

Bayesian regression and Bitcoin

สมาชิกกลุ่ม

นายกฤตพล ผ้ำเจริญ 6401012620161

นายเจษฎา ศรีจุลโพธิ์ 6401012620170

นายศุภกร พลศิริ 6401012620234

นายสิรภพ ห่วงวิไล 6401012630132

งานวิจัยที่นำมาอ้างอิง:

<https://arxiv.org/pdf/1410.1231v1.pdf>

code ที่นำมาอ้างอิง:

<https://github.com/stavros0/bitcoin-price-prediction>

Bitcoin

Bitcoin

Factors affecting the price of Bitcoin

- Market Demand and Supply
- Market Sentiment
- Global Economic Events

Latent Source

An unobservable or hidden variable that influences the observed data.



Bayesian Regression

Bayesian Regression

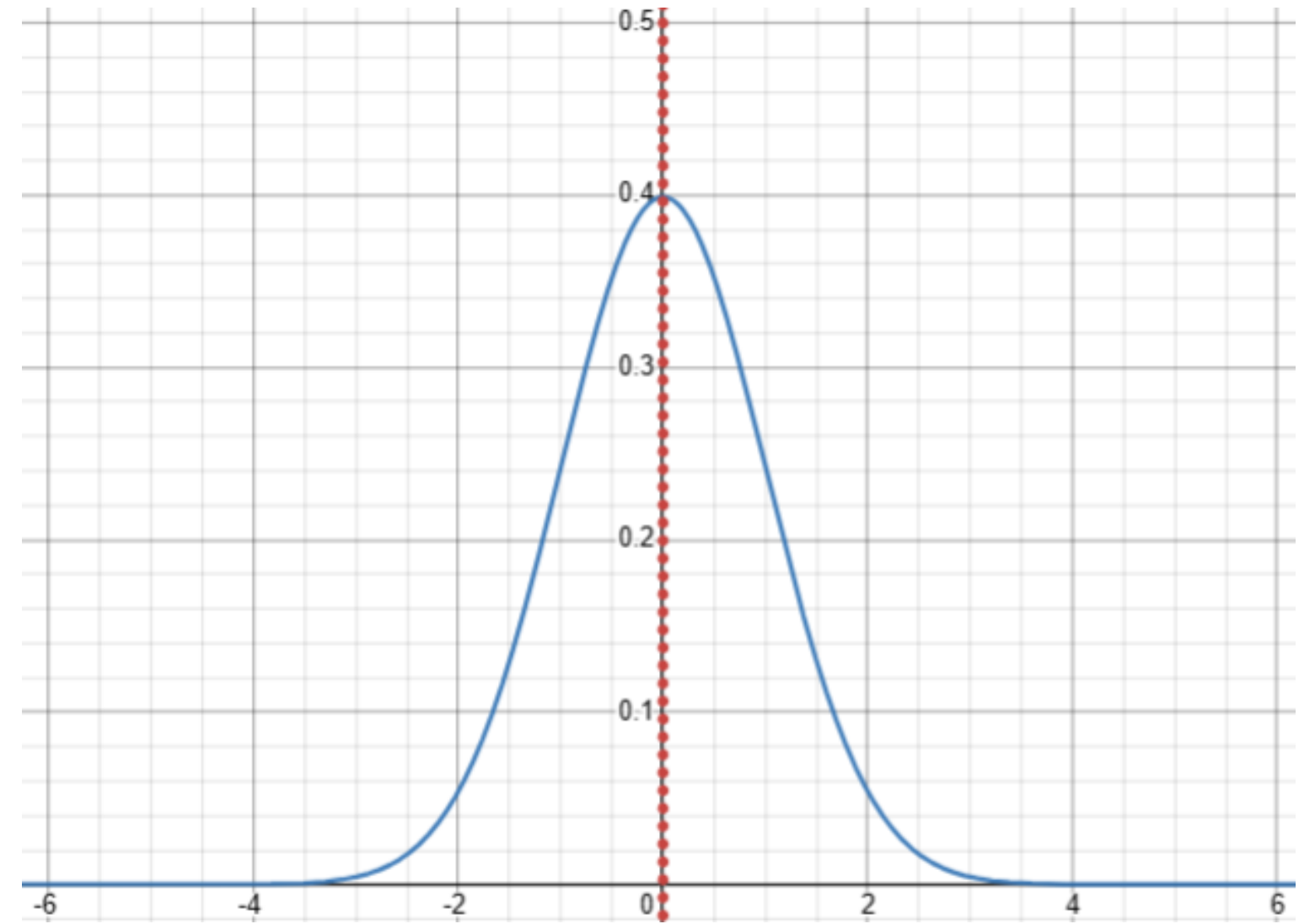
$$y = X\beta + \epsilon$$

$$y \sim N(X\beta, \sigma^2)$$

$$\epsilon \sim N(0, \sigma^2)$$

$$P(\beta|y, X) = \frac{P(y|\beta, X) * P(\beta | X)}{P(y|X)}$$

$$\textit{Posterior} = \frac{\textit{Likelihood} * \textit{Prior}}{\textit{Normalization}}$$



Latent source and Bayesian Regression

K distinct latent source s_1, \dots, s_K

latent probability μ_1, \dots, μ_K

latent distribution P_1, \dots, P_K

Sample index $T \in \{1, \dots, K\}$ with $P(T = k) = \mu_k$ for $1 \leq k \leq K$

$$\begin{aligned} P(y|x) &= \sum_{k=1}^T P(y|x, T = k) P(T = k|x) \\ &\propto \sum_{k=1}^T P(y|x, T = k) P(x|T = k) P(T = k) \\ &= \sum_{k=1}^T P_k(y) P(\epsilon = (x - s_k)) \mu_k \\ &= \sum_{k=1}^T P_k(y) \exp\left(-\frac{1}{2}\|x - s_k\|_2^2\right) \mu_k. \end{aligned} \quad (3)$$

Latent source and Bayesian Regression

K distinct latent source s_1, \dots, s_K

latent source probability μ_1, \dots, μ_K

latent distribution P_1, \dots, P_K

Sample index $1, \dots, K$ $P(T = k) = \mu_k$ for $1 \leq k \leq K$

$$\begin{aligned} P(y|x) &= \sum_{k=1}^T P(y|x, T = k) \mu_k \\ &\propto \sum_{k=1}^T P(y|x, T = k) \mu_k \\ &= \sum_{k=1}^T \exp(-\frac{1}{2}\|x - s_k\|_2^2) \mu_k. \end{aligned}$$

Latent source and Bayesian Regression

Conditional Probability

$$P(y|x) = \sum_{k=1}^T P_k(y) \exp \left(-\frac{1}{2} \|x - s_k\|_2^2 \right) \mu_k \quad (3)$$

Empirical Conditional Probability

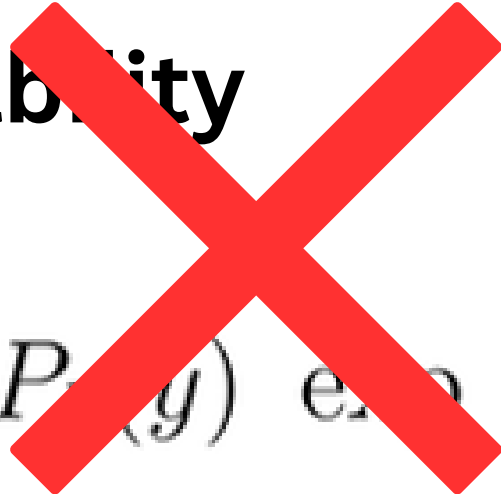
$$P_{emp}(y|x) = \frac{\sum_{i=1}^n 1(y = y_i) \exp \left(-\frac{1}{4} \|x - x_i\|_2^2 \right)}{\exp \left(-\frac{1}{4} \|x - x_i\|_2^2 \right)} \quad (4)$$

Binary Classification by (4)

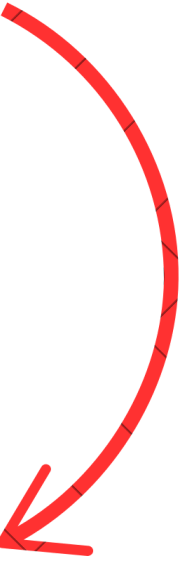
$$\frac{P_{emp}(y = 1|x)}{P_{emp}(y = 0|x)} = \frac{\sum_{i=1}^n 1(y_i = 1) \exp \left(-\frac{1}{4} \|x - x_i\|_2^2 \right)}{\sum_{i=1}^n 1(y_i = 0) \exp \left(-\frac{1}{4} \|x - x_i\|_2^2 \right)} \quad (5)$$

Latent source and Bayesian Regression

Conditional Probability


$$P(y|x) = \sum_{k=1}^T P(y) \exp \left(-\frac{1}{2} \|x - s_k\|_2^2 \right) \mu_k \quad (3)$$

Empirical Conditional Probability

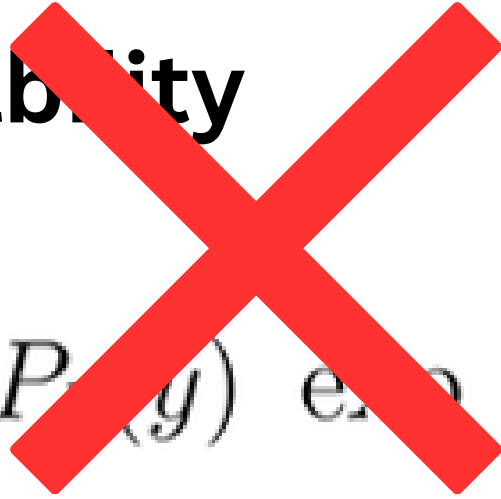

$$P_{emp}(y|x) = \frac{\sum_{i=1}^n 1(y = y_i) \exp \left(-\frac{1}{4} \|x - x_i\|_2^2 \right)}{\exp \left(-\frac{1}{4} \|x - x_i\|_2^2 \right)} \quad (4)$$

Binary Classification by (4)

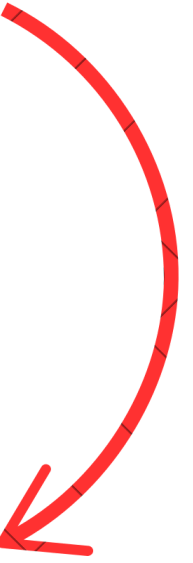
$$\frac{P_{emp}(y = 1|x)}{P_{emp}(y = 0|x)} = \frac{\sum_{i=1}^n 1(y_i = 1) \exp \left(-\frac{1}{4} \|x - x_i\|_2^2 \right)}{\sum_{i=1}^n 1(y_i = 0) \exp \left(-\frac{1}{4} \|x - x_i\|_2^2 \right)} \quad (5)$$

Latent source and Bayesian Regression

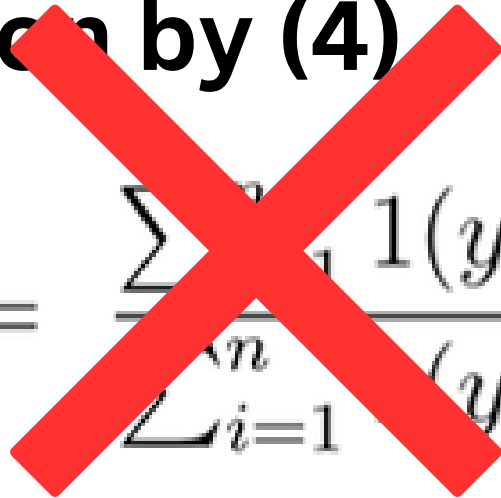
Conditional Probability


$$P(y|x) = \sum_{k=1}^T P(y) \exp \left(-\frac{1}{2} \|x - s_k\|_2^2 \right) \mu_k \quad (3)$$

Empirical Conditional Probability


$$P_{emp}(y|x) = \frac{\sum_{i=1}^n 1(y = y_i) \exp \left(-\frac{1}{4} \|x - x_i\|_2^2 \right)}{\exp \left(-\frac{1}{4} \|x - x_i\|_2^2 \right)} \quad (4)$$

Binary Classification by (4)


$$\frac{P_{emp}(y = 1|x)}{P_{emp}(y = 0|x)} = \frac{\sum_{i=1}^n 1(y_i = 1) \exp \left(-\frac{1}{4} \|x - x_i\|_2^2 \right)}{\sum_{i=1}^n 1(y_i = 0) \exp \left(-\frac{1}{4} \|x - x_i\|_2^2 \right)} \quad (5)$$

Latent source and Bayesian Regression

Empirical Conditional Probability

$$P_{emp}(y|x) = \frac{\sum_{i=1}^n 1(y = y_i) \exp\left(-\frac{1}{4}\|x - x_i\|_2^2\right)}{\exp\left(-\frac{1}{4}\|x - x_i\|_2^2\right)} \quad (4)$$

Estimate Expectation of y

$$E_{emp}[y|x] = \frac{\sum_{i=1}^n y_i \exp\left(-\frac{1}{4}\|x - x_i\|_2^2\right)}{\sum_{i=1}^n \exp\left(-\frac{1}{4}\|x - x_i\|_2^2\right)} \quad (6)$$

Latent source and Bayesian Regression

$$X(x) \in \mathbb{R}^n$$

$$\mathbf{y} \in \mathbb{R}^n \quad \text{with } i\text{th component being } y_i$$

$$X(x)_i = \exp \left(-\frac{1}{4} \|x - x_i\|_2^2 \right) / Z(x)$$

$$Z(x) = \exp \left(-\frac{1}{4} \|x - x_i\|_2^2 \right)$$

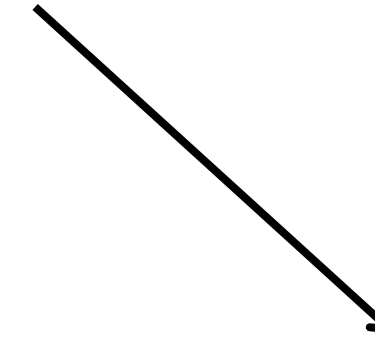
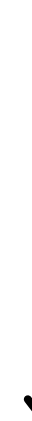
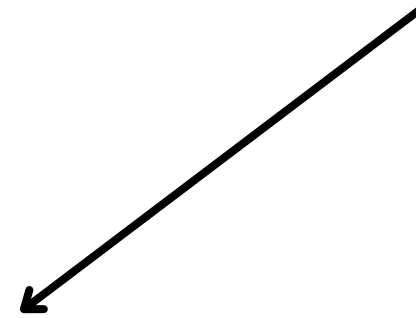
$$\hat{y} \equiv E_{emp}[y|x]$$

$$E_{emp}[y|x] = \frac{\sum_{i=1}^n y_i \exp \left(-\frac{1}{4} \|x - x_i\|_2^2 \right)}{\sum_{i=1}^n \exp \left(-\frac{1}{4} \|x - x_i\|_2^2 \right)}$$

$$\hat{y} = X(x)\mathbf{y}$$

Concept

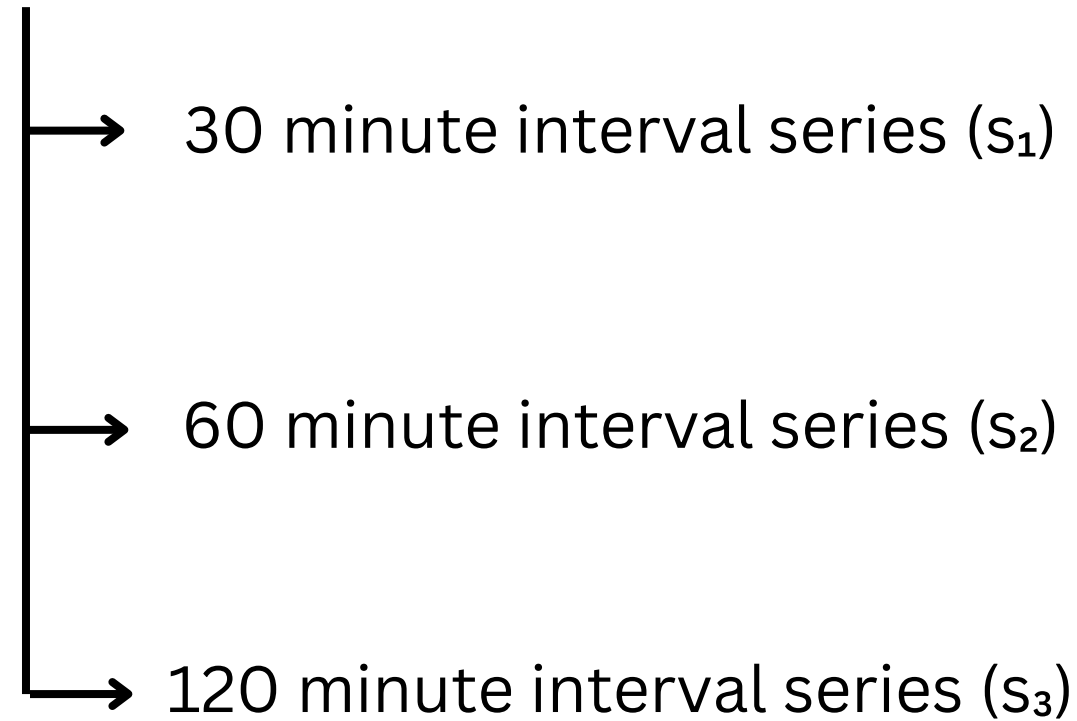
TimeSeriesData(3n)



TimeSeriesData1(n)

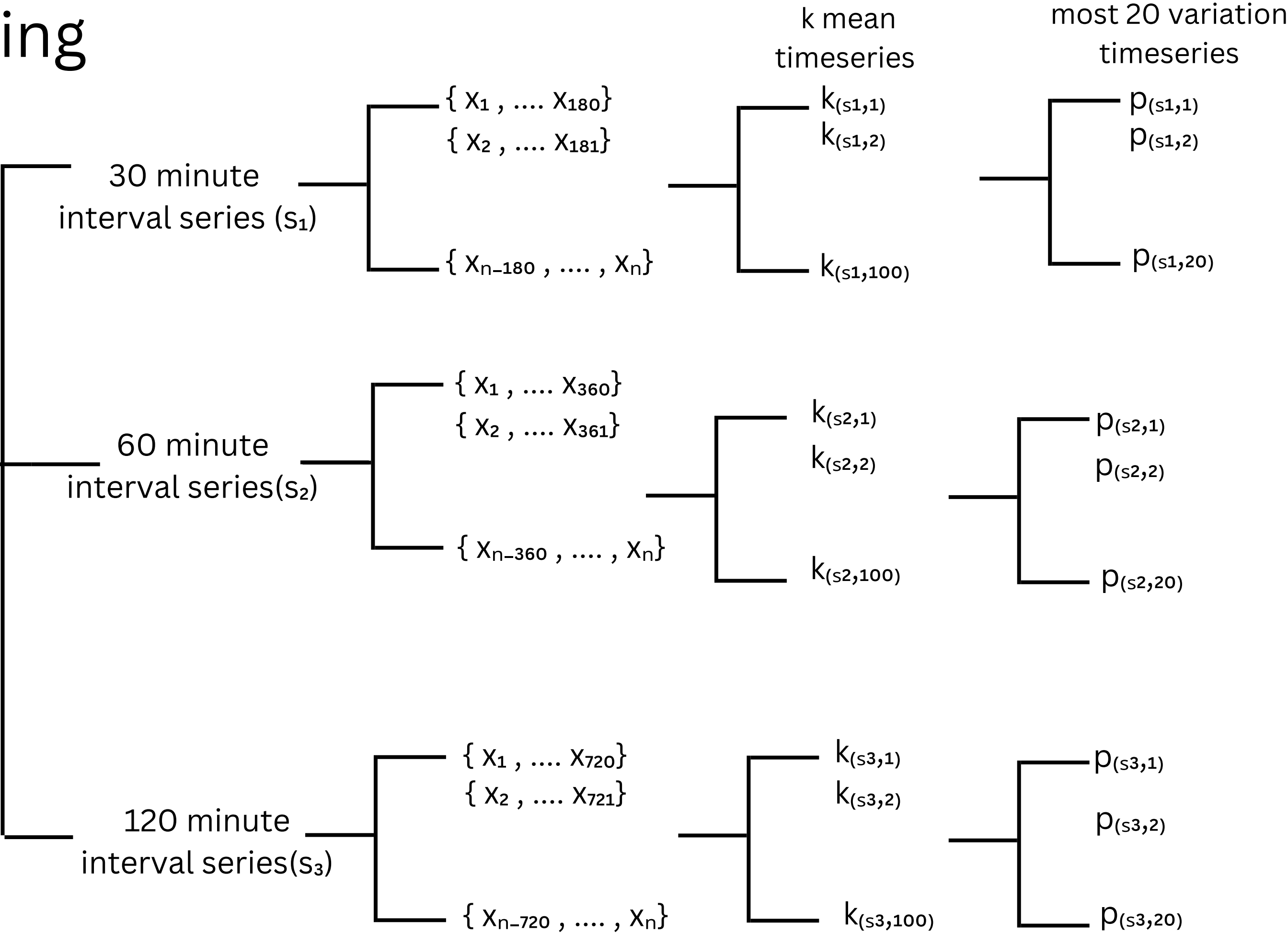
TimeSeriesData2(n)

TimeSeriesData3(n)

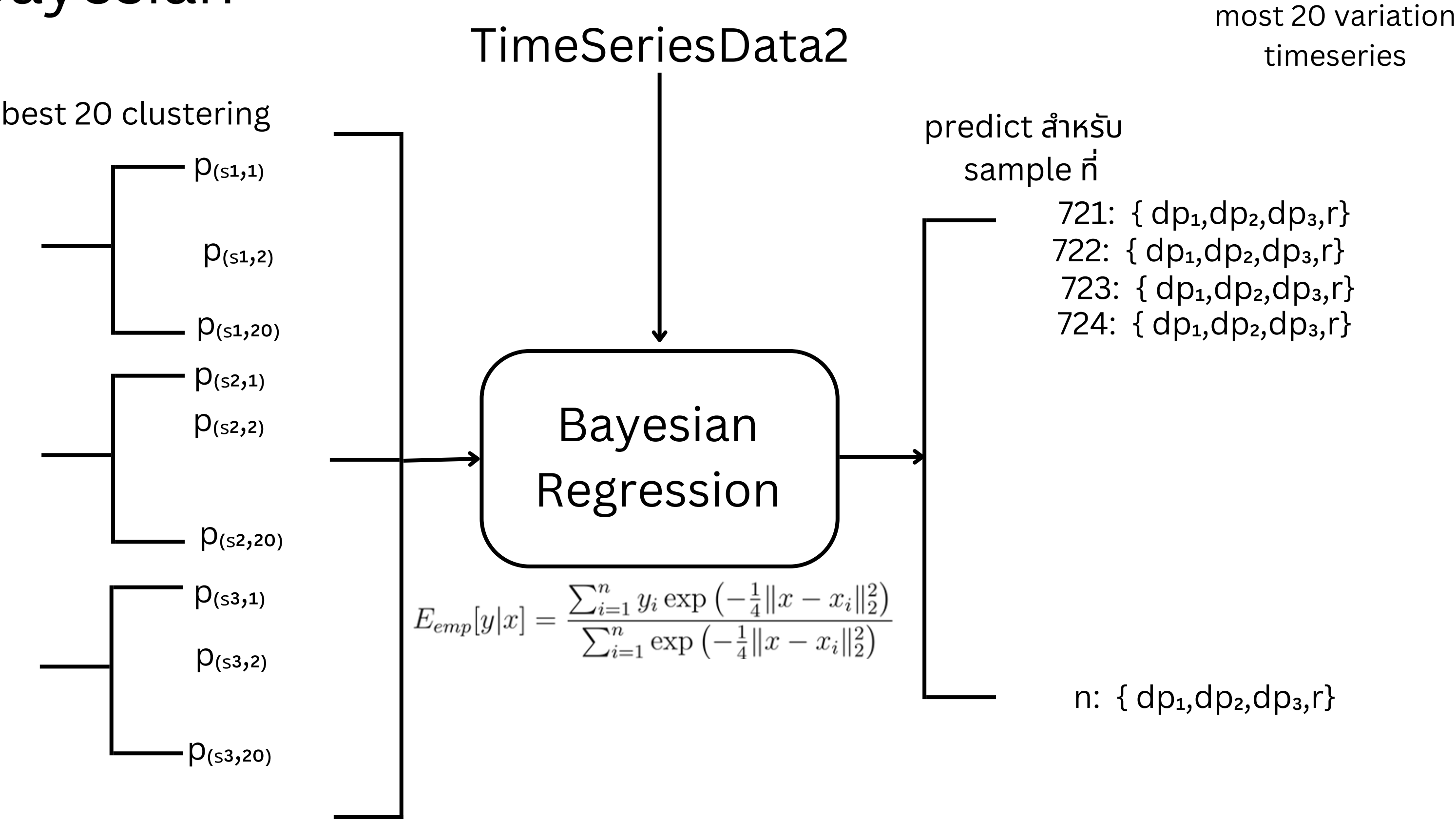


Part1: clustering

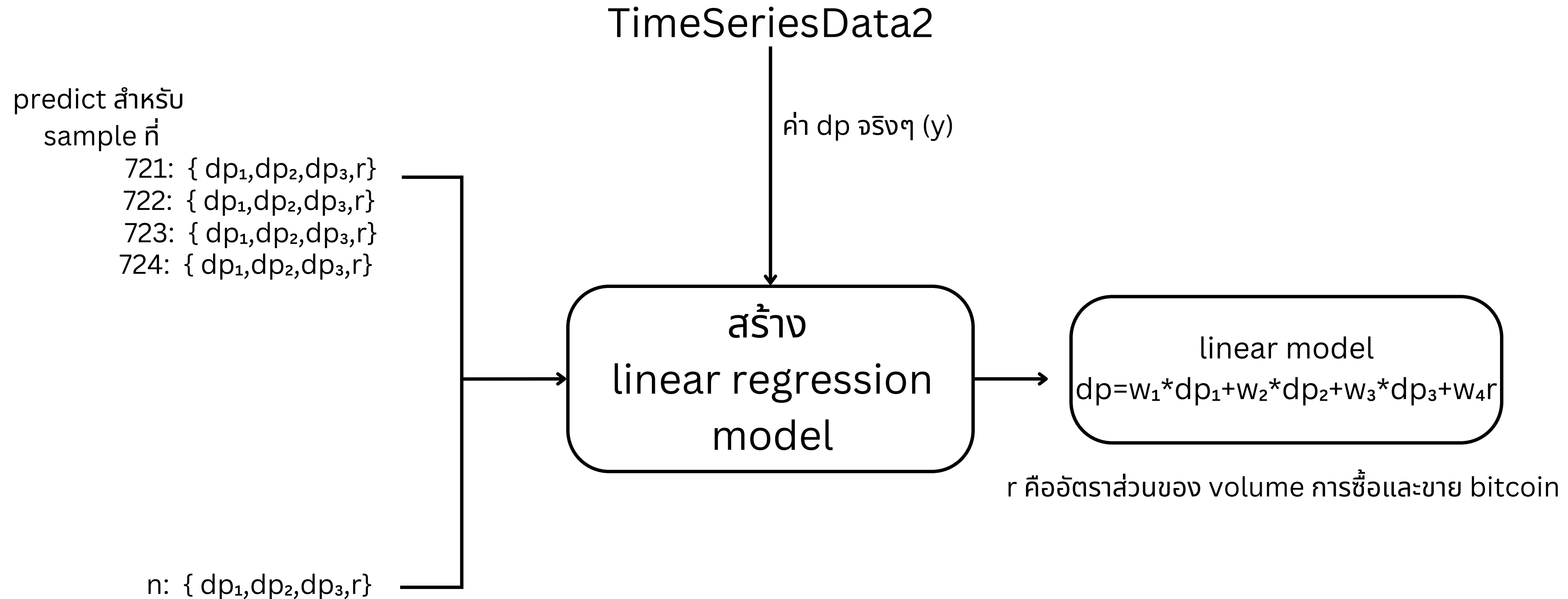
TimeSeriesData1
n sample
(1sample/10secs)
 $\{ X_1 , X_n \}$



Part2: Bayesian



Part3: Linear Construct



Trading Strategy

Trading Strategy

ในแต่ละเวลาเราจะอยู่ใน 1 ใน 3 ตำแหน่งนี้ +1 Bitcoin, 0 Bitcoin, หรือ -1 Bitcoin จากนั้นเราจะทำนายค่าการเปลี่ยนแปลงมูลค่าโดยเฉลี่ยในช่วง 10 วินาทีข้างหน้า ให้เป็น Δp โดยใช้ Bayesian regression และ t คือ threshold

- ถ้า $\Delta p > t$ และ ตำแหน่ง Bitcoin ≤ 0 เราจะซื้อ
- ถ้า $\Delta p < -t$ และ ตำแหน่ง Bitcoin ≥ 0 เราจะขาย
- นอกจากนั้น ไม่ทำอะไร

Trading Strategy



Results

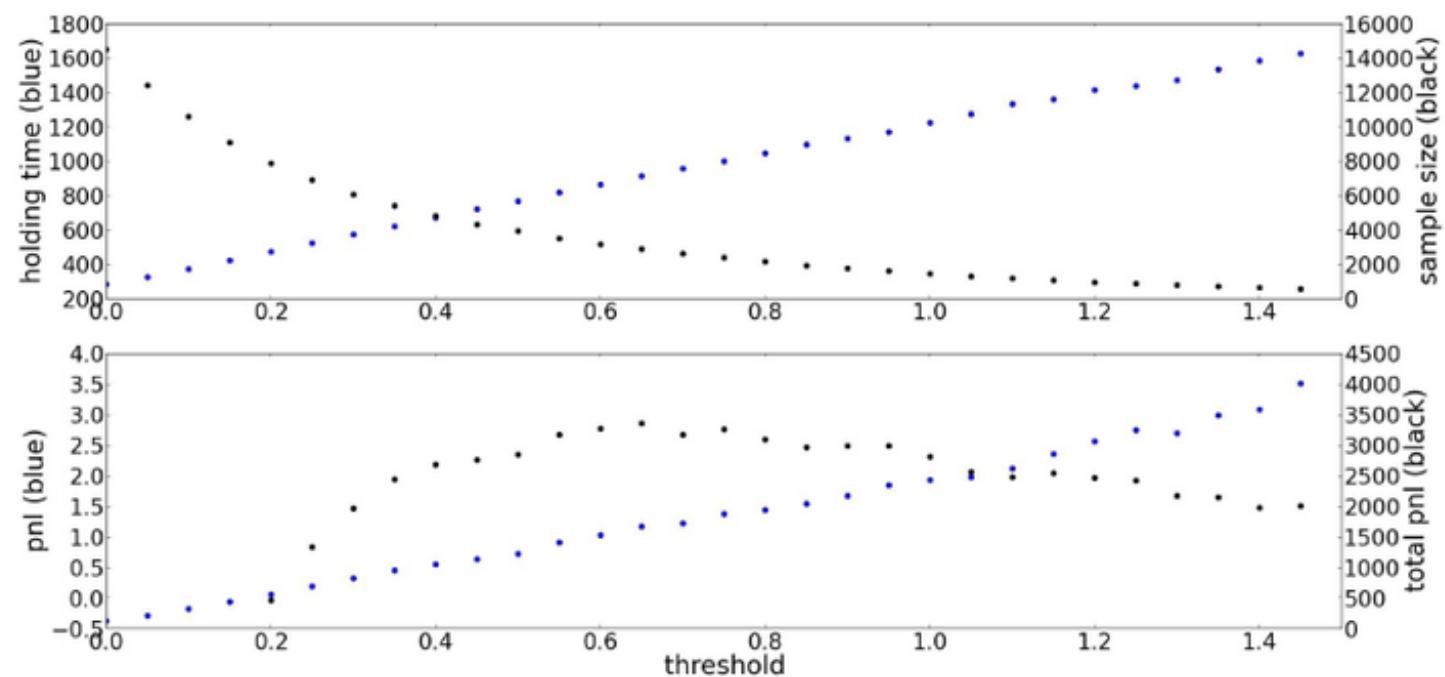


Fig. 1: The effect of different threshold on the number of trades, average holding time and profit

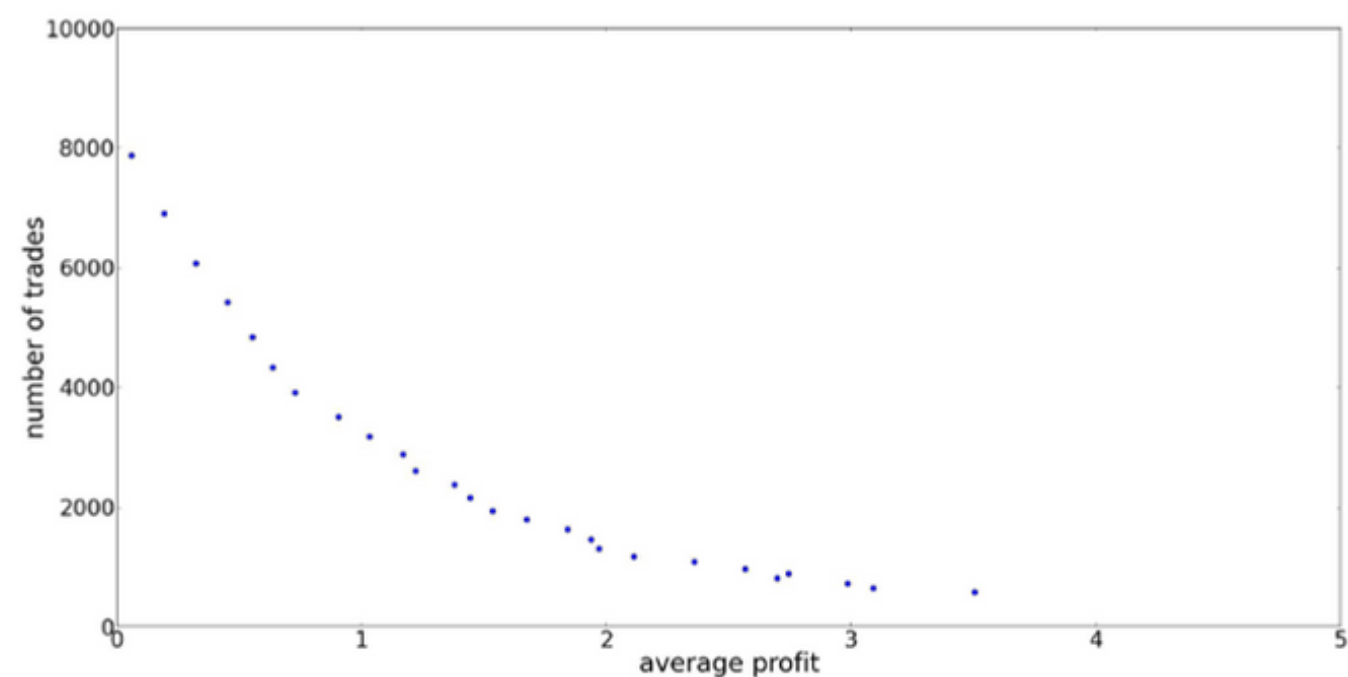
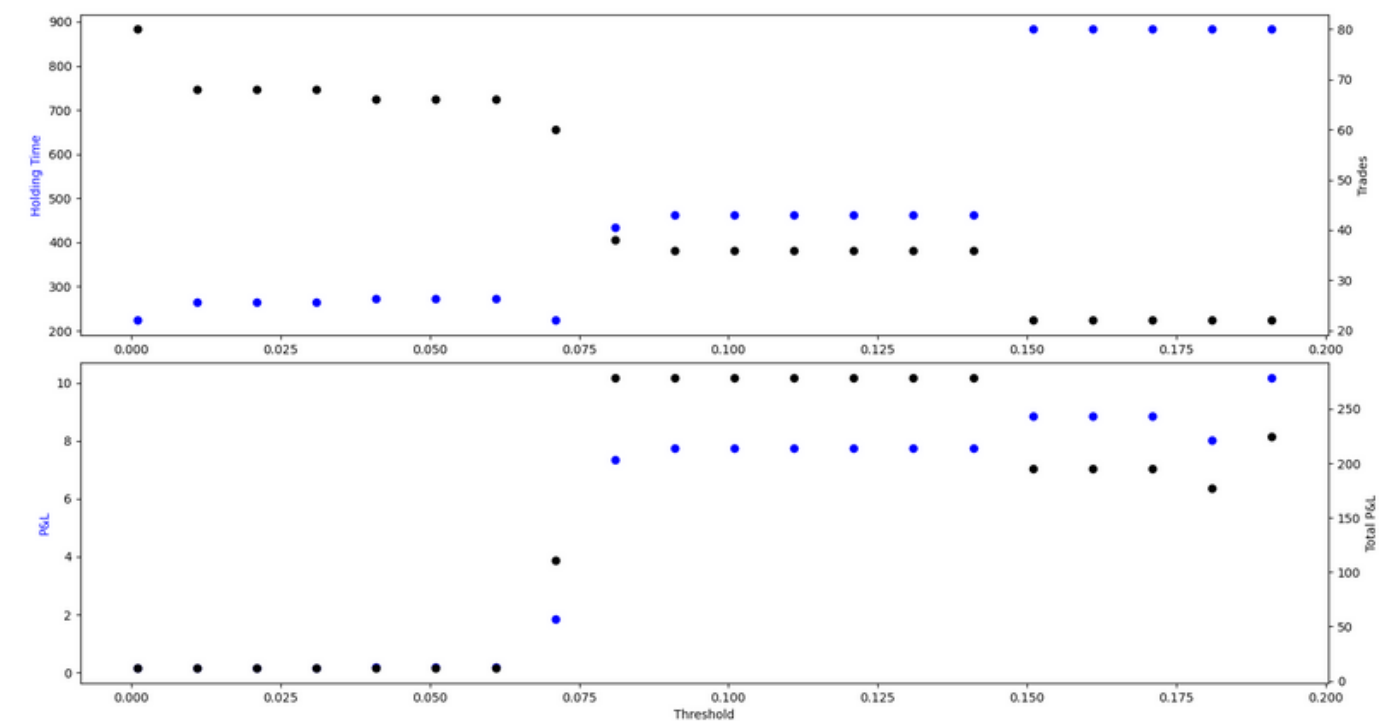
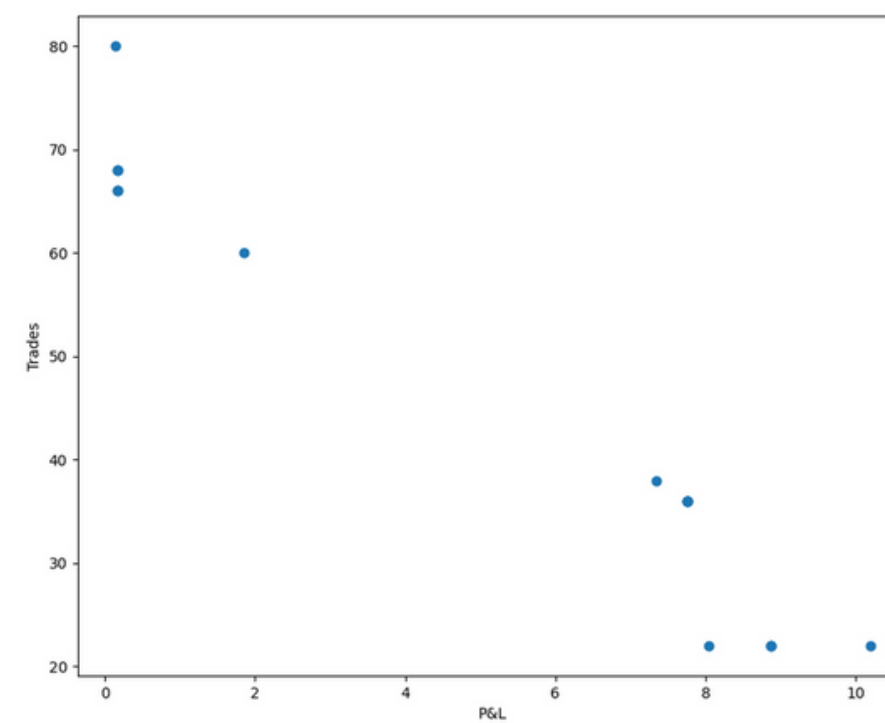


Fig. 2: The inverse relationship between the average profit per trade and the number of trades



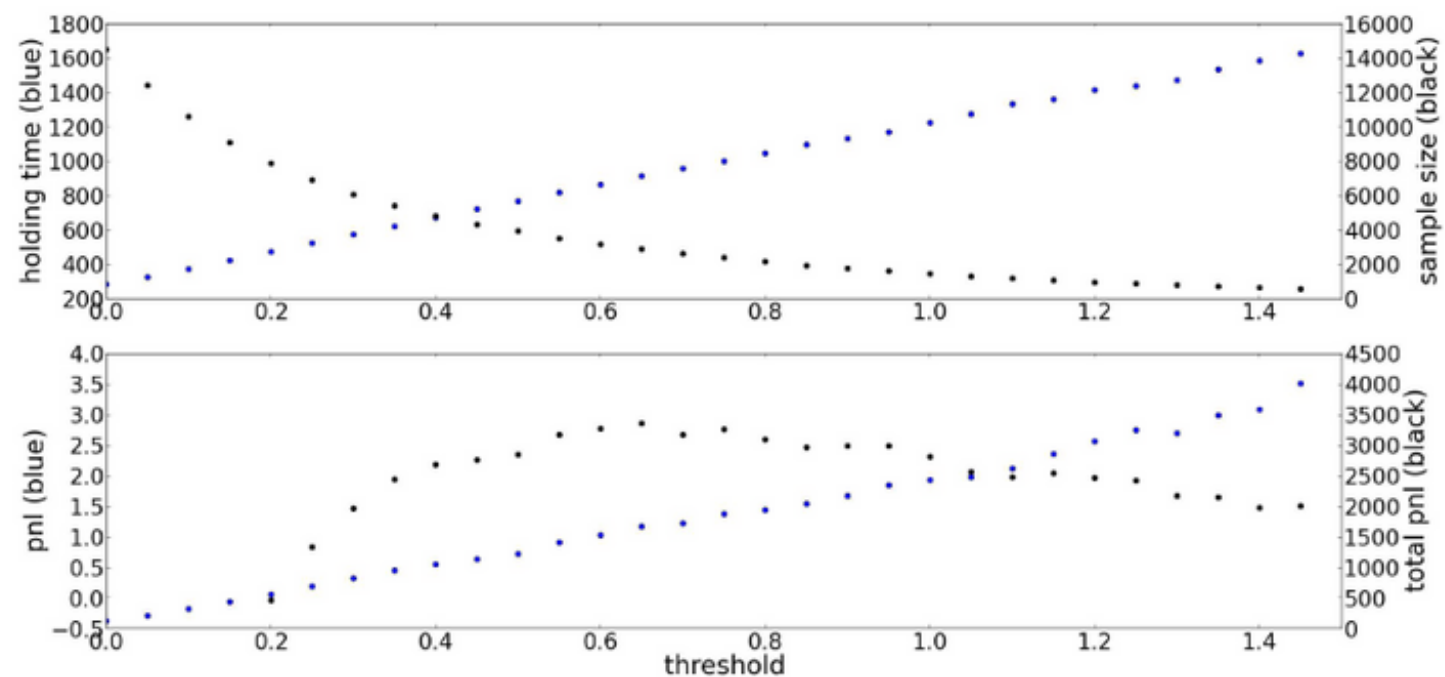


Fig. 1: The effect of different threshold on the number of trades, average holding time and profit

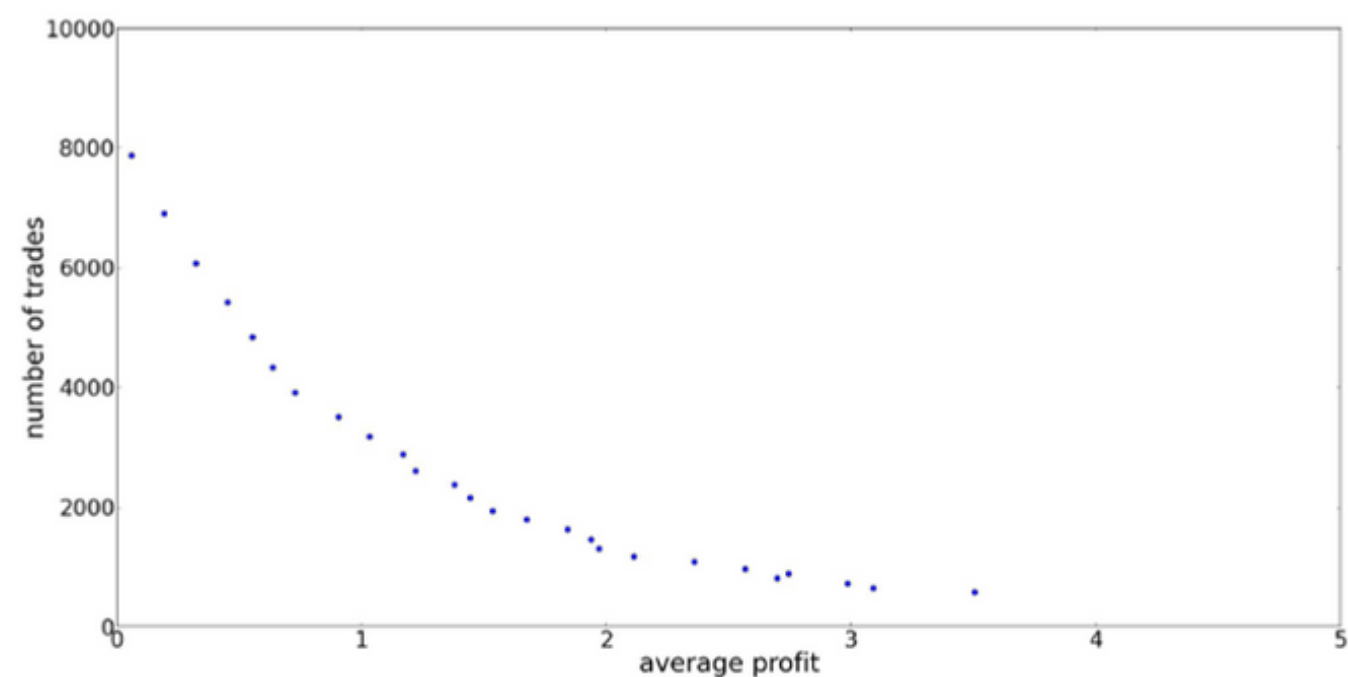
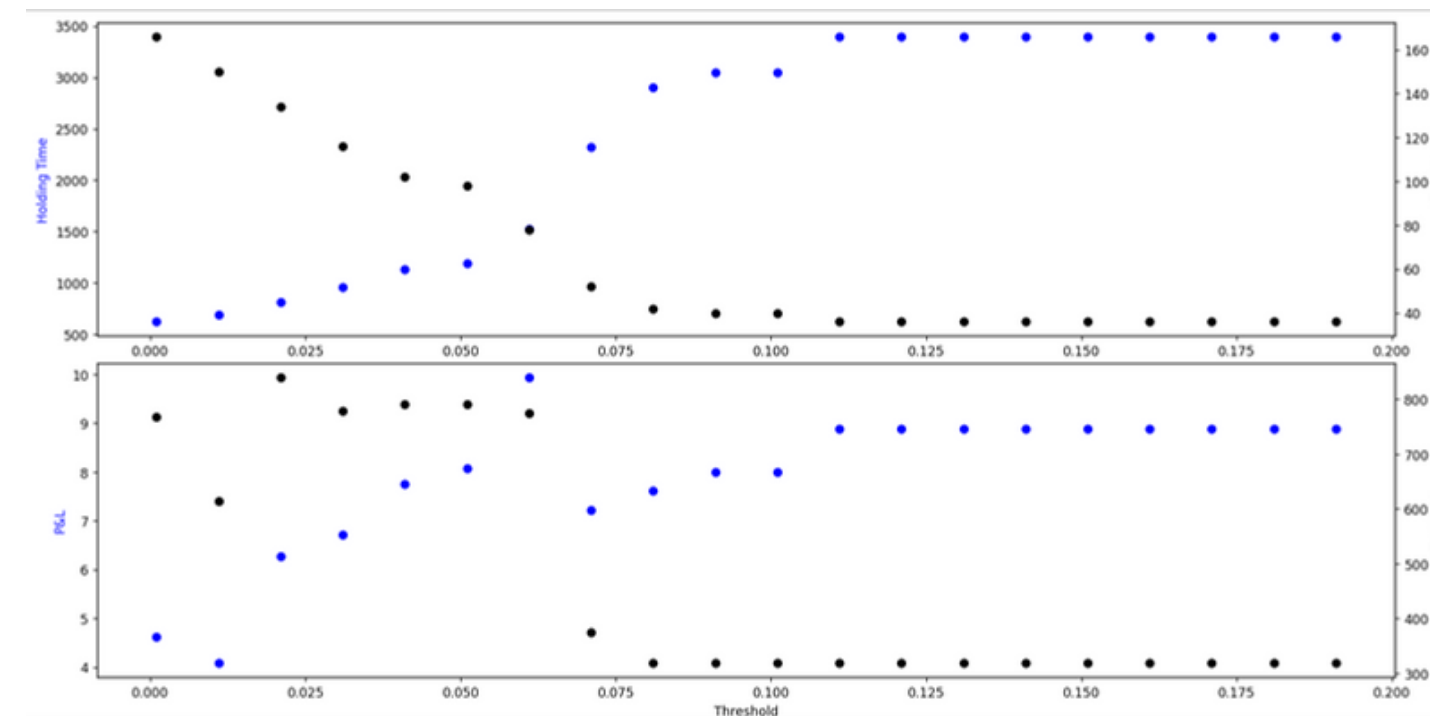
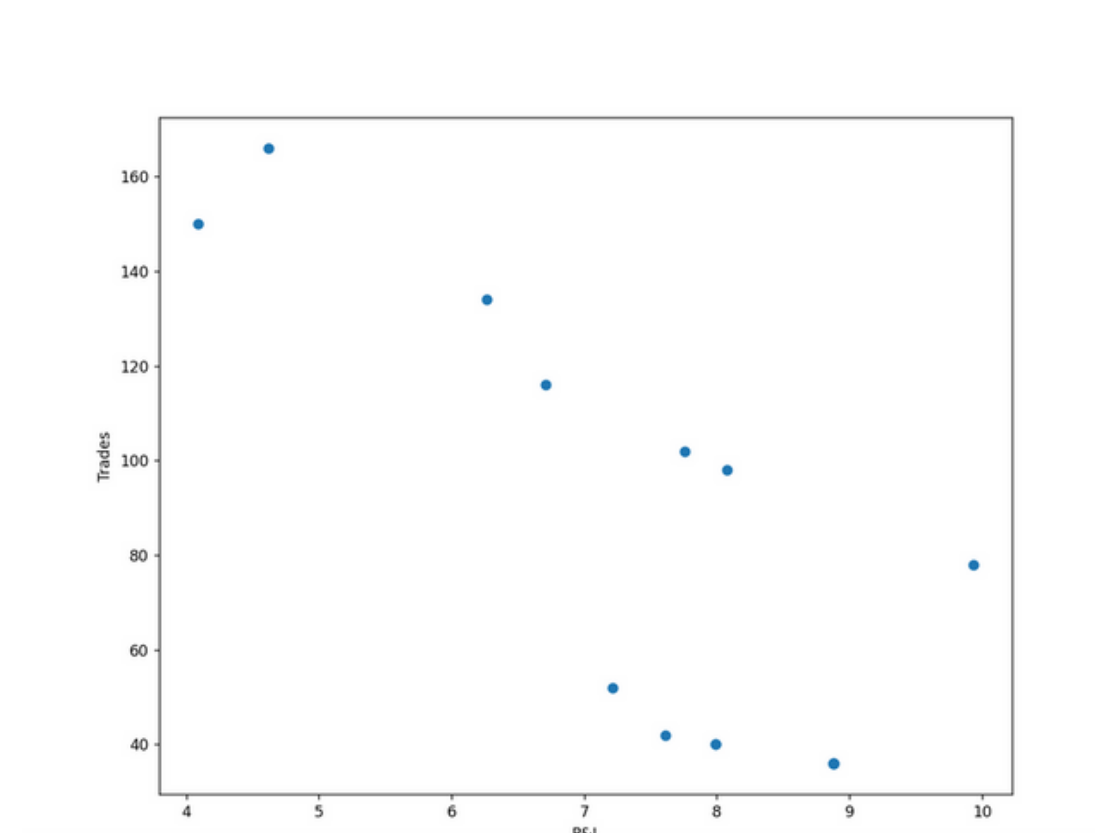


Fig. 2: The inverse relationship between the average profit per trade and the number of trades



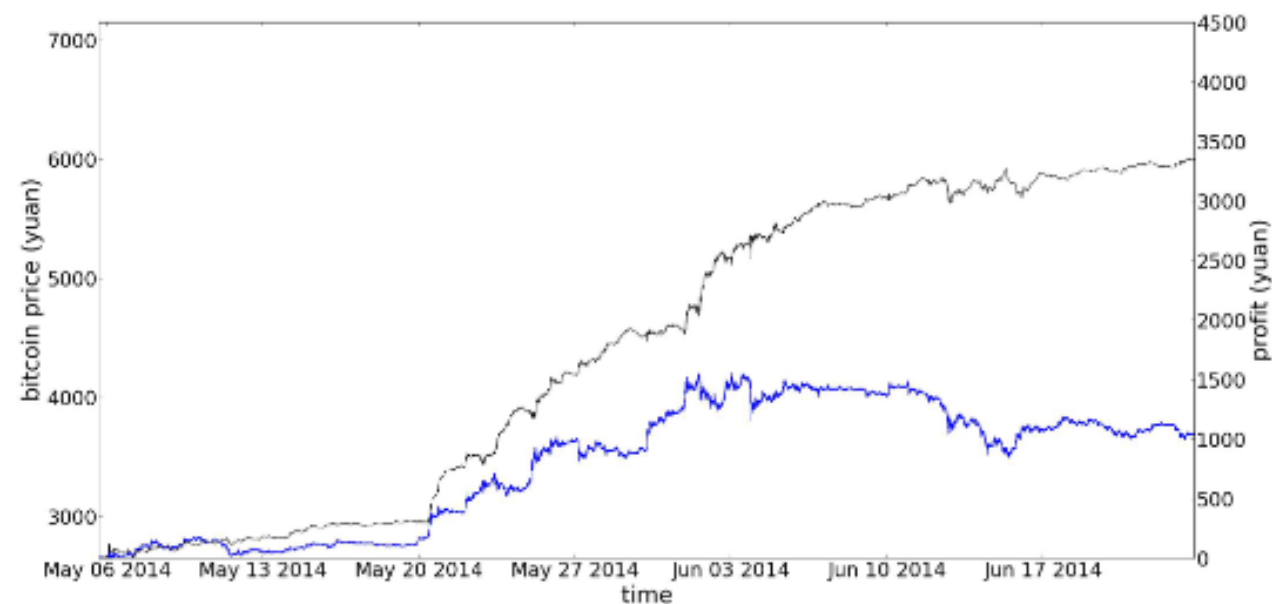
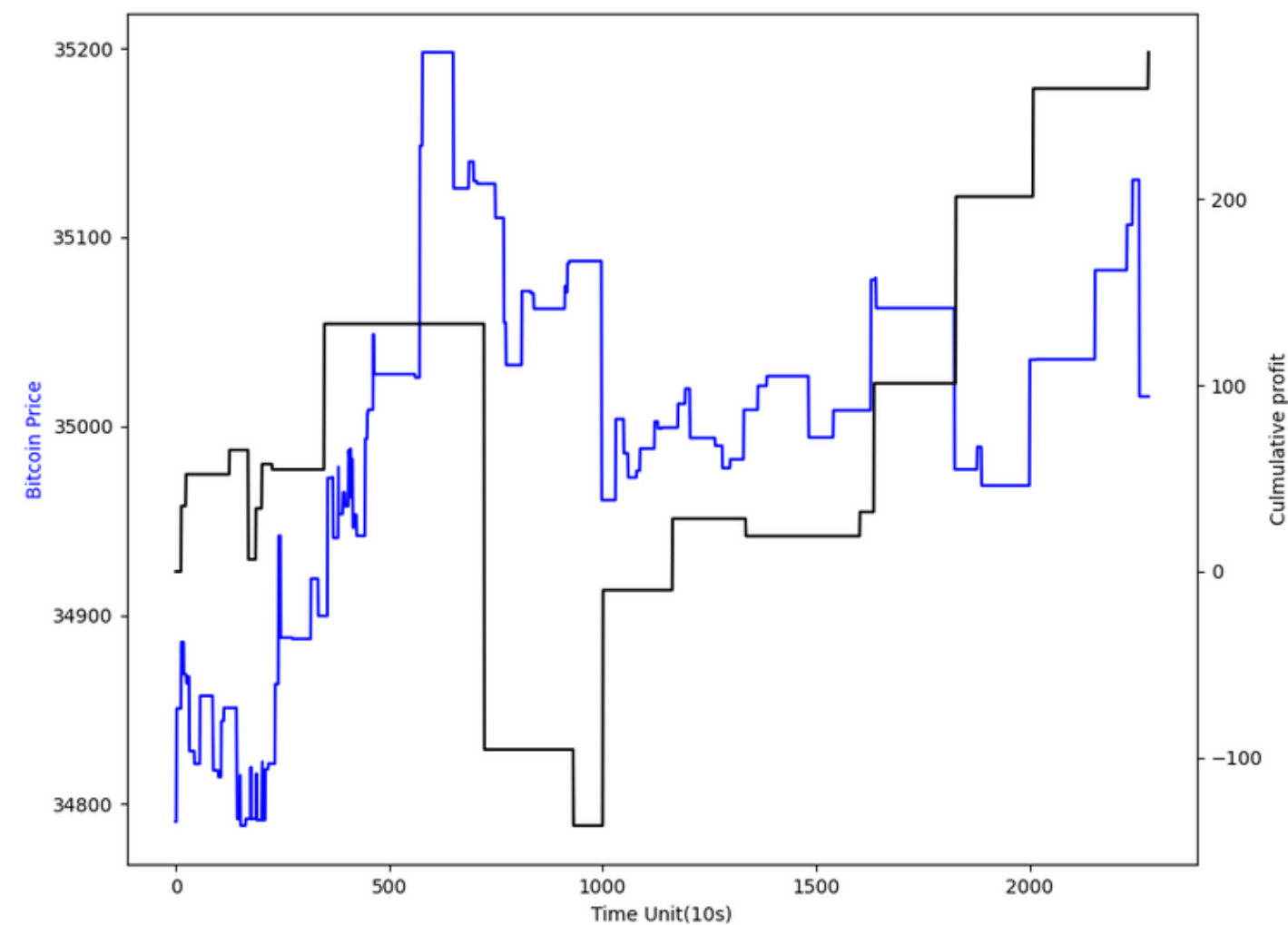


Fig. 3: The figure plots two time-series - the cumulative profit of the strategy starting May 6, 2014 and the price of Bitcoin. The one, that is lower (in blue), corresponds to the price of Bitcoin, while the other corresponds to cumulative profit. The scale of Y-axis on left corresponds to price, while the scale of Y-axis on the right corresponds to cumulative profit.



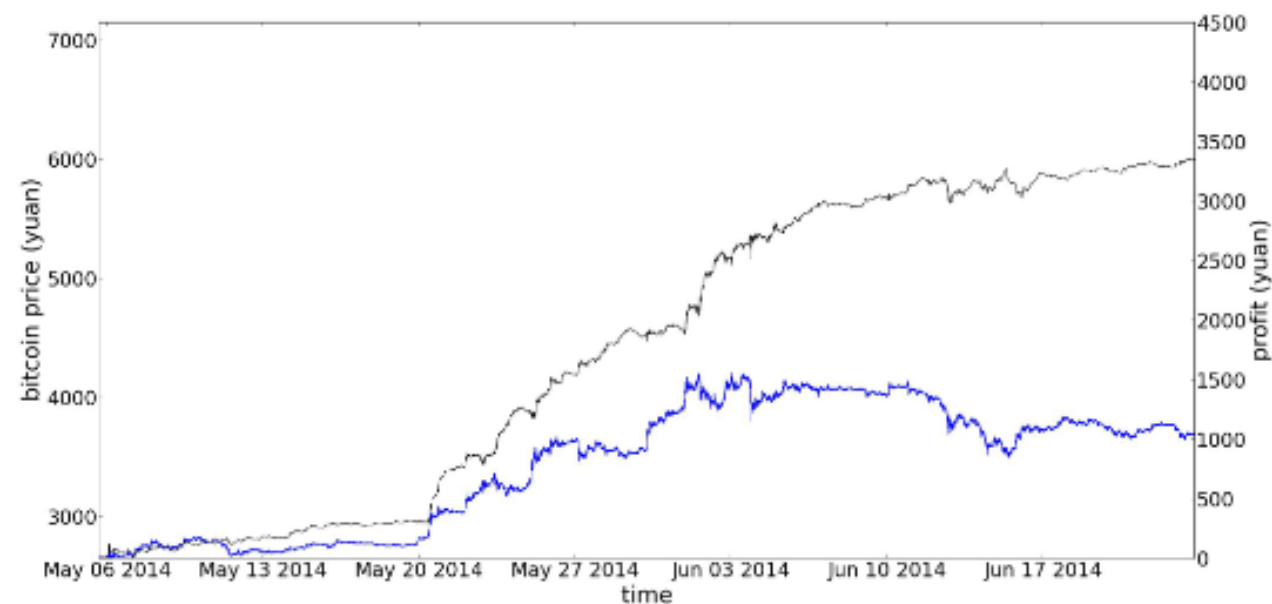
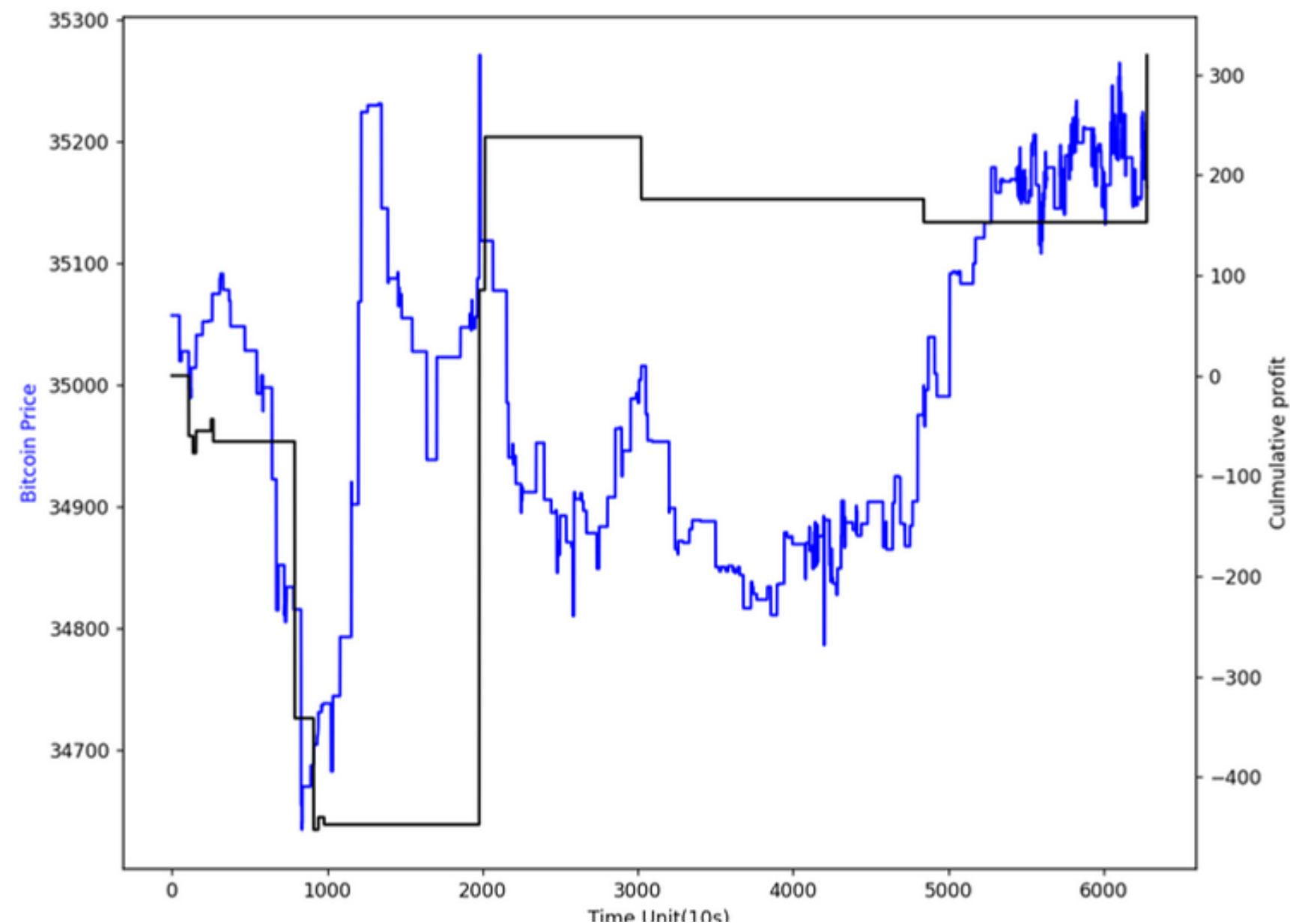


Fig. 3: The figure plots two time-series - the cumulative profit of the strategy starting May 6, 2014 and the price of Bitcoin. The one, that is lower (in blue), corresponds to the price of Bitcoin, while the other corresponds to cumulative profit. The scale of Y -axis on left corresponds to price, while the scale of Y -axis on the right corresponds to cumulative profit.



Pros & Cons

Pros

- ใช้การมีอยู่ของ Pattern โดยที่ไม่ระบุว่า Pattern เป็นอย่างไร
- วิธีทำเข้าใจง่าย
- นำไปใช้กับ Trading Strategy ได้

Cons

- ประสิทธิภาพการ Trade ขึ้นอยู่กับ Threshold
- ใช้ Assumption ว่า Pattern จะส่งผลต่ออนาคตแบบสม่ำเสมอ
- พฤติกรรมตลาดนั้นไม่แน่นอน
- มีความเป็นไปได้สูงที่จะ Overfit

Thank you