ETFs and Their Impact on Bid-Ask Spread
Achyut Gautam
Ramapo College of New Jersey, Mahwah, NJ, 07430

INTRODUCTION
The growth in passive fund investment has been steadily increasing over the last two decades. Within the US, passive funds grew by 138% from 2007-2017 and have about 15% share of total outstanding securities (Sushko et al.). Exchange Traded Fund, commonly known as ETF, is a type of passive fund that tracks a stock index, commodities, bonds, or a basket of assets. At the end of September 2017, global ETF assets totaled to $4.4 trillion with a growth rate 21% from 2015 (Julie et al.). The U.S. is the world’s largest ETF market and as of 2017, the U.S. ETF industry alone had roughly $3 trillion in total industry asset.

Previous literature have found that an increase in ETF trades of a particular stock leads to an increase in trading cost which is captured by an increase in bid-ask spread of the underlying (Israeli et al.). Similarly, stocks that are traded as ETFs reflect higher volatility than otherwise similar securities (Ben-David et al.). Since ETFs are easier to trade and are cost-effective, ETFs encourage passive investment leading to lesser scrutiny from the market. Based on this rationale, the research expands on previous literatures by finding the incremental change in bid-ask spread, a transaction cost, of entire stock market index when volume of ETF tracking the index increases.

THEORETICAL BACKGROUND

The model satisfies five of the seven classical linear assumptions:

1. The regression model is linear in the coefficients and the error term
2. The error term has a population mean of zero
3. All independent variables are uncorrelated with the error term
4. Observations of the error term are uncorrelated with each other
5. No independent variable is a perfect linear function of other explanatory variables
6. The error term has a constant variance (no heteroscedasticity)
7. The error term is normally distributed

The regression model is linear in the coefficients and the error term. The model satisfies five of the seven classical linear assumptions:

\[ Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \epsilon \]

where \( Y \) is the dependent variable (bid-ask spread), \( X_1, X_2, \ldots, X_k \) are independent variables, and \( \epsilon \) is the error term.

REGRESSION RESULTS

The study concludes that there is a strong significant positive relationship between the bid-ask spread of S&P 500 and its ETFs (Table 1) while the same holds true but only for NASDAQ-100 and QQQ (Table 2). TQQQ had relatively small observation, failed the omitted variable test, and led to insignificant result. Thus it was left out from the equation 2. Model 3, which uses Convin and Schultz spread, resulted in either wrong coefficient sign or insignificant result (Table 3).

Unlike the model used by Israeli et al., this study paper excludes control variables such as institutional ownership and number of analysts analyzing the underlying due to lack of data. Incorporating control variables and effectively controlling for noise might have increased the robustness and efficiency of the models, established homoscedasticity and normality assumption, and improved results of the regression model.

CONCLUSION

The study concludes that there is a strong significant positive relationship between the bid-ask spread of S&P 500 and its ETFs (Table 1) while the same holds true but only for NASDAQ-100 and QQQ (Table 2). TQQQ had relatively small observation, failed the omitted variable test, and led to insignificant result. Thus it was left out from the equation 2. Model 3, which uses Convin and Schultz spread, resulted in either wrong coefficient sign or insignificant result (Table 3).

Unlike the model used by Israeli et al., this study paper excludes control variables such as institutional ownership and number of analysts analyzing the underlying due to lack of data. Incorporating control variables and effectively controlling for noise might have increased the robustness and efficiency of the models, established homoscedasticity and normality assumption, and improved results of the regression model.