Exploring Machine Learning Advances in Finance

Guillermo Creus Botella

Tuesday 12th January, 2021

UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH
香港科技大學 THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

Contents

Primer in financial data

Asset log-prices

- Let p_t be the price of an asset at (discrete) time index t
- For modeling purposes, the natural logarithm of prices will be used $y_t = \log(p_t)$ (log-prices)

Figure: S&P 500 log-prices

Asset returns & volatility

- **Linear returns:** $R_t = \frac{p_t - p_{t-1}}{p_{t-1}} = \frac{p_t}{p_{t-1}} - 1$
- **Log-returns:** $r_t = \log\left(\frac{p_t}{p_{t-1}}\right) = y_t - y_{t-1}$

Note that $r_t = \log(1 + R_t)$ and $r_t \approx R_t$ whenever $R_t \approx 0$

- **Volatility:** Measures the variation of returns $\sigma = \sqrt{\text{Var}[r_t]}$

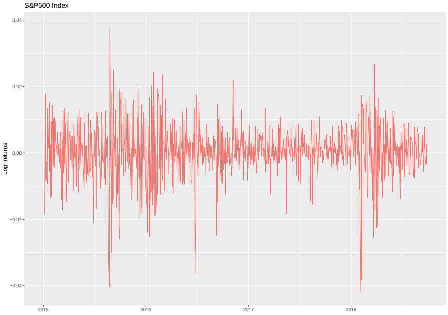


Figure: S&P 500 log-returns

Stylized facts

1. Absence of autocorrelations
2. Heavy tails
3. Gain/loss asymmetry
4. Aggregational Gaussianity
5. Intermittency
6. Volatility clustering

Stylized facts

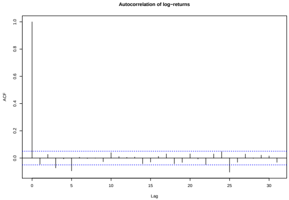
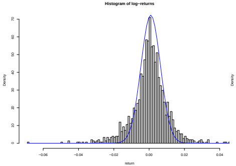
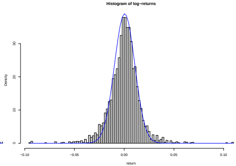


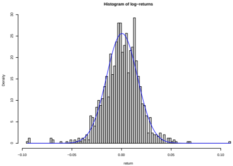
Figure: S&P 500 ACF



(a) Daily



(b) Weekly



(c) Monthly

Figure: S&P 500 log-returns histogram

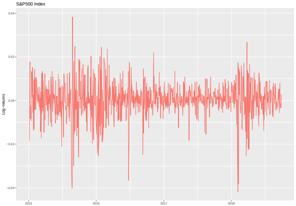


Figure: S&P 500 log-returns

Side of a position

Long position

A long position (or going long on some stock) is the most common way to invest. It just means that you buy an asset and you sell it at some point, expecting to earn a positive return.

Short position

If you short a stock, you first sell a stock that someone has lent you and then try to repurchase it at a lower price to return the stock to the lender. That way, if the **stock goes down in price**, you would **earn a profit** by selling high and buying low.

$$R_{t_1}^{\text{short}} = \frac{\text{profits}}{p_{t_0}} = \frac{p_{t_0} - p_{t_1}}{p_{t_0}} = -R_{t_1}$$

Financial Metrics

Sharpe Ratio

$\text{SR} := \frac{\mathbb{E}[R_t - r_f]}{\sqrt{\text{Var}[R_t - r_f]}}$, representing the reward per unit of risk.

Information Ratio

$\text{IR} := \frac{\mathbb{E}[R_t - R_b]}{\sqrt{\text{Var}[R_t - R_b]}}$, where R_b are the returns of a benchmark.

Drawdown

It measures the relative drop from a historical peak.

$$D(t) := \frac{\text{HWM}(t) - p_t}{\text{HWM}(t)}, \text{ where } \text{HWM}(t) = \max_{1 \le \tau \le t} p_\tau$$

High Frequency Data

- A **buy** (**sell**) order represents the will of a trader to buy (sell) m units of an asset at a price p.

Important concepts:

- **Bid price** (b_t): maximum price out of all the buy orders at time t.
- **Ask price** (a_t): minimum price out of all the sell orders at time t.
- **Mid price:** $m_t = \frac{a_t + b_t}{2}$
- **Volume:** (v_t) number of stocks exchanged.
- **Tick size:** smallest change in price a stock can move (e.g. 0.001$).

A **tick** is defined as $\{t, p_t, b_t, a_t, v_t\}$

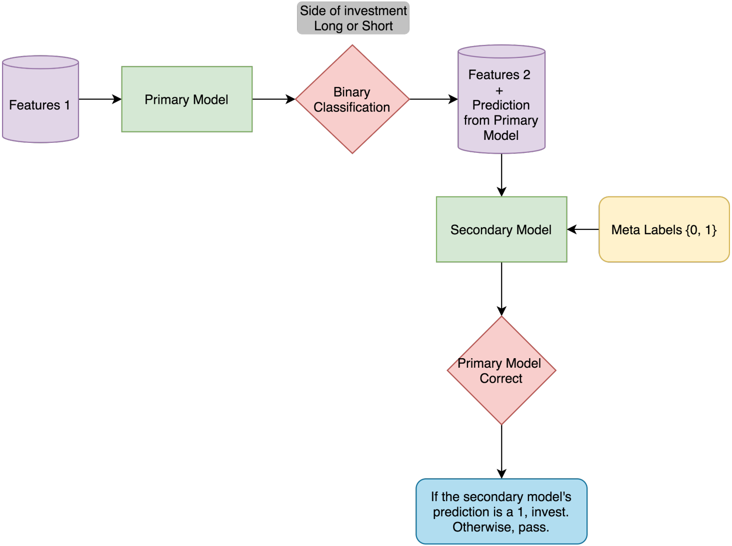
High Frequency Data

Table: Example of tick data (2013-01-02)

t	p_t	b_t	a_t	v_t
08:00:00	67.18	67.18	67.78	125
08:12:56	67.70	67.19	67.70	125
08:12:56	67.75	67.75	67.82	125
08:47:15	67.91	67.21	67.93	150
09:29:09	67.55	67.52	67.55	200
09:29:09	67.56	67.51	67.57	200
09:29:10	67.56	67.51	67.58	200
09:29:10	67.58	67.51	67.57	100
09:29:10	67.58	67.52	67.58	100
09:29:11	67.57	67.52	67.58	200

Meta-labeling

What is meta-labeling?



Binary classification problem

Primary model (M1)

It will predict the side of the investment. The labels will be noted as $y_i^{\text{M1}} \in \{-1, 1\}$ and the predictions as $\widehat{y}_i^{\text{M1}}$

Secondary model (M2)

It will predict whether the primary model was right or not.

The labels will be defined as: $y_i^{\text{M2}} = \begin{cases} 1 & \text{if } y_i^{\text{M1}} = \widehat{y}_i^{\text{M1}} \\ 0 & \text{otherwise} \end{cases}$, while predictions will be noted as $\widehat{y}_i^{\text{M2}}$

Meta-model

M1 + M2. It will **only** open a position, with the side predicted by M1, when M2 determines that M1 is right.

Binary classification problem

Outcomes:

- **1** (Positive): Open a position
- **0** (Negative): Do not open a position

Metrics:

- **Recall** $= \frac{\text{TP}}{\text{TP}+\text{FN}}$
- **Precision** $= \frac{\text{TP}}{\text{TP}+\text{FP}}$
- **F1-Score** $= \frac{2}{\text{recall}^{-1}+\text{precision}^{-1}}$

Possible predictions:

- **TP**: $y_i^{\text{M2}} = 1 = \widehat{y}_i^{\text{M2}}$
- **FP**: $y_i^{\text{M2}} = 0 \neq \widehat{y}_i^{\text{M2}}$
- **TN**: $y_i^{\text{M2}} = 0 = \widehat{y}_i^{\text{M2}}$
- **FN**: $y_i^{\text{M2}} = 1 \neq \widehat{y}_i^{\text{M2}}$

Toy project
Features and labels

The main idea of this project was to determine how meta-labeling works with synthetic data. In this case, 5 features have been used: $\mathbf{X}_1$, $\mathbf{X}_2$, $\mathbf{X}_3$, $\mathbf{X}_4$ and $\mathbf{X}_5$.

- $\mathbf{X}_{k,i} \sim N(\mu_i,\ \sigma^2)$
- $\omega_k = \textbf{sigmoid}\left(\alpha + \sum_{i=1}^{5} \mathbf{X}_{k,i} \cdot \beta_i + \epsilon_k\right)$ where $\epsilon_k \sim N(0,\ \sigma_\epsilon^2)$

The labels are defined as:

$$y_k^{\text{M1}} = \begin{cases} -1 & \text{if } \omega_k < 0.5 \\ 1 & \text{otherwise} \end{cases}$$

In order to **simulate relative abundance and scarcity of data**, models will use different features, e.g.:
M1: $\mathbf{X}_1$,$\mathbf{X}_2$ M2: $\widehat{y}^{\text{M1}}$, $\mathbf{X}_3, \ldots, \mathbf{X}_5$.

To exemplify what meta-labeling does, the following models will be analyzed:

M1: $\mathbf{X}_1, \mathbf{X}_2$ $\quad (N_{\text{M1}} = 2)$

M2: $\widehat{y}^{\text{M1}}, \mathbf{X}_3, \mathbf{X}_4, \mathbf{X}_5$ $\quad (N_{\text{M2}} = 4)$

$\sigma_\epsilon = 0.3$

(a) Primary Model

		Prediction outcome 1	Prediction outcome 0	Total
Actual value	1	TP 143	FN 0	143
	0	FP 57	TN 0	57
	Total	200	0	

(a) Primary Model

(b) Meta-model

		Prediction outcome 1	Prediction outcome 0	Total
Actual value	1	TP 136	FN 7	143
	0	FP 41	TN 16	57
	Total	177	23	

(b) Meta-model

Figure: Confusion Matrices (Test)

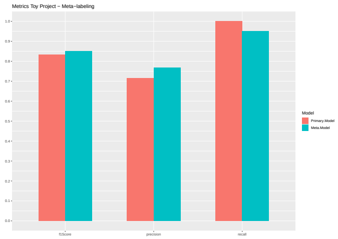
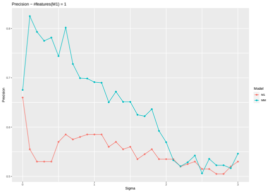


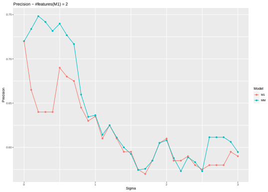
Figure: Toy Project - Metrics of example (Test)

Results

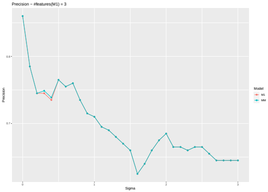
Precision



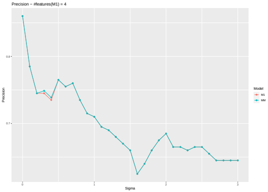
(a) $N_{\mathrm{M1}} = 1$



(b) $N_{\mathrm{M1}} = 2$



(c) $N_{\mathrm{M1}} = 3$

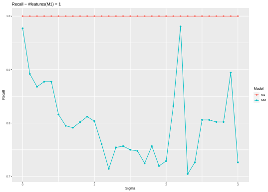


(d) $N_{\mathrm{M1}} = 4$

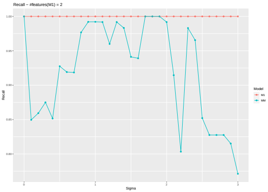
Figure: Precision (Test)

Results

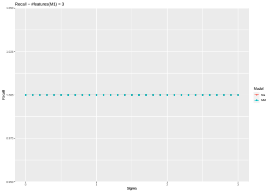
Recall



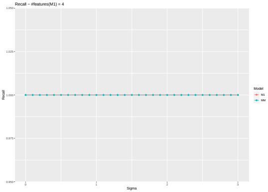
(a) $N_{\text{M1}} = 1$



(b) $N_{\text{M1}} = 2$

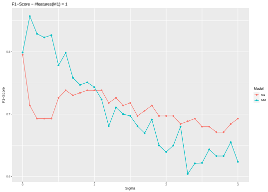


(c) $N_{\text{M1}} = 3$

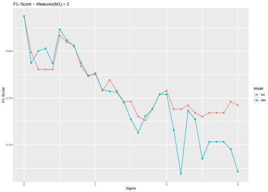


(d) $N_{\text{M1}} = 4$

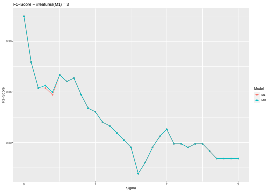
Figure: Recall (Test)



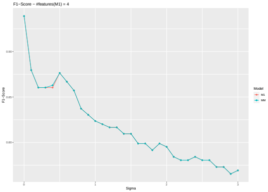
(a) $N_{\text{M1}} = 1$



(b) $N_{\text{M1}} = 2$



(c) $N_{\text{M1}} = 3$



(d) $N_{\text{M1}} = 4$

Figure: F1-Score (Test)

Data

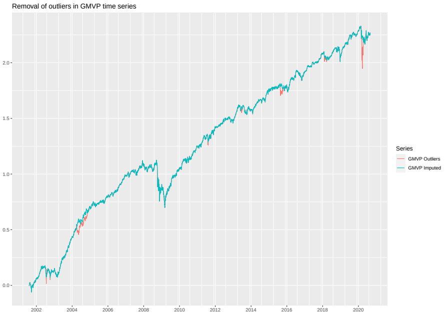
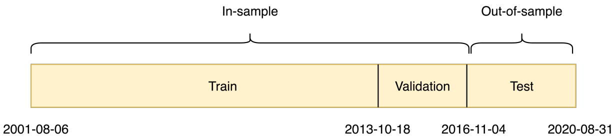


Figure: Imputed log-price time series of the S&P 500 GMVP



Triple barrier method

Symmetric barriers

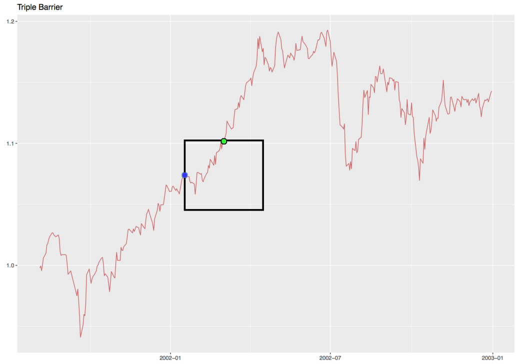


Figure: Symmetric barriers

- **Upper horizontal barrier:** $p_{t_{i,0}}(1 + \mathrm{pt} \cdot \sigma_{t_{i,0}})$
- **Lower horizontal barrier:** $p_{t_{i,0}}(1 - \mathrm{sl} \cdot \sigma_{t_{i,0}})$

where pt $=$ sl $= 2$.

- **Vertical barrier:** 10 days.

$$y_i = \begin{cases} 1 & \text{if } R_i > 0 \\ 0 & \text{otherwise} \end{cases}, \text{ where}$$

$$R_i = (1 - \mathrm{tc})^2 \cdot \left(\frac{p_{t_{i,1}}}{p_{t_{i,0}}} - 1 \right)$$

Triple barrier method

Notation when the side is known

- **Upper horizontal barrier:** $p_{t_{i,0}}(1 + \delta_{+,t_{i,0}})$, where

$$\delta_{+,t_{i,0}} = \begin{cases} \mathsf{pt} \cdot \sigma_{t_{i,0}} & \text{if } \widehat{y}_i^{\mathsf{M1}} = 1 \\ \min(0.5\%,\ \frac{1}{2} \cdot \mathsf{pt} \cdot \sigma_{t_{i,0}}) & \text{otherwise} \end{cases}$$

- **Lower horizontal barrier:** $p_{t_{i,0}}(1 - \delta_{-,t_{i,0}})$, where

$$\delta_{-,t_{i,0}} = \begin{cases} \mathsf{sl} \cdot \sigma_{t_{i,0}} & \text{if } \widehat{y}_i^{\mathsf{M1}} = -1 \\ \min(0.5\%,\ \frac{1}{2} \cdot \mathsf{sl} \cdot \sigma_{t_{i,0}}) & \text{otherwise} \end{cases}$$

- Predicted **Side of the trade** starting on $\boldsymbol{t_{i,0}}$: $\widehat{y}_i^{\mathsf{M1}}$

Triple barrier method

Adaptation when the side is known

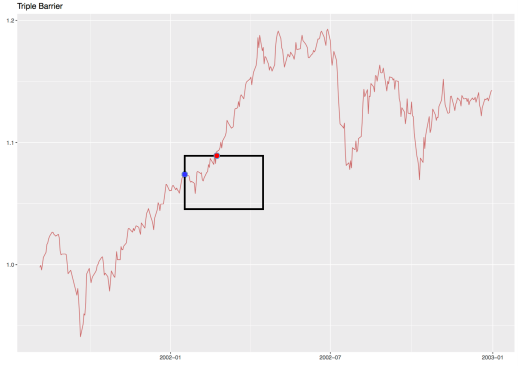


Figure: Triple Barrier Labeling when $y_i^{\text{M1}} = -1$

$$y_i^{\text{M2}} = \begin{cases} 1 & \text{if } R_i > 0 \\ 0 & \text{otherwise} \end{cases}$$

Where

$$R_i = (1 - \text{tc})^2 \cdot \left(\frac{p_{t_{i,1}}}{p_{t_{i,0}}} - 1 \right) \cdot \widehat{y}_i^{\text{M1}}$$

Primary models

MA based

Deterministic moving average (MA) crossovers:

- **Entry points:** CUSUM filter
- **Labels:**

$$y_i^{\text{M1}} = \begin{cases} 1 & \text{if } R_i = (1-\text{tc})^2\left(\frac{p_{t_{i,1}}}{p_{t_{i,0}}} - 1\right) > 0 \\ -1 & \text{otherwise} \end{cases}$$

- **Predictions:**

$$\hat{y}_i^{\text{M1}} = \begin{cases} 1 & \text{if MA}_i < p_{t_{i,0}} \\ -1 & \text{if MA}_i \geq p_{t_{i,0}} \end{cases}$$

ML based

- **Features:**

$\log\left(\frac{\text{MA}_t}{p_t}\right)$, $\frac{\text{CUSUM}_{+,\ t}}{\sigma_t}$, $\frac{\text{CUSUM}_{-,\ t}}{\sigma_t}$, $\textbf{Reset}_{\text{CUSUM}_{+},t}$, $\textbf{Reset}_{\text{CUSUM}_{-},t}$ and EWMSD_t

- **Labels:** $y_i^{\text{M1}} = \begin{cases} 1 & \text{if } R_i > 0 \\ -1 & \text{otherwise} \end{cases}$
- **Predictions:** $\hat{y}_i^{\text{M1}}$

Secondary models

MA based

It will use a random forest with the following **features**:

$\log\left(\frac{\mathsf{MA}_i}{p_{t_{i,0}}}\right)$, $r_{t_{i,0}}$, $\underline{r_{t_{i,0}}} := \left(\prod_{k=0}^{4}(1+r_{t_{i,0-k}})\right) - 1$, $\widehat{y}_i^{\mathsf{M1}}$, $\mathsf{RSI}_{t_{i,0},\ 9}$, $\mathsf{RSI}_{t_{i,0},\ 14}$, $\mathsf{RSI}_{t_{i,0},\ 25}$, $\tilde{\sigma}_{t_{i,0}}$, $\sigma_{t_{i,0},\ 9}$, $\sigma_{t_{i,0},\ 14}$, $\sigma_{t_{i,0},\ 25}$, $\mathsf{ACF}_{t_{i,0},\ 1}$ and $\mathsf{ACF}_{t_{i,0},\ 5}$

With **labels**:

$$y_i^{\mathsf{M2}} = \begin{cases} 1 & \text{if } y_i^{\mathsf{M1}} = \widehat{y}_i^{\mathsf{M1}} \\ 0 & \text{otherwise} \end{cases}$$

ML based

Then only difference with the MA based M2 is the underlying Machine Learning model.

The ML based M2 will use a neural network with a hidden layer (25 fully connected units - Leaky ReLU) and an output layer (Sigmoid).

Hyper-parameter tuning

Results

Coin flip correction

Fractional differentiation

This

Rr

This

This

This

This

Data parsing as bars

This

HFT

Conclusions and future work

Thank you for your attention