High Frequency Autocorrelation in the Returns of the SPY and the QQQ

Scott Davis*

January 21, 2004

Abstract

In this paper I test the random walk hypothesis for high frequency stock market returns of two major index tracking stocks. I use tick by tick quotes from the S&P 500 tracking stock, the SPY, and the NASDAQ 100 tracking stock, the QQQ. These quotes are taken from a four week period in December 2002 and a four week period in March 2003. I analyze regression coefficients to test for significance of autocorrelation. The random walk model is rejected for the highest frequencies, but randomness is upheld for the longer time intervals. I then test to see if the market is efficient by testing the economic significance of the findings of autocorrelation. No significant profits are found, thereby upholding the theory of market efficiency.

This paper was written during an independent study with Professor Bjørn Eraker of Duke University in the fall of 2003. I would like to thank him as well as Professor Ed Tower for advice, assistance, and many helpful comments that lead to the completion of this paper. I would also like to thank Pierre Deboisblanc for technical assistance rendered during the research process.
*Scott Davis is a graduate student in the Ph.D. program in Economics at Vanderbilt University. He can be reached at jonathan.s.davis@vanderbilt.edu

I. Introduction

The random walk model and the theory of weak form market efficiency tell us that trends cannot exist within the market. In the formula:

$$r_t = \alpha + \beta^* r_{t-1} + \varepsilon$$

we should have an insignificant α and β. This paper will look at the existence of trends in high frequency market data. I will test the significance of these trends over varying observation frequencies. Other economists have done similar tests in the past and some of them are cited here, but technology has allowed me to study autocorrelation at high frequencies unlike past studies which had to use lower frequency data. I will study the returns from the S&P 500 tracking stock and the NASDAQ 100 tracking stock from two periods of 4 weeks each, one in December of 2002 and one in March of 2003. I will study the returns from these stocks over 1, 5, 15, 30 and 60 minute intervals to see how frequency can affect the existence of trends. If statistically significant trends do exist, I will test their economic significance to see if a money making opportunity can be found. I will then test for the existence of trends in the differences of returns from these two stocks, or rather:

$$(r_{SPY,t}-r_{QQQ,t}) = \alpha + \beta^*(r_{SPY,t-1}-r_{QQQ,t-1}) + \varepsilon$$

Using the same data as before, an estimated β will be calculated from this equation. This β will be subjected to the same tests for statistical significance and economic significance as in the one security regressions. A statistically significant β will lead to a rejection of the random walk model. A β which leads to profitable trades will imply that the market is not efficient. The goal here is to discover if trends in returns do exist for this high frequency data. If these trends are found then they will be tested for economic significance to see if an opportunity to make money has been found.

II. The Data

For this paper I followed the returns of two major market tracking stocks. I followed the SPY and the QQQ. Index tracking stocks follow a particular index of stocks and post the same returns as that index. A portfolio consisting of every stock in a particular index would have the same return as the index tracking stock, so for simplicity the tracking stock's prices are used. The SPY tracks the S&P 500 index; the QQQ tracks the NASDAQ 100 index. Since these are major stocks with high activity and volume there was no danger of microstructure biases from infrequent trading and bid ask effects causing a bias in the data. This point was raised by Conrad and Kaul in 1988. Also since the stocks had high volume it was possible to find minute by minute returns. This may not have been possible for a thinly traded stock (Lo and MacKinlay 1988).

The data was collected from the Wharton Data Research Service website. Tick by tick data was collected that was then run through a filtering system to make it minute by minute data. This tick by tick data was collected for two separate four week intervals. The first interval was December 9, 2002 to January 3, 2003. This period was slightly shortened by the holidays and has a total of 17 trading days. Therefore there are a total of 6647 data points in the minute by minute data. The second interval was March 10, 2003 to April 4, 2003. This 4 week period had 20 trading days or a total of 7820 data points for the minute by minute data. In further analysis the data was broken down again into 5 minute intervals; therefore the December data had 1329 data points while the March data had 1564 points. Then it was broken into 15 minute intervals, so 443 in December and

521 in March. This was again broken into 30 minute intervals, so 221 for December and 260 for March. And again broken into 60 minute intervals, so 110 for December and 130 for March. This use of multiple time intervals is going to be the foundation for testing the existence of autocorrelation across different time spans. My hypothesis is that randomness will not exist at shorter time intervals, but as the time interval grows, the significance of trends will fall.

III. Trends in Individual Security Returns

I first test for the existence of autocorrelation in the returns of individual securities. The minute by minute returns are derived from the price quotes and then put into this regression equation:

$$r_t = \alpha + \beta * r_{t-1} + \varepsilon$$

The β was then tested for statistical significance to find the statistical significance of the autocorrelation. This same process was done for the returns from both the SPY and the QQQ. This process was also done using not only minute by minute returns but returns over 5, 15, 30 and 60 minute intervals. The β and R^2 from each regression are shown in tables 1 & 2. In every case the α was grossly insignificant so it does not need to be shown.

TABLE 1 **Coefficients from a regression of the returns from period t on the returns from period t-1 for the QQQ. Returns measured over different time intervals.**

	1 minute	5 minutes	15 minutes	30 minutes	60 minutes
Dec.	**-0.256**	**-0.257**	**-0.146**	**-0.147**	-0.079
	(0.012)	**(0.027)**	**(0.047)**	**(0.067)**	(0.094)
March	**-0.079**	-0.016	-0.034	0.116	0.056
	(0.011)	(0.025)	(0.043)	(0.062)	(0.087)
R^2 – Dec	0.066	0.066	0.021	0.022	0.006
R^2 – March	0.006	0.0002	0.001	0.013	0.003

TABLE 2 **Coefficients from a regression of the returns from period t on the returns from period t-1 for the SPY. Returns measured over different time intervals.**

	1 minute	5 minutes	15 minutes	30 minutes	60 minutes
Dec.	**-0.107**	**-0.094**	**-0.11**	0.112	-0.096
	(0.012)	**(0.027)**	**(0.047)**	(0.067)	(0.094)
March	**-0.238**	**-0.079**	-0.06	0.052	0.023
	(0.011)	**(0.025)**	(0.043)	(0.062)	(0.087)
R^2 – Dec.	0.012	0.009	0.012	0.013	0.009
R^2 - March	0.057	0.006	0.004	0.003	0.001

For both the SPY and the QQQ, the significant coefficients are negative. That means that if the SPY has a positive return one minute then we can expect a negative return in the next minute. The shorter time interval leads to a higher percentage autocorrelation, -25% in some cases, but the percentage autocorrelation quickly drops off as the time interval increases. For both stocks in both testing months the random walk model can be easily rejected when the time interval is short, but it becomes harder to reject as the time interval gets longer (see Lo 52) also (Conrad 418).

There is autocorrelation that exists in the returns of the individual securities SPY and QQQ. This autocorrelation is statistically significant when the time interval is short enough. The question of statistical significance must be followed up by a question of economic significance. If we have shown with regression equations that a trend does exist, is it enough for money to be made? Any attempt to make money off this trend must involve making trades and having to adhere to bid prices and ask prices. After all of this, is the trend still significant? Can I take the results from this paper, go to Wall Street, and become a millionaire overnight? The sad answer to this question is no. Using the same data that was used to find the trends, simulated trades were set up that would use the regression results to predict the market. The β in the regressions was always negative. This meant that when the stock had a positive return the last period, we could expect it to have a negative return in the next period. In these trades I would watch for the return in the last period and then make my trade. If the security posted a negative return in the last period then I would buy that security for the next period. Since I was buying I had to accept the lowest ask price in the market. I would hold that security until the end of the time period and then sell it for the highest bid price in the market. If instead the return on the security was positive in the last time period then I would do the opposite. I would sell the security now for the bid price and then buy it back at the end of the time interval for the lowest ask price. From table 1 the securities and time intervals that showed significant autocorrelation were: the 1, 5, 15 minute December SPYs, the 1, 5, 15, and 30 minute December QQQs, the 1 and 5 minute March SPYs, and the 1 minute March QQQ. Tables

3 and 4 show the results from a series of simulated trades over the respective 4 week periods where $1 of stock was bought or shorted each time period and held for one period before sold or bought back.

TABLE 3 Results from the simulated trades involving the QQQ.

	1 minute	5 minutes	15 minutes	30 minutes	60 minutes
Profit December	-$4.21	-$0.07	$0.64	$0.69	$0.80
Profit March	-$6.88	-$2.18	$0.51	$0.83	$0.84

TABLE 4 Results from the simulated trades involving the SPY

	1 minute	5 minutes	15 minutes	30 minutes	60 minutes
Profit December	-$1.31	$0.57	$0.84	$0.86	$0.87
Profit March	-$3.91	-$0.53	$0.77	$0.91	$0.98

Recall from tables 1 & 2 that for both of these stocks, there was less autocorrelation in the March returns than the ones from December. There were few significant regressions for the March data and each does post a significant loss. Therefore one would say that the statistically significant autocorrelations using the March data were not economically significant. The December data is a little different, for there were more periods with statistically significant autocorrelations. A profit was realized using these simulated trades in some of the longer interval trades involving both the SPY and the QQQ. The one concern is that the standard deviation of these profits is much higher than the profits themselves (about $1.40). This means that what little profit that may be realized is just one trade away from a significant loss. The lack of economic significance to the results is consistent with the theory of weak form market efficiency. This regression is fairly simple, and there is no limit to how many investors can profit by predicting the market. No one will earn an economic profit. If for example they predict the stock to go up from a price of $25, then no one would be willing to sell the stock for $25. This future expectation would be reflected in the ask price. A high ask price or a low bid price can eat away profits that can come from this trading strategy. This was the finding of Fama and Blume in 1966. They took the results from an earlier study that showed significant profits from trades based on autocorrelation, but they found that these profits were caused by a bias in not incorporating bid and ask prices. They found, just like I did, that when these bid and ask prices were considered, significant profits do not exist from this type of trading. The results from these regressions are an example of this form of market efficiency. The results may be statistically significant, but since this money making strategy can be learned and used by everyone, it will, and no profits will be realized.

IV. Trends in the differences in Security Returns

The last section looked for statistically and economically significant trends in the returns of individual securities. This section will look for trends in the differences in returns of the market. I will still work with the same two stocks, SPY and QQQ. I will still look at the same time periods, 4 weeks in December and 4 weeks in March. The only change is that now my regression model will be:

$$(r_{SPY,t}-r_{QQQ,t}) = \alpha + \beta^*(r_{SPY,t-1}-r_{QQQ,t-1}) + \varepsilon$$

Therefore my variables in the regression will be by how much the SPY was able to beat out the QQQ in one time period. A negative value for β would mean that if last period the SPY beat the QQQ, the QQQ should catch up in the next period by beating the SPY. A positive β would mean that if in the last period the SPY beat out the QQQ, the SPY should again beat out the QQQ in the next period. Table 5 shows the results from the regressions over the different time intervals.

TABLE 5 Coefficients from a regression of the returns from period t on the returns from period t-1 for the portfolio of a long position on the SPY and a short position on the QQQ. Returns measured over different time intervals.

	1 minute	5 minutes	15 minutes	30 minutes	60 minutes
December	**-0.424**	**-0.391**	**-0.222**	0.043	-0.138
	(0.011)	**(0.025)**	**(0.046)**	(0.067)	(0.092)
March	**-0.397**	**-0.273**	-0.074	0.029	0.031
	(0.010)	**(0.024)**	(0.044)	(0.062)	(0.087)
R^2 December	0.180	0.153	0.050	0.002	0.021
R^2 March	0.157	0.075	0.006	0.001	0.001

The same pattern appears as before, only the absolute value of the autocorrelation is greater in this marginally smarter regression equation. The autocorrelation is significant when the time interval is small, but it quickly loses value as the time interval gets longer. In both months there is around -40% autocorrelation when the time interval is equal to one minute. This means that if the SPY beats the QQQ in one minute, the QQQ will return and beat the SPY in the next minute. As the interval increases this autocorrelation loses statistical significance.

There is a statistically significant trend in the differences of the returns for a small enough a time interval. This shows that the random walk model is disproved for these short intervals. A test of the efficiency of the market will depend on the economic significance of the autocorrelations. I set up a series of simulated trades based on the results from the regressions. This is to test if a money making opportunity exists within these results. If in the last period the SPY beats the QQQ, then the model tells us that the QQQ should beat the SPY in the next period. If I noticed the SPY beat the QQQ then I would buy $1 of QQQ at the ask price and sell $1 of SPY at the bid price. After that period I would sell my QQQ for the bid price, and I would buy back my SPY for the ask price. If instead the QQQ beat the SPY in the last period, I would do the opposite and buy the SPY while selling the QQQ. The results from these simulated trades are shown in table 6.

TABLE 6 Results from simulated trades involving both the SPY and the QQQ.

	1 minute	5 minutes	15 minutes	30 minutes	60 minutes
Profit December	-$8.05	-$1.66	-$0.59	-$0.24	-$0.13
Profit March	-$9.07	-$1.82	-$0.57	-$0.35	-$0.15

As with the simulated trades when just dealing with one security, the bid-ask spreads kill any profit. These trades show that the regressions may yield statistically significant results, but they are not economically significant.

The same simulated trades were attempted using a filter rule similar to the one Fama and Blume used in 1966. A rule was put in place where a trade would only occur when the return from the prior period was big enough. The reasoning here is that when the size of the trigger return is large, the negatively correlated return for the next period should be large enough to have significance over the transaction costs inherent in the bid-ask spread. It is not. Using a series of different filters, the profits over the period would approach and reach zero, but they would never become positive. These are the same results that Fama and Blume came to with their filter rule when they used longer time intervals. This is consistent with the theory of weak form market efficiency. In an efficient market any profits above the expected rate of return will be quickly taken. There can be no easy profits from a simple regression model like this one.

V. Summary and Conclusions

In this paper I wished to test the validity of the random walk model for returns over short time intervals. Other economists have tested for the same thing, but because of the technology available at the time, they were only able to test day to day returns. I looked for significant autocorrelation in the returns of two major market averages and also a portfolio consisting of buying one while shorting the other. I looked at the data over two different four week time spans. I took the tick by tick quotes for the SPY and the QQQ and filtered them to get returns for 5 different time intervals. I wished to see how the randomness of the returns held up under high frequency observations. The empirical results were what I expected. There is statistically significant autocorrelation at high frequencies, but this disperses into randomness as the observations become less frequent. This was the case for the three regressions run, those involving the SPY, those involving the QQQ, and those involving the difference between them. The matter of economic significance is a different story. As a test of the efficiency of the market, I ran simulated trades with the data. The goal was to find a profit from the autocorrelation discovered earlier. This did not happen. The market shows us its efficiency by not allowing an easy profit to be made from these trades.

References

1. Andrew W. Lo and A. Craig MacKinlay. “Stock Market Prices do not Follow Random Walks: Evidence from a Simple Specification Test” *The Review of Financial Studies*,Vol. 1, No. 1 (Spring 1988), 41-66.

2. Jennifer Conrad and Gautam Kaul. “Time-Variation in Expected Returns” *The Journal of Business*, Vol. 61, No. 4 (Oct. 1988), 409-425.

3. Eugene F. Fama and Marshall E. Blume. “Filter Rules and Stock-Market Trading” *The Journal of Business*, Vol. 39, No. 1, Part 2: Supplement on Security Prices (Jan. 1966), 226-241.