the log of Stock Price represents the most appropriate nonlinearity. As stated previously, much of prior literature focuses on stock returns. By logging stock price, we can determine the percentage change in stock price contingent on volume. This effectively represents the change in stock returns, which are defined as the percent change in price of a financial asset or investment (Hayes, 2021). Model 5, presented in Table 3, includes a 2 Stage Least Squares regression with logged Stock Price, while maintaining all previous controls from Model 4, including highlowdiff, Prevday, monthly controls, stock and daily fixed effects, and the day of week instrumental variable. We find a positive percentage increase in stock price (7.400e-09) for an additional trade, statistically significant at the one percent level (standard error of 2.600e-09). When multiplied by the mean trading volume value in Table 1, the percentage change in stock price equates to 3.20%. This indicates a substantial, positive influence of trading volume on stock returns. Coefficients for control variables highlowdiff (.03776876) and Prevday (.01172677) are positive and statistically significant at the one percent level (standard errors of

(.00102136 and .00065871, respectively). All monthly control variables have positive regression coefficients and are statistically significant at the one percent level barring November. Akin to Model 3, November has a positive coefficient (.00264719) with no statistical significance (standard error of .00199356). The month of November appears to have no major discernable impact on stock returns. Ultimately, the results of Model 5 pertaining to the impact of trade volume on stock returns corroborate our previous findings in Model 4 regarding the impact of trade volume on stock price.

All told, we find evidence that trade volume causes a small but notable increase in stock price. These findings extend to the impact of trade volume on stock returns, which is also positive and statistically significant at the one percent level. These findings are consistent with previous literature, including Chen (2012) who finds a positive relationship between volume and stock returns during bull markets; Karpoff (1988) who suggests that costs incurred by short-sellers minimizes the influence of bears during periods of high trading volume; and Bessembinder and Seguin (1992) who discover a positive relationship between volume and volatility in futures trading.

Although our results are in line with prior research, several important caveats apply to our model. First, our dataset lacks any information regarding a relevant control variable: the industry of each stock included in our analysis. Thus, while we can control for company-specific impacts through fixed effects, we cannot account and control for the impacts of certain industries of varying volumes and its impact on stock prices. Informational technologies, financials and real estate stocks are weighted differently within the S&P 500. (Reiff, 2022) Our inability to control for industry-specific effects may have caused us to either overestimate or underestimate the true impact of trade volume on stock price. This is a source of major omitted variable bias.

Secondly, we did not control for any macroeconomic factors that can influence the stock market. Arago and Nieto (2004) found that macroeconomic factors prevail over specific company factors in determining market returns. Our model does not account for these factors outside of through the Prevday control variable, which is a reflection of the general past performance of the stock market. As the period from 2013 to 2018 can be characterized as a bull market, this control variable does capture the effects of certain positive macroeconomic indicators that may influence stock price.

Finally, we assume that changes in our instrumental variable, day of week, has no impact on stock price except through changes in trading volume. Though the F-statistic for the instrumental variable exceeds ten, indicating it is strong, it is still not a perfect instrument. Companies issue earnings reports on certain trading days, which can have a substantial impact on their stock. Shadka (2007) finds a strong negative relationship between expected stock returns and expected firm earnings. Earnings reports may increase trading volume, which permits our model to capture some of this effect. Gillette et al. (1999) finds that trade volume is inversely related to traders' expectations of the standard deviation of dividends forecasts, which lends some credence to our instrument as stock price is only impacted through volume.

Acknowledging these caveats, our results are consistent with prior literature regarding stock returns during bull markets and are robust to minimize the impact of undesired biases. We find that trading volume has a small and somewhat significant impact on both stock price and stock returns.

Conclusion

We find a small, slightly statistically significant increase in stock price when trade volume increases for stocks listed in the S&P 500 index. These results occur after our preliminary results indicated a negative relationship between volume and stock price. We find this positive relationship after controlling for stock performance and volatility, as we implemented fixed effects and a robust instrumental variable to address unobserved heterogeneity in the market. Our findings are consistent with prior literature reviewing the relationship between volume and stock returns during bull markets, which the U.S. stock market between 2013 and 2018 can be characterized as.

As higher volume increases stock price, this presents a potential opportunity for traders to profit during bull markets. Our results are consistent with research by Stickel and Verrecchi (1994), who are unable to deny the existence of arbitrage opportunities in their analysis of volume and stock price. Arbitragers can emerge to take advantage of the disparity in price between the increase volume and the fair market value of a company. The existence of arbitragers disputes the notion that the financial markets are fully efficient. (Downey, 2021) Gebkaa and Wohar (2013) do acknowledge that it may be difficult for traders to take advantage of this relationship. Nonetheless, the actions of these traders will reduce the disparity in price but will increase trade volume on these stocks. Evaluating the difference in the magnitude of price change between the correction of the market inefficiency and the subsequent increase in volume is an area warranting further research.

Overall, increases in trading volume of publicly traded companies listed in the S&P 500 stock index have a positive, though not fully significant, relationship with the corresponding stock price. When accounting for the mean trading volume of a stock, stock price tends to increase by \$2.42 and stock returns increase by 3.20%. Though counteracting our initial research, our analysis is robust in controlling for heterogeneity. This apparent positive connection between trading volume and stock price improves our knowledge of financial market structure. Furthermore, it reveals an opportunity for a volume-based arbitrage trading strategy, disputing a central tenant of the efficient market hypothesis. Finally, it represents an important consideration for public policy makers, as legislation that could impact trade volume has clear bearings on stock prices.

Appendix

Table 1: Descriptive Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
Price	619040	83.044	97.39	1.59	2049
Volume	619040	4321823.4	8693609.5	0	6.182e+08
Open	619029	83.023	97.379	1.62	2044
High	619032	83.778	98.208	1.69	2067.99
Low	619032	82.256	96.507	1.5	2035.11
Highlowdiff	619032	1.522	2.173	-.255	138.26
Prevday	618535	.523	.499	0	1

Table 2: Primary Regressions

Stock Price	(1) First Stage	(2) Controls	(3) Fixed Effects	(4) DoW Instrument
Volume	-1.60e-06*** (2.590e-08)	-9.698e-07*** (1.720e-08)	-5.090e-07*** (1.075e-07)	5.606e-07* (3.273e-07)
Highlowdiff		34.921249*** (.35171411)	7.5999057*** (1.0122101)	7.187223*** (.12977547)
Prevday		2.7617624*** (.15351793)	1.1828226*** (.13263911)	1.2723824*** (.08369697)
February		3.2878162*** (.41319261)	6.3817586*** (.50306153)	6.5160311*** (.20400271)
March		7.9680845*** (.39551458)	4.2436125*** (.46659219)	3.9047188*** (.21998833)
April		5.2664731*** (.39575619)	5.2690293*** (.60144191)	4.9360958*** (.22057438)
May		9.8509729*** (.39705316)	3.1303168*** (.52658681)	2.5182104*** (.27031429)
June		10.165068*** (.40568221)	2.0490714*** (.48008362)	1.5963077*** (.23859348)
July		11.056931***	1.5333182***	.86882919***

		(.40227109)	(.43815368)	(.28147722)
August		9.7191898***	1.3586382***	-.42586915
		(.42969345)	(.35207087)	(.34440017)
September		8.8981954***	1.8359448***	1.2799356***
		(.4014514)	(.34055509)	(.25936144)
October		3.6344436***	-2.531538***	2.1391052***
		(.39148265)	(.23992614)	(.22647024)
November		6.3895765***	.08794807	.61060662**
		(.41258804)	(.21944036)	(.25330551)
December		8.3437717***	.95264986***	1.6565103***
		(.3933017)	(.22517868)	(.28991415)
Constant	89.957527***	26.024646***	75.319132***	70.822962***
	(.17432272)	(.70342683)	(1.6853926)	(1.3845842)
Observations	619040	618529	618529	618529
Adj R^2	.02039084	.62175199	.10410542	.z
F-stat	3828.6506	1444.7762	37.428605	.z
Stock FE	No	No	Yes	Yes
Daily FE	No	No	Yes	Yes
Dow IV	No	No	No	Yes

Notes: Regression coefficients are significant at the one (***), five (**) and ten (*) percent level. Robust standard errors are displayed in parentheses. Month variables are categorical and are included in the 2nd, 3rd and 4th regressions. The sample consists of stock open, close, high, low, volume and date. High-Low Difference is calculated difference between the stock's open and close price on a given trading day. Previous Day's Performance is a binary variable representing whether the stock closed higher than it opened on the preceding trading day.

Table 3: Nonlinearities Regression

LnStockPrice	(5) price
Volume	7.400e-09***
	(2.600e-09)
Highlowdiff	.03776876***
	(.00102136)
Prevday	.01172677***
	(.00065871)
February	.07301438***
	(.00160554)
March	.05128507***
	(.00173135)
April	.05740496***
	(.00173596)
May	.03495897***
	(.00212743)
June	.02485806***
	(.00187778)
July	.01736178***
	(.00221528)
August	-.0152789***
	(.0027105)
September	.02396729***
	(.00204122)
October	.02790584***
	(.00178236)
November	.00264719
	(.00199356)
December	.01509749***
	(.00228168)
Constant	4.0643591***
	(.01089694)
Observations	618529
Adj R^2	.z
F-stat	.z
Stock FE	Yes
Daily FE	Yes
Dow IV	Yes

Notes: Regression coefficients are significant at the one (***), five (**) and ten (*) percent level. Robust standard errors are displayed in parentheses. Month variables are categorical. The sample consists of stock open, close, high, low, volume and date. High-Low Difference is calculated difference between the stock's open and close price on a given trading day. Previous Day's Performance is a binary variable representing whether the stock closed higher than it opened on the preceding trading day. LnStockPrice is the logged value of Stock Price. The coefficient for volume represents the percent change in stock price given an additional trade of a company's stock.

References

Aragó, V., & Nieto, L. (2005). Heteroskedasticity in the returns of the main world stock exchange indices: volume versus GARCH effects. *Journal of International Financial Markets, Institutions and Money*, *15*(3), 271–284. https://doi.org/10.1016/j.intfin.2004.06.001

Bajzik, J. (2021). Trading volume and stock returns: A meta-analysis. *International Review of Financial Analysis*, *78*, 101923. https://doi.org/10.1016/j.irfa.2021.101923

Bessembinder, H., & Seguin, P. J. (1992). Futures-Trading Activity and Stock Price Volatility. *The Journal of Finance*, *47*(5), 2015–2034. https://doi.org/10.1111/j.1540-6261.1992.tb04695.x

Chen, S.-S. (2012). Revisiting the empirical linkages between stock returns and trading volume. *Journal of Banking & Finance*, *36*(6), 1781–1788. https://doi.org/10.1016/j.jbankfin.2012.02.003

Chordia, T., Subrahmanyam, A., & Anshuman, V. R. R. (2000). Trading Activity and Expected Stock Returns. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.204488

Downey, L. (2021, March 25). Efficient market hypothesis (EMH) definition (G. Scott, Ed.). Investopedia. https://www.investopedia.com/terms/e/efficientmarkethypothesis.asp

Gallant, A. R., Rossi, P. E., & Tauchen, G. (1992). Stock Prices and Volume. *Review of Financial Studies*, *5*(2), 199–242. https://doi.org/10.1093/rfs/5.2.199

Gebka, B., & Wohar, M. E. (2013). Causality between trading volume and returns: Evidence from quantile regressions. *International Review of Economics & Finance*, *27*, 144–159. https://doi.org/10.1016/j.iref.2012.09.009

Gillette, A. B., Stevens, D. E., Watts, S. G., & Williams, A. W. (1999). Price and Volume Reactions to Public Information Releases: An Experimental Approach Incorporating Traders' Subjective Beliefs. *Contemporary Accounting Research*, *16*(3), 437–479. https://doi.org/10.1111/j.1911-3846.1999.tb00590.x

Hayes, A. (2021, May 29). *Return* (G. Scott, Ed.). Investopedia. https://www.investopedia.com/terms/r/return.asp

Karpoff, J. M. (1988). COSTLY SHORT SALES AND THE CORRELATION OF RETURNS WITH VOLUME. *Journal of Financial Research*, *11*(3), 173–188. https://doi.org/10.1111/j.1475-6803.1988.tb00080.x

Kenton, W. (2021, March 23). *Understanding S&P 500 Index – Standard & Poor's 500 Index* (M. Boyle, Ed.). Investopedia. https://www.investopedia.com/terms/s/sp500.asp

Nugent, C. (2018). *S&P 500 stock data*. Www.kaggle.com. https://www.kaggle.com/datasets/camnugent/sandp500

Reiff, N. (2022, April 5). Where to Find a List of the Stocks in the S&P 500 (C. Stapleton & V. Velasquez, Eds.). Investopedia. https://www.investopedia.com/ask/answers/08/find-stocks-in-sp500.asp

Ruhani, F., Aminul Islam, Md., Salha Tunku Ahmad, T., & Ruhul Quddus, M. (2018). Effects of Financial Market Variables on Stock Prices: A Review of the Literature. *Journal of Modern Accounting and Auditing*, *14*(11). https://doi.org/10.17265/1548-6583/2018.11.002

Saad, L. (2019, September 13). *What Percentage of Americans Owns Stock?* Gallup.com; Gallup. https://news.gallup.com/poll/266807/percentage-americans-owns-stock.aspx

Sadka, G. (2007). Understanding Stock Price Volatility: The Role of Earnings. *Journal of Accounting Research*, *45*(1), 199–228. https://doi.org/10.1111/j.1475-679x.2006.00230.x

Shen, C.-H., & Wang, L.-R. (1998). Daily serial correlation, trading volume and price limits: Evidence from the Taiwan stock market. *Pacific-Basin Finance Journal*, *6*(3-4), 251–273. https://doi.org/10.1016/s0927-538x(98)00011-0

Stickel, S. E., & Verrecchia, R. E. (1994). Evidence that Trading Volume Sustains Stock Price Changes. *Financial Analysts Journal*, *50*(6), 57–67. https://doi.org/10.2469/faj.v50.n6.57

Tatom, J. A. (2021, July 6). *Biden's capital gains tax increase is more unproductive misdirection*. The Hill. https://thehill.com/opinion/finance/561274-bidens-capital-gains-tax-increase-is-more-unproductive-misdirection/#:~:text=Whenever%20the%20capital%20gains%20tax

Tetlock, P. C. (2007). Giving Content to Investor Sentiment: The Role of Media in the Stock Market. *The Journal of Finance*, *62*(3), 1139–1168. https://doi.org/10.1111/j.1540-6261.2007.01232.x

Twin, A. (2022, January 2). *Volume of Trade Definition*. Investopedia. https://www.investopedia.com/terms/v/volumeoftrade.asp

Weber, P., & Rosenow, B. (2006). Large stock price changes: volume or liquidity? *Quantitative Finance*, *6*(1), 7–14. https://doi.org/10.1080/14697680500168008