WXML Final Report: Mathematics of Exchanged-Traded Fund

Tim Leung, Jack Simonson, Chaoyang Jiang, Zach Dingels

Fall 2016

1 Introduction

The focus of our project is on researching the returns dynamics of leveraged exchange traded funds (LETFs) and estimating empirical leverage ratios. Our research team focused on commodity ETFs, all of which track oil indices. First, we calculated and compared daily returns of each ETF and its reference index. We then modeled the returns as an Ornstein−Uhlenbeck process and used maximum likelihood estimation to estimate mean-reversion speed, long-run mean, and volatility parameters. We then fit our data to an autoregressive – AR(1) – model and used regression to estimate the autocorrelation parameter. Next, we fit our returns data to a well-known distribution. Lastly, we estimated the empirical leverage ratio of each LETF and compared these estimates with their theoretical ratios.

1.1 Price Ratio

We first compute daily simple returns of each ETF and reference index and then compute the price ratio. Let P_t denote the reference index and price of ETF on day t.

$$R_t^{index} = \frac{P_t}{P_{t-1}} - 1$$

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

$$Ratio_t = \frac{R_t^{ETF}}{R_t^{index}}$$

The motivation behind finding the returns ratio for each ETF and its reference index is that if their prices move proportionately, then the ratio would fluctuate tightly around the advertised leverage ratio. If the returns ratio isn't close to the ideal leverage ratio, then there exists tracking error.

1.2 Issues

The our results showed that the returns ratios centered around the leverage ratio for each ETF, but there existed large spikes in our time series plots.

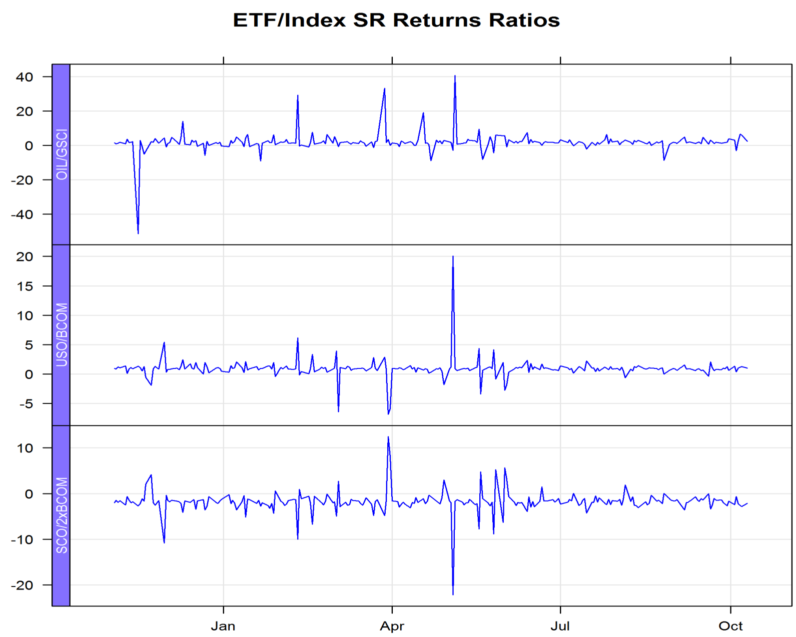


Figure 1: Time series of returns ratios for OIL, SCO, and USO ETFS. OIL tracks the SP GSCI Crude Oil Total Returns index 1:1, while USO and SCO track the Bloomberg WTI Oil Subindex 1:1 and -2:1, respectively.

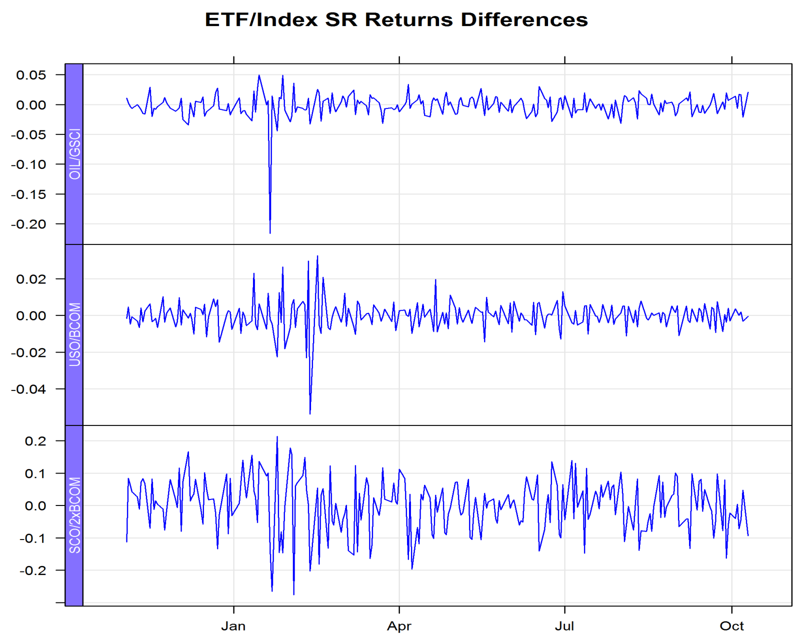
These spikes indicate that there are factors affecting the returns of the ETFs and indices that lie outside of the daily pricing. A better method is to examine returns differences, which will help eliminate pricing abnormalities and give us a better idea of the tracking fidelity between each ETF and its reference index.

1.3 Daily Return Difference

We calculated the returns differences

$$Difference_t = R_t^{ETF} - \beta R_t^{index}$$

where β is the leverage ratio. We then plotted the time series for these results and noticed a significant improvement in our data.



The mean of daily return difference is much tighter fluctuation about 0. Since we are using scaled differences, we expect to see our data fluctuate about a mean of 0. Figure 2 shows that the outlying data points are closer to the expected value by at least an order of magnitude.

2 OU Process

In theory, if the returns of an ETF and index differ, then the ETF will correct in pricing. The divergence will continue to happen, but we expect to see the returns differences return to their mean after they diverge. To test the mean reversion, we modeled our data as an Ornstein−Uhlenbeck process.

An Ornstein−Uhlenbeck process satisfies the stochastic differential equation (SDE)

$$dX_t = \theta(\mu - X_t)dt + \sigma dW_t$$

where $\mu > 0$ is the speed of mean-reversion, $\theta \in \mathrm{R}$ is the long-run mean, $\sigma > 0$ is the volatility parameter. We want to estimate μ in order to see how quickly the price differences correct.

2.1 Issues

We used maximum likelihood estimation to estimate parameters of OU model, but it resulted in complex values for μ and θ. Seeing as an OU process is a continuous time process, and our data are observed in discrete time, we decided to estimate the degree of time dependence between returns as an AR(1) model, which is roughly the discrete-time equivalent of an OU process.

2.2 AR(1) Model

A standard AR(1) model is given by

$$X_t = c + \varphi_1 X_{t-1} + \varepsilon_t \Leftrightarrow X_{t+1} - X_t = (1-p)(\frac{c}{1-p} - X_t + \epsilon_t)$$

where ρ is the coefficient of correlation that indicates the level of time dependence between X_t and X_{t-1}. To estimate ρ, we used a least squares estimate

$$\rho = \frac{\sum_{k=2}^{n} X_k X_{k-1}}{\sum_{k=2}^{n} X_{k-1}^2}$$

2.3 Results

ETF	OIL	USO	SCO	USL	DWTI	UWTI	UCO	DBO	DTO
$\hat{\rho}$	−0.078	−0.527	−0.065	−0.252	−0.511	−0.517	0.013	−0.499	−0.038

Unofrtunately there doesn't appear to be any strong correlation between AR(1) estimates and the LETF ratios. Both leveraged and unleveraged, even ones that track the same index, have disparate levels of time dependence. With more than 1-lag, the time dependence drops below the $|0.2|$

level of signifcance, but for the 1-lag case between X_t and X_{t-1} there appears to be significant differences in time dependence. One important result from our estimates is that all but one of our AR(1) coefficient estimates are negative, indicating that returns have a negative time dependence and this reinforces our claim that returns data has mean-reversion properties. The returns correct once they diverge from the mean, and continue to do so over time.

3 Fit to a Distribution

We have calculated the daily return difference of ETF and reference index. And now we want to fit the histograms to well-known distributions and find which one fit the best.

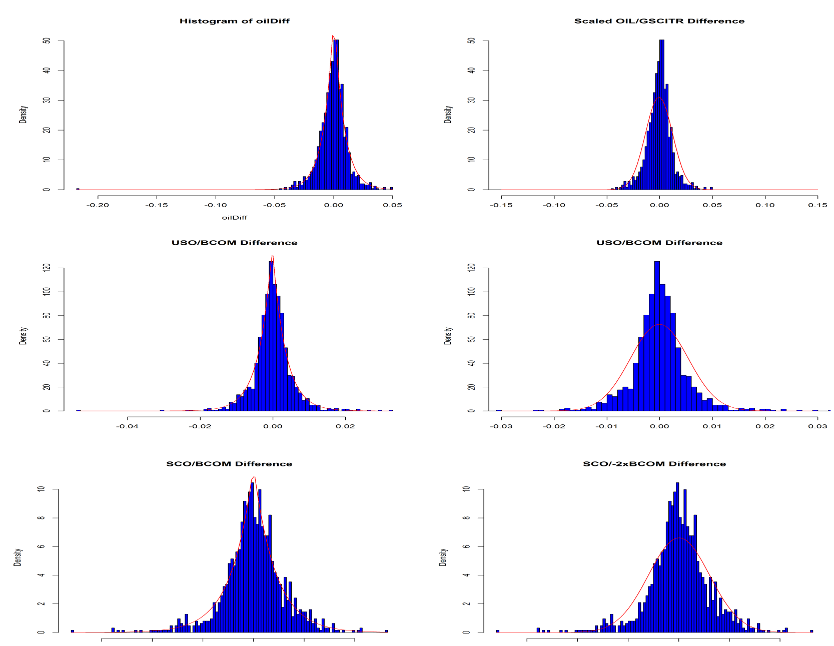


Figure 2: Here we have fit our data to a double exponential distribution (left) and a double exponential distribution (right).

3.1 Results

Our results are best captured by a double exponential distribution. We see clustering of the majority of our data close to 0, indicating that the majority of our returns differences are close to zero, both positive and negative. With the exception of a few outliers, we see that this supports our claim that returns differences have mean-reversion properties.

3.2 Theoretical

One of our goals was to estimate the leverage ratio for each LEFT. We did this using Tim Leung and Marco Santoli's method from "Leveraged Exchange-Traded Funds: Price Dynamics and Options Valuation". Here is a brief over view of the mathematics behind their method.

The log return for day k of a LEFT is given by

$$\log \frac{L_{t+\Delta k}}{L_t} = \beta \log \frac{S_{t+\Delta k}}{S_t} + \theta V_t^{(k)} + ((1-\beta)r - f)k\Delta t \tag{1}$$

where L is the value of the LEFT, S is the value of the reference index, $V_t^{(k)}$ is the realized variance, r is the interest rate on the borrowing amount and f is a small expense fee.

Due to $\log \frac{S_{t+\Delta k}}{S_t}$ and θ being strongly dependent we can't use a linear model to estimate the leverage ratio. Theoretically, θ has a value of $\frac{\beta(1-\beta)}{2}$. This leads to an issue of colineartiy, where our estimates for volatility coefficient $\hat{\theta}$ are heavily dependent on our esitmates for the leverage ratio $\hat{\beta}$. To help eliminate this issue, we used Tim and Marco's method of solving the optimization problem

$$\min_{\beta \in \mathbb{R}} \sum_{i=1}^{n} (y_i - f_i(\beta))^2 \tag{2}$$

where y_i are the empirical log returns of the LETF and $f_i(\beta)$ is the theoretical log return defined by

$$\begin{aligned} f_i(\beta) &= \beta x_i + \frac{\beta(1-\beta)}{2} v_i + ((1-\beta)r - f)\Delta T \\ &= \beta(x_i - r\Delta T) - \frac{\beta(1-\beta)}{2} v_i + (r - f)\Delta T \end{aligned}$$

where x_i and v_i are the log return and realized variance of the reference index. This will allow us to accurately estimate leverage rations and then use these to find $\hat{\theta}$ estimates.

From the first-order optimality condition we can find the leverage ratio

$$\sum_{i=1}^{n}(y_i - f_i(\beta))(x_i - r\Delta T - \beta v_i + \frac{1}{2}v_i) = 0 \tag{3}$$

And by expansion we have

$$\begin{aligned}
&\sum_{i=1}^{n}(y_i - f_i(\beta))(x_i - r\Delta T - \beta v_i + \frac{1}{2}v_i) && (4)\\
=&\sum_{i=1}^{n}(y_i - f_i(\beta))(x_i - r\Delta T - \beta v_i + \frac{1}{2}v_i) && (5)\\
=&\sum_{i=1}^{n}(y_i - \beta(x_i - r\Delta T) + \frac{\beta^2}{2}v_i - \frac{\beta}{2}v_i - (r-f)\Delta T)(x_i - r\Delta T - \beta v_i + \frac{1}{2}v_i) && (6)\\
=&\sum_{i=1}^{n}(y_i - (r-f)\Delta T + \beta(x_i - r\Delta T + \frac{v_i}{2}) + \frac{\beta^2}{2}v_i)(x_i - r\Delta T - \beta v_i + \frac{1}{2}v_i) && (7)\\
=&(-\sum_{i=1}^{n}\frac{v_i^2}{2})\beta^3 && (8)\\
&+(\sum_{i=1}^{n}\frac{3}{2}(x_i - r\Delta T)v_i + v_i^2)\beta^2 && (9)\\
&+(\sum_{i=1}^{n}-((x_i - r\Delta T) + \frac{1}{2}v_i)^2 + v_i((r-f)\Delta T - y_i))\beta && (10)\\
&+(\sum_{i=1}^{n}(y_i - (r-f)\Delta T)((x_i - r\Delta T) + \frac{1}{2}v_i)) && (11)\\
& && (12)
\end{aligned}$$

which reduces to

$$A\beta^3 + B\beta^2 + C\beta + D = 0. \tag{13}$$

where

$$A = -\sum_{i=1}^{n} \frac{v_i^2}{2} \tag{14}$$

$$B = \sum_{i=1}^{n} \frac{3}{2}(x_i - r\Delta T)v_i + v_i^2 \tag{15}$$

$$C = \sum_{i=1}^{n} -((x_i - r\Delta T) + \frac{1}{2}v_i)^2 + v_i((r-f)\Delta T - y_i) \tag{16}$$

$$D = \sum_{i=1}^{n}(y_i - (r-f)\Delta T)((x_i - r\Delta T) + \frac{1}{2}v_i) \tag{17}$$

$$\tag{18}$$

and dividing by A we have

$$\beta^3 + b\beta^2 + c\beta + d = 0 \tag{19}$$

By the well-known Cardano's method for cubic polynomials, the explicit solutions when the discriminate is negative are given by

$$\beta_1 = u_0 + u_1 - \frac{b}{3}, \beta_{2,3} = -\frac{1}{2}(u_0 + u_1) \pm \frac{i\sqrt{3}}{2}(u_0 - u_1) - \frac{b}{3} \tag{20}$$

where

$$u_i = \sqrt{-\frac{p}{2} + (-1)^i \sqrt{\frac{p^2}{4} + \frac{q^3}{27}}} \tag{21}$$

$$p = \frac{2b^3 - 9bc + 27d}{a}, q = \frac{3c - b^2}{3} \tag{22}$$

if the discriminate is equal to 0 then we have 3 roots with 2 of them being equal.

$$-2\sqrt{-\frac{q}{3}} - \frac{b}{3}, \sqrt{-\frac{q}{3}} - \frac{b}{3}, \sqrt{-\frac{q}{3}} - \frac{b}{3} \qquad \text{if } p > 0 \tag{23}$$

$$2\sqrt{-\frac{q}{3}} - \frac{b}{3}, -\sqrt{-\frac{q}{3}} - \frac{b}{3}, -\sqrt{-\frac{q}{3}} - \frac{b}{3} \qquad \text{if } p < 0 \tag{24}$$

$$0, 0, 0 \qquad \text{if } p = 0 \tag{25}$$

if the discriminate is greater than 0

$$\beta_n = 2\sqrt{-\frac{q}{3}}\cos(\frac{\gamma}{3} + \frac{2n\pi}{3}), n = 0, 1, 2 \tag{26}$$

where

$$\gamma = \cos^{-1}\sqrt{\frac{p^2/4}{-q^3/27}} \tag{27}$$

This is the general idea behind Tim and Marco's idea and what we used to create leverage ratio estimates.

4 Findings

We used MATLAB to get estimates via linear regression and Leung and Santoli's method.

ETF	β	θ	$\hat{\beta}_{cub}$	$\hat{\beta}_{reg}$	$\hat{\theta}_{cub}$	$\hat{\theta}_{reg}$
OIL	1	0	−0.018	0.010	−0.009	−7.341
USO	−2	−3	0.255	0.265	0.095	−1.695
SCO	1	0	0.068	0.057	0.032	2.261
USL	−3	−6	0.019	0.038	0.009	2.752
$DWTI$	3	−3	−0.148	−0.138	−0.085	−1.598
$UWTI$	2	−1	0.270	0.273	0.099	−0.557
UCO	1	0	0.481	0.493	0.125	−2.256
DBO	−2	−3	0.971	0.964	0.014	0.988
DTO	1	0	−2.706	−2.775	−5.105	3.526

Unfortunately, we were unable to recreate their results. We were able to recreate their results for equity LETFs, but we could not do so for our commodity LETFs.

5 Conclusions

We can reasonably conclude that ETF returns are able to be modeled as a stochastic process with mean-reversion. Unfortunately we couldn't recreate the results from Leung and Santoli, but our hypothesis is that the empirical leverage rations and volatility coefficients are less accurate for volatile commodity ETFs than for more stable equity ETFs.