VIX Future Statistical Trading Strategy

Zhilin Liu, Yi Shan, Zetian Sun

December 24, 2019

Contents

1 Introduction

The purpose of this project is to create a strategy to trade the Volatility Index (VIX) futures, and in brief, the strategy trades only close-to-maturity VIX futures (VX). In order to accomplish this, five time series models (ARIMA, ARIMA-GARCH, Heston, Ornstein–Uhlenbeck process,ARIMA-GJRGARCH) are tested first. After training these models with historical data of VIX, the validation process uses AIC as the main criteria to do the comparison. Finally this project decides to simulate VIX with Ornstein–Uhlenbeck model (O-U) that gives the highest accuracy on fitting VIX.

For a specific VX contract, this project applies the chosen O-U model and Monte-Carlo simulation on VIX, using the VX data a week (five trading days) before the maturity to predict the price of VIX at the maturity. For instance, if the contract matures on Dec 24^{th} 2019, this strategy will train an O-U model with VIX data before Dec 17^{th} (a week before Dec 24^{th}) and use this specific O-U model to randomly simulate 3.2×10^6 paths to predict the value of VIX on Dec 24^{th}. The expected value of the 3.2×10^6 simulated results is thus the expected future price.

With the analysis and forecasting result above, if the current future price is significantly higher than the expected future price at maturity, the strategy in this paper proposes to open a short position; if the the current future price is significantly lower than the expected future price, the strategy opens a long position. This specific trading strategy generates an annualized return of 54.3%, based on the simulation.

2 Basic Assumptions and Notations

2.1 Assumptions

- Assume sufficient market liquidity, because of lacking high frequency data. The problem of this assumption will be analyzed in Section 5.3.
- Assume zero transaction cost. because of the low-frequency trading nature of this strategy, this assumption is safe.
- This trading strategy has trivial impact on the market (will not affect VIX/VX price).
- Impact of risk-free return and dividend can be ignored, since this strategy will hold asset(s) or simulate VIX price for only five trading days.
- Time series of volatility, VIX in specific, is a(n) (integrated) weak-stationary process that has a constant long-term expectation and mean-reversion property.

2.2 Notations

In this project, $\{X_t\}$ is a discrete time series symbolize VIX, and $\forall t \in \mathbb{Z}^+$:

- X_t = VIX at time t

- $\nabla X_t = X_t - X_{t-1}$ is the first difference operation
- i.i.d. stands for independent and identically distributed

3 VIX Prediction Models

3.1 Introduction to VIX

To make it concise, VIX is the weighted sum of the implied volatility of the options on SPY (ETF that tracks the price of S&P500); therefore, VIX is simply the implied volatility of S&P500. In this case, it can be safely assumed that VIX, or its first difference, is a weak-stationary process; moreover, as a kind of volatility, VIX has mean-reversion property, and it will come back to its mean once in a while. There are different potential methods to model the time series of VIX, but the core foundation is ARIMA, GARCH and mean-reversion. This project will explain ARIMA, ARIMA-GARCH, ARIMA-GJRGARCH, Heston and O-U process, and compare the AIC of these models on fitting VIX.

3.2 Models

3.2.1 ARIMA

The most fundamental and straightforward model for $\{X_t\}$ is ARIMA(1,1,1).

$$\nabla X_t = \alpha \nabla X_{t-1} + Z_t + \beta Z_{t-1},$$

where $\{Z_t\}$ is white noise, and Z_1, Z_2, Z_3 ... are i.i.d. standard normal distribution with mean 0 and standard deviation 1.

Blow is the ACF (Figure 1) and PACF (Figure 2) of $\{X_t\}$, and PACF justifies that the process is an AR(1) process.

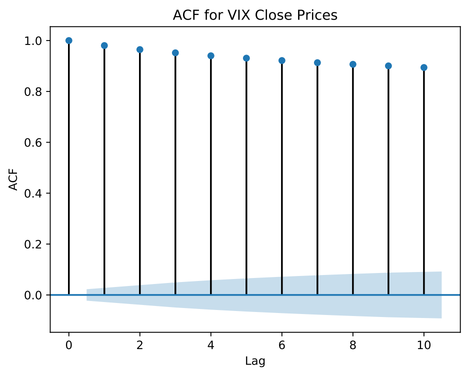

Figure 1: ACF for VIX

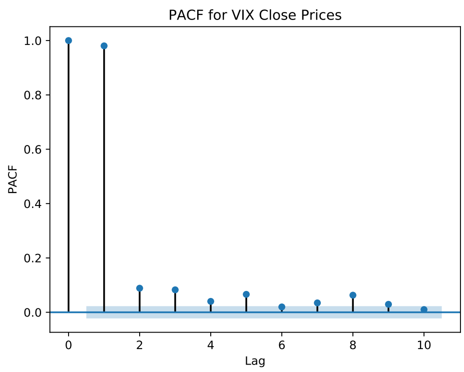

Figure 2: PACF for VIX

Also, comparing the graphs of $\{X_t\}$ (Figure 3) and $\{\nabla X_t\}$ (Figure 4), clearly the first difference graph looks more stationary than $\{X_t\}$. $\{X_t\}$ tends to have dramatic reaction sometimes, although most part of $\{X_t\}$ tends vibrate among 10 to 30. On the other hand, the graph of $\{\nabla X_t\}$ looks more like a stationary process with mean 0. In this case, VIX is probably a Integrated-1 process, and ARIMA(1,1,q) is a potential model to be considered.

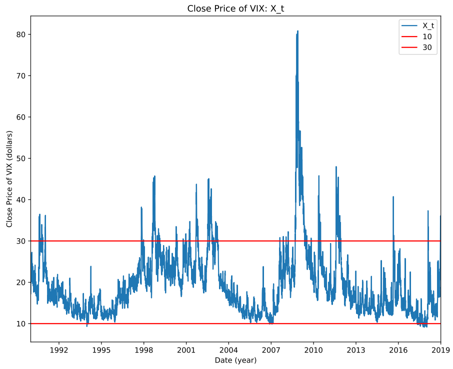

Figure 3: VIX Close Price

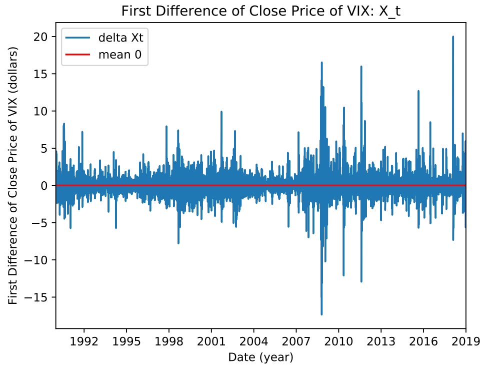

Figure 4: 1^{st} Diff of VIX Close Price

3.2.2 ARIMA-GARCH(1,1,1)

The second model, ARIMA-GARCH(1,1,1),is based on ARIMA(1,1,1) and brings GARCH model into account:

$$\nabla X_t = \beta_0 + \beta_1 \nabla X_{t-1} + a_t + \theta a_{t-1}$$

where

$$a_t = \sigma_t \varepsilon_t \qquad \sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \alpha_2 \sigma_{t-1}^2$$

and $\varepsilon_1, \varepsilon_2, \varepsilon_3$ ... are i.i.d. standard normal with mean 0 and standard deviation 1.

This model can be considered as two part. First is the auto-regression part on the integrated time series:

$$\nabla X_t = \beta_0 + \beta_1 \nabla X_{t-1}$$

which models the expected value of ∇X_t. The second part of the model, moving-average part on $\{a_t\}$, is then the residuals of the first part. GARCH model is applied on the residual a_t in order to account the volatility σ_t into consideration. That is why this model is an extension of the classical ARIMA model.

3.2.3 ARIMA-GJRGARCH

ARIMA-GJRGARCH is another potential model evaluated in this paper. This model is extended from the GARCH model, instead of simply defining σ_t as a symmetry process such as:

$$\sigma_{GARCH_t}^2 = \alpha_0 + \alpha_1 a_{GARCH_{t-1}}^2 + \alpha_2 \sigma_{GARCH_{t-1}}^2$$

The above GARCH model, the sign of a_t does not affect the model, but in reality, market has a more dramatic reaction to drops. GJR model captures this market psychology, and its volatility σ follows the below equation:

$$\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \gamma_1 I_{a_{t-1}} a_{t-1}^2 + \alpha_2 \sigma_{t-1}^2$$

where $I_{a_{t-1}}$ is an indicator function defined as:

$$I_{a_{t-1}} = \begin{cases} 1 & \text{if } a_{\text{t-1}} \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

since $\gamma_1 \geq 0$, σ_t will have a relatively more dramatic reaction if a_{t-1} is positive.

The GJR model for the volatility of VIX is different from GJR for the volatility of stocks. Stock volatility has larger sensitivity to stock decrease, but volatility of VIX has larger sensitivity to dramatic increase in VIX, since VIX has its fame as fear gauge, and it increases significantly when S&P500 decreases during bear market.

3.2.4 Heston Model

The Heston approach is a totally different approach comparing to ARIMA, and it tries to fit VIX into the following stochastic differential equation:

$$dX_t = \lambda(\mu - X_t)dt + \sigma \cdot \sqrt{X_t} dW_t$$

or, in discrete form:

$$\nabla X_t = \lambda(\mu - X_{t-1}) + \sigma \cdot \sqrt{X_t} \cdot Z_t$$

where $\{Z_t\}$ is white noise, and Z_1, Z_2, Z_3 ... are i.i.d. standard normal distribution with mean 0 and standard deviation 1.

The main contribution of this model is to capture the mean-reversion property of VIX. With mean μ:

$$\mathbb{E}[\nabla X_t] = \begin{cases} \text{positive} & \text{if } \mu \geq \text{X}_{\text{t-1}} \\ \text{zero or negative} & \text{otherwise} \end{cases}$$

In other words, if X_{t-1} is below the average μ, ∇X_t is expected to be positive and X_t is increasing in this time period. Vice versa. This explains the mean-reversion property perfectly.

3.2.5 O-U Process

The O-U model of VIX is an alternative of Heston model. Instead of dealing with X_t, O-U model on VIX models $\ln X_t$:

$$d \ln X_t = \lambda(\mu - \ln X_t)dt + \sigma dW_t$$

which is equivalent to the discrete form:

$$\nabla \ln X_t = \lambda(\mu - \ln X_{t-1}) + \sigma Z_t$$

and $\{Z_t\}$ i.i.d. white noise as usual.

Looking at the graph of $\{X_t\}$ (Figure 5) and $\{\ln X_t\}$ (Figure 6), the second graph looks clearly more stationary (vibrating around the mean without dramatically large events) than the first one. This explains that while modeling with $\{\ln X_t\}$, the O-U model assumes a constant volatility, but in the Heston Model, the volatility changes regarding to the value of $\sqrt{X_t}$.

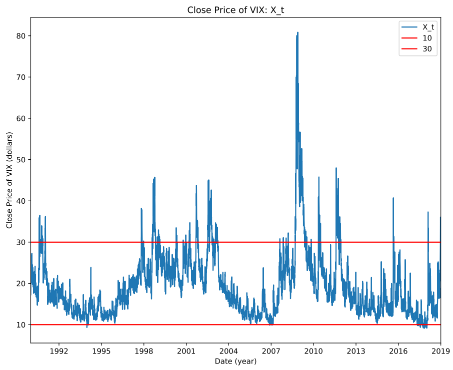

Figure 5: VIX Close Price

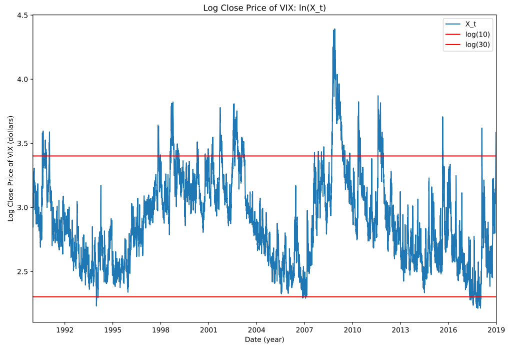

Figure 6: Log VIX Close Price

3.3 Model Training, Selection and Validation

Applying Maximum Log-Likelihood approach, this project first trains the five models with the VIX data from January 1990 - December 2013. For each of the model, the likelihood probability density $L(X)$ below can be derived from Baye's Formula:

$$
\begin{aligned}
L(X_n, X_{n-1}, X_{n-2}...X_0) &= L(X_n, X_{n-1}, X_{n-2}...X_1|X_0) \cdot L(X_0) \\
&= L(X_n, X_{n-1}, X_{n-2}...X_2|X_1, X_0) \cdot L(X_1|X_0) \cdot L(X_0) \\
&= L(X_n|X_{n-1}, X_{n-2}...X_0) \cdot L(X_{n-1}|X_{n-2}, X_{n-3}...X_0)... \\
&\cdot L(X_1|X_0) \cdot L(X_0)
\end{aligned}
$$

Using O-U Process as an example. since $\forall$ t $\in \{1, 2, 3...n\}$

$$\nabla \ln X_t \sim \mathcal{N}(\lambda(\mu - \ln X_{t-1}), \sigma^2)$$

$$
\begin{aligned}
L_{O-U}(X_t|X_{t-1}, X_{t-2}...X_0) &= L_{O-U}(X_t|X_{t-1}) \\
&= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\nabla \ln X_t - \lambda(\mu - \ln X_{t-1}))^2}{2\sigma^2}\right)
\end{aligned}
$$

Therefore:

$$
\begin{aligned}
&\log L_{O-U}(\lambda, \mu, \sigma|X_n, X_{n-1}, X_{n-2}...X_0) \\
&\qquad = -\frac{n}{2}\log 2\pi - n \cdot \log \sigma - \frac{1}{2}\sum_{i=0}^{n} \frac{(\nabla \ln X_t - \lambda(\mu - \ln X_{t-1}))^2}{\sigma^2}
\end{aligned}
$$

Maximizing L is equivalent to minimizing $-\log L$. Through this maximum likelihood approach, the parameters of each models can be trained.

These five trained model then are applied on data from January 2014 to December 2016 for validation purpose. The AIC of both training and validation sets of all five models are attached (Figure 7):

AIC	ARIMA(1,1,1)	ARIMA(3, 1, 3)	ARIMA-GARCH	ARIMA-GJRGARCH	Heston	OU process
Train	19239.6	19205.3	43062.8	41774.7	9622.37	-24619
Validation	4416.34	6160.85	11883.1	11519.3	2912.53	-6102.5

Figure 7: AIC for Five Models

The Python ARIMA package automatically fit $\{X_t\}$ into ARIMA(3,1,3) model, but comparing to ARIMA(1,1,1) model, they have similar AIC for the training set, but ARIMA(1,1,1) clearly has a smaller AIC for the validation set. This suggests that ARIMA(3,1,3) is over-fitting, and as discussed in Section 3.2.1, ARIMA(1,1,1) is probably the better ARIMA model to fit VIX.

For similar reasons, ARIMA-GARCH and ARIMA-GJRGARCH also might have over-fitting issue. It is questionable whether GARCH family model is a good option to model the volatility of VIX based on empirical data. Since VIX is already the volatility of S&P500, the volatility of VIX is supposed to have a somehow relationship with S&P500, but not necessarily follows GARCH formulas.

Finally, Heston and O-U process gives the best performance based on AIC, this is probably because both of these models capture the mean-reversion property of VIX better than ARIMA. In this case, this project decideds to use O-U model to simulate VIX.

4 VIX Future Trading Strategy

4.1 VX Pricing

VX is the future contract of VIX. VX pricing is different from the traditional futures. Because of law of one price, traditional future is normally priced as:

$$F_t = F_0 \cdot e^{(r-q)t}$$

where r is the risk-free return, q is the dividend yield and t is time to maturity. While pricing tradition future, people are assuming that the value of the underlying asset will not change besides the effect of interest rate and dividend.

On the other hand, VIX future cannot be simply priced as the above equation. Because of the mean-reversion property of VIX, VIX tends to increase if the current VIX value is below the mean, and vice versa. A legit pricing model for VX has to capture this mean-reversion property.

Recall the O-U Model of VIX and back operator:

$$\begin{aligned}
\nabla \ln X_t &= \lambda(\mu - \ln X_{t-1}) + \sigma Z_t \\
\ln X_{t+1} &= \nabla \ln X_{t+1} + \ln X_t
\end{aligned}$$

Combining the above equations, it is possible to predict the price of VIX using Euler method:

$$\ln X_{t+1} = \lambda\mu + (1-\lambda)\ln X_t + \sigma Z_t$$

which is equivalent to:

$$X_{t+1} = e^{\lambda\mu + (1-\lambda)\ln X_t + \sigma Z_t}$$

Since Z_t is the white noise, Monte-Carlo simulation can simulate the distribution of $X_{n+t}|X_0, X_1, X_2 ... X_t$ as in figure 8, where the green line is the expectation and the grey shadow is its distribution. Let

$$F_{n,t} = \mathbb{E}[X_{t+n}|X_0, X_1, X_2 ... X_t]$$

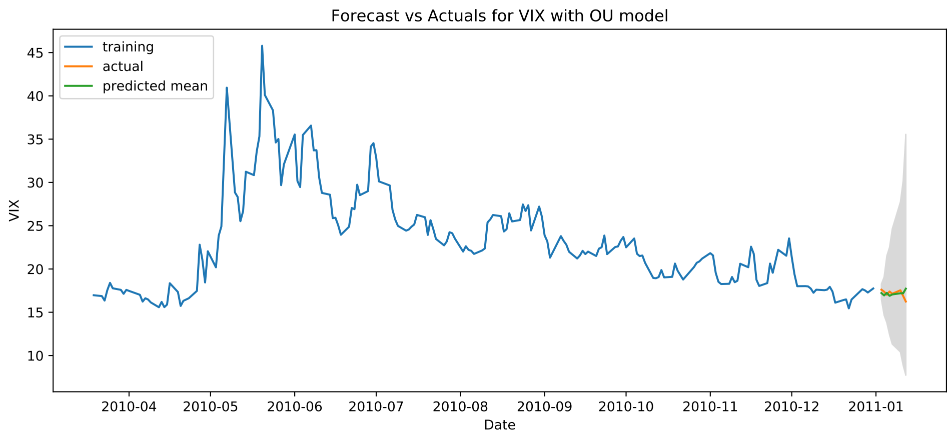

Figure 8: O-U Process Prediction

$F_{n,t}$ is the theoretical future price (the green line in Figure 8) for VIX matures in n days at time t, based on Monte-Carlo simulation and O-U model. Detailed steps and example of the simulation will be given in Section 4.3.

4.2 Trading Strategy

The rest of the strategy is straightforward. At time t, assume a VX contract matures in n days (close-to-maturity) is traded at price F:

$$\begin{cases} \text{open a long position} & \text{if } F \leq F_{n,t} - p\sigma_t \\ \text{open a short position} & \text{if } F \geq F_{n,t} + p\sigma_t \\ \text{no action} & \text{otherwise} \end{cases}$$

where

$$\sigma_t = \sqrt{var(X_{t+n}|X_0, X_1, X_2 ... X_t)}$$

is the standard deviation of $F_{n,t}$ generated by Monte-Carlo. p is the threshold value evaluated by validation.

In brief, if currently the future is traded p times σ_t above the theoretical price - $F_{n,t}$, the strategy suggests to open a short position; if the future is traded p times σ_t below $F_{n,t}$, then open a long position; otherwise, no action is needed. Once the position is started, this strategy will hold the future to maturity. n is suggested to be a small number (close to maturity future), so the Monte-Carlo simulation can be more accurate.

4.3 Strategy Simulation

This simulation uses VX future daily data from 1990 to 2013 to train the O-U model, and 2014 to 2016 to train and validate the threshold value p in the trading strategy. Finally, data from 2017 to 2019 are used to test the PnL of the strategy.

VX have futures mature on every Wednesday. This simulation considers purchasing future contracts five days before maturity. In other words, $n = 5$. For example, V19-VX40 contract matures on October 2^{nd} 2019. On September

25^{th} 2019 (five working days before the maturity), V19-VX40 is traded at $F = \$16.575$. Apply the trained O-U model and Euler method on \$16.575, 20 possible prices of V19-VX40 can be randomly simulated for September 26^{th} (four working days before maturity). Each of the simulated 20 possible prices on the 26^{th} can again, generate 20 paths for the 27^{th} (3 working days before maturity). This process continues and eventually $20^5 = 3.2 \times 10^6$ potential prices will be generated for October 2^{nd}. The tree figure below explains this simulation (Figure 9):

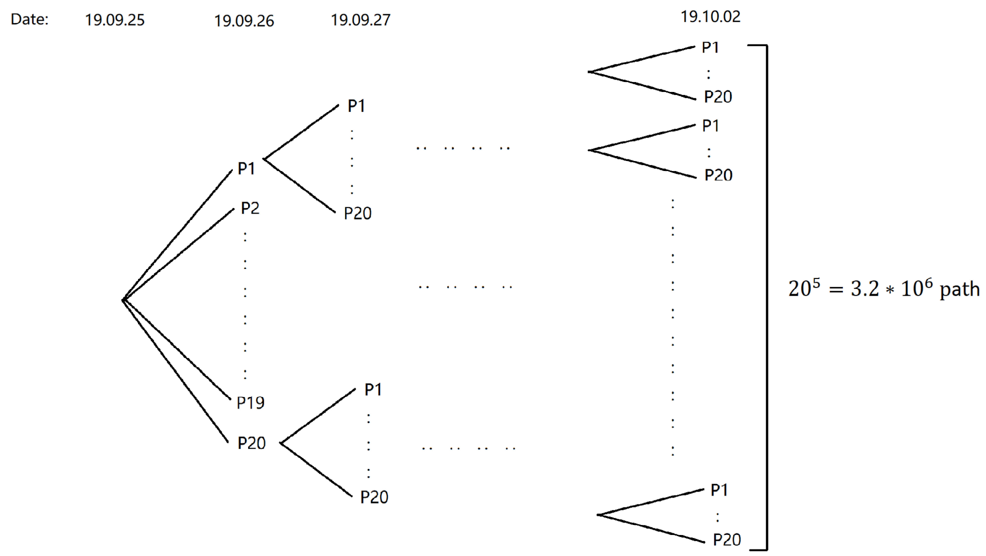

Figure 9: Simulation

The average of these 3.2×10^6 paths is $F_{5,Sep25} = \$19.573$, with a standard deviation $\sigma_{Sep25} = 2.234$. In other words, at September 25^{th}, a future contract matures in 5 days should worth \$19.573.

Based on the validation set, the threshold $p = 0.04$ gives the maximum profit. Since:

$$F = 16.575 \leq F_{5,Sep25} - p\sigma_{Sep25} = 19.573 - .04 \times 2.234$$

On September 25^{th} 2019, this simulation will long V19-VX40 contract, and hold this contract until its maturity. At the maturity, this contract is settled at 19.45, and a profit of $19.45 - 16.575 = \$2.875$ per contract has been made from this trading.

Applying this strategy through 2017-2019, a total profit of 361.2% can be made in three years, which is equivalent to an annualized profit of 53.4% Attached is the PnL graph (Figure 10):

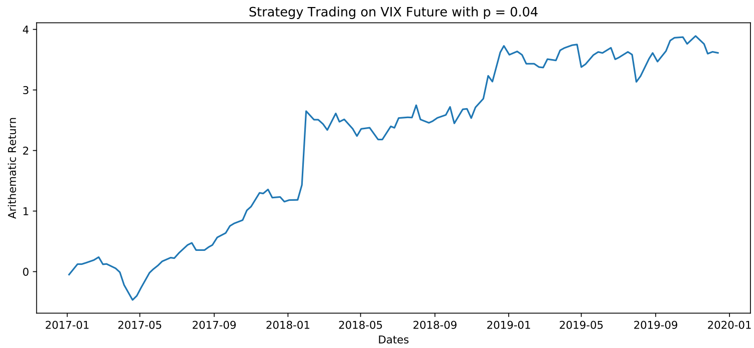

Figure 10: PnL

This strategy is indeed recommendable, but the annualized 53.4% profit too theoretical. its limitation will be discussed in more detail in the following Weakness Analysis section.

5 Conclusion

5.1 Strength and Weakness Analysis

Since the mean-reversion property of VIX is almost guaranteed, in long-term, this strategy is profitable in high confidence level. Because of high volatility of VIX, this strategy can bring significant profit to investors.

However, the simulation above is too good to be true, and there are several weaknesses need to be considered:

- This project only uses daily data, due to accessibility. High frequency data might improve the accuracy of VIX model.
- In the very beginning, sufficient market liquidity is assumed, but it is not true in the real world. VX has a remarkable liquidity, but it does not guarantee that each placed order can be filled.
- Dramatic events can barely be predicted by statistical model. About half of the profit in the simulation comes from sudden increase in VIX. The unpredictability of these events will cause huge trouble in trading. Another hedging strategy to avoid risk of dramatic events can significantly improve this strategy.

5.2 Conclusion

This strategy has its potential, but more detailed tests need to be performed if high frequency data is available. Developing another hedging strategy to avoid the risk of unforeseeable dramatic increase in VIX can significantly decrease the volatility of this strategy.

6 Bibliography

- Ahoniemi K.: Modeling and Forecasting the VIX Index, July 17, 2008
- Cboe Historical Data of VX futures https://markets.cboe.com/us/futures/market_statistics/historical_data/
- Detemple J., Osakwe C.: The valuation of volatility options. European Finance Review, 2000, 4, 21-50.
- Donninger, Chrilly, Modeling and Trading the VIX and VSTOXX with Futures, Options and the VXX, May 22, 2015
- Grünbichler A., Longstaff F.: Valuing futures and options on volatility. May 1995
- Smith W.: On the Simulation and Estimation of the Mean-Reverting O-U Process. Especially as Applied to Commodities Markets and Modeling. Version 1.01, February 2010