Long Term Performance of Leveraged ETFs[1]

Lei Lu
School of Finance
Shanghai University of Finance and Economics

Jun Wang
Department of Economics and Finance
Baruch College

Ge Zhang
School of Business
Long Island University

[1] Lu is from School of Finance, Shanghai University of Finance and Economics, Shanghai, China, 200433, email: lu.lei@mail.shufe.edu.cn. Wang is from Department of Economics and Finance, Zicklin School of Business, Baruch College, One Bernard Baruch Way, Box B10-225, New York, NY 10010, email: jun.wang@baruch.cuny.edu. Zhang is from Long Island University, 700H School of Business, 1 University Plaza, Brooklyn, NY 11201, email: ge.zhang@liu.edu.

Long Term Performance of Leveraged ETFs

ABSTRACT

In this paper, we study leveraged ETFs, in particular, Ultra ETFs and UltraShort ETFs from the ProShares family. These Ultra (UltraShort) ETFs are designed to provide twice (twice the opposite) of the performance of the benchmark on a daily basis. We focus on the relation between long term performance of leveraged ETFs and benchmarks. Our results show that over holding periods no greater than one month, an investor can safely assume that the Ultra (UltraShort) ETF would provide twice the return (twice the negative return) of the underlying benchmark. Over the holding period of one quarter, the UltraShort ETFs can deviate from twice the negative returns of the benchmark. For Ultra ETFs, this deviation occurs when the holding period is one year. Finally, we show that the long term performance of the leveraged ETFs is negatively impacted by the quadratic variation and the auto-variation during the period, with auto-variation being the more dominant factor.

JEL Classifications: G1

Keywords: Leveraged ETFs; Bootstrapping

1. Introduction

Leveraged Exchange Traded Funds (ETFs) are designed to provide more than 100 percent of the exposure of the benchmark. Since July 2006, ProShares has launched a series of Ultra ETFs and UltraShort ETFs. The Ultra ETFs are designed to double the daily returns of the underlying index while the UltraShort ETFs are designed to provide twice the inverse performance of the underlying index on a daily basis. These leveraged ETFs become popular trading instruments and their daily trading volumes are quite high.[1] In November 2008, Direxion even introduces triple leveraged ETFs and these triple ETFs immediately gain acceptance from the market.

Given the large trading volumes of these leveraged ETFs, it is tempting to use them as part of the investment portfolio. While the focus of these ETFs is to double or triple the daily movement of the benchmark, for a long term investor, it is crucial to understand the long term performance of these leveraged ETFs. Do they deliver the same times of the long term performance of the benchmark? If the answer is yes, then an investor may gain the same long term exposure as holding the benchmark by holding a small amount of the leveraged ETFs. And if an investor wants to make a long-term bet against a benchmark, he or she may use the inverse leveraged ETFs without engaging in any short selling activities. However, there is some anecdotal evidence that the long term performance of leveraged ETFs may be problematic. Lauricella writes in his Wall Street Journal article[2]:

> Many of these funds promise to deliver twice the return of an underlying stock or bond index – or move twice as much in the opposite direction. So with the

[1] According to Yahoo Finance, the average daily trading volume of Ultra QQQ and UltraShort QQQ in the three months before the end of 2008 is 41 million and 34 million.
[2] Tom Lauricella, "ETF Math Lesson: Leverage Can Produce Unexpected Return," Wall Street Journal, January 4, 2009.

> Standard & Poor's 500-stock index down 38.5% in 2008, a double-leveraged fund designed to profit when the S&P 500 falls would be up 77%, right? Wrong. The UltraShort S&P500 ProShares rose 61%. Even more confusing, the ProShares fund designed to return twice the opposite of the Dow Jones U.S. Real Estate Index was down 50% for 2008, while the index was also down, by 43%.

Note that this discrepancy in the expectation of the long-term performance of the leveraged ETFs is not a flaw in the design. It is more a result of the misperception of the investors. Actually, ProShares in its own document "Understanding ProShare's Long Term Performance" states that:

> In the end, ProShares ETFs are designed to accomplish their objectives on a daily basis. As a result, you shouldn't expect ProShares to provide 200%, -200% or -100% of index performance over longer periods.

However, the ProShares document and other comments in blogs and newspapers only contain a warning about the long term performance of leveraged ETFs, they do not address the issue of how leveraged ETFs perform against the benchmarks over longer periods since the inception. This paper fills this void. First we use a two-day example to demonstrate the relation between leveraged ETF and the benchmark. We show why the inverse leveraged ETF may deviate more from the benchmark over longer periods than the leveraged ETF. Next we study four common benchmark indices and their corresponding double ETFs and inverse double ETFs from the ProShares family. We investigate the performance over holding periods of two days, one week, one month, one quarter, and one year. Furthermore, we employ stationary bootstrap method to generate more samples of the data and to understand the range of the long term performance relation of the leveraged ETF with the benchmark.

Our results show that over holding periods no greater than one month, an investor can safely assume that the double (inverse double) ETF would provide twice the return (twice the negative return) of the underlying benchmark. As the holding period gets longer, one needs to be cautious about the long term relation. Over the holding period of one quarter, the relation between the inverse double ETF and the benchmark breaks down, that is, the one-quarter performance of the inverse double ETF can be far from twice the negative one-quarter performance of the benchmark. On the other hand, the relation between the double ETF and the benchmark appears to break down at the one-year holding period. Overall our results caution against the use of leveraged ETFs as long term investment substitutes for long or short position of the benchmark indices.

The other contribution we make is to illustrate the factors that contribute to the divergence of long-term return of leveraged ETFs from simple multiples of underlying benchmark. Using a continuous time framework, we show that the divergence is related to the quadratic variation over the holding period. Because the quadratic variation is linearly increasing in the length of the holding period, the long-term performance of both double and double inverse ETFs tend to underperform the simple multiples of underlying benchmark over long holding period. That is, no matter which direction the benchmark index moves, it is not optimal for investors to hold leveraged ETFs over long term. Our empirical analysis also shows that the divergence is also negatively affected by auto-variation, the 1-day auto-covariance of daily returns multiplying by the length of period. For the four benchmark indices we study, auto-variation appears to be a more important factor than quadratic variation to explain the divergence between the long-term leveraged ETF return and the simple multiple of the underlying benchmark.

A number of researchers have studied exchange traded funds. Svetina and Wahal (2008) study the competition among ETFs. Engle and Sarkar (2008) focus on the pricing of ETF with respect to the underlying index to track. Ackert and Tian (2008) compare the performance of U.S. and country ETFs. None of these works studies the leveraged ETFs, which is the focus of this study. Our paper complements these studies by providing a systematic analysis of the long term relation between leverage ETFs and their corresponding benchmark ETFs.

This paper is organized as follows: Section 2 studies the basics of leveraged ETF and the corresponding relation between long-run performances of leveraged ETFs and benchmarks. Section 3 presents the data and results, and Section 4 concludes the paper.

2. Basics

Define r_t^B as daily return of the benchmark index at date *t*. Define r_t^D and r_t^I as daily return of the double ETF and the inverse double ETF at date *t*. If the double ETF and inverse double ETF perform exactly as they are designed, that is, doubling (or inverse doubling) the daily return of the underlying benchmark index, then $r_t^D=2\ r_t^B$, and $r_t^I=-2\ r_t^B$.

Let R_{tn}^B be the *n*-day cumulative return of the benchmark index starting at date *t*. Let R_{tn}^D and R_{tn}^I as *n*-day cumulative return of the double ETF and the inverse double ETF starting at date *t*. Then

$$R_{tn}^B=\prod_{i=0}^{n-1}(1+r_{t+i}^B)\ \ -1,$$

$$R_{tn}^D=\prod_{i=0}^{n-1}(1+r_{t+i}^D)\ \ -1,$$

$$R_{tn}^I=\prod_{i=0}^{n-1}(1+r_{t+i}^I)\ \ -1.$$

To check the relation between the long-term returns of leveraged ETF and the long-term returns of the benchmark index, it is easy to consider two-day returns first.

Given n is two, then $R_{t2}^{B} = (1+r_t^{B})(1+r_{t+1}^{B}) - 1 = r_t^{B} + r_{t+1}^{B} + r_t^{B} r_{t+1}^{B}$. The two-day return of the double ETF is $R_{t2}^{D} = (1+2r_t^{B})(1+2r_{t+1}^{B}) - 1 = 2r_t^{B} + 2r_{t+1}^{B} + 4r_t^{B} r_{t+1}^{B}$, or in terms of R_{t2}^{B},

$$R_{t2}^{D} = 2R_{t2}^{B} + 2r_t^{B} r_{t+1}^{B}.$$

The two-day return of the inverse double ETF is $R_{t2}^{I} = (1-2r_t^{B})(1-2r_{t+1}^{B}) - 1 = -2r_t^{B} - 2r_{t+1}^{B} + 4r_t^{B} r_{t+1}^{B}$, or in terms of R_{t2}^{B},

$$R_{t2}^{I} = -2R_{t2}^{B} + 6r_t^{B} r_{t+1}^{B}.$$

Note that although neither the double ETF nor the inverse double ETF delivers twice (or twice the negative) of the long-term return of the underlying benchmark, the inverse double ETF is further off its target. The two-day return of the double ETF is twice the two-day return of the benchmark plus two times $r_t^{B} r_{t+1}^{B}$, while the two-day return of the inverse double ETF is the twice of the negative two-day return of the benchmark plus six times $r_t^{B} r_{t+1}^{B}$.

In Table 1, we construct six cases of daily returns to see the different effect on the two-day return of the double ETF and the inverse double ETF. In the first two cases, the two-day cumulative return of the benchmark is 20%. In the first case, this 20% return is achieved by two consecutive positive daily returns of 10% and 9.09%, and in the second case, this 20% return is achieved by a positive daily return of 30% followed by a negative return of -7.60%. The two-day cumulative return of the double ETF is 41.82% in the first case and 35.38% in the second case. On the other hand, the two-day cumulative return of the inverse double ETF is -34.55% in the first case and -53.85% in the second case. Since the cumulative return of the benchmark is constructed to be 20%, a naïve investor would consider the two-day return of the double (inverse double) ETF to be 40% (-40%).

However, the actual two-day return of the double ETF and the inverse double ETF is considerably different. Especially in the second case, the inverse double ETF has a two-day return that is -13.85% lower.

In the next two cases, the two-day cumulative return of the benchmark is -20%. In the third case, this -20% return is realized by two consecutive negative daily returns of -10% and -11.11%, and in the fourth case, this -20% return is realized by a negative daily return of -30% followed by a positive return of 14.29%. The two-day cumulative return of the double ETF is -37.78% in the third case and -48.57% in the fourth case. On the other hand, the two-day cumulative return of the inverse double ETF is 46.67% in the third case and 14.29% in the fourth case. Since the cumulative return of the benchmark is constructed to be -20%, a naïve investor would consider the two-day return of the double (inverse double) ETF to be -40% (40%). In these two cases, the actual two-day return of the double ETF and the inverse double ETF is considerably different, especially in the fourth case, where the inverse double ETF has a two-day return that is 25.71% lower than the naïve estimate.

In the last two cases, the two-day cumulative return of the benchmark is zero. In the fifth case, this zero return is realized by consecutive daily returns of -10% and 11.11%, and in the sixth case, this zero return is realized by a daily return of -30% followed by a positive return of 23.08%. Given the two-day cumulative return is zero, a naïve investor would expect the two-day return of the double (inverse double) ETF to be zero. However, the two-day cumulative return of the double ETF is -2.22% in the fifth case and -13.85% in the sixth case. On the other hand, the two-day cumulative return of the inverse double ETF is -6.67% in the fifth case and -41.54% in the sixth case. Again, the cumulative return of inverse double ETF is further away from the naïve estimate.

Although the above derivation works for the two-day return of the double ETF or the inverse double ETF, it also works for the long-term return of two periods as long as the double (inverse double) ETF return in each period doubles (doubles the negative of) the return of the benchmark in the period. Intuitively, if the underlying benchmark has a large negative return over a long period and during this period the benchmark drops to a much lower level initially and rebounds significantly later, the performance of the inverse double ETF would be significantly different from twice the negative return expected by a naïve investor. This effect can explain the performance to Ultra Short Real Estate ETF cited in the introduction to some extent.

Finally, the sequence of the returns in the two periods does not matter. As long as the product of the two returns is negative, the two-day return of the double (inverse double) ETF would be lower than double (negative double) of the return of the underlying benchmark. If the product of the two returns is positive, then the two-day return of the double (inverse double) ETF would be higher than double (negative double) of the return of the underlying benchmark index.

While it is possible to explicitly derive the relation between n-day return of the leveraged ETF and n-day return of the underlying benchmark, the mathematics quickly becomes quite cumbersome and difficult to draw any economic intuition. In the appendix, we provide an analysis in a continuous time framework, and we show that the long term performance of leveraged ETF is negatively correlated with the quadratic variation of underlying index. The bigger the quadratic variation, the more the leveraged ETF tracks below the simple multiple of the underlying index over long term. In the next section, we study the long-term performance of the leveraged ETF empirically.

3. Results

We study four benchmark indices and their corresponding double ETFs and inverse double ETFs. The four benchmark indices are Dow Jones Industrial Average, S&P 500 index, Nasdaq 100 index, and Russell 2000 index. We use the ETFs that track the corresponding indices and have large trading volume as our proxies of benchmarks. The Diamond Trust Series 1 (ticker: DIA), S&P Deposit Receipts (ticker: SPY), PowerShares QQQ (ticker: QQQQ), and iShare Russell 2000 Index (ticker: IWM) measure the returns for holding Dow Jones Industrial Average, S&P 500 index, Nasdaq 100 index, and Russell 2000 index, respectively. According to Engle and Sarkar (2008), domestic ETFs track the underlying indices very well and the premiums/discounts are small and highly transient. Using these ETFs as the benchmarks allows us to make a fair comparison with leveraged ETFs as an investor can potentially trade either the benchmark ETFs or the leveraged ETFs to gain the same desired investment exposure.

The leveraged ETFs in this study are all in the ProShares family. For each benchmark ETF, we have the corresponding double ETF and inverse double ETF. For Dow Jones Industrial Average, the double ETF is ProShares Ultra Dow 30 (ticker: DDM) and the inverse double ETF is ProShares UltraShort Dow 30 (ticker: DXD). For S&P 500 index, the double ETF is ProShares Ultra S&P 500 (ticker: SSO) and the inverse double ETF is ProShares UltraShort S&P 500 (ticker: SDS). For Nasdaq 100 index, the double ETF is ProShares Ultra QQQ (ticker: QLD) and the inverse double ETF is ProShares UltraShort QQQ (ticker: QID). For Russell 2000 index, the double ETF is ProShares

Ultra Russell 2000 (ticker: UWM) and the inverse double ETF is ProShares UltraShort Russell 2000 (ticker: TWM).

We choose these four indices because they are widely used as benchmarks in the asset management and their corresponding double ETFs and inverse double ETFs have huge trading volumes. For example, on December 15, 2008, the last day of our sample period, the trading volumes for Ultra QQQ and UltraShort QQQ are 27.8 million and 19.8 million, respectively.

The data period for S&P 500 index and Nasdaq 100 index starts from July 13, 2006 and the data period of Dow Jones Industrial Average and Russell 2000 index starts from January 25, 2007. The starting points of the study are determined by the introduction of the corresponding double ETF and inverse double ETF (the double and inverse double have different inception dates, while we use the most recent one as the starting date). We end our sample period on December 15, 2008. The daily returns for the benchmark ETFs and leveraged ETFs are obtained from the CRSP up to the end of 2007 and the remaining data are augmented from Yahoo Finance. Table 2 reports some summary statistics for the daily returns.

Next, we regress the daily returns of the leveraged ETFs (both double and inverse double) on the benchmark ETFs and a constant intercept term. This way we can assess whether the leveraged ETFs performs as the fund management company attempts to achieve, i.e., to double the (negative) daily return of the benchmark index for the Ultra (UltraShort) ETFs. From the results in Table 3, Ultra Russell 2000 and UltraShort Russell 2000 track the underlying benchmark very well. The coefficients of the benchmark (2.02 for Ultra Russell 2000 and -2.01 for UltraShort) are not different from two or minus two statistically. For the remaining three groups, although the coefficients on the benchmark are different from two or minus two by any statistical significance, they are actually reasonably close to two or minus two. For double ETFs, Ultra Dow 30 is the worst one with coefficient of 1.78 and followed by Ultra S&P 500 (1.85) and Ultra QQQ (1.95). The same ranking holds also for UltraShort ETFs. In all of these regressions, the intercept coefficients are never significantly different from zero and the R squares are all above 0.94. Overall, these double and inverse double ETFs deliver the promised performance to a large extent.

Leveraged ETF is a great trading instrument for traders. Since long-term investors care about the long-term return and volatility, we need to study these characteristics of leveraged ETFs to assess the suitability to long-term investors. Given that we have a relatively short time series and our goal is to assess the long-term amplification capability of leveraged ETFs, we use overlapping data to calculate long-term returns. For example, when we study the performance of five-day returns, we first calculate the five-day returns of the benchmark, double ETF and inverse double ETF starting from the beginning of the time period. Then we move to one day after the beginning and calculate five-day returns again. These new five-day returns become the second observation of the long-term performance study. Note that these two sets of five-day returns are calculated from the same four days plus a different (first day or) fifth day. This way, we only lose four observations from the daily return series when we study the five-day returns. In this manner, we study the long-term performance for two days, five days, 21 days, 63 days and 252 days. These periods correspond to holding the assets overnight, a week, a month, a quarter and a year, respectively. To show an example of long term returns, Figures 1 and 2 are scatter plots of Ultra QQQ (ticker: QLD) and UltraShort QQQ (ticker: QID) against the benchmark returns (ticker: QQQQ) over different holding periods.

For each holding period, we study the return and volatility of the leveraged ETFs with respect to the benchmark ETFs. In particular, we regress the holding-period returns of Ultra ETF and UltraShort ETF on the returns of benchmark with the same holding-period. Given that there is no prior reason of a constant term and the fact that the constant terms of the one-day regressions are close to zero, we do not include the constants in the regressions.[3] In Table 4, we report the beta coefficients and R-square of these regressions. In addition, we also calculate a measure of volatility effect of the leveraged ETFs. The volatility multiplier (VOLM) is the standard deviation of holding-period returns of the leveraged ETF divided by the standard deviation of the holding-period returns of the underlying benchmark ETF. If this volatility multiplier is two, then investing in the leveraged ETF provides investors exactly twice of the volatility as investing in the underlying benchmark. Otherwise, the volatility exposure would be different. In this way we can compare the risk of investing in the leveraged ETF relative to investing in the underlying benchmark.

Table 4 demonstrates that, for Dow Jones Industrial Average and S&P 500, the leveraged ETFs behave very close to double (or double the negative of) the underlying benchmark holding-period returns when the holding-period is two days or five days. Starting at the one month period, the UltraShort ETFs move away from doubling the negative of the underlying benchmark by a large margin, while this occurs at one year holding period for the Ultra ETFs. The volatility multiplier shows the same pattern as the beta coefficient.

[3] The inclusion of a constant term does not change our results qualitatively.

For Nasdaq 100 and Russell 2000 indices, the leveraged ETFs tend to double (or double the negative of) the underlying benchmark holding-period returns for periods of two days, five days, one month and one quarter. It is only at the one-year holding period that the Ultra and UltraShort ETFs do not track the underlying benchmark proportionally. Consistent with our analysis in the previous section, double inverse ETFs deviate further away from doubling the negative of benchmark returns as the holding period gets longer.

The results in Table 4 apply to the data period in this study. The fact that we use overlapping long-term returns to investigate the relation between the long-term performance of leveraged ETFs and benchmarks may lead to some bias. To further assess this relation, we adopt the bootstrapping method. The bootstrapping method has been widely used in finance as a powerful tool to resample the distribution of the actual data.[4] Politis and Romano (1994) propose the stationary bootstrapping which is applicable to strictly stationary and weakly dependent time series to generate the stationary pseudo-time series. The algorithm of stationary bootstrapping is based on resampling the overlapping blocks with random length, where the length of each block follows a geometric distribution with probability, p. This means that the average length of these blocks is $1/p$. A large value of p is appropriate for data with weak dependence, while a smaller value of p is appropriate for data with more dependence.

Sullivan, Timmermann and White (1999) employ this method to evaluate technical trading rules, and choose 0.1 for the value of p. Leger, Politis and Romano (1992) discuss the choice of p for the stationary bootstrap and mimic the actual data with

[4] See Sullivan, Timmermann and White (1999); Byun and Rozeff (2003); Wang (2003); Kosowski, Timmermann, Wermers, and White (2006), etc.

500 replications and p=0.05. In our paper, we follow the algorithm of Politis and Romano (1994) to set p=0.05, which corresponds to the average block length of 20.

Each time we create a sample of daily returns of the benchmark and the corresponding leveraged ETFs, we compute the beta coefficients and the volatility multipliers for different holding periods the same way as the original sample. Altogether we generate 1,000 samples from bootstrapping and we have a distribution of betas and volatility multipliers. In Table 5, we report the mean and the 95 percentile confidence interval for the beta coefficients.[5] From the table, we can see that investors can not reasonably expect to double the returns of the benchmark over any periods longer than a quarter with double ETFs. For example, the 95 percentile confidence interval for beta between Ultra Dow 30 and Dow Jones Industrial Average at one year holding period is from 0.9 to 2.6. This result means that, if an investor holds Ultra Dow 30 for a year, he or she may have 95 percent probability to gain a return equal to somewhere from 0.9 times to 2.6 times of the Dow Jones return over the same year. The same pattern holds for the other Ultra ETFs.

For UltraShort ETFs, the breakdown occurs at shorter holding period. At holding period of one quarter, UltraShort Dow 30 may generate somewhere from -2.2 to -1.2 times of the Dow Jones return over the same quarter. For one year holding period, the range becomes much wider. Overall, the wider ranges for betas of UltraShort ETFs are consistent with the simple two-period results in the previous section. In general, all the results demonstrate that the long-term performance of leveraged ETFs do not follow the long-term performance of benchmarks in the same way as the daily returns do.

[5] Bootstrap results for volatility multiplier have the similar pattern as results for betas. Details are available from the authors upon request.

From the continuous time analysis in Appendix, it appears the long term performance of double or inverse double ETFs is negatively related to the quadratic variation (σ^2 T). We study this effect by regressing the n-day return of the leveraged ETF on the n-day return of the benchmark and the quadratic variation. We compute the quadratic variation by multiplying the variance estimate from the daily returns in the n-day period by the number of days, n. Table 6 reports these regression results. Clearly, the coefficients of quadratic variation are all negative and significant. That is, the larger the volatility of the underlying benchmark, the more likely that the leveraged ETFs would underperform the simple multiple of the underlying indices.

Finally, considering our two-period examples, we conjecture that auto-variation should also contribute to the long-term performance. To study this effect, we regress the n-day return of the leveraged ETF on the n-day return of the benchmark, the quadratic variation, and the auto-variation. We compute the auto-variation by multiplying the 1-day auto-covariance estimate from the daily returns in the n-day period by the number of days, n. Table 7 reports these results. We find that the coefficients of auto-variation are all negative and highly significant. In addition, the impact of auto-variation appears to be stronger that that of quadratic variation. With the inclusion of auto-variation, the beta coefficients on the underlying benchmark become closer to the expected value, two or minus two. Overall, auto-variation is a significant contribution to the long-term performance of leveraged ETFs. Note that both quadratic variation and auto-variation are the product of variance or auto-variance and the number of days in the period. For any stationary process, the variance or auto-variance may be quite stable. The longer the holding period, the greater the number of days. Hence leveraged ETFs, no matter on the long side or the short side, tend to lose out to the simple multiple of the underlying benchmark over the long run, because both variance and auto-variance are positive.

4. Conclusion

In this paper, we study leveraged ETFs that provide more than 100 percent exposure in either long or short direction of underlying benchmark. In particular, we study the long term performance of four groups of Ultra ETFs and UltraShort ETFs from the ProShares family. These Ultra (UltraShort) ETFs are designed to provide twice (twice the opposite of) the performance of the benchmark on a daily basis. Over longer holding periods, the same relation between leveraged ETFs and the benchmark do not necessarily hold. First we use a two-day example to demonstrate the relation between leveraged ETFs and the benchmark. We show why the UltraShort ETF may deviate more from the benchmark over longer periods than the Ultra ETF. Next we investigate the performance of these four groups over holding periods of two days, one week, one month, one quarter, and one year. Furthermore, we employ stationary bootstrap method to generate more samples of the data and to understand the parameter range of the long term performance relation of the leveraged ETFs.

Our results show that over holding periods of one month or less, an investor can safely assume that the double (inverse double) ETF would provide twice the return (twice the negative return) of the underlying benchmark. As the holding period gets longer, one needs to be cautious about the long term relation. Over the holding period of one quarter, the relation between the inverse double ETF and the benchmark breaks down, that is, the one-quarter performance of the inverse double ETF can be far from twice the negative one-quarter performance of the benchmark. On the other hand, the relation between the double ETF and the benchmark appears to break down at the one-year holding period. Overall our results show that leveraged ETFs are not long term substitutes for long or short position of the benchmark indices. Finally, we show that the long term performance of the leveraged ETFs is negatively impacted by the quadratic variation and the auto-variation during the period, with auto-variation being the more dominant factor.

While we only study four groups of benchmark and leveraged ETFs, our method can be easily applied to other leveraged ETFs to assess at what holding period an investor can use the leveraged ETF as a substitute for long or short position of the benchmark. The other interesting research avenue is to investigate the trading of these leveraged ETFs and their contribution to the overall market dynamics.

Appendix

In this appendix, we study the long term performance of leveraged ETFs in the continuous time framework. Let S be the stock index price, then the standard geometric Brownian motion assumption gives,

$$dS/S = \mu\, dt + \sigma\, dW$$

where μ is the instantaneous return, σ is the volatility and W is a standardized Brownian motion. Then the expected continuously compounded long term return between t=0 and T can be written as $R_T^B = \mu T - \sigma^2 T/2$. If the leveraged ETF tracks the underlying index continuously and perfectly. Let Z be the leveraged ETF price and the multiple over the underlying index be N. Then,

$$dZ/Z = N\, dS/S = N\mu\, dt + N\sigma\, dW.$$

Then the expected continuously compounded long term return of the leveraged ETF between t=0 and T can be written as $R_T^L = N\mu T - N^2\sigma^2 T/2$. Comparing R_T^L to the N multiple of R_T^B, we have,

$$R_T^L = N\, R_T^B - [(N^2 - N)/2]\ \sigma^2 T$$

Note the difference between the long term return of leveraged ETF and the N multiple of the underlying index is $-[(N^2 - N)/2]\ \sigma^2 T$, where $\sigma^2 T$ is the quadratic variation of the Brownian process between 0 and T, and the coefficient $-[(N^2 - N)/2]$ is always negative for leveraged ETFs. If N is 2, then the coefficient is -1; if N is -2, then the coefficient is -3. Again the inverse leveraged ETF deviates more from the N multiple of the underlying index. In the same way, we can derive the coefficients for triple leveraged ETF and triple inversely ETF.

References

Ackert, L. and Y. Tian, 2008, “Arbitrage, Liquidity, and the Valuation of Exchange Traded Funds,” *Financial Markets, Institutions & Instruments,* Vol. 17, Issue 5, pp. 331-362.

Byun, J. and M. S. Rozeff, 2003, “Long-run Performance after Stock Splits: 1927 to 1996”, *Journal of Finance* 58, 1063-1085.

Engle, R. and D. Sarkar, 2008, “Pricing Exchange Traded Funds”, working paper, NYU.

Kosowski, R., A. Timmermann, R. Wermers, and H. White, 2006, “Can Mutual Fund “Stars” Really Pick Stocks? New Evidence from a Bootstrap Analysis”, *Journal of Finance* 61, 2551-2595.

Leger, C., Politis, D. and J. Romano, 1992, “”Bootstrap Technology and Applications”, *Technometrics* 34, 378-398.

Politis, D. and J. Romano, 1994, “The Stationary Bootstrap”, *Journal of the American Statistical Association* 89, 1303-1313

Sullivan, R., A. Timmermann, and H. White, 1999, “Data-Snooping, Technical Trading Rule Performance, and Bootstrap”, *Journal of Finance* 54, 1647-1691.

Svetina, M. and S. Wahal, 2008, “Exchange Traded Funds: Performance and Competition,” working paper, Arizona State University.

Wang, K., 2003, “Asset Pricing with Conditioning Information: A New Test”, *Journal of Finance* 58, 161-196

Table 1 Hypothetical two-day performance of double ETF and inverse double ETF versus the benchmark

This table provides four hypothetical scenarios. r_t^B and r_{t+1}^B are daily returns of the benchmark at date t and $t+1$. R_{t2}^B is the two-day return of the benchmark for the period, t and $t+1$. The daily return of the double ETF is twice the daily return of the benchmark. R_{t2}^D is the two-day return of the double ETF for the period, t and $t+1$. The daily return of the inverse double ETF is twice the negative daily return of the benchmark. R_{t2}^I is the two-day return of the inverse double ETF for the period, t and $t+1$. $r_t^B\ r_{t+1}^B$ is the cross product of r_t^B and r_{t+1}^B.

Case	r_t^B	r_{t+1}^B	R_{t2}^B	R_{t2}^D	R_{t2}^I	$r_t^B\ r_{t+1}^B$
1	10.00%	9.09%	20.00%	41.82%	-34.55%	0.91%
2	30.00%	-7.69%	20.00%	35.38%	-53.85%	-2.31%
3	-10.00%	-11.11%	-20.00%	-37.78%	46.67%	1.11%
4	-30.00%	14.29%	-20.00%	-48.57%	14.29%	-4.29%
5	-10.00%	11.11%	0.00%	-2.22%	-6.67%	-1.11%
6	30.00%	-23.08%	0.00%	-13.85%	-41.54%	-6.92%

Table 2 Summary statistics of daily returns

This table has four panels, corresponding to four benchmark indices. Each panel reports summary statistics of daily returns for the benchmark ETF (the first row), the double ETF (the second row), and the inverse double ETF (the third row).

Ticker	Mean	Std Dev	Skewness	Kurtosis	Max	Min
Panel A: Dow Jones Industrial Average (Jan. 25, 2007 - Dec. 15, 2008)						
DIA	-0.0002	0.0170	1.0597	19.2391	0.1356	-0.0939
DDM	-0.0008	0.0312	0.1367	12.5839	0.2259	-0.1597
DXD	0.0005	0.0314	-0.3843	12.8628	0.1682	-0.2221
Panel B: S&P 500 index (July. 13, 2006 - Dec. 15, 2008)						
SPY	-0.0004	0.0179	0.5061	17.2471	0.1452	-0.0984
SSO	-0.0011	0.0336	0.1512	13.1391	0.2241	-0.1729
SDS	0.0009	0.0341	-0.0992	12.4218	0.1807	-0.2244
Panel C: NASDAQ 100 Index (July. 13, 2006 - Dec. 15, 2008)						
QQQQ	-0.0002	0.0180	0.2767	10.8406	0.1216	-0.0896
QLD	-0.0008	0.0359	0.1771	11.5456	0.2459	-0.1930
QID	0.0005	0.0359	-0.1243	11.0054	0.1995	-0.2298
Panel D: Russell 2000 index (Jan. 25, 2007 - Dec. 15, 2008)						
IWM	-0.0009	0.0226	-0.2907	6.6535	0.0864	-0.1124
UWM	-0.0019	0.0462	-0.2389	7.3364	0.2120	-0.2394
TWM	0.0019	0.0461	0.2786	7.4551	0.2466	-0.2097

Table 3 Regression results of daily returns of double ETF and inverse double ETF on benchmark ETF daily returns

This table has four panels, corresponding to four benchmark indices. In each panel, the first row is the regression of daily returns of double ETF on the daily returns of the benchmark ETF and a constant, and the second row is the regression of daily returns of inverse double ETF on the daily returns of the benchmark ETF and a constant. T-statistics are in parentheses. T-statistics for Intercept are testing whether the constant term is different from zero. T-statistics for Benchmark are testing whether the coefficient is different from two (double ETF) or minus two (inverse double ETF).

Ultra ETF	Intercept	Benchmark	R^2
Panel A: Dow Jones Industrial Average			
DDM	-0.0004 (-1.3050)	1.7778 (-12.2930)	0.9410
DXD	-0.00005 (-0.1879)	-1.8143 (12.8197)	0.9628
Panel B: S&P 500 index			
SSO	-0.0004 (-1.5790)	1.8515 (-10.9197)	0.9683
SDS	0.0001 (0.5682)	-1.8807 (9.2839)	0.9725
Panel C: NASDAQ 100 Index			
QLD	-0.0004 (-1.4826)	1.9488 (-3.1972)	0.9607
QID	0.0002 (0.5664)	-1.9526 (3.0820)	0.9637
Panel D: Russell 2000 index			
UWM	-0.0002 (-0.4973)	2.0216 (1.5362)	0.9776
TWM	0.0002 (0.5298)	-2.0174 (-1.3029)	0.9798

Table 4 Relation between long-term performance of leveraged ETF and long-term performance of benchmark ETF.

This table has four panels, corresponding to four benchmark indices. In each panel, the first (second) half is the regression of n-day returns of the (inverse) double ETF on n-day returns of the benchmark ETF and no constant. T-statistics for testing whether the coefficient is different from (minus) two are in parentheses. VOLM is the ratio between the standard deviation of n-day returns of the (inverse) double ETF and the standard deviation of n-day returns of the benchmark ETF.

		2 days	**5 days**	**21 days**	**63 days**	**252 days**
			Panel A: Dow Jones Industrial Average			
DDM	Beta	1.8561	1.9107	1.9518	1.9592	1.7652
		(-9.8164)	(-6.0116)	(-3.0422)	(-2.6378)	(-9.3344)
	R^2	0.9635	0.9648	0.9628	0.9671	0.9325
	VOLM	1.8898	1.9409	1.9622	1.9366	1.6650
DXD	Beta	-1.8807	-1.9843	-1.7215	-1.4289	-1.6607
		(9.2552)	(0.9463)	(11.5786)	(26.5485)	(16.5150)
	R^2	0.9723	0.9596	0.8971	0.8900	0.9483
	VOLM	1.9075	2.0274	1.8324	1.5315	1.7056
			Panel B: S&P 500 index			
SSO	Beta	1.8732	1.9367	1.9602	1.9148	1.8274
		(-10.8147)	(-5.7056)	(-3.4253)	(-7.8581)	(-7.5289)
	R^2	0.9768	0.9806	0.9798	0.9828	0.9468
	VOLM	1.8939	1.9504	1.9522	1.8808	1.6697
SDS	Beta	-1.9069	-2.0015	-1.7486	-1.5931	-1.8921
		(7.3694)	(-0.0819)	(11.0528)	(22.7831)	(4.8205)
	R^2	0.9741	0.9506	0.9096	0.9358	0.9524
	VOLM	1.9321	2.0554	1.8508	1.6550	1.9278
			Panel C: NASDAQ 100 Index			
QLD	Beta	1.9660	1.9691	1.9701	1.8641	1.6399
		(-2.5516)	(-2.6801)	(-2.7389)	(-11.3643)	(-13.7415)
	R^2	0.9730	0.9798	0.9823	0.9780	0.9165
	VOLM	1.9923	1.9861	1.9744	1.8694	1.6779
QID	Beta	-1.9874	-2.0255	-1.9532	-2.0565	-1.8699
		(0.9816)	(-1.6224)	(2.3722)	(-3.0045)	(3.9216)
	R^2	0.9752	0.9650	0.9435	0.9564	0.8990
	VOLM	2.0124	2.0622	2.0125	2.0924	1.9749
			Panel D: Russell 2000 index			
UWM	Beta	2.0418	2.0065	1.9448	1.8603	1.9714
		(3.5283)	(0.6826)	(-5.8299)	(-13.4376)	(-1.2632)
	R^2	0.9843	0.9896	0.9894	0.9873	0.9714
	VOLM	2.0573	2.0135	1.9260	1.7956	1.4557
TWM	Beta	-2.0270	-2.0999	-2.0498	-1.9898	-1.6815
		(-2.0416)	(-4.7302)	(-1.6966)	(0.4118)	(8.6165)
	R^2	0.9802	0.9547	0.9148	0.9397	0.9029
	VOLM	2.0476	2.1524	2.1839	2.1258	2.3277

Table 5 Stationary bootstrap results for Beta

This table has four panels, corresponding to four benchmark indices. For each index, we generate 1000 samples using stationary bootstrap with probability of 0.05 or averaged length = 20. For each sample, we estimate the beta, the coefficient between n-day returns of the (inverse) double ETF and the n-day returns of the benchmark ETF. The mean, 2.5 percentile, and 97.5 percentile of the 1000 betas are reported. The regressions for one-day returns include the constant term, while the regressions for n-day *(n>1)* returns do not have the constant term.

		1 day	**2 days**	**5 days**	**21 days**	**63 days**	**252 days**
			Panel A: Dow Jones Industrial Average				
DDM	Mean	1.8055	1.8658	1.9097	1.9326	1.9407	1.9114
	2.5 Percentile	1.6754	1.7717	1.8309	1.7925	1.6944	0.9182
	97.5 Percentile	1.9704	1.9687	1.9945	2.0291	2.1415	2.6765
DXD	Mean	-1.8390	-1.8887	-1.9674	-1.8474	-1.7078	-1.4620
	2.5 Percentile	-1.9878	-1.9833	-2.0374	-2.0889	-2.1747	-2.6469
	97.5 Percentile	-1.7104	-1.8025	-1.8836	-1.6297	-1.2069	-0.0737
			Panel B: S&P 500 index				
SSO	Mean	1.8675	1.8856	1.9353	1.9438	1.9431	1.9118
	2.5 Percentile	1.7769	1.8116	1.8794	1.8470	1.7450	1.1539
	97.5 Percentile	1.9917	1.9945	2.0035	2.0377	2.1445	2.5524
SDS	Mean	-1.8955	-1.9155	-1.9865	-1.8636	-1.7659	-1.6372
	2.5 Percentile	-2.0057	-2.0035	-2.0801	-2.1337	-2.2964	-2.9675
	97.5 Percentile	-1.8120	-1.8590	-1.8926	-1.6569	-1.3051	-0.3785
			Panel C: NASDAQ 100 Index				
QLD	Mean	1.9426	1.9603	1.9689	1.9700	1.9397	1.8149
	2.5 Percentile	1.8747	1.9018	1.9280	1.9055	1.7797	0.9930
	97.5 Percentile	1.9976	2.0050	2.0112	2.0405	2.1076	2.5142
QID	Mean	-1.9498	-1.9800	-2.0097	-1.9674	-1.9140	-1.7229
	2.5 Percentile	-1.9965	-2.0243	-2.0820	-2.1698	-2.4980	-3.3423
	97.5 Percentile	-1.8940	-1.9235	-1.9375	-1.7796	-1.5012	-0.5867
			Panel D: Russell 2000 Index				
UWM	Mean	2.0184	2.0378	2.0055	1.9540	1.9342	1.8880
	2.5 Percentile	1.9741	1.9896	1.9611	1.8747	1.7471	1.3292
	97.5 Percentile	2.0551	2.0818	2.0493	2.0442	2.1464	2.5779
TWM	Mean	-2.0144	-2.0236	-2.0807	-2.0839	-2.0160	-1.8542
	2.5 Percentile	-2.0414	-2.0618	-2.2222	-2.4450	-2.8300	-4.1790
	97.5 Percentile	-1.9772	-1.9789	-1.9505	-1.8081	-1.3801	0.1970

Table 6 Regression results of long-term performance of leveraged ETF on long-term performance of benchmark ETF and quadratic variation.

This table has four panels, corresponding to four benchmark indices. In each panel, the first (second) half is the regression of n-day returns of the (inverse) double ETF on n-day returns of the benchmark ETF (Beta), n-day quadratic variation (QV), and no constant. Quadratic variation is the product of the variance of daily returns in the n-day period and n. T-statistics are in parentheses. For Beta, the test is whether the coefficient is different from (minus) two, and for QV, the test is whether the coefficient is different from zero.

		2 days	5 days	21 days	63 days	252 days
Panel A: Dow Jones Industrial Average						
DDM	Beta	1.8655	1.9244	1.9935	2.0121	1.9144
		(-9.0406)	(-5.2508)	(-0.4201)	(0.7413)	(-11.4347)
	QV	-0.929	-0.8184	-0.7254	-0.7047	-1.1517
		(-4.8804)	(-8.1137)	(-14.7283)	(-18.9124)	(-81.0794)
	R^2	0.9638	0.9697	0.9761	0.9825	0.9971
DXD	Beta	-1.8832	-1.9214	-1.681	-1.4203	-1.3956
		(8.7568)	(5.4095)	(15.4987)	(24.0016)	(38.7257)
	QV	-1.6174	-1.6989	-1.364	-0.8162	-0.5016
		(-9.4777)	(-16.6873)	(-20.7571)	(-14.8648)	(-16.9368)
	R^2	0.9706	0.9669	0.9233	0.8874	0.9616
Panel B: S&P 500 index						
SSO	Beta	1.8783	1.9589	2.0384	2.0811	2.0149
		(-10.4821)	(-3.9476)	(3.5785)	(7.5731)	(2.4159)
	QV	-0.937	-0.7907	-0.6708	-0.5781	-0.9721
		(-6.4114)	(-11.1236)	(-19.0691)	(-22.0380)	(-83.7777)
	R^2	0.9781	0.9848	0.9895	0.9932	0.9984
SDS	Beta	-1.9037	-1.9253	-1.6297	-1.3612	-1.3299
		(7.2625)	(4.9027)	(18.2910)	(30.4415)	(46.6243)
	QV	-1.6338	-1.6914	-1.1746	-0.5649	-0.2308
		(-9.7860)	(-16.2608)	(-17.7080)	(-10.9890)	(-8.5360)
	R^2	0.9716	0.9642	0.9257	0.9185	0.9711
Panel C: NASDAQ 100 Index						
QLD	Beta	1.9779	1.9962	2.0564	2.0579	1.9892
		(-1.6331)	(-0.3294)	(5.7221)	(6.4243)	(-2.2130)
	QV	-1.0528	-1.0202	-0.9084	-0.95	-1.425
		(-5.1450)	(-10.0692)	(-19.7992)	(-29.3026)	(- 133.8263)
	R^2	0.9732	0.9818	0.9898	0.9929	0.9984
QID	Beta	-1.9782	-1.9622	-1.7752	-1.6378	-1.5222
		(1.5297)	(2.7742)	(14.4014)	(29.0221)	(56.4262)
	QV	-1.8383	-1.8198	-1.4608	-0.9863	-0.85
		(8.5254)	(-15.1218)	(-20.1155)	(-21.9691)	(-46.1359)
	R^2	0.9697	0.9721	0.9593	0.9725	0.9891

Panel D: Russell 2000 index						
UWM	Beta	2.0505	2.0422	2.0911	2.1637	2.1863
		(4.2553)	(4.9837)	(12.4860)	(26.5298)	(20.4034)
	QV	-0.6606	-0.6854	-0.5836	-0.4437	-0.5733
		(-4.8202)	(-11.7226)	(-23.8258)	(-30.0836)	(-39.2599)
	R^2	0.9849	0.9926	0.9963	0.9983	0.9992
TWM	Beta	-2.0207	-2.0171	-1.7808	-1.5129	-1.2439
		(-1.5085)	(-1.2711)	(11.7667)	(26.9515)	(22.1377)
	QV	-1.8013	-1.8007	-1.3576	-0.7436	-0.3143
		(-11.3748)	(-19.3389)	(-21.7004)	(-17.2114)	(-5.7544)
	R^2	0.9789	0.9797	0.9576	0.9591	0.9428

Table 7 Regression results of long-term performance of leveraged ETF on long-term performance of benchmark ETF, quadratic variation and 1-day auto-variation. This table has four panels, corresponding to four benchmark indices. In each panel, the first (second) half is the regression of n-day returns of the (inverse) double ETF on n-day returns of the benchmark ETF (Beta), n-day quadratic variation (QV), n-day auto-variation (AutoV), and no constant. Quadratic variation is the product of the variance of daily returns in the n-day period and n. Auto-variation is the product of 1-day auto-covariance of daily returns in the n-day period and n. T-statistics are in parentheses. For Beta, the test is whether the coefficient is different from (minus) two, and for QV and AutoV, the test is whether the coefficient is different from zero.

		5 days	**21 days**	**63 days**	**252 days**
		Panel A: Dow Jones Industrial Average			
DDM	Beta	1.9373	1.9594	1.943	1.7775
		(-4.4313)	(-2.6371)	(-2.8115)	(-34.1693)
	QV	-0.4204	-0.4235	-0.5558	-0.8417
		(-3.5675)	(-6.8815)	(-12.3076)	(-61.0332)
	AutoV	-0.5723	-0.5618	-0.4805	-1.1051
		(-6.0776)	(-7.6058)	(-5.5373)	(-27.6194)
	R^2	0.9714	0.9782	0.9835	0.9991
DXD	Beta	-1.897	-1.7977	-1.8322	-1.7277
		(7.8413)	(14.6860)	(15.3498)	(41.0673)
	QV	-0.947	-0.3313	0.0709	0.2504
		(-8.6543)	(-6.0178)	(2.9100)	(17.8337)
	AutoV	-1.0813	-1.9219	-2.8628	-2.6808
		(-12.3667)	(-29.0849)	(-61.1441)	(-65.7938)
	R^2	0.9736	0.9686	0.9857	0.9971
		Panel B: S&P 500 index			
SSO	Beta	1.965	1.9781	1.9929	1.8688
		(-3.4790)	(-2.2063)	(-0.4951)	(-23.2658)
	QV	-0.4703	-0.2926	-0.4287	-0.7002
		(5.6700)	(-7.5390)	(-14.2502)	(-66.4782)
	AutoV	-0.4717	-0.7921	-0.5761	-1.0671
		(-6.8549)	(-15.2173)	(-8.6447)	(-31.3677)
	R^2	0.9859	0.9925	0.9941	0.9996
SDS	Beta	-1.907	-1.7945	-1.8215	-1.7031
		(7.3333)	(14.6110)	(15.7971)	(34.0344)
	QV	-0.7179	-0.1411	0.2151	0.4638
		(-6.8693)	(-2.5652)	(9.0609)	(28.4654)
	AutoV	-1.4331	-2.1645	-3.0071	-2.7256
		(-16.5305)	(-29.3349)	(-57.1760)	(-51.7900)
	R^2	0.9754	0.9699	0.9884	0.9966

Table 7 (continued)

		5 days	21 days	63 days	252 days
		Panel C: NASDAQ 100 Index			
QLD	Beta	1.9984	2.0307	2.0415	1.8663
		(-0.1408)	(2.8981)	(3.1785)	(-14.0250)
	QV	-0.7499	-0.6181	-0.8586	-0.8208
		(5.9269)	(-9.1669)	(-13.8890)	(-18.8377)
	AutoV	-0.4056	-0.5527	-0.2201	-1.327
		(-3.5092)	(-5.7493)	(-1.7349)	(-14.1388)
	R^2	0.9822	0.9904	0.9929	0.999
QID	Beta	-1.9494	-1.9367	-1.925	-1.8399
		(5.2157)	(9.1998)	(11.9255)	(17.5717)
	QV	-0.2256	0.3623	0.6131	0.7121
		(-2.0906)	(8.2801)	(20.5807)	(17.1033)
	AutoV	-2.3921	-3.4708	-3.8524	-3.4308
		(-24.2634)	(-55.6343)	(-63.0137)	(-38.2593)
	R^2	0.9859	0.9935	0.9967	0.9979
		Panel D: Russell 2000 index			
UWM	Beta	2.043	2.0315	2.0497	2.0385
		(5.2235)	(5.0720)	(6.0529)	(2.7571)
	QV	-0.4052	-0.167	-0.2076	-0.3879
		(-5.2278)	(-5.9409)	(-11.5734)	(-20.5961)
	AutoV	-0.4447	-0.8789	-0.8005	-0.7476
		(-5.3168)	(-19.3855)	(-16.9708)	(-12.2905)
	R^2	0.993	0.998	0.999	0.9995
TWM	Beta	-2.0131	-1.9936	-1.9856	-2.0306
		(-1.4838)	(0.8936)	(1.1043)	(-1.0708)
	QV	-0.385	0.1303	0.2359	0.672
		(-4.6393)	(4.0072)	(8.2960)	(17.4369)
	AutoV	-2.247	-3.1388	-3.3205	-3.9769
		(-25.0931)	(-59.8527)	(-44.3921)	(-31.9504)
	R^2	0.9913	0.9953	0.993	0.9898

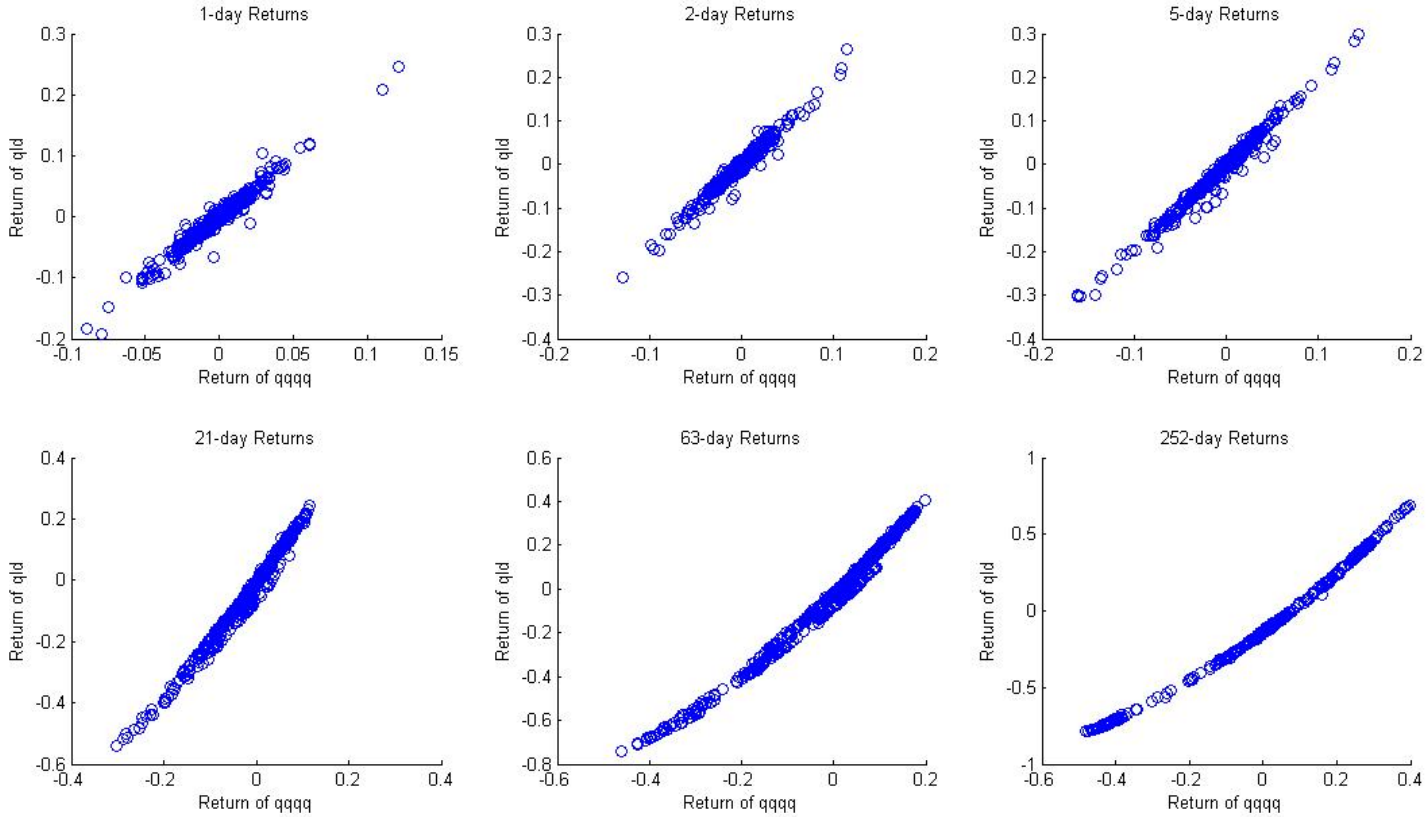

Figure 1 Returns of Ultra QQQ (Ticker: QLD) versus returns of PowerShares QQQ (Ticker: QQQQ) over different holding periods. The sample period is from July 13, 2006 to December 15, 2008. The multi-day returns are all calculated using overlapped data.

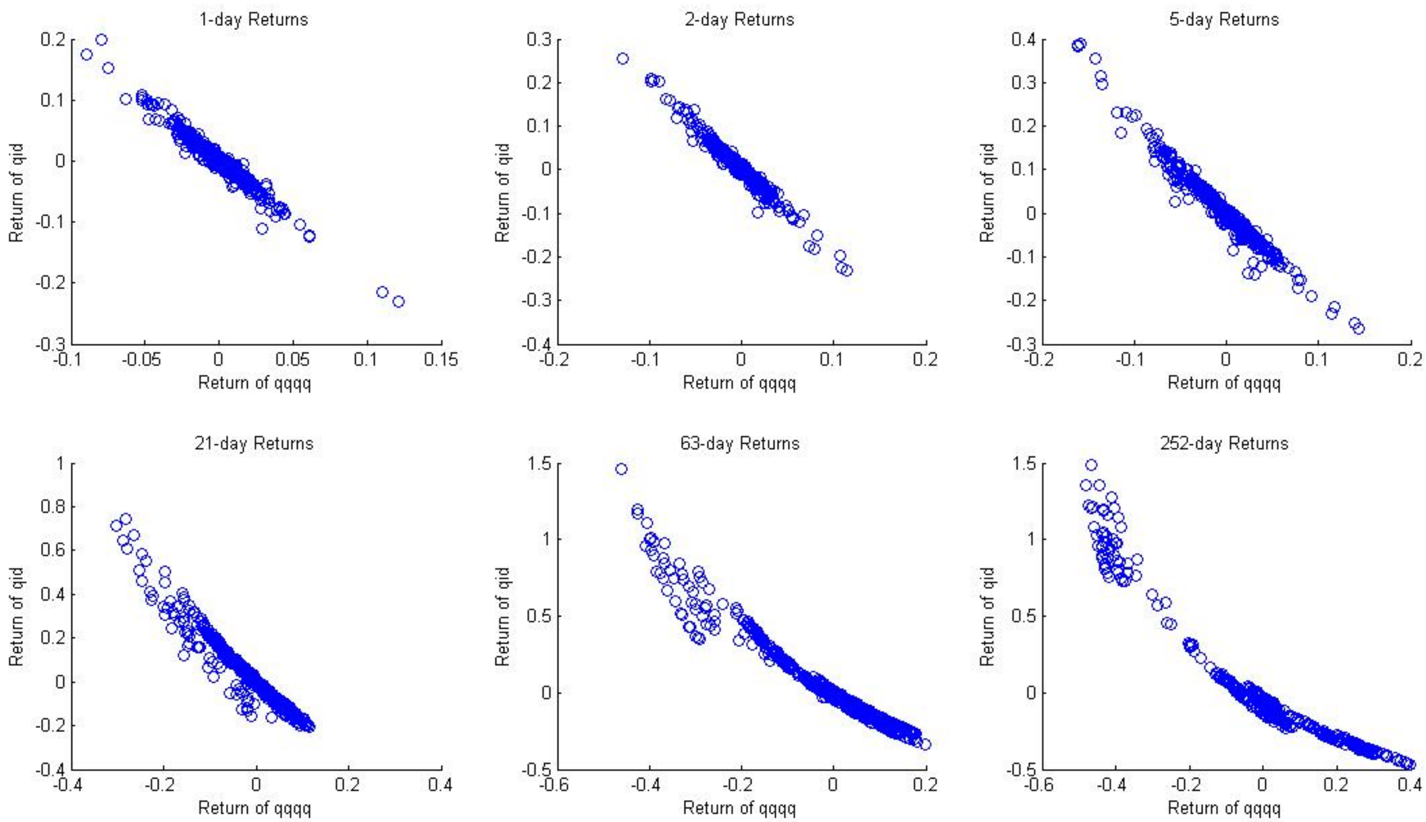

Figure 2 Returns of UltraShort QQQ (Ticker: QID) versus returns of PowerShares QQQ (Ticker: QQQQ) over different holding periods. The sample period is from July 13, 2006 to December 15, 2008. The multi-day returns are all calculated using overlapped data.