

# Review on Nonparametric Estimation of Financial Risk

## Nonparametric Estimation of Conditional VaR and Expected Shortfall

Young-geun Kim<sup>2,1,\*</sup>

20 Dec, 2019

### Abstract

Value at Risk (VaR) is one of many risk measures for financial assets. Expected shortfall is the other one. I investigate nonparametric estimation methods concerning to these two. Especially, I consider conditional information such as exogenous variable or past observed returns. Thus, our problem becomes estimating the conditional VaR (CVaR) and conditional expected shortfall (CES). It is widely known that local linear fitting involves in linear smoother. In this local linear scheme, we replace the smoother with weighted nadaraya watson kernel. This will estimate the conditional probability density. Given this estimator, we call the estimated CVaR and CES by weighted double kernel local linear estimator. It can also be shown that these estimators follow normal distribution asymptotically.

<sup>1</sup> Sungkyunkwan University

<sup>2</sup> Department of Statistics

\* Correspondence: Young-geun Kim <dudrms33@g.skku.edu>

## 1 Introduction

### 1.1 Concepts of Financial Risk

Consider any loss distribution. See Figure 1. This is the probability distribution of given time horizon. Value at risk (VaR) is the quantile for the right tail probability.

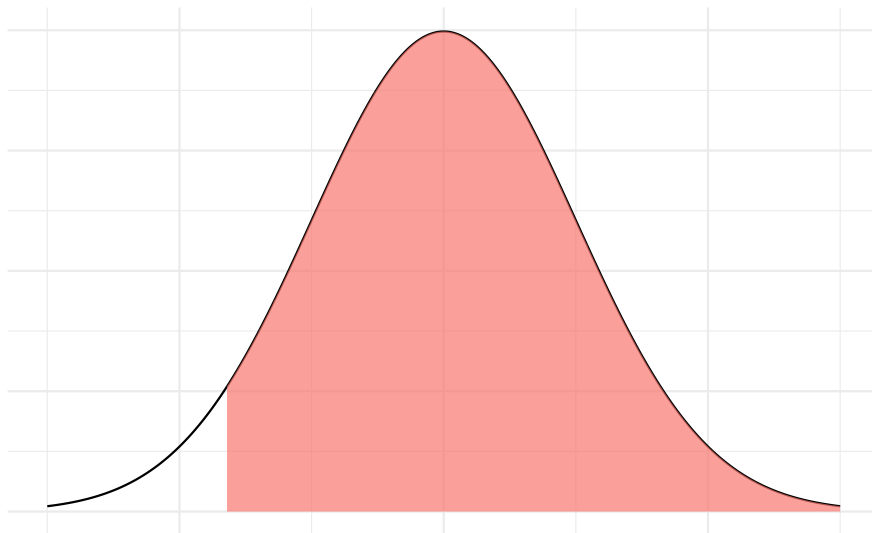


Figure 1: Loss Distribution - Can the financial institution still be in business after a catastrophic event?

Why do economists care about this kind of measures? Both financial institutions and regulatory committee should analyze the risk of their interesting portfolio. Tsay (2010) interprets this in the two viewpoint. For financial institutions, VaR can be read as maximal loss of a financial position during a given time period for a given probability. It leads to meaning the measure of loss under normal market conditions. For regulatory committee, on the other hand, it can be read as minimal loss under extraordinary market circumstances.

As we can see in the Figure 1, VaR is defined using the probability distribution of loss. Let  $p$  be the right tail probability, the red area in the figure. Let  $l$  be the time horizon, let  $L(l)$  be the loss function of the asset from  $t$  to  $t + l$ , and let  $F_l$  be the cdf of  $L(l)$ . Then

$$p = P[L(l) \geq VaR] \quad (1)$$

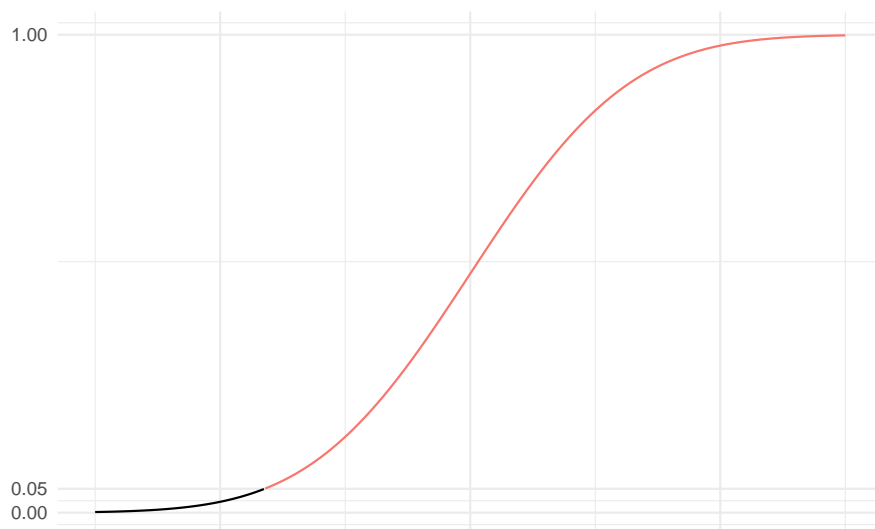


Figure 2: CDF of Loss

See Figure 2. VaR can be computed by finding the  $p$ -th quantile.

$$VaR = \inf \{x: F_l(x) \geq 1 - p\} \quad (2)$$

## 1.2 Calculating VaR

There are many ways to get VaR in parametric way. Equation (2) shows that VaR is the quantile in each time horizon. It is natural to employ quantile regression to get this quantile value.

This report consists of three parts. In Section 2, we covered basic concepts about CVaR and CES. In Section 3, I reviewed the paper (Cai and Wang 2008) briefly. Theoretical parts were mostly skipped. In Section 4, we applied the method in both simulation setting and real data.

## 2 Background

### 2.1 Expected Shortfall

The reason why many authors only cover these two value-at-risk or expected shortfall that I will explain. Value at risk is famous and its concept is simple. In market, however, sometimes two portfolios can be merged. When this happens, the risk measure should not be greater than the sum of each. This is called subadditivity. VaR does not satisfy this property. It is possible for VaR to underestimate the actual loss. Expected loss is

the one satisfying the condition. Expected shortfall is defined by the expected value of loss function if the VaR is exceeded.

$$ES = E[L(l) \mid L(l) \geq VaR] \quad (3)$$

In other words, while VaR cares the maximal loss in the right tail probability part, saying 0.95, ES cares average loss in the remaining part, 0.05.

## 2.2 Return

We prefer to use log return data. Let  $\{P_t\}$  be the price series. Then the log return is defined by  $\{Y_t := \ln \frac{P_t}{P_{t-1}}\}$ . Loss occurs when the return  $\{P_t - P_{t-1}\}$  are negative, so we should use the negative returns or negative log returns.

Consider Taylor expansion for this log function. For any  $x_0 > 0$ ,

$$\ln x \approx \ln x_0 + \frac{1}{x_0}(x - x_0) \quad (4)$$

Write  $x = x_2$ ,  $x_0 = x_1$ . Then

$$\ln \frac{x_2}{x_1} \approx \frac{x_2}{x_1} - 1 = \frac{x_2 - x_1}{x_1} \quad (5)$$

Observe that the log return approximates to the change rate. Cai and Wang (2008) used the following value in a real example part.

$$-100Y_{t+1} = -100 \ln \frac{P_{t+1}}{P_t}$$

which approximates to percentage loss.

## 2.3 Conditional Information

In econometrics, conditional information are always researchers' interests. For example, we are interested in the exogenous variables like economic or market variables. When we look at return data, we can condition past observed returns.

Let  $\{Y_t\}$  be stationary log returns and let  $\{X_t\}$  be conditional information series.

**Definition 2.1** (Conditional Value-at-Risk). Let  $F(y \mid x)$  be the conditional cdf of  $Y_t$  given  $X_t = x$  and let  $S(y \mid x) := 1 - F(y \mid x)$ . Then Conditional VaR is

$$\nu_p(x) := S^{-1}(p \mid x)$$

When formulating the conditional expected shortfall, we just add the term  $X_t = x$  in Equation (3). Let  $B \equiv \{\omega: Y_t(\omega) \mid X_t = x \geq \nu_p(x)\} \in \mathcal{B}$ . Then

$$\begin{aligned}
\mu_p(x) &= E[Y_t \mid Y_t \geq \nu_p(x), X_t = x] \\
&= \frac{1}{P(B)} \int_B Y_t dP \\
&= \frac{1}{P(Y_t \geq \nu_p(x) \mid X_t = x)} \int_{\nu_p(x)}^{\infty} y f(y \mid x) dy \\
&= \frac{1}{p} \int_{\nu_p(x)}^{\infty} y f(y \mid x) dy
\end{aligned} \tag{6}$$

### 3 Nonparametric Estimation

The workflow of estimating is very simple. It uses the formulation of each CVaR and CES. Just put estimating notation in Equation (6).

$$\hat{\mu}_p(x) = \frac{1}{p} \int_{\hat{\nu}_p(x)}^{\infty} y \hat{f}(y \mid x) dy \tag{7}$$

Observe  $\hat{\mu}_p(x)$  and  $\hat{f}(y \mid x)$ . After estimating these two, we will plug the above integration and get the result. Then we can summarize the process as follows.

1. Estimate the conditional pdf  $\hat{f}(y \mid x)$ .
2. Estimate the conditional cdf  $\hat{F}(y \mid x)$ .
3. Invert the conditional cdf, and get  $\hat{\nu}_p(x)$ .
4. Plugging, and get  $\hat{\mu}_p(x)$ .

#### 3.1 Double Kernel Local Linear

Consider any symmetric kernel  $K_h(\cdot)$ . Then by Taylor expansion,

$$\begin{aligned}
E[K_{h_0}(y - Y_t) \mid X_t = x] &= K_{h_0} * f_{y|x}(y) \\
&= f(y \mid x) + \frac{h_0^2}{2} \mu_2(K) f^{(2)}(y \mid x) + o(h_0^2)
\end{aligned} \tag{8}$$

where  $\mu_j(K) = \int_{\mathbb{R}} u^j K(u) du$ . Thus, we now have a smoothing problem

$$f(y \mid x) \approx E[K_{h_0}(y - Y_t) \mid X_t = x] \tag{9}$$

Different with the other usual smoothing setting, the response becomes  $Y_t^*(y) \equiv K_{h_0}(y - Y_t)$ . Given this, implement local linear least squares.

$$\hat{f}(y \mid x) = \operatorname{argmin}_{\alpha(x), \beta(x)} \sum_{t=1}^n W_h(x - X_t) [Y_t^*(y) - \alpha(x) - \beta(x)(X_t - x)]^2 \tag{10}$$

This is called double local linear in that the problem involves in the two kernel,  $K_{h_0}$  and  $W_h$ . Recall that the local linear estimate is equivalent to the weighted least squares. Let  $\mathbf{Y}_y^* = (Y_1(y), \dots, Y_n(y))^T \in \mathbb{R}^n$ , let  $\mathbf{b}_x(x_t) := (1, x_t - x)^T \in \mathbb{R}^2$ , let  $\mathbf{b}_x(x) = \mathbf{e}_1 := (1, 0)^T$ , let  $X_x := (\mathbf{b}_x(x_i)^T) \in \mathbb{R}^{n \times 2}$ , and let  $W_x := \operatorname{diag}(W_h(x - X_j)) \in \mathbb{R}^{n \times n}$ . Then the local linear solution is given by  $\hat{f}_l = \hat{\alpha}$ ,

$$\begin{aligned}
\hat{f}_{ll}(y | x) &= \mathbf{e}_1^T (X_x^T W_x X_x)^{-1} X_x^T W_x \mathbf{Y}_y^* \\
&= \mathbf{l}(x)^T \mathbf{Y}_y^* \\
&\equiv \sum_{t=1}^n l_t(x) Y_t^*(y)
\end{aligned} \tag{11}$$

Cai and Wang (2008) provides the exact form of each element by matrix calculation in the paper. This linear form of pdf easily gives its cdf. This LL estimator has some good properties such as differentiability. The conditional cdf can be calculated by

$$\begin{aligned}
\hat{F}_{ll}(y | x) &= \int_{-\infty}^y \hat{f}_{ll}(y | x) dy \\
&= \sum_{t=1}^n l_t(x) G_{h_0}(y - Y_t)
\end{aligned} \tag{12}$$

where  $G(\cdot)$  is the cdf of  $K(\cdot)$ . Since it is cdf, so it must be  $\hat{F}_{ll} \in [0, 1]$  and monotone increasing. Double local linear, however, does not guarantee these properties.

### 3.2 Weighted Nadaraya Watson

Since the above estimator cannot give a desirable cdf to us, we consider another method, weighted nadaraya watson suggested by Cai (2001). The form is similar to the nadaraya watson estimator.

$$\hat{F}_{NW}(y | x) = \sum_{t=1}^n W_{c,t}(x, \lambda) I(Y_t \leq y) \tag{13}$$

where the WNW weights given by

$$W_{c,t}(x, h) = \frac{p_t(x) W_h(x - X_t)}{\sum_{i=1}^n p_i(x) W_h(x - X_i)} \tag{14}$$

The term “weighted” is due to the weight for each kernel.  $\{p_t(x)\}$  is chosen by

$$p_t(x) = \frac{1}{n [1 + \lambda(X_t - x) W_h(x - X_i)]} \geq 0 \tag{15}$$

We should find  $\lambda$  maximizing the empirical log likelihood function

$$L_n(\lambda) = - \sum_{t=1}^n \ln [1 + \lambda(X_t - x) W_h(x - X_i)] \tag{16}$$

Cai (2001) recommended employing Newton-Raphson to find it.

### 3.3 Weighted Double Kernel Local Linear

Weighted Nadaraya Watson estimator for conditional cdf (Equation (13)) can fill the defect of local linear (Equation (12)). In Equation (11), Cai and Wang (2008) can replace the linear smoother with WNW weights. This is called the weighted double kernel local linear estimator.

$$\hat{f}_{cai}(y | x) = \sum_{t=1}^n W_{c,t}(x, h) Y_t^*(y) \quad (17)$$

It proceeds in a similar way that the conditional cdf can be estimated by

$$\hat{F}_{cai}(y | x) = \sum_{t=1}^n W_{c,t}(x, h) G_{h_0}(y - Y_t) \quad (18)$$

The next step is estimating  $\hat{\nu}_p(x)$  and  $\hat{\mu}_p(x)$  one by one. Let  $\hat{S}_{cai}(y | x) = 1 - \hat{F}_{cai}(y | x)$ . By inverting this function, we can get the CVaR function.

$$\hat{\nu}_p^{(cai)}(x) = \hat{S}_{cai}^{-1}(p | x) \quad (19)$$

When inverting it, I used Equation (2). Finally, we plug  $\hat{\nu}_p(x)$  and  $\hat{f}_{cai}(y | x)$  into Equation (7). Then

$$\hat{\mu}_p(x) = \frac{1}{p} \sum_{t=1}^n W_{c,t}(x, h) [Y_t \bar{G}_{h_0}(\hat{\nu}_p(x) - Y_t) + h G_{1,h_0}(\hat{\nu}_p(x) - Y_t)] \quad (20)$$

where  $\bar{G}(u) = 1 - G(u)$  and  $G_1(u) = \int_u^\infty v K(v) dv$ .

### 3.4 Asymptotic Normality

Cai and Wang (2008) has shown that each estimator follows Normal distribution at both interior and boundary point. Since the formulation is quite complicated, I provide Monte Carlo simulation setting here. Consider AR(1)-GARCH(0,1) model as in the paper.<sup>1</sup> This setting continues to the next simulation section.

**Example 3.1** (AR(1)-GARCH(0,1) Model). With  $X_t = Y_{t-1}$ ,

$$\begin{cases} Y_t = 0.01 + 0.62X_t + \sigma_t \epsilon_t \\ \sigma_t^2 = 0.15 + 0.65\sigma_{t-1}^2 \\ \epsilon_t \sim N(0, 1) \end{cases}$$

Monte Carlo Samples were generate by the following procedure.

1. For fixed  $x$ ,
2. Generate GARCH(0, 1):  $(\sigma_t, \epsilon_t)$
3. Generate  $Y_t$  using AR(1) for each  $X_t = Y_{t-1}$
4. AR(1):  $Y_t = 0.01 + 0.62Y_{t-1} + \sigma_t \epsilon_t$

---

<sup>1</sup>Authors wrote that the model was ARCH(1), but many textbooks including Tsay (2010) specify that model as GARCH(0,1).

### 3.4.1 Interior

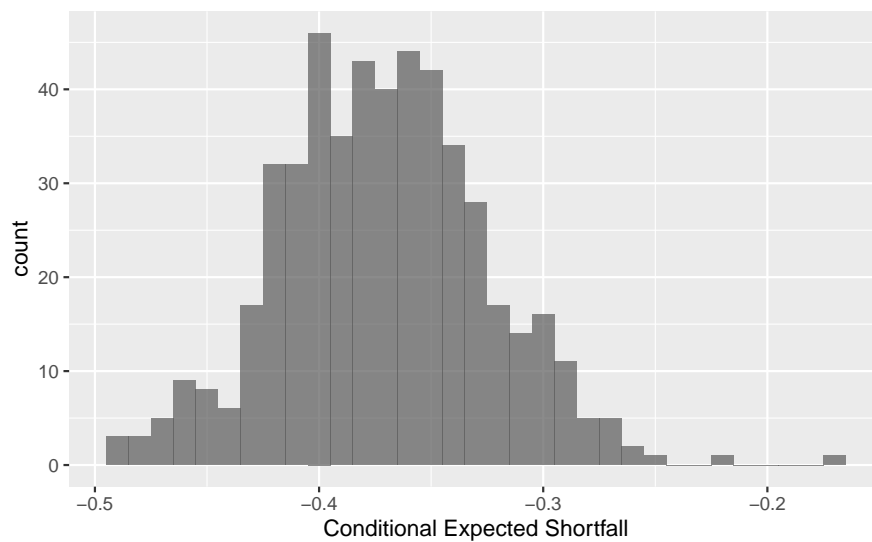


Figure 3: Asymptotic normality for  $\hat{\mu}_p(x)$  in interior point

In Figure 3, we can see that  $\hat{\mu}_p(x)$  follows the normal distribution approximately at  $x = -0.734$ .

### 3.4.2 Boundary

Cai and Wang (2008) proved at the left boundary point  $x = ch$ ,  $0 < c < 1$ .

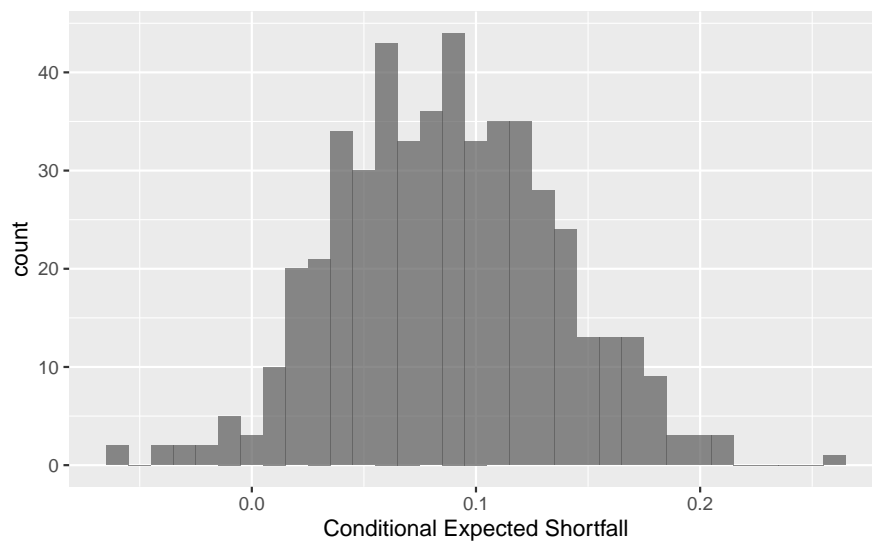


Figure 4: Asymptotic normality for  $\hat{\mu}_p(x)$  at left boundary

Similarly, Figure 4 presents asymptotic normality of WDKLL estimator of CES at the left boundary point.

## 4 Experiments

### 4.1 Simulation

Now we try another simulation under Example 3.1. This model implies the true CVaR and CES. Since  $\epsilon_t \sim N(0, 1)$ ,

$$\sigma_t \epsilon_t \sim N(0, \sigma_t^2) \quad (21)$$

By construction,

$$Y_t | X_t \sim N(0.01 + 0.62X_t, \sigma_t^2) \quad (22)$$

Given this conditional pdf, we can compute cdf using statistical package such as **R**. The procedure after this is same with before. In this Monte Carlo simulation setting, we want to simulate the expected prediction error between the true  $\mu_p(x)$  and  $\hat{\mu}_p(x)$ . Absolute loss was used following Cai and Wang (2008). Figure 5 shows the structure of MC simulation. After training each model using each MC sample, we test it in the independent test set. Individual samples and a test set are generated as follows.

1. For fixed  $x_t$  (pre-determined grid points)
2. Generate GARCH(0, 1):  $(\sigma_t, \epsilon_t)$
3. Generate  $Y_t$  using AR(1) for each  $X_t = Y_{t-1}$
4. AR(1):  $Y_t = 0.01 + 0.62Y_{t-1} + \sigma_t \epsilon_t$

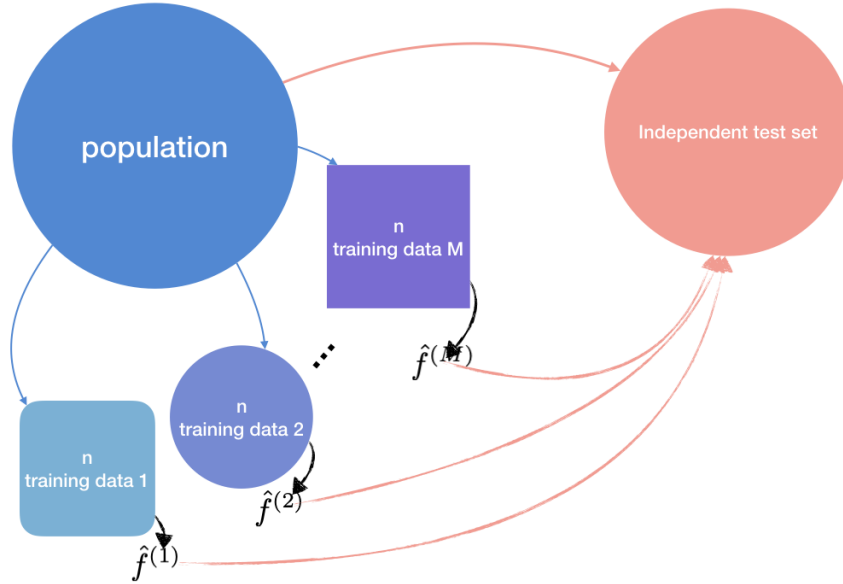


Figure 5: Simulating Expected prediction error

See Figure 6. WDKLL estimator produces much less expected prediction error in each bandwidth than usual Nadaraya Watson estimator.



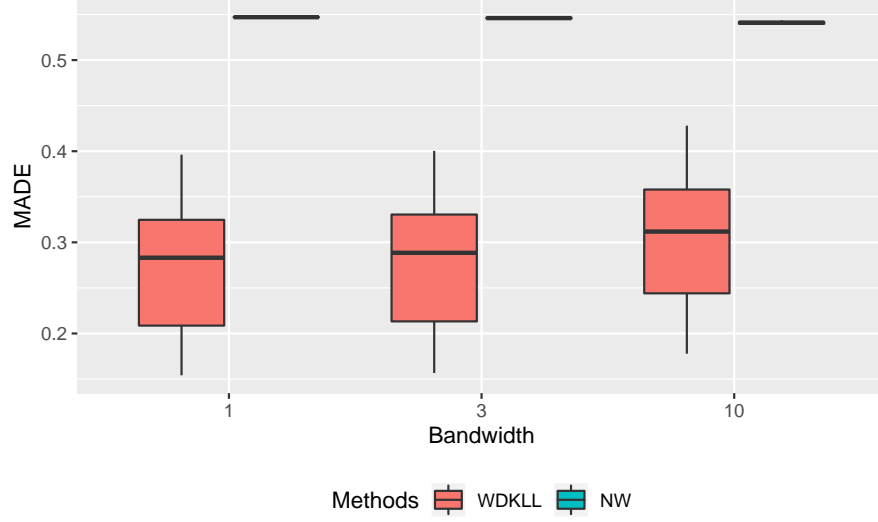


Figure 6: WDKLL versus Nadaraya-Watson

## 4.2 Data Analysis

We can implement CVaR and CES to measure risk of Bitcoin. We analyze 607 daily prices of bitcoin from 2018 – 04 – 05 to 2019 – 12 – 12.



Figure 7: Bitcoin price in USD

We use  $y_t := -100 \ln \frac{P_t}{P_{t-1}}$  as daily return to approximate percentage loss.

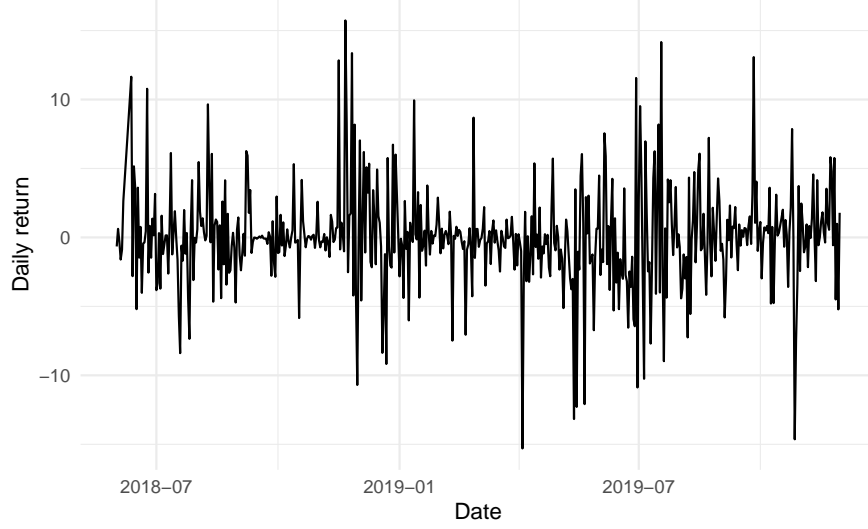


Figure 8: Daily return of bitcoin price

Figure 8 is the daily return  $\{y_t\}$ . We now try to estimate using this  $y_t$  given the first lagged variable  $x_t = y_{t-1}$ . Look at Figure 9. It is the plot of CVaR for each conditional information  $x_t = y_{t-1}$ . CVaR which is the maximal loss under 0.95 probability is increasing given the last day return. How about the other 0.05 case?

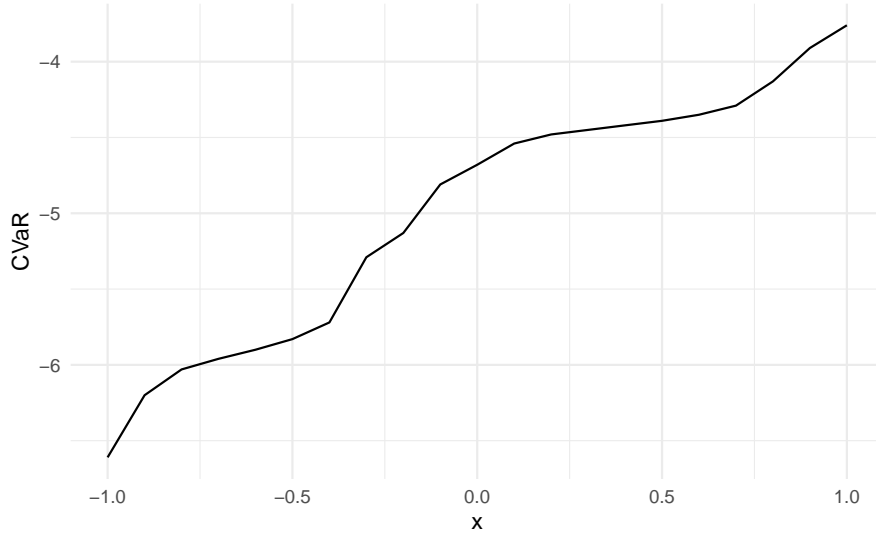


Figure 9: Conditional Value-at-Risk given each lagged variable value

See Figure 10. This is the conditional expected shortfall. Recall that we are dealing with negative returns. Thus, positive  $x_t$  means the loss. We can expect that its price is more likely to fall if there were a loss within the last day than if there was a same amount of positive return.<sup>2</sup>

<sup>2</sup>This interpretation have followed the one by Cai and Wang (2008)

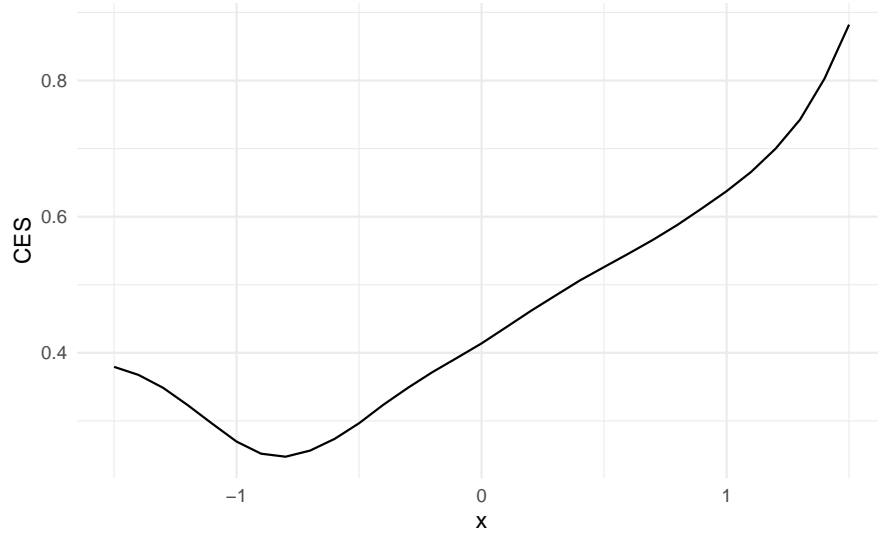


Figure 10: Conditional Expected Shortfall given each lagged variable value

## 5 Conclusion

### 5.1 Discussion

## References

- Cai, Zongwu. 2001. “Weighted Nadaraya–Watson Regression Estimation.” *Statistics & Probability Letters* 51 (3): 307–18.
- Cai, Zongwu, and Xian Wang. 2008. “Nonparametric estimation of conditional VaR and expected shortfall.” *Journal of Econometrics* 147 (1): 120–30. <https://doi.org/10.1016/j.jeconom.2008.09.005>.
- Tsay, Ruey S. 2010. *Analysis of Financial Time Series*. John Wiley & Sons.