STAT430: Machine Learning for Financial Data

LarryHua.com/teaching

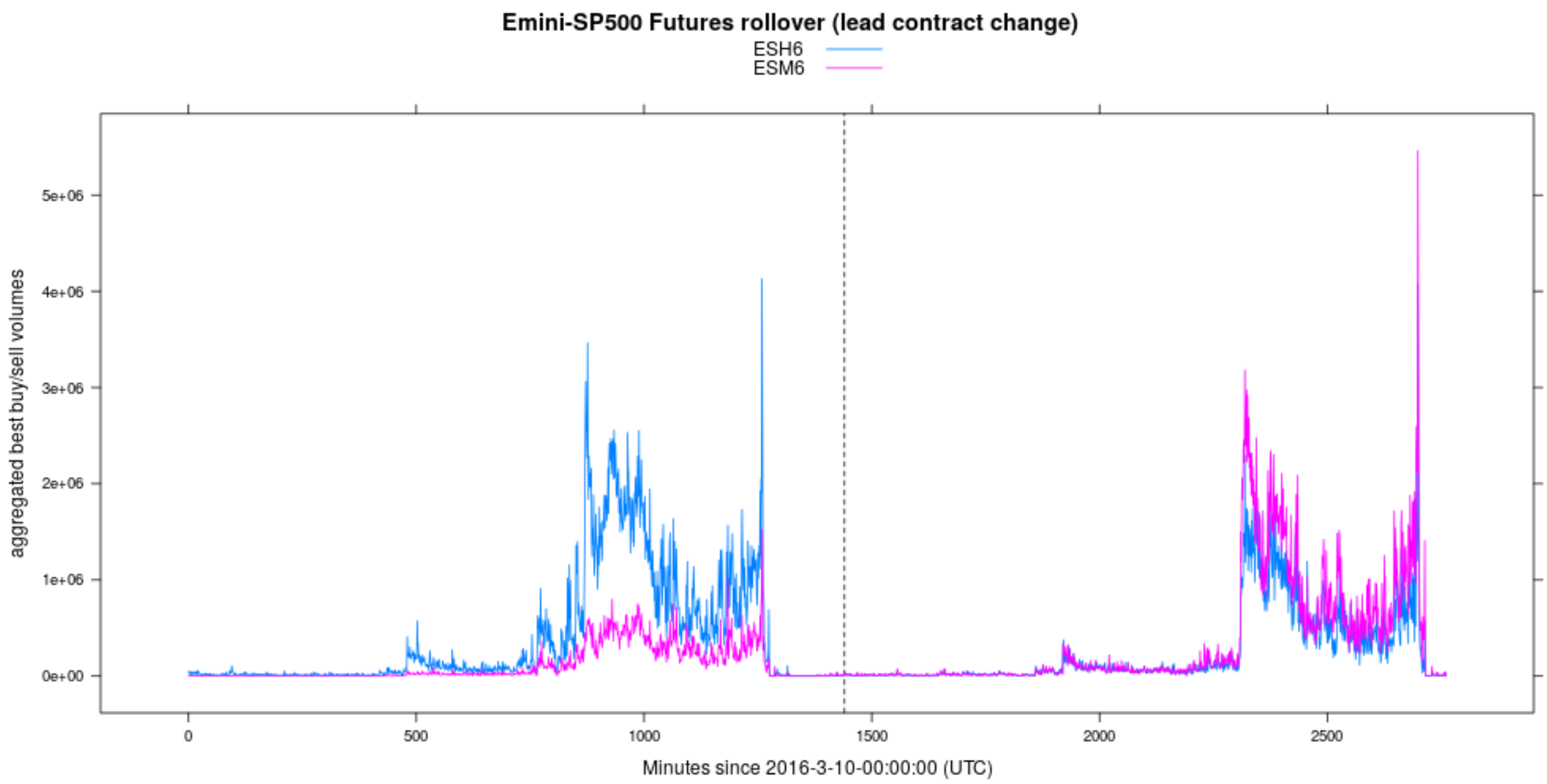
Spring 2019

Futures

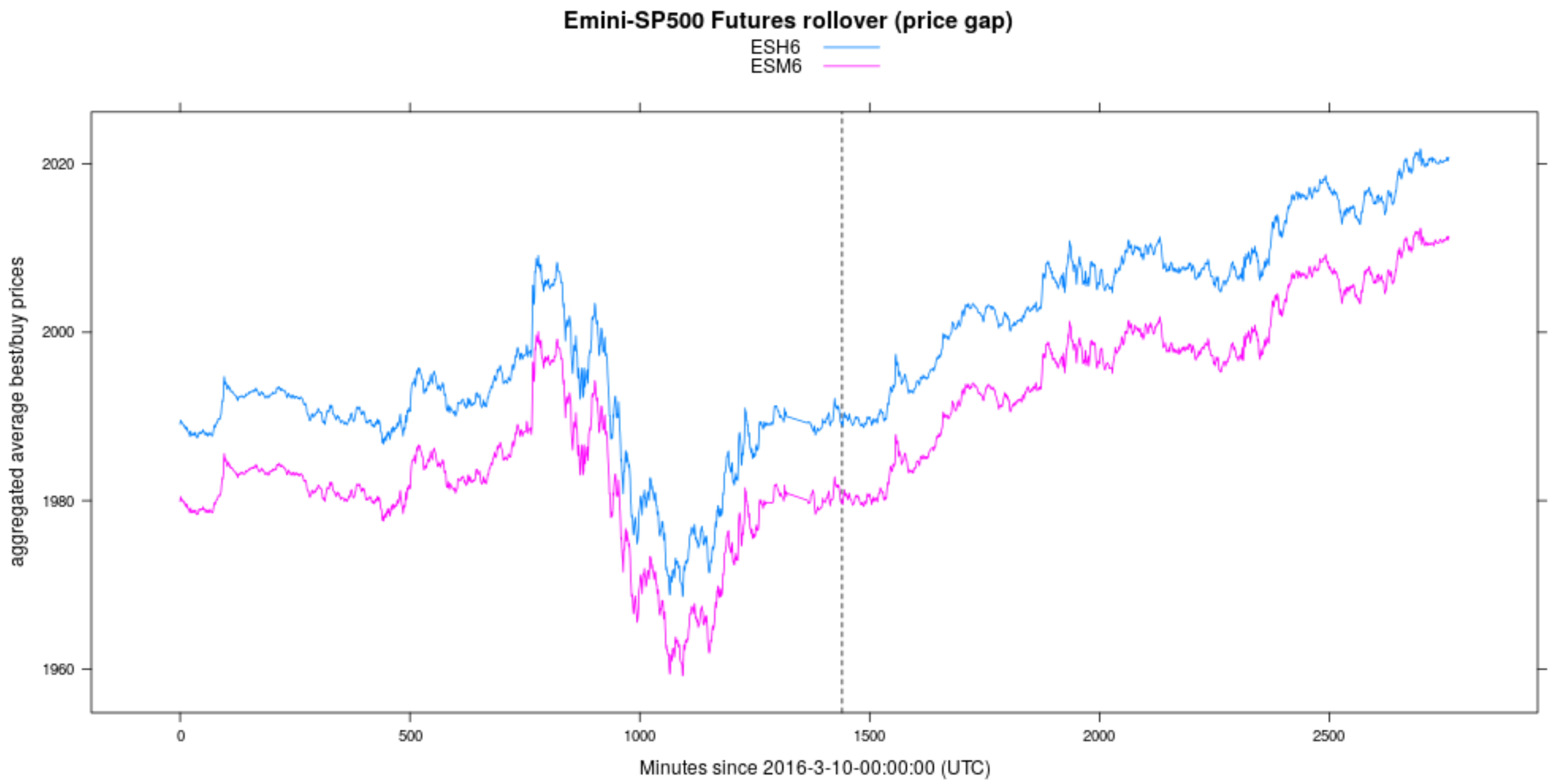
Futures

- Basic concepts
  - Watch videos from CME
  - Naming rules: Base contract + expiration month + expiration year
    - eg: ESH6, Emini-SP500, March 2016
    - Month symbols: 1F, 2G, 3H, 4J, 5K, 6M, 7N, 8Q, 9U, 10V, 11X, 12Z
- Peek into the data.
  - use `fmlr::read_algoseek_futures_fullDepth()`

Futures rollover - lead contract



Futures rollover - price adjustment



Futures rollover - Toy examples

- Negative rolled prices are not present
- Simply fill in the price gaps, either backward or forward

```
price_expiringContract_priorRolloverDay <- x <- c(1, 2, 3, 2, 3, 4)
price_theNextExpiringContract_rolloverDay <- y <- c(5, 6, 5, 6, 7, 7)
gap <- head(y,1) - tail(x,1)
(Price_rolled <- c(x+gap,y))
```

```
##  [1] 2 3 4 3 4 5 5 6 5 6 7 7
```

Futures rollover - Toy examples

- Negative rolled prices are present
- Match the tail-head prices only and then keep the return values

```
price_expiringContract_priorRolloverDay <- x <- c(1, 2, 3, 2, 3, 4)
price_theNextExpiringContract_rolloverDay <- y <- c(2, 3, 4, 3, 2, 3)
gap <- head(y,1) - tail(x,1)
(price_rolled <- c(x+gap,y))
```

```
##  [1] -1  0  1  0  1  2  2  3  4  3  2  3
```

```
return_rolled <- diff(price_rolled) / c(x, y[-length(y)])
(c(1, cumprod(1+return_rolled)))
```

```
##  [1] 1 2 3 2 3 4 4 6 8 6 4 6
```

- Note that this method can also be applied when negative returns are not present, and these two methods are essentially different!

Labeling

Labeling

- Motivation: create response variables for supervised learning
- The Fixed-Time Horizon Method
  - Features matrix X with I rows, each row (X_i) is the features sampled from some bars with index $t = 1, \dots, T$, where $I \leq T$
    - For each X_i (a row vector), assign a label y_i; e.g.,

$$y_i = \begin{cases} -1 & \text{if } r_{t_{i,0},t_{i,0}+h} < -\tau \\ 0 & \text{if } |r_{t_{i,0},t_{i,0}+h}| \leq \tau \\ 1 & \text{if } r_{t_{i,0},t_{i,0}+h} > \tau \end{cases}$$

    - In statistical language: features = independent variables/predictors; label = response variable
    - Questions: think about features that might be useful?

The fixed-time horizon method - limitations

- The h is fixed, and time bars do not exhibit good statistical properties
- The threshold τ is fixed regardless of the observed volatility
- Better ways
  - A varying threshold based on a rolling exponentially weighted standard deviation of returns
  - Volume or dollar bars might lead to stable volatilities (note: different markets)
- But all have a critical limitation: stop-loss limit is omitted

The triple-barrier method

- Motivation: prepare labels that are directly related to trading practice
- **Stop order** (i.e., stop-loss order) is always used to protect from losing positions
  - Stop order: an order to buy or sell a stock once the price of the stock reaches the stop price
    - When the stop price is reached, a stop order becomes a market order
    - Investors use a sell stop order to limit a loss or to protect a profit on a long position
    - Investors use a buy stop order to limit a loss or to protect a profit on a short position

The triple-barrier method

- Path-dependent labeling technique
- Two horizontal barriers and one vertical barrier
- The two horizontal barriers are defined by profit-taking (pt) and stop-loss (sl) limits, which are a dynamic function of estimated volatility
- The third barrier is the expiration limit, defined as the number of bars elapsed
- Label Y: upper touched first: 1; lower touched first: −1; vertical touched first: the sign of the return or 0
- See FIGURE 3.1 of AFML

Configuration triplet [pt,sl,t1]

- If we use 1 and 0 to indicate whether pt/sl/t1 is applied or not, then there are some commonly used cases:
- Three useful configurations:
  - $[1, 1, 1]$: Want to realize a profit, but a maximum tolerance for losses and a holding period.
  - $[0, 1, 1]$: Want to exit after a number of bars, unless stopped-out.
  - $[1, 1, 0]$: Want to take a profit as long as we are not stopped-out.

Configuration triplet [pt,sl,t1]

- Three less realistic configurations:
  - $[0, 0, 1]$: The fixed-time horizon method. It may still be useful when applied to volume-, dollar-, or information-driven bars, and multiple forecasts are updated within the horizon.
  - $[1, 0, 1]$: A position is held until profit taking or the maximum holding period.
  - $[1, 0, 0]$: A position is held until profit taking.

The triple-barrier method - implementation

- The horizontal barriers can be asymmetric, which is controlled by `ptSl`

```
#' @param x: time series to be labeled
#' @param events: dataframe, has the following two columns:
#'                 t0: event's start time index
#'                 t1: event's end time index;
#'                     Inf: no vertical barrier;
#'                 trgt: unit absolute return
#'                       used to set up upper and lower barrier
#'                 side: 0: no side; 1: up; -1: down
#' @param ptSl: two multipliers for upper and lower barriers

label_meta <- function(x, events, ptSl){...}
```

The triple-barrier method - implementation

- Initialize the events

```
unit_bar <- bar_unit(dat, unit=6e9)
i_CUSUM <- istar_CUSUM(unit_bar$C, h=1000)
n_Event <- length(i_CUSUM)

events <- data.frame(t0=i_CUSUM+1, t1 = i_CUSUM+100,
                     trgt = rep(0.002, n_Event),
                     side=rep(0,n_Event))
x <- unit_bar$C
ptSl <- c(1,1)
out0 <- label_meta(x, events, ptSl)
```

The triple-barrier method - implementation

- For each row of the feature matrix

```
t0 <- events$t0
t1 <- events$t1
trgt <- events$trgt
side <- events$side
u <- ptSl[1]
l <- ptSl[2]
```

The triple-barrier method - implementation

- For symmetric upper and lower barriers

```
if(i_side==0)
{
  up <- i_trgt*u
  lo <- i_trgt*l
  isup <- (i_x/i_x[1]-1) >= up
  islo <- -(i_x/i_x[1]-1) >= lo
  T_up <- ifelse(sum(isup)>0, min(which(isup)), Inf)
  T_lo <- ifelse(sum(islo)>0, min(which(islo)), Inf)
}

...

ret <- i_x[min(T_up, T_lo, length(i_x))] / i_x[1] - 1
rst <- c(T_up, T_lo, i_t1, ret)
```

Meta-labeling

- A primary model for learning only the side of the bet
    - e.g., moving average crossing strategy
- Given the side of the bet, learn the size of the bet
- Quantamental investment research

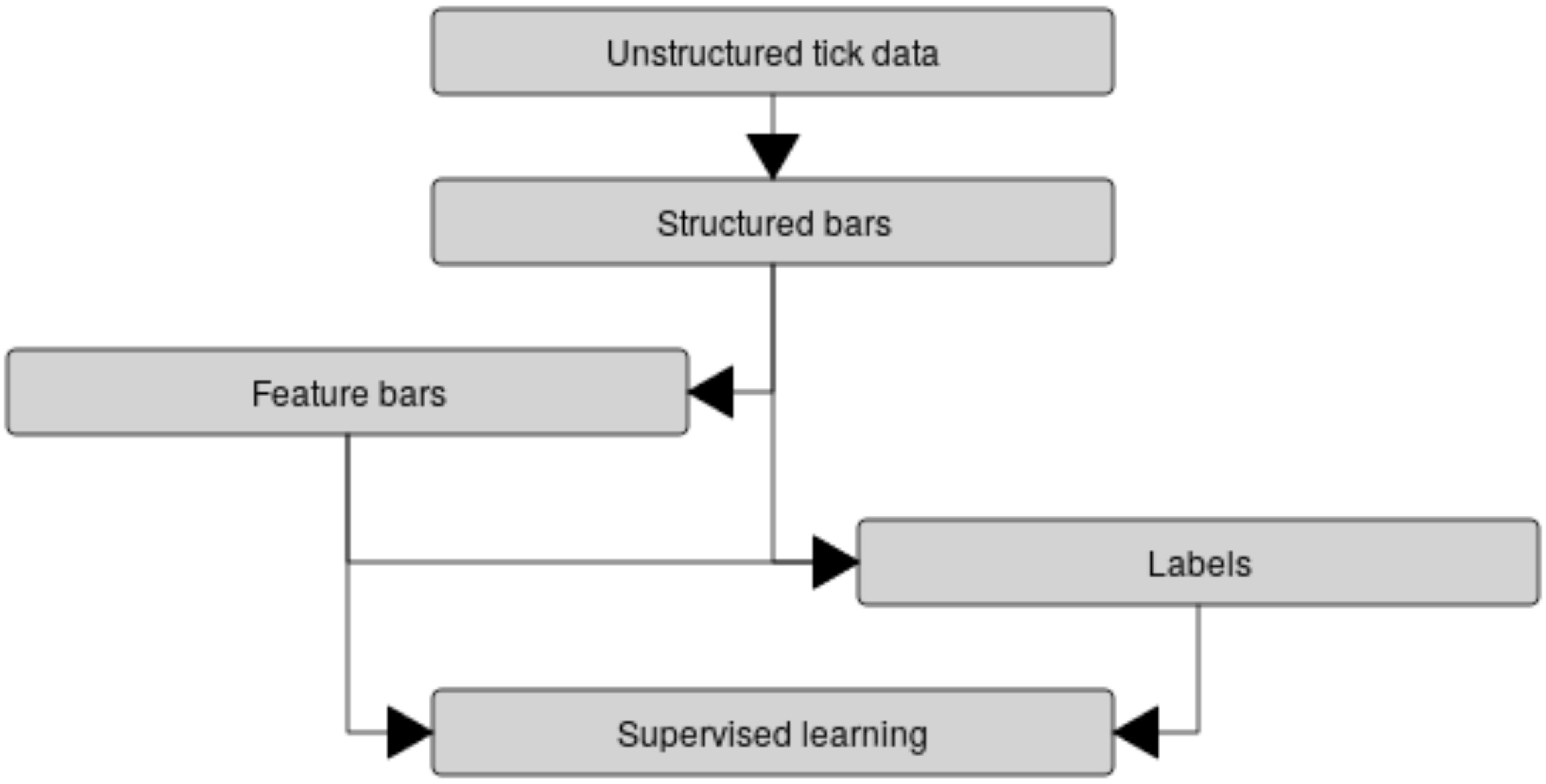
Meta-labeling

- Re-use the **label_meta** function

```
...

else if(i_side == 1)
       {
         up <- i_trgt*u
         isup <- (i_x/i_x[1]-1) >= up
         T_up <- ifelse(sum(isup)>0, min(which(isup)), Inf)
         T_lo <- Inf
}
...
```

Data process flowchart



- [Try R]
- [Back to Course Scheduler]