Accounting for Volatility Decay in Time Series Models for Leveraged Exchange Traded Funds*

Ahmed Abdou
Columbia University
ahmed.abdou@columbia.edu

June 15, 2017

Abstract

Leverage Exchange Traded Funds (LETF's) returns tend to deviate from their underlying assets' multiple returns as their holding period increase, a phenomenon known as volatility decay. Algebraically, it is shown that volatility decay is intensified for inverse leveraged funds and as the leverage multiplier increases. The paper uses a novel approach to account for volatility decay. The ARIMA model ability to forecast future returns is tested for three major indexes and is shown to provide more accurate estimates for S&P500. The returns of S&P500 and its corresponding LETFs are fitted to an Autoregressive Integrated Moving Average (ARIMA) model. Theoretically, the constant of the ARIMA model and the variance of their Gaussian errors captures the volatility decay effect. Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models provide more flexibility in modeling conditional variance that is non-stationary. The theoretical results are verified empirically, and the constant of the fitted model captures the intensity of the decay and its direction.

*Second draft. First draft: November 28, 2016

1 Introduction

Exchange Traded Funds (ETFs) provide exposure to a commodity, bond, or index that is not easily accessible to investors. The market for ETFs has expanded significantly since ETFs appeared 23 years ago. The global ETF assets under management doubled from \$1,463 billion in December 2010 to \$2,959 billion in December 2015. The ETF market is expected to continue to grow over the next 5 years reaching \$7 trillion dollars by 2020. The US has the largest share of this market, which is still growing with a 23% increase expected over the next 5 years[1].

Leveraged ETFs provide a multiple of the daily return of the ETFs, making them more attractive to investors aiming to multiply their returns. LETFs rebalance daily, providing a multiple of the daily return of the index being tracked, but not a mutliple of the cumulative return if held for more than one day, a phenomenon known as volatility decay. Therefore, LETFs' returns are largely affected by the holding period and the volatility of the underlying asset. Bruno et al. (2014) showed that the longer the holding period and the higher the volatility of the underlying asset, the further the LETF is from its target return . However, LETFs can still be held short term, especially in periods of small volatlility and large momentum, or they can be used to construct static portfolios to take advantage of the volatility decay (see Leung and Santoli (2012, 2016)).

The inability of LETFs to track their underlying asset cumulative return is attributed to their daily rebalancing. Consequences of daily rebalancing have been intensively studied in the literature. Cheng and Madhavan (2009) suggested a unified framework explaining the return dynamics of leveraged and inverse ETFs. Tang and Xu (2013) attempts at explaining deviations that remain after netting out the compounding effect. Guo and Leung (2015) and Leung and Ward (2015) studied the price dynamics of LETFs based on commodities, ranging from oil and gas to gold and silver. Leung and Park (2017) examined the long-term growth rates of LETFs under many different models, including those with stochastic volatility and stochastic interest rate. Bruno et al. (2014) proposed constructing LETFs that balance monthly to mitigate the volatility decay effect caused by daily rebalancing. Avellaneda and Zhang (2010) used a second-degree taylor series approximation to illustrate the dependence on the reference asset's realized variance in the price dynamics of an LETF. This path dependence plays a crucial role in the valuation of options written on LETFs, as studied by Leung and

[1]ETFs: A roadmap to growth. www.pwc.com/us/etf

Sircar (2015) and Leung et al. (2016). In this paper we quantify the LETFs' variance using time series discrete models. The time series models give a novel approach to address the volatility decay, one of the daily re-balancing consequences

Autoregressive Integrated Moving Average (ARIMA) models are standard in time series forecasting. They work by fitting a model to the historical data available, which can be used to forecast future values. The popularity of ARIMA models stems from Box and Jenkins (1990) simple methodology in determining the parameters of the model. The error terms are also assumed to follow a white noise process that is Gaussian and the data has to be stationary. The failure of any of these assumptions greatly affects the performance of the model and results in poor predictions. We use the variance of the ARIMA model as a proxy for estimating the variance of the ETFs and their corresponding leveraged funds.

ARIMA models major drawback is their linearity as most financial time series are non-linear. Engle (2001) also highlights that ARIMA models assume the expected values of their errors squared to be constant at any given point in time-homoskedasticity. According to Mandelbrot (1997), financial data usually suffer from volatility clustering, "large changes tend to be followed by large changes, of either sign, and small changes tend to be followed by small changes." Engle (1982) developed the autoregressive conditional heteroskedastic (ARCH) models to overcome this implausible assumption. Bollerslev (1986) generalized Engle's model by incorporating the past conditional variance observations into the current variance equation developing the generalized autoregressive conditional heteroskedasticity (GARCH) model. Modeling the conditional variance can give a better confidence interval for volatility. We use a GARCH(1,1) model to examine the conditional variance of LETFs, and check their convergence to the unconditional values produced by the ARIMA models

The design of Leveraged ETFs suggests a higher variance than their underlying assets, which is confirmed by examining their ARIMA models' variances. The results confirm the volatility decay phenomenon i.e. The higher the variance, the more is the loss in the LETF value. They also indicate that inverse leveraged funds decay more than leveraged funds with the same leverage ratio. The paper starts by explaining the phenomenon of volatility decay using a modified example of Leung and Santoli(2012). The second section searches for the optimal time frame for out of sample forecasting. The third section examines the relationship between the variance and beta of several LETFs, and the last section fits GARCH(1,1) model to estimate the conditional variance of LETFs.

Day	ETF	%-Change	+2x LETF	%-Change	-2x LETF	%-Change
0	100		100		100	
1	102	2.00%	104	4.00%	96	-4.00%
2	100	-1.96%	99.92	-3.92%	99.76	3.92%
3	102	2.00%	103.92	4.00%	95.77	-4.00%
4	100	-1.96%	99.84	-3.92%	99.53	3.92%
5	102	2.00%	103.76	4.00%	95.55	-4.00%
6	100	-1.96%	99.69	-3.92%	99.30	3.92%

Table 1: The table shows a modified example of Leung and Santoli (2016). The underlying ETF is mean reverting to a constant level equal to 100$. The inverse fund decays at a faster rate compared to the leveraged fund. It lost 0.70$ after six days compared to only 0.31$ for the leveraged fund.

2 Price Dynamics of LETFs

Leveraged ETFs are designed to replicate the multiple returns of an underlying asset or index. Leung and Santoli (2016) Leung and Santoli (2012) describe a benchmark model for the LETF value on day n with an initial value L_o

$$L_t = L_o \prod_{j=1}^{t} (1 + \beta R_j) \tag{1}$$

where R_j is the daily return of the underlying asset and β is the leverage ratio. It is also equivalent to express the LETF value by

$$L_t = L_{t-1}(1 + \beta R_j) \tag{2}$$

and the LETF's return by

$$R_t^{\beta} = \frac{L_t}{L_{t-1}} - 1 = \beta R_t \tag{3}$$

Equation (3) shows that the target return of an LETF is the return of the underlying ETF multiplied by the leverage multiple β on day t. LETFs fail to match the target return if they are being held for periods of more than one day. The following example is a modified version of the Leung and Santoli (2012) example, illustrating this volatility decay for an ETF that is mean reverting. The ETF value remained the same on day 6, indicating a net change of 0%, the +2x LETF shows a net loss of 0.31% and the -2x

LETF shows a net loss of 0.7%. The loss in value is present during all days except day 1. For example, on day 3 the ETF has a net profit of 2%, while the +2x LETF has a net gain of 3.92% which is less than twice the ETF net gain, and the -2x LETF has a net loss of 4.13%, more than twice the net loss in value. It might be intuitive that a 2x LETF will have an equal and opposite net change to a -2x LETF which is true for Day 1. However, it is not true for longer holding periods of time.

Indeed, the higher the volatility of the ETF, the more the LETF experiences loss of value. In the previous example, setting x to 4% instead of 2% will result in a higher loss in the LETF value. For example, on day 6, the +2x LETF value will be 99.08 compared to 99.69, and the -2x LETF value will be 97.26 compared to 99.30. Changing the ETF volatility from 2% to 4% increased the net loss from 0.31% to 0.72% for the +2x LETF which is more than double the increase in volatility. Similarly, the -2x LETF net loss went from 0.7% to 2.74%, almost 4 times the loss in value for doubling the volatility.

In order to have a better understanding of the different patterns of volatility decay for different leverage ratios, it is possible to write down the algebraic equations describing their movement. If x is the positive %-change on Day 1, the negative %-change required on Day 2 for the ETF value to return to its initial figure is $y = \frac{x}{1+x}$. The LETF value on day t for a leverage ratio that is equal to +2 and -2 becomes

$$L_t^{(+2x)} = L_o(1 - \frac{2x^2}{1+x})^{\frac{t-j}{2}}(1+2x)^j \tag{4}$$

$$L_t^{(-2x)} = L_o(1 - \frac{6x^2}{1+x})^{\frac{t-j}{2}}(1-2x)^j \tag{5}$$

where t is the number of holding days and j equals 0 for even values of t and 1 otherwise.

Equations (4) & (5) explain the lower net value of the inverse leveraged ETF. On even days ($j = 0$), equations (4) & (5) are reduced to their first term, but the first term of equation (4) is always bigger than the first term of equation (5). On odd days ($j = 1$), the second term in both equations amplifies the difference in the LETF value. Therefore, the net value of the +2x LETF will always be higher than that of the -2x LETF.

Equations (4) & (5) shows an interesting pattern if derived for the the leverage ratios 3 and -3,

$$L_t^{(+3x)} = L_o(1 - \frac{6x^2}{1+x})^{\frac{t-j}{2}}(1+3x)^j \tag{6}$$

$$L_t^{(-3x)} = L_o(1 - \frac{12x^2}{1+x})^{\frac{t-j}{2}}(1-3x)^j \tag{7}$$

the similarity between equations (5) and (6) becomes evident. Hence, the values of the -2x LETF and +3x LETF are equal on even days ($j = 0$). The same relationship exists between the -1x LETF and the +2x LETF.

It is also possible to visualize the phenomenon of volatility decay for different leverage ratios. Figure (1) shows the decay in values of three different LETFs with β's equal to +2,-2 & +3. The LETFs are tracking a mean reverting ETF with an initial value of 100 for a holding period of 100 days.

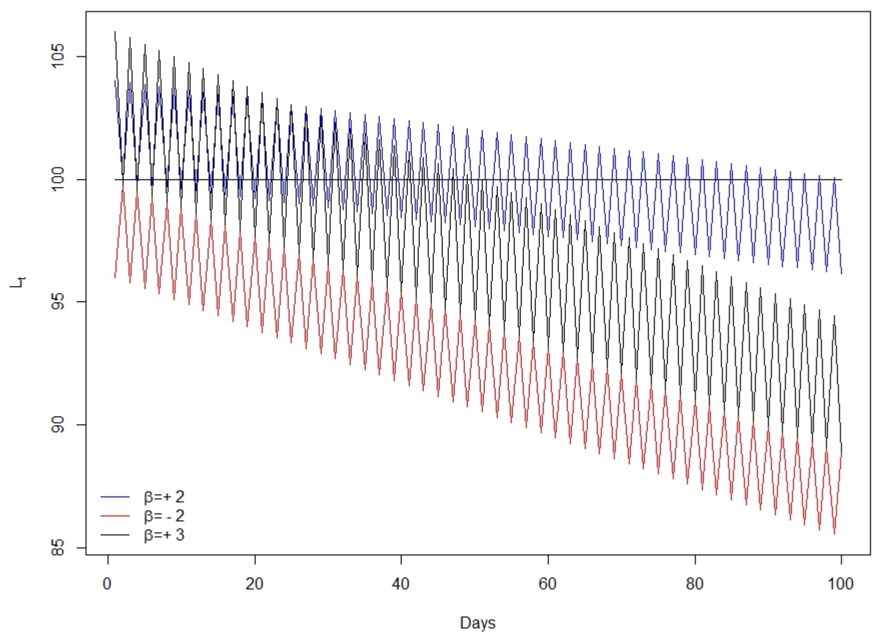

Figure 1: The value of the LETF L_t for three different β's tracking a mean reverting ETF to a constant level that is equal to 100. The blue +2x LETF is losing value at a rate slower than the black +3x LETF indicating that the higher the leverage index, the more is the loss in value. The red -2x LETF is also decaying faster than the blue +2x LETF and at the same rate as the black +3x LETF indicating that inverse leveraged funds tend to lose more value.

Figure (1) shows that the +3xLETF (black) and the -2xLETF (red) have the same values on even days. The +3x and +2x LETFs experience volatility decay at a rate slower than the -2x LETF. The +3x LETF outperforms the return of the +2x LETF at first, but decays faster as the holding period increases.

Equation (8) generalizes the volatility decay equations for a mean reverting ETF.

$$L_t^{(\beta x)} = L_o(1 + \frac{x^2(\beta - \beta^2)}{1+x})^{\frac{t-j}{2}}(1 + \beta x)^j \tag{8}$$

The term $\beta - \beta^2$ explains the the non linear volatility decay when the leverage ratio changes. For $\beta = -2$ and $\beta = 3$, the term $\beta - \beta^2 = -6$, similarly for $\beta = -1$ and $\beta = 2$, the term $\beta - \beta^2 = -2$ explaining the reason behind the same values of the -2xLETF and +3xLETF, and the -1xLETF and +2xLETF on even days. Cheng and Madhavan (2009) observed the same relationship when calculating the amount by which the exposure of the total return swaps that need to be adjusted or re-hegded.

The above analysis shows that for a mean reverting ETF, the Leveraged ETF will suffer from a loss in value regardless of the multiplier direction. The higher the leverage ratio of the LETF and the volatility of the tracked ETF, the greater is the volatility decay. The simulation highlights the inadequacy of daily re-balanced leveraged and inverse leveraged ETFs for buy and hold investors.

3 Box and Jenkins Methodology

Autoregressive Moving Average (ARMA) models assume that the present value of a variable is a linear function of its past values, equation (2) shows a high correlation between the present and past values of LETFs, so it seems intuitive to model them using ARMA models. Let $\{S_t\}_{t=0}^{\infty}$ be a stochastic process that follows an ARMA(p,q) process, then it has the following form

$$S_t = c + \sum_{i=1}^{p} \phi_i S_{t-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t \quad \varepsilon_t \sim WN(0, \sigma^2) \tag{9}$$

where ϕ_i $(i = 1, ..., p)$ are the autoregressive coefficients AR(p), θ_j $(j = 1, ..., q)$ are the moving average coefficients MA(q), and ε_t are the error terms which are assumed to follow a white noise process that is Gaussian with zero mean and σ^2 variance. Box and Jenkins (1990) outline a three steps methodology to estimate the ARMA model parameters.

The first step is identifying and selecting the model order. ARMA(p,q) models require the time series data to be stationary. If the data is not stationary, it is possible to remove the trend by subtracting the present observation from the previous ones resulting in an ARIMA(p,d,q) model, where d is the number of differences computed in order to make the data stationary. An example of taking the first difference is $\Delta S_t = S_t - S_{t-1}$ and the corresponding ARIMA(p,1,q) will be

$$\Delta S_t = c + \sum_{i=1}^{p} \phi_i \Delta S_{t-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t \quad \varepsilon_t \sim WN(0, \sigma^2) \tag{10}$$

Box and Jenkins (1990) propose plotting the autocorrelation and partial autocorrelation functions of the data to identify the ARMA model orders.

The second step is estimating the model parameters, it is usually done using an optimization technique such as the maximum likelihood. The last step is diagnostic checking which tests the goodness of fit of the model. Enders (1995) highlights the fact that R^2 and the average of the residual sum of squares are standard goodness of fit measures in ordinary least squares, but they fail to penalize the addition of more parameters as the fit necessarily improves as more parameters are included in the model. Akaike information criterion (AIC) and Bayesian information criterion (BIC) are more appropriate measures, since they penalize the inclusion of more parameters. The model with the least AIC and/or BIC is considered to be the one with the best fit.

In order to evaluate the out of sample performance of the fitted models, three different measures of performance are calculated:

Mean Error (ME)

$$\frac{1}{n}\sum_{i=1}^{n}(S_i - \hat{S}_i)$$

Mean Absolute Error (MAE)

$$\frac{1}{n}\sum_{i=1}^{n}|S_i - \hat{S}_i|$$

Root Mean Square Error (MSE)

$$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(S_i - \hat{S}_i)^2}$$

where S_i are the actual values, $\hat{S}_i$ are the values forecasted by the model and n is the number of the out of sample predictions.

It is important to note that the goodness of fit of the model is determined by the in sample data. For example, in an ARIMA model specification, the lagged order (p) and the number of moving average terms (q) is selected based on the least AIC/BIC criterion that is calculated using the in sample data. On the other hand, the measures of performance depend on the out of sample data. The best fitted models tend to have a smaller forecasting error, but in some cases, alternative model specifications may provide a smaller forecasting error.

4 Three Market Indexes

Market indexes provide exposure to different sectors of the market, while shielding from idiosyncratic risk. The paper is studying three main market indexes and some of the funds tracking their performance. The indexes are S&P500 ($\hat{\text{G}}$PSC), Russell 2000($\hat{\text{R}}$UT) and NASDAQ 100($\hat{\text{N}}$DX). They reflect the performance of large-cap companies, small-cap companies, and the technology sector, respectively. Enders (1995) highlights the fact that time series discrete modeling is based on the assumption that there is a data generating process that has been going for an infinite time. Therefore, it might seem that by including more historical data, the fitted model will be more representative of the data generating process.

In order to assess the validity of this assumption, an ARIMA model is fitted to the three indexes using all data since their inception until April 13$^{\text{th}}$, 2016. The performance of the fitted ARIMA model is evaluated for the next 100 data points i.e. the period from April 14$^{\text{th}}$until September 2$^{\text{nd}}$, 2016. Table 2 reports the best fitted ARIMA models and their corresponding out of sample performance for the three indexes in addition to QQQ; QQQ is an ETF that tracks NASDAQ-100. S&P500 ARIMA model has the smallest out of sample forecast errors among the three major indexes.

The forecast error of the QQQ ETF is much smaller than any other index, the result could be attributed to its smaller data relative to the other indexes, contradicting Enders (1995) assumption. In order to further probe the enhanced ability of ARIMA models to forecast with smaller data sets, ARIMA(p,d,q) models are fitted to S&P500 data during varying time frames, from September 2$^{\text{nd}}$, 2000 ,2009, 2010 , 2011, 2012 until April 13$^{\text{th}}$, 2016. Table 3 reports the out of sample performance of the best fitted

	S&P500	RUT	NDX	QQQ
ARIMA(p,d,q)	(1,1,1)	(1,1,1)	(1,1,1)	(2,1,5)
Mean Error	31.98	38.59	-27.59	0.29
Mean Absolute Error	44.58	45.77	139.15	3.58
Root Mean Square Error	56.19	57.36	160.15	4.24

Table 2: Best fitted ARIMA models for S&P500, Russel, NASDAQ-1000 Indexes, and QQQ ETF. The S&P500 data is from March 4^{th}, 1957 until April 13^{th}, 2016;The RUSSEL 2000 data is from September 10^{th}, 1987 until April 13^{th}, 2016; The NASDAQ-100 data is from October 1^{st}, 1985 until April 13^{th}, 2016; The QQQ data is from March 10^{th}, 1999 until April 13^{th}, 2016. The measures of performance evaluate the forecasting error for 100 points ahead in time i.e. until September 2^{nd}, 2016.

	2000	2009	2010	2011	2012
ARIMA(p,d,q)	(1,1,1)	(2,1,2)	(2,1,2)	(0,1,0)	(0,1,0)
Mean Error	38.37	23.59	20.88	-5.35	-3.52
Mean Absolute Error	49.09	37.01	35.31	27.86	28.50
Root Mean Square Error	62.48	45.45	43.07	34.02	34.34

Table 3: Best fitted ARIMA models for S&P500 during five different time frames and their corresponding measures of performance for 100 multi-step ahead forecasts.

ARIMA models for the five different time frames. The forecasts are for 100 points ahead in time i.e. until September 2^{nd}, 2016.

The measures of performance are minimized when a random walk with a drift is fitted to the time series data in the period between 2011 and 2016. An explanation to the diminished forecasting accuracy of the ARIMA models when more data is used could be attributed to the major economic shocks such as the 1973 Oil Shock, Dot-com bubble,...etc. More specifically, the housing bubble provides an explanation for the diminished power of the ARIMA model forecasts when the period between 2007-2009 is included in the data used to fit the model. The forecasts' accuracy suggests that around 2011, the effect of the housing crisis is negligible.However, the data is not enough to generate coefficients for the ARIMA model. Therefore, the fitting of all subsequent discrete time series models in this paper will be based on the observations in the period from September 2^{nd},2010 until April 13^{th}, 2016 as the process can be represented by an ARIMA(p,d,q) model and the

MAE is minimized.

5 LETF's ARIMA Return Forecasting

The arithmetic return of any stock is calculated by

$$R_t = \frac{S_t}{S_{t-1}} - 1$$

The time series return data of any stock is stationary, it is easy to see why if the above equation is expressed as

$$R_t = \frac{\Delta S_t}{S_{t-1}}$$

where ΔS_t is the first difference of the stock prices. Therefore, ARMA(p,q) models are sufficient for fitting the return time series data. In the first section, the LETF benchmark equation (2) showed that it is possible to express the return of the LETF as a multiple of the return of the underlying stock or index.

If the return of the underlying stock is assumed to follow an ARMA(p,q) process,

$$R_t = c + \sum_{i=1}^{p} \phi_i R_{t-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t \quad \varepsilon_t \sim WN(0, \sigma^2) \tag{11}$$

then, the return of the LETF can be expressed as follow:

$$R_t^\beta = \beta c + \sum_{i=1}^{p} \phi_i R_{t-i}^\beta + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t \quad \varepsilon_t \sim WN(0, \beta^2\sigma^2) \tag{12}$$

It is important to note that the coefficients of the ARMA(p,q) model remained the same. The differences between the R_t and R_t^β models are in the constant and error's variance. The constant changed from c to βc, consequently changing the expected value of the return $E[R_t]$ to

$$E[R_t^\beta] = \frac{\beta c}{1 - \phi_1 - \phi_2 - ... - \phi_p}$$

Moreover, the variance of the White Noise changed from σ^2 to $\beta^2\sigma^2$.

In order to verify the validity of the returns models, it is imperative to test them using market data. Table 1 shows the LETFs that will be modeled and their issuers. Mirror funds are defined to be two LETFs providing equal but directionally opposite beta and have the same issuer Henderson and Buetow (2014). For example, by looking at Table(1) it is clear that (UPRO,SPXU), (SDS,SSO), (SPXL,SPXS) and (SCO, UCO) are mirror funds. The mirror funds enable us to confirm the fact that inverse LETFs decay more than their leveraged counterparts netting out other effects. S&P500 provides the best exposure to the market, leading multiple funds to issue Leveraged ETFs that track its performance and its LETFs are the most traded in their categories. Table (2) presents the results of fitting the $R_t^{S\&P500}$, R_t^{SH}, R_t^{SSO}, R_t^{SSO*} R_t^{SDS},, R_t^{UPRO}, R_t^{SPXU}, R_t^{SPXL}, R_t^{SPXl*}, and R_t^{SPXS} to ARMA(p,q) models for the period between September, 2$^{\text{nd}}$ 2010 to April 13$^{\text{th}}$,2016.

Index	Issuer	(-1x)ETF	(-2x)ETF	(+2x)ETF	(-3x)ETF	(+3x)ETF
S&P500	Proshares	SH	SDS	SSO	UPRO	SPXU
	Direxion				SPXL	SPXS
NASDAQ-100	Proshares				SQQQ	TQQQ
Russel 2000	Direxion				TZA	TNA

Table 4: The ticker symbols of the different inverse and leveraged ETFs and the names of the issuing funds

The number of coefficients of the ARMA models were determined using Akaike Information Criterion (AIC). The underlying asset S&P500 showed the least AIC when fitted to an ARMA(5,3) model. We expect its LETFs to have the same number of coefficients(p,q) and their values to be considerably close to their underlying asset coefficients, which is the case for all LETFs except SSO and SPXL where an ARMA(5,0) and an ARMA(1,1) showed a better fit. For the sake of discussion, the returns of both LETFs were fitted to an ARMA(5,3). The coefficients of the ARMA(5,3) models are all significant except the AR(1), AR(4) & MA(1) coefficients, their standard errors are big and are not significantly different from zero. The variances of the different ARMA models agree with the quadratic relation that was implied by the ARMA models equations (9) & (10)

$$\sigma^2_{LETF} \approx \beta^2 \sigma^2_{ETF} \tag{13}$$

It is possible to check the relationship above by plotting the corresponding betas of the LETF and their variances as in Figure 2. It is obvious that the relationship is quadratic as was determined in the analysis above.

Model	μ	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5	θ_1	θ_2	θ_3	$\sigma^2(10^{-3})$
$R_t^{S\&P500}$	0.0005	-0.0611	-0.4867	0.5053	0.0239	-0.135	0.0137	0.5253	-0.6265	0.0912
	(0.0002)	(0.1349)	(0.1021)	(0.1448)	(0.0297)	(0.028)	(0.1351)	(0.0939)	(0.1351)	
R_t^{SH}	-0.0006	-0.0567	-0.4903	0.5052	0.0202	-0.1344	0.0146	0.5278	-0.6233	0.0906
	(0.0002)	(0.1358)	(0.1020)	(0.1455)	(0.0295)	(0.0280)	(0.1360)	(0.0945)	(0.1362)	
R_t^{SSO}	0.0011	-0.0432	0.0193	-0.0730	-0.0145	-0.1030				0.3655
	(0.0004)	(0.0265	(0.0265)	(0.0264)	(0.0265)	(0.0265)				
R_t^{SSO*}	0.0011	-0.0476	-0.4796	0.5186	0.0206	-0.1328	0.0084	0.5172	-0.6351	0.3598
	(0.0004)	(0.1317)	(0.0991)	(0.1408)	(0.0293)	(0.0281)	(0.1316)	(0.0920)	(0.1316)	
R_t^{SDS}	-0.0012	-0.0530	-0.4832	0.5125	0.0215	-0.1352	0.0132	0.5226	-0.6304	0.3594
	(0.0004)	(0.1324)	(0.0991)	(0.1409)	(0.0292)	(0.0278)	(0.1325)	(0.0918)	(0.1317)	
R_t^{UPRO}	0.0017	-0.0376	-0.4774	0.5259	0.0201	-0.1303	0.0001	0.5129	-0.6385	0.8075
	(0.0006)	(0.1325)	(0.1002)	(0.1424)	(0.0295)	(0.0284)	(0.1324)	(0.0933)	(0.1331)	
R_t^{SPXU}	-0.0018	-0.0528	-0.4906	0.5100	0.0175	-0.1310	0.0153	0.5252	-0.6246	0.8081
	(0.0006)	(0.1359)	(0.1017)	(0.1453)	(0.0293)	(0.0282)	(0.1361)	(0.0949)	(0.1363)	
R_t^{SPXL}	0.0016	-0.8557					0.8069			0.8383
	(0.0007)	(0.0666)					(0.0753)			
$R_t^{SPXL^*}$	0.0016	-0.0640	-0.4943	0.4934	0.0164	-0.1389	0.0302	0.5355	-0.6094	0.8176
	(0.0006)	(0.1335)	(0.0996)	(0.1421)	(0.0288)	(0.0280)	(0.1337)	(0.0931)	(0.1344)	
R_t^{SPXS}	-0.0018	-0.0675	-0.4924	0.4922	0.0197	-0.139	0.0284	0.5327	-0.6109	0.8248
	(0.0006)	(0.1332)	(0.1000)	(0.1420)	(0.0291)	(0.028)	(0.1333)	(0.0929)	(0.1337)	

Table 5: The ARMA(p,q) models are fitted using the Box-Jenkins methodology, the number of coefficients (p,q) is determined using akaike information criterion (AIC) except for $SPXL^*$ & SSO^* where the AIC criterion favored different models. The number of coefficients for $SPXL^*$ & SSO^* is the same as the previous models for illustration purposes. The values in the parenthesis are the standard error.

In order to have more insight about the LETFs return dynamics. It is imperative to compare two LETFs which have two different issuers but similar betas such as (UPRO,SPXU) & (SPXL,SPXS). They are both mirror funds which suggests that they should suffer from an equal and opposite loss in value, which is not the case. UPRO and SPXL expected returns are 0.0017 & 0.0016 compared to an expected return of -0.0018 for both SPXU & SPXL. The expected returns of the inverse funds are always less than the leveraged funds regardless of having the same issuer or not. The finding confirms that inverse funds will always miss their target returns by margins bigger than the leveraged counterparts.The above analysis confirms the ability of the ARMA(p,q) models to capture the returns dynamics of the LETF by only knowing the model for the underlying stock or index.

Tang and Xu (2013) suggest that all LETFs fail to achieve their target returns, the discrepancy is usually caused by the expense ratio charged by the funds managing the LETFs in addition to other factors such as management tracking error, market frictions, or inefficiency. Henderson and Buetow (2014) baseline regression model suggests that leverage inverse funds exhibit returns below the benchmark model and the leveraged funds exhibit positive excess returns. Our results confirm their findings, the expected return of S&P500 is 0.05% suggesting -0.15%, -0.1% , 0.1%,& 0.15% benchmark returns for -3,-2,+2,+3 leverage index, respectively. However, Table 2 shows -0.18%, -0.12%, 0.11%,0.17% for the same leverage index i.e. leverage inverse funds outperform their benchmark model while inverse funds underperform.

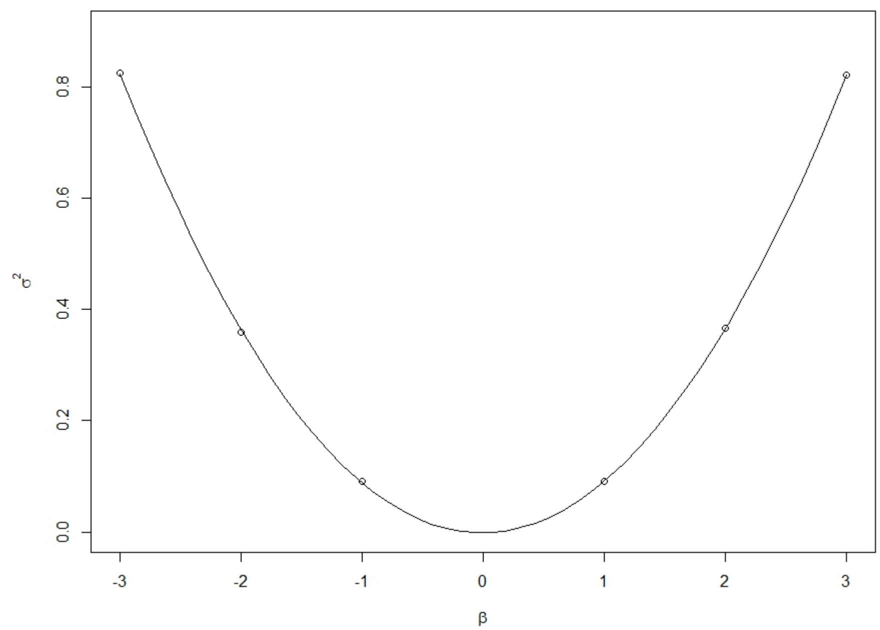

Figure 2: A curve is fitted to the ARIMA models' variances of S&P500, SH, SSO, SDS,SPXL and SPXS returns. A quadratic relationship exists between the leverage index β and the models' variances

6 Volatility Estimation LETF Returns

ARIMA discrete time models are standard in forecasting, but their linearity is a major drawback when dealing with financial time series experiencing volatility clustering. If the error terms are different for some points or ranges of the data than for others, the data is said to suffer from heteroskedasticity. Engle (2001) explains that heteroskedasticity does not afffect the regression coefficients, but the standard errors and confidence intervals will give a false sense of precision. In order to better model the variance of LETF returns, the conditional variance is examined by fitting the data residuals to a GARCH(1,1) model. GARCH(1,1) models are parsimonious and are more accurate in modeling non-stationary variance. GARCH(1,1) allows us to calculate the conditional variance which is more accurate than the unconditional variance produced by the ARMA models.
Suppose R_t's error ε_t follows a GARCH(1,1) process

$$\varepsilon_t = v_t\sqrt{h_t} \quad h_t = \omega + \alpha_1\varepsilon_{t-1} + \beta_1 h_{t-1} \tag{14}$$

$$where \quad v_t \sim (0,1)$$

Multiplying the above equation by β, it is straightforward to see that R_t^{β}'s error ε_t will follow a GARCH(1,1) process

$$\begin{aligned}\varepsilon_t &= \beta v_t\sqrt{h_t} \quad h_t = \omega + \alpha_1\varepsilon_{t-1} + \beta_1 h_{t-1} \\ &= v_t\sqrt{\beta^2 h_t} \quad \beta^2 h_t = \beta^2\omega + \alpha_1\varepsilon_{t-1} + \beta_1 h_{t-1}\end{aligned} \tag{15}$$

$$where \quad v_t \sim (0,1)$$

and the conditional variance will be

$$\begin{aligned}E_{t-1}[\varepsilon_t^2] &= E_{t-1}[v_t^2]E_{t-1}[\beta^2\omega + \alpha_1\varepsilon_{t-1} + \beta_1 h_{t-1}] \\ &= E_{t-1}[\beta^2\omega + \alpha_1\varepsilon_{t-1} + \beta_1 h_{t-1}] \\ &= \beta^2\omega + \alpha_1\varepsilon_{t-1} + \beta_1 h_{t-1}\end{aligned} \tag{16}$$

Table 3 presents the results of fitting the errors' of the returns of S&P500, SH, SSO, SDS, UPRO, SPXU, SPXL and SPXS to a GARCH(1,1) model for the period between September, 2^{nd} 2010 to April 13^{th},2016
It is obvious that the constant variance assumption of the ARIMA models is not realistic. The coefficients of the GARCH(1,1) model for all LETFs are significant. The relationship in (15) $\sigma^2_{LETF} \approx \beta^2\sigma^2_{ETF}$ is accounted for by

Model	$\omega(10^-3)$	α_1	β_1	DoF
S&P500	0.0042806	0.15724	0.80049	6
	$(3.743)^2$	(5.352)	(25.102)	(5.105)
SH	0.0042142	0.16643	0.79257	6
	(3.875)	(5.601)	(25.490)	(4,976)
SSO	0.016931	0.16509	0.79456	6
	(3.805)	(5.441)	(25.049)	(5.295)
SDS	0.016805	0.16423	0.79522	6
	(3.809)	(5.524)	(25.469)	(5.150)
UPRO	0.037756	0.16705	0.79322	6
	(3.864)	(5.580)	(25.662)	(5.301)
SPXU	0.037788	0.16699	0.79270	6
	(3.882)	(5.616)	(25.770)	(5.143)
SPXL	0.037625	0.16546	0.79446	6
	(3.850)	(5.573)	(25.659)	(5.132)
SPXS	0.038107	0.16855	0.79221	6
	(3.825)	(5.497)	(25.098)	(5.264)

Table 6: GARCH(1,1) models are fitted for the variance of the returns of different LETFs. The errors are assumed to have a t-distribution and the last column indicate their degrees of freedom. The values in the parenthesis are the t-statistic

.

the first term $\beta^2\omega$ in equation (14). It is clear that the quadratic relationship is preserved regardless of the model that is employed as it was determined in the analysis above.

References

Avellaneda, M. and Zhang, S. (2010). Path-dependence of leveraged ETF returns. *SIAM Journal on Financial Mathematics*, 1(1):586–603.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3):307–327.

Box, G. E. P. and Jenkins, G. (1990). *Time Series Analysis, Forecasting and Control*. Holden-Day.

Bruno, S., Chincarini, L., and Whitelaw, R. (2014). An alternative method to construct levered indexes. *Journal of index investing*, 5(2):34–47.

Cheng, M. and Madhavan, A. (2009). The dynamics of leveraged and inverse exchange-traded funds. *Journal of Investment Management*, 4.

Enders, W. (1995). *Applied econometric time series*. Wiley series in probability and mathematical statistics. Probability and mathematical statistics. Wiley.

Engle, R. (2001). Garch 101: The use of ARCH/GARCH models in applied econometrics. *The Journal of Economic Perspectives*, 15(4):157–168.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the Econometric Society*, pages 987–1007.

Guo, K. and Leung, T. (2015). Understanding the tracking errors of commodity leveraged ETFs. In Aid, R., Ludkovski, M., and Sircar, R., editors, *Commodities, Energy and Environmental Finance, Fields Institute Communications*, pages 39–63. Springer.

Hamilton, J. (1994). *Time Series Analysis*. Princeton University Press.

Henderson, B. J. and Buetow, G. W. (2014). The performance of leveraged and inverse leveraged exchange traded funds. *Journal of investment management*, pages 69–93.

Leung, T., Lorig, M., and Pascucci, A. (2016). Leveraged ETF implied volatilities from ETF dynamics. *Mathematical Finance*. to appear.

Leung, T. and Park, H. (2017). Long-term growth rate of expected utility for leveraged ETFs: Martingale extraction approach. *International Journal of Theoretical and Applied Finance*. to appear.

Leung, T. and Santoli, M. (2012). Leveraged ETFs: Admissible leverage and risk horizon. *Journal of Investment Strategies*, 2(1):39–61.

Leung, T. and Santoli, M. (2016). *Leveraged Exchange-Traded Funds: Price Dynamics and Options Valuation*. SpringerBriefs in Quantitative Finance. Springer.

Leung, T. and Sircar, R. (2015). Implied volatility of leveraged ETF options. *Applied Mathematical Finance*, 22(2):162–188.

Leung, T. and Ward, B. (2015). The golden target: Analyzing the tracking performance of leveraged gold ETFs. *Studies in Economics and Finance*, 32(3):278–297.

Mandelbrot, B. B. (1997). The variation of certain speculative prices. In *Fractals and Scaling in Finance*, pages 371–418. Springer.

Shumway, R. and Stoffer, D. (2000). *Time Series Analysis and Its Applications*. Springer texts in statistics. Springer.

Tang, H. and Xu, X. E. (2013). Solving the return deviation conundrum of leveraged exchange-traded funds. *Journal of Financial and Quantitative Analysis*, 48(1):309–342.