# Advanced Python for Finance

Lecture 4

## Lecture 4

- Optimization of List Appends
- Target Functions
- Tolerance Functions
- VWAP
- Factor based approaches
- A simple predictive model: the Lee Ready Tick Test

# Appending to DataFrames and Lists: A Comparison

See L4.1 Append Operations.ipynb

# TWAP and VWAP: Target Functions

# **Target Functions**

Both TWAP and VWAP require you to know what your target % complete is at any given time.

Create a cumulative volume function such that:

Target % complete at time t = f(t)

TWAP: even distribution over n bins, e.g. 1/390 \* t total shares to each bin

Cumulative TWAP:

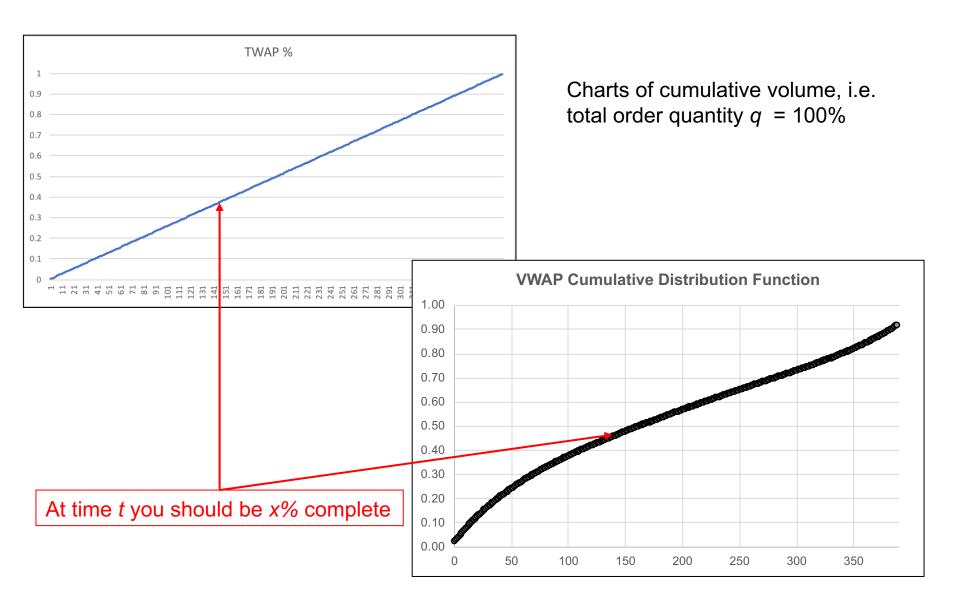
Target % complete at time t = bin t / total bins \* total sharese.g. for 1 minute bins, an order to buy 100000 shares at minute 39 should be (39 / 390) \* 100000 = 10000 complete.

Cumulative VWAP : Target % complete at time t = something like

Target % =  $c_1t + c_2t^2 + c_3t^3 + c_4t^4 + c_5t^5$ 

More on this later...

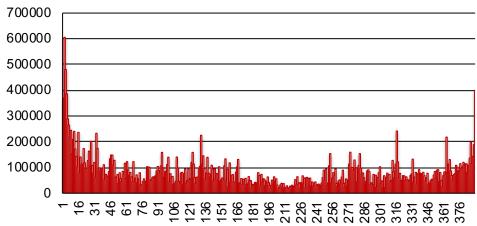
# Target Function – Visual Intuition



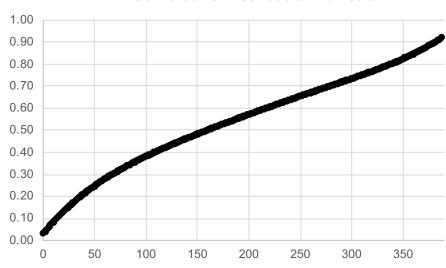
## VWAP: Volume Distribution Profile

- 1. The distribution of volume over the course of a trading day in the US is quite stable.
- 2. It can be used to estimate the proportion of volume that will trade *on average* in a given segment of the day.
- 3. Using the profile as a guide, you can apply other techniques to improve order placement and volume prediction





#### **AAPL Cumulative Distribution Function**



# Building a Volume Profile - Data

BarTp=T							
				LAST_PRI			
	OPEN	HIGH I	LOW	CE	NUMBER_TICKS	VOLUME	VALUI
9/15/16 9:30	113.86	114.25	113.49	114.09	6,141	2,531,717	288,265,40
9/15/16 9:31	114.099	114.35	114	114.1099	5,341	1,390,148	158,717,05
9/15/16 9:32	114.1	114.28	113.72	113.7519	3,637	948,012	108,137,63
9/15/16 9:33	113.8	114.01	113.74	113.91	2,827	640,697	72,981,64
9/15/16 9:34	113.9001	114.2	113.84	114.17	2,418	580,885	66,263,00
9/15/16 9:35	114.17	114.4	114.17	114.252	3,742	858,753	98,137,39
9/15/16 9:36	114.26	114.47	114.21	114.435	2,945	744,893	85,186,47
9/15/16 9:37	114.445	114.65	114.43	114.554	3,938	939,103	107,570,56
9/15/16 9:38	114.569	114.73	114.51	114.61	3,243	773,059	88,624,21
9/15/16 9:39	114.62	114.62	114.3	114.3799	2,071	529,009	60,537,70
9/15/16 9:40	114.37	114.51	114.33	114.36	1,833	450,583	51,565,11
9/15/16 9:41	114.35	114.4	114.1799	114.35	2,638	668,989	76,453,02
9/15/16 9:42	114.36	114.43	114.34	114.34	1,544	356,846	40,817,20
9/15/16 9:43	114.34	114.36	114.19	114.35	2,083	427,851	48,893,00
9/15/16 9:44	114.345	114.36	114.13	114.26	2,261	434,212	49,595,81
9/15/16 9:45	114.26	114.26	114	114.01	2,174	536,700	61,239,47



- 1. Decide on data requirements
  - Bar size (1 minute in this example)
  - Number of days of data (often a rolling 30 day window)
- 2. Acquire historical data
  - Open, High, Low, Close and Volume
- 3. Prepare the data
  - Convert timestamp to bin (1-390 for US Stock Market)
  - Calculate Cumulative Volume
  - Normalize Cumulative Volume -> Total volume = 1%
- 4. Fit the data via regression or some other method

ate		bin	VOLUME	CUM_VOL	PCT
	9/15/16 9:30	1	2,531,717	2,531,717	0.03
	9/15/16 9:31	2	1,390,148	3,921,865	0.05
	9/15/16 9:32	3	948,012	4,869,877	0.06
	9/15/16 9:33	4	640,697	5,510,574	0.07
	9/15/16 9:34	5	580,885	6,091,459	0.07
	9/15/16 9:35	6	858,753	6,950,212	0.08
	9/15/16 9:36	7	744,893	7,695,105	0.09
	9/15/16 9:37	8	939,103	8,634,208	0.1
	9/15/16 9:38	9	773,059	9,407,267	0.11
	9/15/16 9:39	10	529,009	9,936,276	0.12
	9/15/16 9:40	11	450,583	10,386,859	0.13
	9/15/16 9:41	12	668,989	11,055,848	0.13
	9/15/16 9:42	13	356,846	11,412,694	0.14
	9/15/16 9:43	14	427,851	11,840,545	0.14
	9/15/16 9:44	15	434,212	12,274,757	0.15

# Another Approach: Bar-wise Averaging

## Rather than fit a function, create a normalized average curve

Given a selected set of minute bar data:

- 1. Calculate cumulative "target % complete" as in Method 1
- 2. Calculate bin-wise average "target % complete" over the population period

# A Basic VWAP Algo

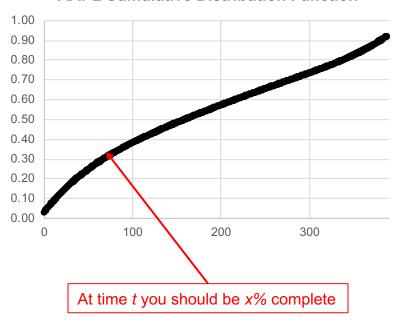
#### Order Placement

- Target Volume = volume\_at\_time(q, t) where q
   total order quantity and t = time segment
- 2. Target order size at time t = Target volume at time t+n Executed volume
- 3. Example Primitive Order Placement Strategy
  - Define Shortfall => Target volume at time t –
     Executed Volume
  - Threshold value is the tolerance for being off target volume distribution
  - Passive if Shortfall <= threshold value</li>
  - Aggressive if Shortfall > threshold value
- 4. Advanced versions may leverage non-displayed orders, etc. depending on size of shortfall.

Target the Volume Weighted Average Price  $p_{VWAP}$ :

$$p_{VWAP} = \frac{\sum (quantity_{@price} \times price)}{total\ quantity}$$

#### **AAPL Cumulative Distribution Function**



# VWAP Profile and Algo

- L4.2 VWAP Profile.ipynb
- L4.3 Preprocessing Tick Data.ipynb
- L4.4 VWAP with Tick Data v1.ipynb

## Tolerance Factors

## **Tolerance Factors**

We have now established the concept of target functions.

The next step is to consider how we will allow ourselves to vary from the deterministic target function over time in order to take advantage of market fluctuations, short term forecasts, etc.

We will define the concept of **tolerance factor**.

## **Tolerance Factors**

### Establishing a Tolerance factor:

- How far can you get behind (or ahead)
- Think of this factor as a proxy for risk tolerance

#### Simple intuition:

- Tolerance is high = patient. Capture spread, optimize for cost
- Tolerance is low = aggressive. Stay close to target, pay spread. Risk averse

#### Threshold model A:

- Calculate a threshold value, when you are behind by some amount > threshold value then trade aggressively until amount behind < threshold</li>
- 2. Threshold value = % behind% behind = (1- actual) / target

# A Simple VWAP with Tolerance

L4.5 VWAP with Tick Data v1.5.ipynb

Lee-Ready: Trade Identification

## Trade Identification

### **Inferring Trade Direction from Intraday Data**

Charles M. C. Lee; Mark J. Ready The Journal of Finance, Vol. 46, No. 2 (Jun., 1991), 733-746.

- All Trades are a match between one or more buy orders and one or more sell orders.
- Refer back to our concepts of passive (providing) and aggressive (taking)
- The liquidity taking order will cause the the trade to occur
- The question: can we identify "directionality" of trades, i.e. did a buy or a sell order trigger the trade?
- If we can identify buys and sells from trade data then we can potentially use that information in our analyses and trading signals

## Trade Identification

We will consider three general cases

- 1. Identifying Trades without Quotes
- 2. Identifying Trades with Quotes
- 3. Trades inside the Spread

## Trade Identification without Quotes

- We often may have only trades, not quotes
- In these cases, we can derive information from the trades themselves
- Let's classify trades relative to the trades <u>immediately</u> <u>preceding</u> them by two methods:
  - Tick Test
  - Reverse Tick Test

## Tick Test

Look at trade price compared with the immediately preceding trade:

- 1. The trade is an **uptick** if the price is *higher* than the previous trade.
- 2. The trade is a **downtick** if the price is *lower* than the previous trade.
- 3. The trade is a **zero-uptick** if the previous trade was at the same price, but the previous price change was an *uptick*.
- 4. The trade is a **zero-downtick** if the previous trade was the same price, but the previous price change was a *downtick*.

#### Classification

A trade is a **buy** if it occurred on an uptick or a zero-uptick

A trade is a **sell** if it occurred on an downtick or zero-downtick

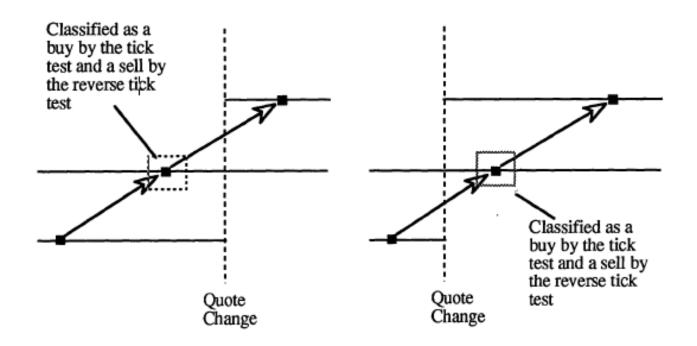
## Reverse Tick Test

The reverse tick test compared to the immediately **following** trade.

- 1. If the following trade is at a *higher* price, the current trade is classified as a **sell**.
- 2. If the following trade is at a lower price, the current trade is classified as a **buy**.

## Tick Test and Reverse Tick Test

Classification of Trades in Trending Markets with Quote Changes



## Tick Test and Reverse Tick Test

Table I

# A Comparison of the Tick Test and the Reverse Tick Test When Identification of the Prevailing Quote is Unambiguous

For a sample of 150 NYSE firms during 1988, this table reports the percentage of trades classified as buys and sells by the tick test and the reverse tick test when the identity of the prevailing quote is unambiguous. A trade is classified as a buy (sell) by the tick test if the prior price change is positive (negative). A trade is classified as a buy by the reverse tick test if the next price change is negative (positive). A prevailing quote is considered unambiguously identified if the most recent quote revision occurred more than 5 seconds before the trade. This situation applied to 90.9 percent of the classified (nonopening) trades in the sample.

Classification Based on the Prevailing Quote	Tick Test Classification	Reverse Tick Test Classification
Buy-Above Ask	98.8% buy	91.3% buy
Buy—At Ask	92.1% buy	80.0% buy
Inside the Spread	52.4% sell	52.0% buy
Sell—At Bid	90.2% sell	79.4% sell
Sell—Below Bid	98.7%  sell	94.5%  sell

## Trade Identification with Quotes

We may have quotes, but which ones are the right ones to use?

Lee-Ready asks: are quotes at the time of a trade the <u>cause</u> or the <u>result</u> of a trade?

Using their *isolated trades*, they established that if a quote is *less than 5 seconds old*, it was probably caused by the trade being examined.

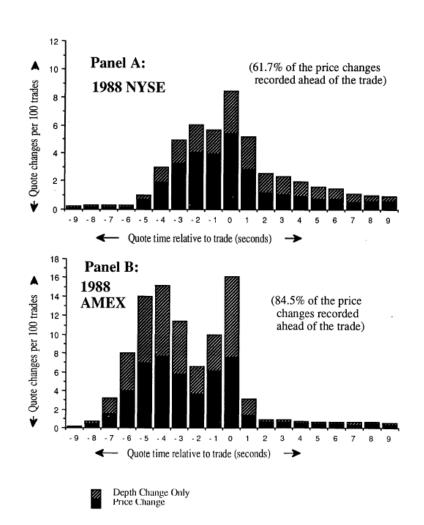


Figure 2. The frequency of quote revisions around isolated trades. The frequency of quote revisions in a 19-second window centered on isolated trades for 150 NYSE firms and all AMEX firms during 1988. An isolated trade is defined as the first trade after 11:00 a.m. but before 2:30 p.m., with no other trades within a 2-minute window centered on the trade. Both quote revisions which reflect price changes and quote revisions which affect depth changes only are shown.

# Trades Inside the Spread

### Proposed options

- Location relative to quote ("bidness" and "askness")
  - A trade is a buy if the price is closer to the ask
  - A trade is a sell if the price is closer to the bid
  - Trades at midpoint are discarded

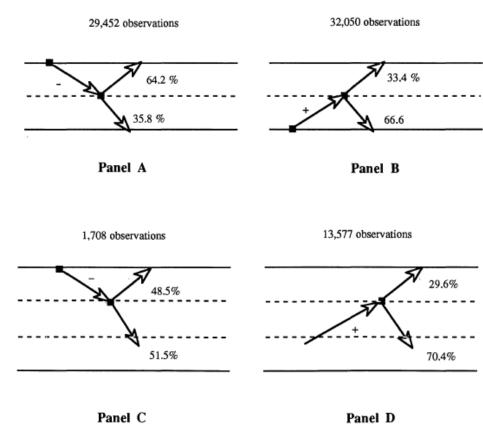


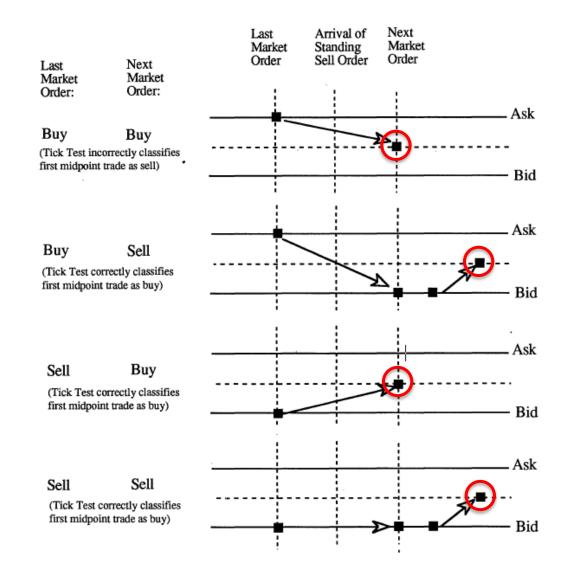
Figure 3. Frequency of price reversals for trades inside the spread. For a sample of 150 NYSE firms during 1988, these figures report the frequency of trades at the midpoint of a 1/4 spread (Panels A and B) and at the upper inside point of a 3/8 spread (Panels C and D). Panels A and C show the number of trades which occurred on a downtick and the proportion of uptick and downtick movements in the next price change. Panels B and D show the number of trades which occurred on an uptick and the proportion of uptick and downtick movements in the next price change.

Source: Lee, Ready (1991) (NOTE: ¼ spread = \$0.25, 3/8 spread = \$0.375)

# Trades at Midpoint

According to the scenarios put forth in the paper, the tick test is effective at classifying trades inside the spread as well.

Figure 4. Classifying the first midpoint trade after the arrival of a standing sell order. Assume constant bid and ask prices and a ¼ spread. Further assume market buy and sell orders follow independent Poisson processes with the same arrival rate. Standing buy and sell orders also follow independent Poisson processes with a lower average arrival rate than market orders. The four equally likely patterns of market orders immediately before and after the arrival of a standing order are depicted above (buy/buy, buy/sell, sell/buy, and sell/sell). In three of the four cases, the tick test classification for the first midpoint trade is correct.



# Summary of the Lee-Ready Algorithm

- 1. If quotes are available, use them:
  - a) If quotes are < 5 seconds (!) old, use previous quote
  - b) Trades at the bid are sells
  - c) Trades at the offer are **buys**
- 2. If trade is inside the spread but not at midpoint, classify by proximity to the bid or offer, apply same logic as #1
- 3. If trade is at midpoint or no quotes are available, use the tick test.

See the original paper for more detail

# Limitations of Lee Ready

## **Changing Market Structure**

- Narrower spreads
- Decimalization (2001)
- The original paper's 5-second window!
- Replication with isolated trades
- Market fragmentation
  - Non-displayed liquidity
  - Latency considerations

# Signal Factor Example: Tick Test

How do we use the theoretical value in trading decisions?

- 1. Consider a simplified conceptual version of a tick test signal  $T_i$
- 2. Derive a signal from autocorrelation of trade directions:
  - State 1: Next tick predicted up
  - State 2: Next tick predicted down
  - State 3: Next tick indeterminate
- 3. Assume for the moment we want to trade immediately based on that signal
- 4. For a stock j, define price  $s_i$  to be the midpoint within the current quote.

We can then express our three states as:

- State 1: Next tick predicted up (buy): FV = the offer price = midpoint + spread/2
- State 2: Next tick predicted down (sell): FV = the bid price = midpoint spread/2
- O State 3: Next tick indeterminate: FV = midpoint

# Measuring Autocorrelation

Let's look at the autocorrelation of the tick data:

- 1. Load Tick Data, (just trades)
- 2. Apply Lee-Ready (tick test) to do up/down
  - Use shift() operator to obtain previous row
  - Calculate difference
  - Use numpy sign() operator to retrieve the sign for each record
  - O Apply from tick 2 to n
- 3. Calculate autocorrelation for n lags
  - Use autocorr(lag=n) for specific lags
- 4. Examine and Plot

# Signal Factor Implementation: Tick Test

Once again, we can express this the signal as a continuous function:

$$FV = midpoint_j + b_{jF}F_{tick}$$

Where

 $b_{iF}$  = loading of the tick signal factor

 $F_{tick}$  = tick signal factor described as buy: 1, sell: -1

Calculate the exponential moving average of the tick value, something like this:

$$F_n = \frac{1}{\tilde{\tau}_w} f_n + w_n F_{n-1}$$

Where the weighting factor is defined as a function of a volume window  $\tilde{\tau}_w$ :

$$W_n = e^{-(t_n - t_{n-1})/\tilde{\tau}_W}$$

In practice you would calibrate the decay factor once and store the value.

 $f_n$  is by construction in {-1,1}, so multiply by  $\frac{spread_{avg}}{2}$  to scale the factor.

Source: Derived from Michael Sotiropoulos BAML Lecture on trading models

# **Determining Volume Window**

For the volume window, we want trade time, not clock time.

During simulation we can count trades tick by tick and construct a rolling window of *n* trades.

But if you want to convert to clock time:

$$\tilde{\tau}_{w} = max \left( \tau_{min}, \left( \tau_{max}, \frac{T_{day}}{N_{trd}} n \right) \right)$$

#### Where:

 $\tau_{min}$ ,  $\tau_{max}$  are user defined min and max values

 $T_{day}$  = length of the trading day (390 minutes for the US stock exchanges)

 $N_{trd}$  = avg number of trades per day

n = target number of trades to include

Source: Derived from Michael Sotiropoulos presentation on trading models

# Theoretical Values and Multi-factor Price Models

# Rules and Multiple Signals

Assume we are buying and consider if we have a model with three signals

#### Signal 1

#### Signal 2

```
if momentum_signal > momentum_threshold: # trade aggressively
    buy at offer or sell at bid
```

#### Signal 3

```
if reversion_signal >= reversion_threshold: # trade passively
    buy at bid
```

## Combinations of Factors

Before we get to linear combinations, let's consider multiple factors in a simple "voting" mechanism.

#### Example Assumptions:

- 1. Consider a VWAP **buy** order
- 2. We can only buy on the bid (passive) or the offer (aggressive). Signal must be +0.5 to be aggressive
- 3. Assume equal weighting of all factors

Case	Schedule	Momentum	Reversion	FV adjustment
1	On schedule: 0	Tick up: +0.5	Neutral: 0	+0.5 aggressive
2	Very behind: +0.5	Tick down: -0.5	Neutral: 0	0.0: passive
3	Very behind: +0.5	Tick down: -0.5	Down: -0.5	-0.5 passive
4	A little behind: +0.2	Tick up (weaker) +0.2	Up (weak): +0.1	+0.5 aggressive

This gets even more interesting for non-execution strategies with bi-directional signals!

## Multiple Signals and Theoretical Values

We need to be able to combine <u>multiple factors</u> into one expression of <u>a desire to trade at a given price</u> (passive, aggressive, or anywhere in between)

Applicable to any signal, not just tick-level / microstructure

#### Tolerance Factors: Model B

- Apply adjustment as a continuous factor
  - Combines with other factors
  - Can support concept of "ahead" as well as "behind" (i.e. if my short term signals are strong, can I trade ahead to some degree?)
- Use % ahead/behind to adjust target price continuously
  - $\circ$  Start with a mid or "micro" price  $P_m$  = midpoint of spread
  - $\circ$  Apply "catchup" factor  $F_c$  such that  $0.0 < F_c < (Q_{ask} Q_{bid})$ :
  - If I'm buying, and not behind, my catchup factor is 0. buy at bid
  - If I'm buying and sufficiently behind, let's say my catchup factor is 1.0.
     Buy at offer
  - o If I'm only slightly behind, F<sub>c</sub> might be 0.2
  - O Buy Order Placement: "Fair" price  $P_f = P_m ((Q_{ask} Q_{bid}) / 2) + F_c$
  - $\circ F_c = k(Q_{ask} Q_{bid})$
  - o k = c \* (target actual) / total shares
  - c = calibration factor tune based on empirical performance

# **Arbitrage Pricing Theory**

Consider Arbitrage Pricing Theory (APT)

(Ross, 1976), A decent intro at: <a href="https://en.m.wikipedia.org/wiki/Arbitrage">https://en.m.wikipedia.org/wiki/Arbitrage</a> pricing theory

A stock's return can be expressed as a linear combination of multiple factors:

$$r_j = a_j + b_{j1}F_1 + b_{j2}F_2 + \dots + b_{jn}F_n + \epsilon_j$$

#### Where:

 $a_i$  = constant factor for stock j

 $F_k$  = factor

 $b_{jk}$ = stock j's sensitivity to factor  $F_k$ 

### Fair Value

How does this help us?

We can add arbitrarily many factors in linear combination to arrive at a predicted return and a "fair value" or "FV" of a stock.

$$FV = s_j + r_j$$

Factors can be

- signals such as tick test, order imbalance
- risk-based
- or anything else you can think of

#### Fair Value Factors - Intuition

When we think something is over priced we sell:

Lower FV versus the market

When we think something is under priced, we buy: Raise FV so stock looks undervalued by the market

## Signal Factor Example: Tick Test

How do we use the theoretical value in trading decisions?

- 1. Consider a simplified conceptual version of a tick test signal  $T_i$
- 2. Derive a signal from autocorrelation of trade directions:
  - State 1: Next tick predicted up
  - State 2: Next tick predicted down
  - State 3: Next tick indeterminate
- 3. Assume for the moment we want to trade immediately based on that signal
- 4. For a stock j, define price sj to be the midpoint within the current quote.

We can then express our three states as:

- State 1: Next tick predicted up (buy): FV = the offer price = midpoint + spread/2
- State 2: Next tick predicted down (sell): FV = the bid price = midpoint spread/2
- State 3: Next tick indeterminate: FV = midpoint

# Signal Factor Example: Tick Test

More generally, we can express this the signal as a continuous function:

$$FV = midpoint_j + b_{jF}F_j$$

Where

 $b_{iF}$  = loading of the tick signal factor

 $F_i$  = tick signal factor described in spread units in the range [-0.5..0.5]

- O State 1: Next tick predicted up (buy): FV = the offer price = midpoint + 0.5
- State 2: Next tick predicted down (sell): FV = the bid price = midpoint − 0.5
- State 3: Next tick indeterminate: FV = midpoint + 0
- States between 1 and 2:  $FV = midpoint + F_i$  spread units

## Risk-based Factor Example: Position

Consider a proprietary trading strategy in which a variety of signals are working.

In the absence of information you would prefer to *not* have a position:

- 1. If you are long you would prefer to be flat (sell to close)
- 2. If you are short, you would prefer to be flat (buy to close)

## Risk-Based Factor Example: Position

We can again express this the signal as a continuous function:

$$FV = midpoint_j - b_{jF}F_j$$

Where

 $b_{iF}$  = loading of the position factor

 $F_i$  = risk adjustment factor described in spread units in the range [-0.5..0.5]

- $\circ$  State 1: Long (sell): FV = the bid price = midpoint 0.5
- State 2: Short (buy): FV = the offer price = midpoint + 0.5
- State 3: 0 position: FV = midpoint + 0
- States between 1 and 2: FV = midpoint +  $F_j$  spread units

## Risk-Based Factor Example: Schedule

Consider a schedule based strategy.

"Risk" is defined as the difference between your current position and your target position. The greater the difference (i.e. the greater behind), the greater your risk

#### If you are buying:

- 1. If you are behind, you want to buy more: raise FV
- 2. If you are ahead, you want to buy less: lower FV
- 3. If you are on schedule, you are neutral. FV at midpoint

# Risk-Based Factor Example: Schedule

And again we can again express this the signal as a continuous function:

$$FV = midpoint_j + b_{jF}F_j$$

Where

 $b_{iF}$  = loading of the tick signal factor

 $F_i$  = schedule factor described in spread units in some range, e.g. [-0.5..0.5]

- State 1: behind (buying): FV = the offer price = midpoint + 0.5 # trade now
- State 2: ahead (buying): FV = the bid price = midpoint − 0.5 # slow down
- State 3: on schedule: FV = midpoint + 0 # no impact
- O States between 1 and 2:  $FV = midpoint + F_i$  spread units
- Reverse signs for selling

## Putting it together for VWAP

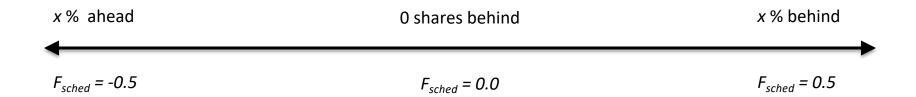
$$FV_j = midpoint_j + b_j F_{tick} + b_j F_{imbal} + b_j F_{schedule}$$

Where for stock *j*:

 $F_k$  = factor for each signal or adjustment

 $b_{jk}$ = stock j's sensitivity to factor  $F_k$ 

### VWAP v2 Schedule Factor Construction



$$F_{schedule} = \frac{q_{pct\_behind}}{q_{max\_behind}} \times \frac{spread_{avg}}{2}$$

**Example 1**: If max\_behind = 0.02 and average spread = \$0.05, then:

At 0.02 behind: 
$$\frac{0.02}{0.02} \times \frac{0.05}{2} = 0.025$$

**Example 2**: If max\_behind = 0.02 and average spread = \$0.05, then:

At 0.01 behind: 
$$\frac{0.01}{0.02} \times \frac{0.05}{2} = 0.0125$$

#### VWAP v2 Fair Value Calculation

$$FV_j = midpoint_j + b_j F_{schedule}$$

#### Where:

 $F_k$  = factors for each signal or adjustment

 $b_{jk}$ = stock j's sensitivity to factor  $F_k$ 

#### **Example 1 continued:**

if bid = 100, offer = 100.05, max\_behind = 0.02 and avg\_spread = 0.05

midpoint = 100.025

 $F_{schedule} = 0.025$ 

FV = 100.025 + 0.025 = 100.05: this is >= offer price, so aggress

What happens if current spread is < or > average spread??

#### VWAP v2 Order Placement

$$FV_j = midpoint_j + b_j F_{schedule}$$

It's time to place our order. Even though our FV should now reflect our "behindness", we still may not necessarily want to place the order at that price in call cases.

Why?

Simple placement logic for order\_side = BUY:

```
if FV >= offer:
    order_price = offer
else # FV is not sufficiently high to warrant an aggressive order
    order_price = bid
```