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

1. Introduction
2. Primer in financial data
3. Meta-labeling
4. Fractional differentiation
5. Data parsing as bars
6. Conclusions

1 Introduction

2 Primer in financial data

3 Meta-labeling

4 Fractional differentiation

5 Data parsing as bars

6 Conclusions

Introduction

This work will be centered around three novel techniques proposed by Marcos López de Prado in his book *Advances in Financial Machine Learning*:

- **Meta-labeling**
- **Fractional differentiation**
- **Data parsing as bars**

These advances will be **independently** analyzed to ascertain if they deliver **better forecasts** or **risk-adjusted returns** in a stock market context.

1 Introduction

2 Primer in financial data

3 Meta-labeling

4 Fractional differentiation

5 Data parsing as bars

6 Conclusions

Notation

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

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

- **Volume:** $v_t =$ number of stocks exchanged
- **Dollar volume:** $d_t = v_t \cdot p_t$

Data

Figure: Price and volume chart

S&P 500

- **Meta-labeling:** Daily data
- **Fractional differentiation:** Open, High, Low, Close (OHLC)
- **Data bars:** Tick data

Stationary time series

y_t presents a **positive trend** $\Rightarrow$ Non-stationary. To solve that, **log-returns** are computed: $r_t = (1 - B)y_t$, where B is the **backshift** operator: $By_t = y_{t-1}$.

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.

Financial Metrics

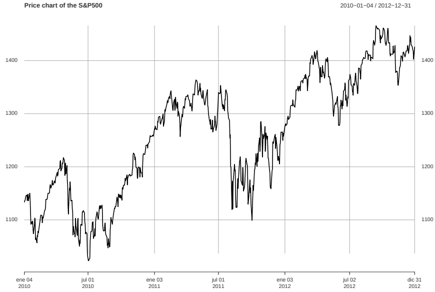
Sharpe Ratio: It represents the **reward per unit of risk**.

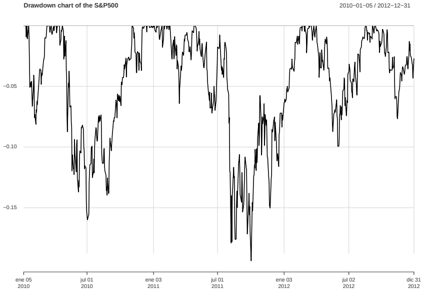
$$\text{SR} := \frac{\mathbb{E}[R_t - r_f]}{\sqrt{\text{Var}[R_t - r_f]}}$$

Drawdown: It measures the **relative drop** from a **historical peak**.

$$D(t) := \frac{\text{HWM}(t) - p_t}{\text{HWM}(t)}$$

where $\text{HWM}(t) = \max_{1 \leq \tau \leq t} p_\tau$





1 Introduction

2 Primer in financial data

3 Meta-labeling

4 Fractional differentiation

5 Data parsing as bars

6 Conclusions

What is meta-labeling?

Primary model (M1)

Binary classifier that will **predict** the **side of the investment**. In this work, two different primary models will be explored:

- Moving average (MA) based
- Machine Learning (ML) based

Secondary model (M2)

Binary classifier that **predicts** whether the **primary model** was **right or not**. The meta-labels will be defined as: $y_i^{\text{M2}} = \begin{cases} 1 & \text{if } y_i^{\text{M1}} = \widehat{y}_i^{\text{M1}} \\ 0 & \text{otherwise} \end{cases}$

Meta-model

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

Why should meta-labeling be used?

- Exogenous model that can work on top of a fundamental approach (it avoids the ML **black box** stigma).
- Enables more **sophisticated strategies** by decoupling side from size.
- **Avoids overfitting** by giving the ability to pass.

Labeling in financial time series

Triple Barrier Method

- **Horizontal barriers:** Dynamic levels that depend on the 10 day rolling volatility. They can be symmetric or not.
- **Vertical barrier:** Set as a fixed time horizon. In this case, 10 days.

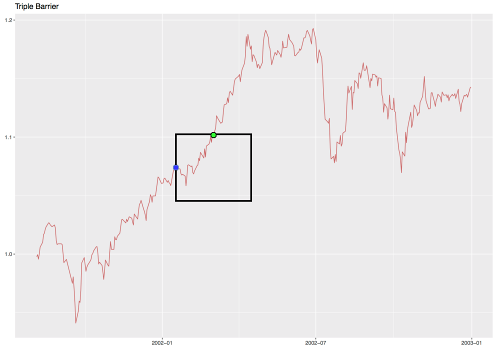


Figure: Symmetrical horizontal barriers

Train data set: one will be able to "see the future" and train the algorithms accordingly.

Test data set: one will try to "predict the future" and performance will be assessed.

Results

MA based

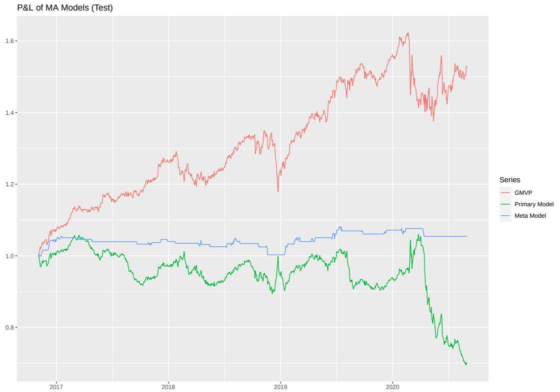


Table: MA based metrics in the Test data set

Model	Max. Drawdown	Sharpe Ratio
Buy & Hold	15.20%	0.97
Primary model	34.41%	-0.75
Meta-model	5.02%	0.49

Results
ML based

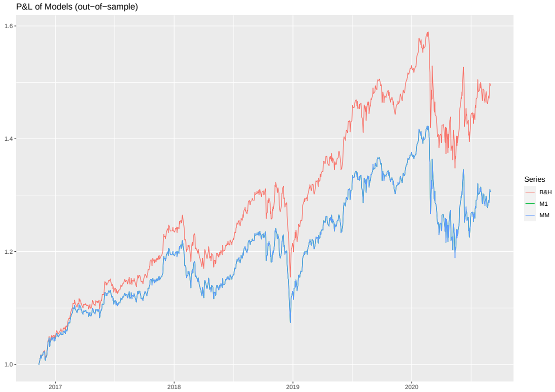


Table: ML based metrics in the Test data set

Model	Max. Drawdown	Sharpe Ratio
Buy & Hold	15.20%	0.97
Primary model	16.42%	0.66
Meta-model	16.42%	0.66

Coin flip correction

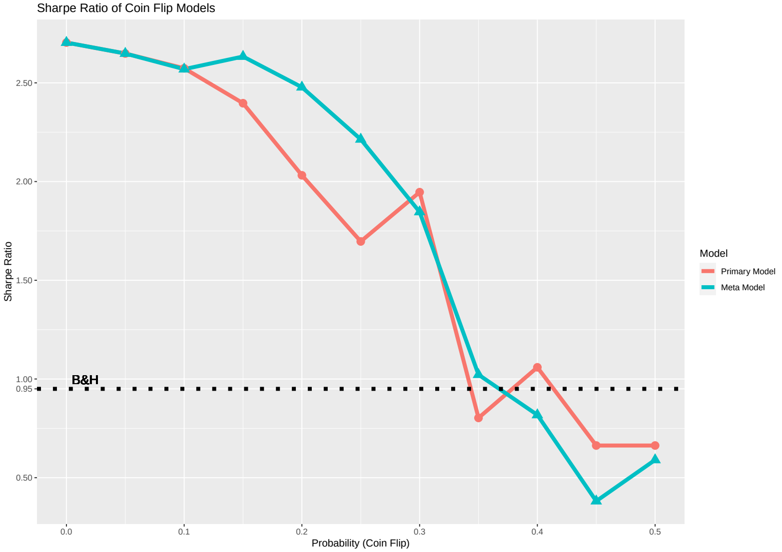
In an attempt to create **better (but artificial) primary models**, they will use a **new feature** F:

$$F_i = (1 - 2 \cdot S_i) \cdot y_i^{\mathsf{M1}}$$

Where:

- $S_i \sim Be(p)$ is the r.v. in charge of swapping the label y_i^{M1}
- $p = \mathsf{Pr}(S_i = 1)$
- $y_i^{\mathsf{M1}} \in \{-1, +1\}$ is the **label** representing the **side**.

Coin flip correction
Results



1 Introduction

2 Primer in financial data

3 Meta-labeling

4 Fractional differentiation

5 Data parsing as bars

6 Conclusions

Fractional differentiation

Stationarity of $y_t \Rightarrow$ **Log-returns** $r_t = (1-B)y_t = y_t - By_t = y_t - y_{t-1}$

Is part of the **signal** lost?

Fractionally differentiated time series: $x_t^d = (1-B)^d \, y_t$, where $d \in (0,1)$ and

$$(1-B)^d = \sum_{k=0}^{\infty} \binom{d}{k} (-B)^k = \sum_{k=0}^{\infty} B^k \underbrace{(-1)^k \prod_{i=0}^{k-1} \frac{d-i}{k-i}}_{=: w_k}$$

FFD method: $(1-B)^d = \sum_{k=0}^{\infty} w_k B^k \approx \sum_{k=0}^{l^*} w_k B^k = \phi_d(B)$

$$x_t^d = \phi_d(B) y_t$$

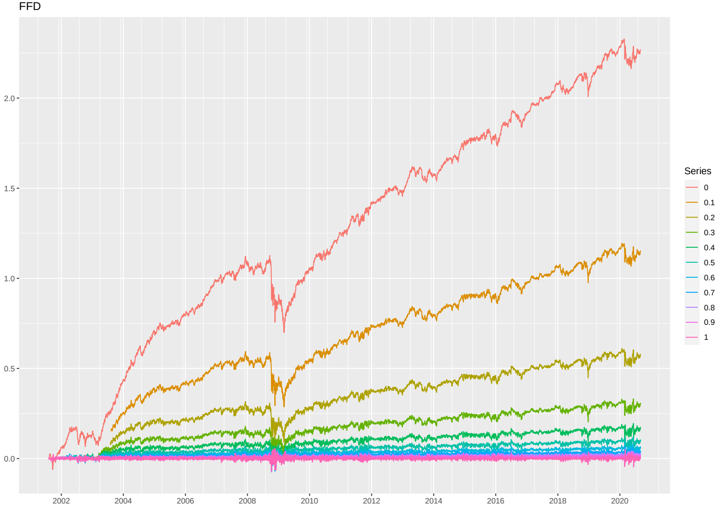


Figure: Memory of differentiated time series $\left(\tau = 10^{-4}\right)$

FFD method

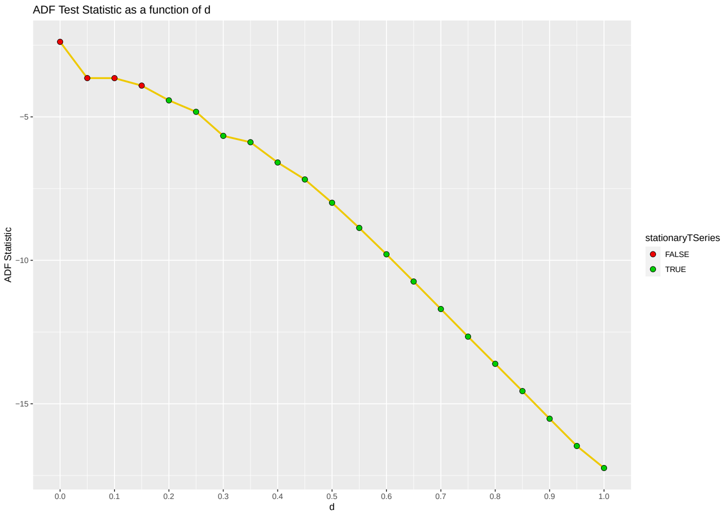


Figure: Stationary test statistic as a function of d

Models

Goal: Determine if **fractionally differentiated** features can give **better** 1-day **forecasts**.

- **Naive** model (benchmark): $\widehat{y}_{t+1}^{\text{Close}} = y_t^{\text{Close}}$
- **FFD** model: Use the FFD method with d^*, the minimum d such that it passes the ADF test.
- **Returns** model: Use fully differentiated features $(d = 1) \Rightarrow$ log-returns.

Features: y_t^{Open}, y_t^{High}, y_t^{Low} and y_t^{Close}

Target: y_{t+1}^{Close}

Metrics

Let's suppose that the time series y_t has T observations and the Test data set starts at $t = n_0$. Then, one can define the following:

- **Errors:** $e_k := y_k - \widehat{y}_k$, where $k \in \{n_0, \ldots, T\}$
- **Median of errors:** $\widetilde{e} := \text{median}(e_k)$

RMSE

$$\text{RMSE} = \sqrt{\frac{\sum_{k=n_0}^{T} (e_k)^2}{T - n_0 + 1}}$$

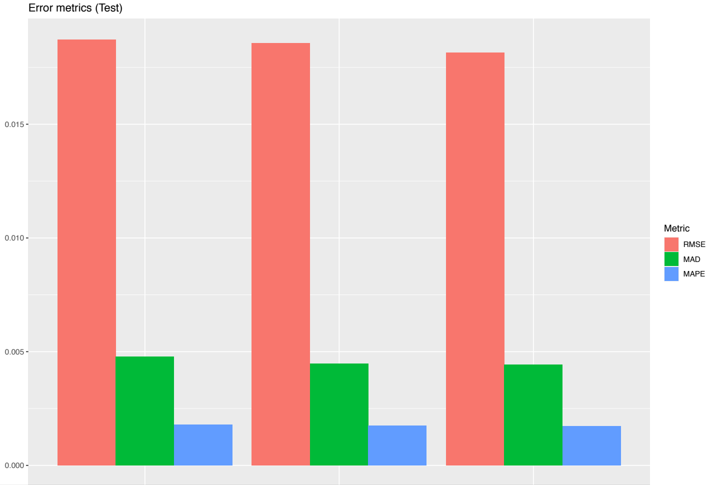
MAPE

$$\text{MAPE} = \frac{1}{T - n_0 + 1} \cdot \sum_{k=n_0}^{T} \left| \frac{e_k}{y_k} \right|$$

MAD

$$\text{MAD} = \text{median}(e_k - \widetilde{e})$$

Results



	FFD model	Naive model	Returns model
RMSE $(\times 10^{-2})$	1.87 (-0.78%)	1.86	1.82 (2.27%)
MAPE $(\times 10^{-3})$	1.79 (-1.91%)	1.76	1.72 (2.24%)
MAD $(\times 10^{-3})$	4.78 (-6.85%)	4.47	4.44 (0.73%)

1 Introduction

2 Primer in financial data

3 Meta-labeling

4 Fractional differentiation

5 Data parsing as bars

6 Conclusions

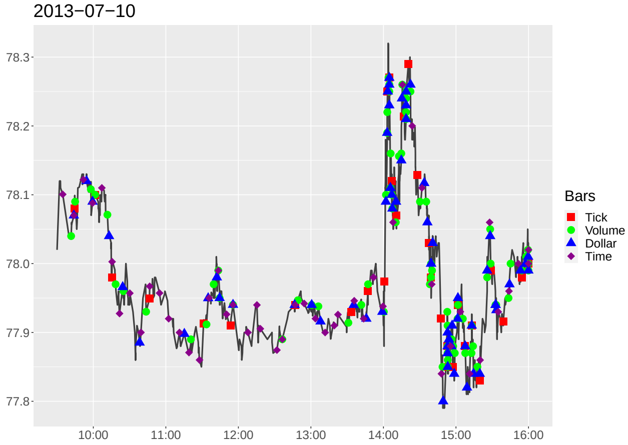
Data bars

Data:

Tick data of IVE (S&P 500 ETF) from 2013

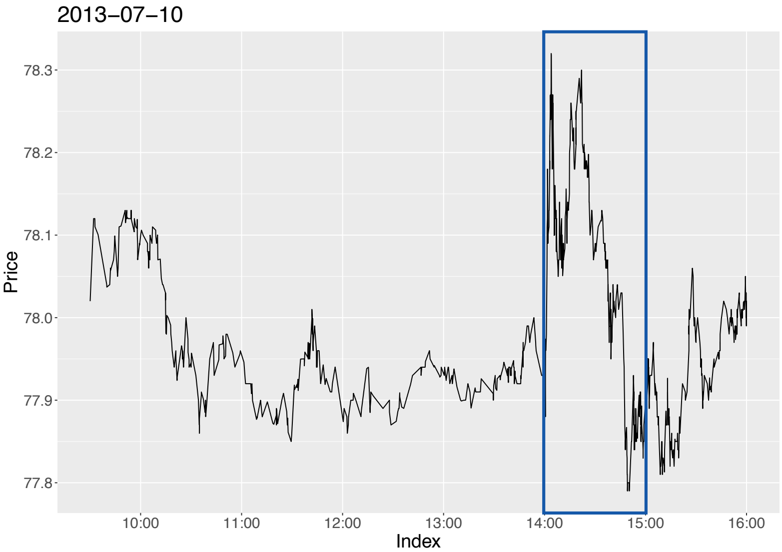
Types of bars:

- Time
- Volume
- Dollar volume
- Tick

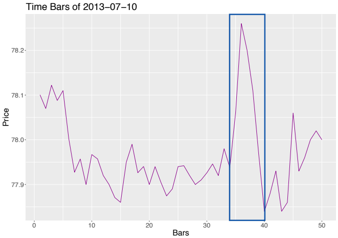


Sampling

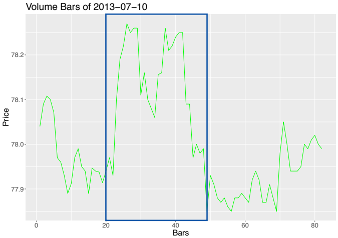
Example day



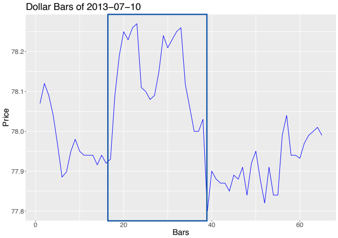
Sampling



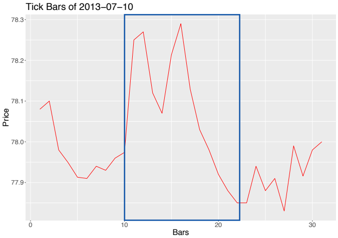
(a) Time bars



(b) Volume bars



(c) Dollar bars



(d) Tick bars

Models

When the different samples have been gathered, **log-returns will be computed for every type of bar**. Then, every type of bar will be **fitted an Autoregressive model** - AR(p):

$$r_k = c + \sum_{i=1}^{p} \theta_i r_{k-i} + \epsilon_k$$

Note that **time bars** will be the **benchmark**, since these bars are the predominant sampling technique.

	Time	Volume	Dollar	Tick
Lag (p)	2	1	10	0
MAPE	2.041	1.130	1.134	1.188
Improvement	0%	44.64%	44.43%	41.82%

1 Introduction

2 Primer in financial data

3 Meta-labeling

4 Fractional differentiation

5 Data parsing as bars

6 Conclusions

Conclusions and future work

Conclusions:

- Efficient market (daily data) $\Rightarrow$ Noise $\Rightarrow$ "Garbage in, garbage out"
- Low signal-to-noise ratio.
- These techniques are not plug-and-play solutions. In fact, we are not dealing with a matter of what, but when.

Future work:

- Develop ML models that take into account alternative financial data.
- Explore High Frequency Trading strategies.

Thank you for your attention