



US Equity Trade and Quote Second Bar Extended Guide

version 1.0 (Jan 2024)



algoseek | the market data company

We provide research market data for machine learning and quantitative trading



CONTACT US

We are here to help you do great things with our market and reference data. For questions, feedback, and other concerns, you may reach our team of experts using the following contact information:

algoseek customer support

✉ support@algoseek.com

☎ (+1) 646 583 1832

algoseek sales

✉ sales@algoseek.com

☎ (+1) 646 583 1832



TABLE OF CONTENTS

INTRODUCTION	4
DATA SOURCE	4
STANDARD AND NO-FINRA/TRF VERSIONS	4
MINUTE BAR CALCULATIONS	5
DATA ORGANIZATION AND FILE FORMAT	7
APPENDIX A. FREQUENTLY ASKED QUESTIONS	15
APPENDIX B. BAR CALCULATIONS FROM TRADE AND QUOTE EVENTS	16



INTRODUCTION

algoseek Trade and Quote (TAQ) Second Bar Extended data is built from top-of-book intraday quotes and trades for all listed stocks, ETNs, ETFs, ADRs, and funds from 15+ US exchanges and marketplaces.

With 89 data fields, algoseek TAQ Second Bar datasets are the most comprehensive and detailed TAQ second bar products in the financial industry. They are designed for quantitative trading, backtesting, machine learning, and other advanced applications.

Data files are in CSV (Comma-Separated Values) format. An individual CSV file is created for each active ticker on each trading day, and these data files are arranged in a flat-file database by date and then by ticker.

DATA SOURCE

algoseek TAQ Second Bar datasets are built from “as-is” tick data collected from live SIP feed algoseek’s co-located ticker plant servers in Equinix NY2 and NY4 data centers, connected with 10Gb fiber for low latency.

The Securities Information Processor (SIP) includes Tape A and Tape B covered by the Consolidated Tape Association (CTA) plan and Tape C covered by the Unlisted Trading Privileges (UTP) plan. The SIP links the US markets by processing and consolidating all protected bid/ask quotes and trades from every trading venue into a single and easily consumable data feed.

The SIP disseminates and calculates critical regulatory information, including the National Best Bid and Offer (NBBO) and Limit Up Limit Down (LULD) price bands, among other important regulatory information such as short sale restrictions and regulatory halts. In the highly fragmented world of US equities, the SIP is an easy way for people to get a view of the current state of the market.

FINRA TRF AND ODD LOT TRADES

FINRA TRF

Equity trades are executed on Public Exchanges (e.g., NASDAQ, BATS, NYSE, ARCA, etc.) and off the public exchanges in Dark Pools, Broker-Dealer internal crossing, and Block Trades.

Regulation National Market System (NMS) requires all trades to be reported. There are currently three FINRA Trade Reporting Facilities (TRF) affiliated with registered national



securities exchanges and provide FINRA members with a mechanism for reporting transactions affected otherwise than on an exchange.

Regulation NMS allows up to 10 seconds after the Trade execution time for the trade report to be sent to an exchange's TRF for publication. The delay can result in TRF Trade reports printed on the market data feed being out of the current NBBO.

Round Lot and Odd Lot

A round lot (or board lot) is a normal unit of trading of a security, which currently is 100 shares of stock in the US. Any quantity less than 100 shares is referred to as an odd lot. Odd lots are not subject to the Regulation NMS rules requiring execution to be within the current NBBO. Broker-dealers send odd lots to the exchange paying the most rebate per share and not the best execution price. *Odd lot executions can create unrealistic high/low trade prices in an OHLC bar.*

SECOND BAR CALCULATIONS

Aggregated NBBO Bid/Ask Size

The Bid/Ask size in a bar field is the total aggregate from all the exchanges that have a matching price to the NBBO Bid/Ask price.

Event-based Bars

algoseek's equity TAQ second bars are event-based, which means a bar is only created when there was at least one event during the bar period. The three types of events (bid, ask, and trade) don't always happen together during a bar period. For example, there could be three bids, but neither trades nor asks for a second. If there is only one bid, or ask, or trade, during the bar period, the event fields will have the same values. For example:

```
OpenBidPrice = HighBidPrice = LowBidPrice = CloseBidPrice
```

```
OpenBidSize = HighBidSize = LowBidSize = CloseBidSize
```

If there are no events, then there will not be any bars for these seconds.

Quote Price Filter

When trading and quoting activities are inactive, for example, during extended trading hours or with an illiquid stock, bid prices can be extremely low, and ask prices can be extremely high. An exchange can also send a bad price, for example, a stock has a bid of \$12.05 then an exchange sends a bid of \$212.05.

To make TAQ Second Bar datasets usable for illiquid stocks and ETFs/ETNs, algoseek filters out extreme quotes by the following two criteria:

```
Bid price < (0.05 x average price of last 10 days)
```



Ask price > (10 x average price of last 10 days)

Separating Exchange and Finra Volume

The bars have trade volume separated into fields:

Volume: Trades done on the listed exchanges

FinraVolume: Trades done in Dark Pools, internally by Broker-Dealers, or on an Over-the-Counter (OTC) market reporting to FINRA

The Volume data is separated to make it easy to understand the trading in a bar period for either the public-listed exchanges or private non-public trading.

DATA ORGANIZATION AND FILE FORMAT

algoseek provides Equity market data in plain text CSV files. The first row of CSV file is a fixed header and then rows of data corresponding to individual bars. By default, data is organized into one file per symbol per trading day. For example, all trade and quote bars for ticker AAPL on Mar 3, 2020, are stored in one CSV file.

Due to the large data size, CSV files are gzip-compressed (having a csv.gz extension) with a compression ratio of about 8:1.

Table 1 (below) provides the name, base event, default value, brief description, and data type for each data field (column) in the Equity TAQ Second Bar Extended CSV file.

Table column “Missing” indicates a default behavior in case the data field value is not present or cannot be calculated. The column value “Never” means that the data field value is always present.

Table column “Base Event” indicates what type of events are included for data field calculation. Quote: bid/ask event, Trade-X: trades on the exchange, Trade-F: trades on FINRA/TRF, Trade: trades on both exchange and FINRA/TRF.

Table 1: CSV File Fields Schema

Field	Base Event	Type (Format)	Missing	Description
Date	-	string (yyyymmdd)	Never	Trading date in yyyymmdd format
Ticker	-	string	Never	Symbol name
TimeBarStart	-	string (time)	Never	Start time of the bar. The format is HH:MM:SS



US Equity Trade and Quote Second Bar Extended Guide

OpenBarTime	Quote	string (timestamp)	Never	Open time of the bar, for example: 11:03:00.000000000
OpenBidPrice	Quote	decimal	Never	NBBO Bid Price as of bar Open, (e.g., current price as of bar start)
OpenBidSize	Quote	integer	Never	Total Size from all exchanges with OpenBidPrice
OpenAskPrice	Quote	decimal	Never	NBBO Ask Price as of bar open (e.g., current price as of bar start)
OpenAskSize	Quote	integer	Never	Total Size from all Exchanges with NBBO OpenAskPrice
FirstTradeTime	Trade	string (timestamp)	Blank	Time of the first trade
FirstTradePrice	Trade	decimal	Blank	Price of the first trade
FirstTradeSize	Trade	integer	Blank	Number of shares of the first trade
HighBidTime	Quote	string (timestamp)	Never	Time of highest NBBO bid price
HighBidPrice	Quote	decimal	Never	Highest NBBO bid price
HighBidSize	Quote	integer	Never	Total size from all exchanges with HighBidPrice
HighAskTime	Quote	string (timestamp)	Never	Time of highest NBBO ask price
HighAskPrice	Quote	decimal	Never	Highest NBBO ask price
HighAskSize	Quote	integer	Never	Total size from all exchanges with HighAskPrice
HighTradeTime	Trade	string (timestamp)	Blank	Time of the highest trade
HighTradePrice	Trade	decimal	Blank	Price of the highest trade
HighTradeSize	Trade	integer	Blank	Number of shares of the highest trade
LowBidTime	Quote	string (timestamp)	Never	Time of the lowest bid
LowBidPrice	Quote	decimal	Never	Lowest NBBO bid price of a bar
LowBidSize	Quote	integer	Never	Total Size from all exchanges with LowBidPrice



US Equity Trade and Quote Second Bar Extended Guide

LowAskTime	Quote	string (timestamp)	Never	Time of the lowest ask
LowAskPrice	Quote	decimal	Never	Lowest NBBO Ask price of a bar
LowAskSize	Quote	integer	Never	Total size from all exchanges with LowAskPrice
LowTradeTime	Trade	string (timestamp)	Blank	Time of the lowest trade
LowTradePrice	Trade	decimal	Blank	Price of the lowest trade
LowTradeSize	Trade	integer	Blank	Number of shares of the lowest trade
CloseBarTime	Quote	string (timestamp)	Never	Close time of the bar, for example: 11:03:59.999999999
CloseBidPrice	Quote	decimal	Never	NBBO Bid Price at bar Close
CloseBidSize	Quote	integer	Never	Total Size from all Exchange with CloseBidPrice
CloseAskPrice	Quote	decimal	Never	NBBO Ask Price at bar Close
CloseAskSize	Quote	integer	Never	Total Size from all Exchange with CloseAskPrice
LastTradeTime	Trade	string (timestamp)	Blank	Time of the last Trade
LastTradePrice	Trade	decimal	Blank	Price of last Trade
LastTradeSize	Trade	integer	Blank	Number of shares of last trade
MinSpread	Quote	decimal	Never	Minimum Bid-Ask spread size. This may be 0 if the market was crossed during the bar. If there is a negative spread due to a back quote, make it zero
MaxSpread	Quote	decimal	Never	Maximum NBBO Bid-Ask spread in a bar
CancelSize	Trade	integer	Blank	Total shares canceled
VolumeWeight Price	Trade-X	decimal	Blank	Trade Volume weighted average price excluding FINRA/TRF Trades. For FINRA-reported trades see the field "FinraVolumeWeightPrice". Note: Blank if no trades. Excludes FINRA-reported trades.
NBBOQuoteCount	Quote	integer	0	Number of Bid and Ask NNBO quotes during the bar period



TradeAtBid	Quote Trade	integer	0	Sum of trade volume that occurred at or below the bid (a trade reported/ printed late can be below the current bid)
TradeAtBidMid	Quote Trade	integer	0	Sum of trade volume that occurred between the bid and the midpoint: TradeAtBidMid = (Trade Price > NBBO Bid) & (Trade Price < NBBO Mid)
TradeAtMid	Quote Trade	integer	0	Sum of trade volume that occurred at mid. TradePrice = NBBO MidPoint
TradeAtMidAsk	Quote Trade	integer	0	Sum of ask volume that occurred between the mid and ask. TradeAtMidAsk = (Trade Price > NBBO Mid) & (Trade Price < NBBO Ask)
TradeAtAsk	Quote Trade	integer	0	Sum of trade volume that occurred at or above the Ask
TradeAtCrossOrLocked	Quote Trade	integer	0	Sum of trade volume for the bar when NBBO is locked or crossed. Locked is Bid = Ask Crossed is Bid > Ask
Volume	Trade-X	integer	0	Total number of shares traded Excluding FINRA/TRF reported trades, see field "FinraVolume" for FINRA trades. TotalVolume = Volume + FinraVolume
TotalTrades	Trade	integer	0	Total number of trades
FinraVolume	Trade-F	integer	0	Number of shares traded reported by FINRA/TRF. Trades reported by FINRA are from broker-dealer internalization, dark pools, over-the-counter, etc. FINRA trades represent volume that is hidden or not publicly available to trade
FinraVolume WeightPrice	Trade-F	decimal	Blank	FINRA Trade Volume weighted average price. Trades reported by FINRA are from broker-dealer internalization, dark pools, over-the-counter, etc. FINRA trades represent volume that is hidden or not publicly available to trade.
UptickVolume	Trade	integer	0	Total number of shares traded with upticks during the bar. Uptick = (Trade Price > Last Trade Price)



DowntickVolume	Trade	integer	0	Total number of shares traded with downticks during the bar. Downtick = (Trade Price < Last Trade Price)
RepeatUptick Volume	Trade	integer	0	Total number of shares where trade price is the same (repeated) and last price change was up during the bar. Repeat Uptick = (Trade Price == Last Trade Price) & (Last Tick Direction == Up)
RepeatDowntick Volume	Trade	integer	0	Total number of shares where trade price is the same (repeated) and last price change was down during the bar. Repeat Downtick = (Trade Price == Last Trade Price) & (Last Tick Direction == Down)
UnknownTick Volume	Trade	integer	0	When the first trade of the day takes place, the tick direction is “unknown” as there is no previous trade to compare it to. This field is the volume of the first trade after 4 AM and acts as an initiation value for the tick volume directions.
TradeToMidVol Weight	Quote Trade-X	decimal	Blank	Indicator for the bar period showing the sum difference between each trade’s price and NBBO midpoint at the time of the trade-weighted by volume. It returns a positive or negative number indicating buying or selling pressure. Note: Blank if no Trades. FINRA-reported trades are not included
TradeToMidVol WeightRelative	Quote Trade-X	decimal	Blank	Indicator for the bar period showing the sum difference between each trade’s price and NBBO midpoint at the time of the trade relative to the spread and weighted by volume. It returns a positive or negative number indicating buying or selling pressure. Note: Blank if no trades. FINRA-reported trades are not included.
TimeWeightBid	Quote	decimal	Blank	Time-weighted average price of National Best Bid during the bar period
TimeWeightAsk	Quote	decimal	Blank	Time-weighted average price of National Best Ask during the bar period



US Equity Trade and Quote Second Bar Extended Guide

OddLotTradeCount	Trade-X	integer	Blank	Number of OddLot trades during bar period.
OddLotTotalShares	Trade-X	integer	Blank	Total number of Odd Lot shares traded during bar period.
TotalVolume	Trade	integer	Blank	Total number of shares traded during the bar period from both Exchanges and off-exchange FINRA/TRF trades.
TotalQuoteCount	Quote	integer	Blank	Total count top-of-book Bid and Ask from public exchanges for bar period.
TotalVolumeWeightPrice	Trade	integer	Blank	Trade Volume weighted average price for shares traded during bar period from both on the Exchanges and off-exchange FINRA/TRF trades.
TimeWeightSpread	Quote	decimal	Blank	Spread during bar time weighted by time for each spread. TimeWeightSpread= sum(spread x spread_duration) / total_duration.
SpreadValidTime	Quote	integer	Blank	Total number of milliseconds during bar time that the spread was defined as valid for use in fields requiring a spread calculation. See below "Spread Validation Rules"
ExchangeTradeCount	Trade-X	integer	Blank	Total number of trades on public exchanges for bar period.
FinraTradeCount	Trade-F	integer	Blank	Total number of FINRA/TRF trades for bar period.
ExchangesBidCount	Quote	integer	Blank	Number of Bids from top-of-book of all Public exchanges. Shows the number of times the Bid changed for all public exchanges.
ExchangesAskCount	Quote	integer	Blank	Number of Asks from top-of-book of all Public exchanges. Shows the number of times Asks changed for all public exchanges.
VolumeWeightSpread	Quote Trade	decimal	Blank	Average bid/ask spread weighted by volumes of shares traded during the spread period. VolumeWeightSpread=sum((bid:ask spread) x (number of shares traded at this spread)) / total_volume



TimeWeightBid Size	Quote	decimal	Never	<p>The time-weighted average size of National Best Bid during the bar period.</p> <p>$\text{TimeWeightBidSize} = \frac{\sum(\text{bid_size} \times \text{bid_duration})}{\text{total_duration}}$</p> <p>See the Bar Notes for the calculation</p>
TimeWeightAsk Size	Quote	decimal	Never	<p>Time-weighted average size of National Best Ask during bar period.</p> <p>$\text{TimeWeightAskSize} = \frac{\sum(\text{ask_size} \times \text{ask_duration})}{\text{total_duration}}$</p> <p>See the Bar Notes for the calculation</p>
TradeAtBidCount	Trade	integer	Blank	Sum of a number of trades that occurred at or below the bid (a trade reported/printed late can be below the current bid)
TradeAtBidMid Count	Trade	integer	Blank	<p>Sum of the number of trades that occurred between the bid and the midpoint:</p> <p>$\text{TradeAtBidMidCount} = (\text{Trade Price} > \text{NBBO Bid}) \ \& \ (\text{Trade Price} < \text{NBBO Mid})$</p>
TradeAtMidCount	Trade	integer	Blank	<p>Sum of the number of trades that occurred at mid.</p> <p>$\text{TradePrice} = \text{NBBO MidPoint}$</p>
TradeAtMidAsk Count	Trade	integer	Blank	<p>Sum of the number of trades that occurred between the mid and ask.</p> <p>$\text{TradeAtMidAsk} = (\text{Trade Price} > \text{NBBO Mid}) \ \& \ (\text{Trade Price} < \text{NBBO Ask})$</p>
TradeAtAskCount	Trade	integer	Blank	Sum of a number of trades that occurred at or above the Ask
TradeAtCrossOrLockedCount	Trade	integer	Blank	<p>Sum of the number of trades for the bar when NBBO is locked or crossed.</p> <p>Locked is Bid = Ask Crossed is Bid > Ask</p>
PriorReference PriceTradeCount	Trade	Integer	Blank	Number of trades during bar period with condition flag "tPriorReferencePrice," a sale condition that identifies a trade based on a price at a prior point in time.



US Equity Trade and Quote Second Bar Extended Guide

PriorReference PriceTradeShar es	Trade-X	integer	Blank	Number of trades during bar period with condition flag "tPriorReferencePrice," a sale condition that identifies a trade based on a price at a prior point in time.
VolumeWeight PriceExcludeP RP	Trade	decimal	Blank	VWAP price of all trades, but excluding with condition flag "tPriorReferencePrice."
VolumeWeight SpreadExclude PRP	Quote Trade	decimal	Blank	VWAP of Bid/Ask spread weighted by trade volume during spread period, but excluding trades with condition flag "tPriorReferencePrice."
RelativeSpread Average	Quote Trade	decimal	Blank	The Relative Spread is the Bid/Ask spread relative to the midpoint price at time t for a trade. It shows how wide the spread is compared to the price. For each minute, the average of the Relative Spreads for each trade is calculated. See below "RelativeSpreadAverage"
TradeCumulDis tributionToBid	Quote Trade	string	Blank	Cumulative distribution volume of Trade price relative to the Bid during the bar period with 0 being trade at Bid and 1 being trade at Ask. Cumulative distribution created with percentage probabilities of 0:0.05:0.1:0.20:0.40:0.60:0.80:0.90:0.95:1. See below "TradeCumulDistributionToBid"
RetailTRFBuyS ize	Trade Quote	string	Blank	Estimated number of shares that are Buy retail order flow. Retail trades are identified using TRF trades executed sub-penny within a specific range. See "RetailBuySize" below.
RetailTRFSellS ize	Trade Quote	string	Blank	Estimated number of shares that are Sell retail order flow. Retail trades are identified using TRF trades executed sub-penny within a specific range. See "RetailSellSize" below.

Time Range

The TAQ Second Bar datasets cover the entire trading day from the start of pre-market trading to the end of after-hours trading (ET time):



Pre-Market Hours: 04:00:00 to 09:30:00 (excluding)

Market Hours: 09:30:00 to 16:00:00 (excluding)

Post-Market Hours: 16:00:00 to 20:00:00

Note: Occasionally second bars are extended several minutes past 20:00.

Market Holidays and Early Closes

The stock market is closed for trading on most US holidays. For reference, algoseek publishes a list of historical holidays which is available at

s3://us-equity-market-holidays/holidays.csv (direct download link:
<https://us-equity-market-holidays.s3.amazonaws.com/holidays.csv>).

Markets sometimes close early at 13:00:00 on the day before holidays such as Independence Day and Thanksgiving. You can download algoseek's early close date and time list from AWS S3 storage at

s3://us-equity-market-holidays/earlycloses.csv (or use a direct link
us-equity-market-holidays.s3.amazonaws.com/earlycloses.csv).

Timestamp

The event timestamp has a nanosecond resolution, and the time zone is ET. The timestamp field takes the format of HH:MM:SS.mmmuuunnn, for example, 09:31:01.723317846, where

HH: Hour

MM: Minute

SS: Seconds

mmm: Milliseconds

uuu: Microseconds

nnn: Nanoseconds

Before 2016 events were published with millisecond timestamps (HH:MM:SS.mmm format). For example, 09:32:00.321.

Timestamps in Excel. Excel fails when importing timestamp fields as Excel automatically tries to convert milliseconds and nanoseconds to Excel time format. When importing timestamp, you can import as Text fields instead.



Bar Notes

Time Bar Start Format: One-second bar 13:03:01 is from time greater than 13:03:01 to less than 13:03:02.

Empty Fields: an empty field has no value and is “Blank.” For example, FirstTradeTime and there are no trades during the bar period. The field “Volume” measuring the total number of shares traded in a bar will be “0” if there are no Trades. Look at the “Missing” column above for each field.

Time-Weighted Fields: In time-weighted fields, milliseconds are used as this is a consistent time across the whole dataset.

No Bid/Ask/Trade OHLC: There may not be a change in the NBBO or an actual trade during a bar timeframe. For example, there can be a bar with OHLC Bid/Ask but no Trade OHLC.

Single Event: For bars with only one trade, one NBBO bid or one NBBO ask then Open/High/Low/Close price, size, and time will be the same.

Spread Validation Rules: For fields that include spreads (for example, VolumeWeightSpread), the spread needs to be excluded when it is clearly incorrect.

Pre-Market: Exclude when a bid or ask is further away than 30% of the midpoint. For example, if stock is 100 then

$$\text{bid} \geq (0.7 * 100) = 70 \text{ and } \text{Ask} \leq (1.3 * 100) = 130$$

Regular Market: Exclude when a bid or ask is further away than 10% of the midpoint. For example, if stock is 100 then

$$\text{bid} \geq (0.9 * 100) = 90 \text{ and } \text{Ask} \leq (1.1 * 100) = 110$$

Use dynamic calculation to move from Pre-Market to Regular Market. When 3 or more Bid/Ask spreads are within 10% of the midpoint after the start of the Regular market (9:30:00), continue to use 10% of the midpoint. After 20 NBBO updates (consider one NBBO update to be two rows with an update for Bid and Ask), move to 10%.

Post Market: Exclude when a bid or ask is further away than 30% of the midpoint. Start immediately after market Close (4 pm or 1 pm for half-days)

Always exclude spread for

Cross/Locked: If NBBO is crossed/locked, then exclude.

Missing Bid/Ask: If an NBBO update is missing a Bid/Ask, then ignore the update.

The field “SpreadValidTime” has the total milliseconds for each bar period showing the total number of milliseconds that meet valid criteria based on the requirements listed here.



VolumeWeightPrice and **FinraVolumeWeightPrice**: volume-weighted price and FINRA Volume-weighted price are calculated as a dollar volume sum of all trades divided by the total number of shares traded

$$\text{sum}(\text{Trade_Shares} \times \text{Trade_Price}) / \text{sum}(\text{Trade_Shares})$$

For the “VolumeWeightPrice” column FINRA trades are excluded and only FINRA Trades are included for “FinraVolumeWeightPrice”.

TradeToMidVolWeight, **TradeToMidVolWeightRelative**: volume-weighted trade to the midpoint is calculated as the following sum over all trades during the bar

$$\frac{\text{sum}(\text{Trade_Shares} \times (\text{Trade_Price} - \text{NBBOMidpoint}))}{\text{sum}(\text{Trade_Shares})}$$

Similarly, volume-weighted relative trade to the midpoint

$$\frac{\text{sum}(\text{Trade_Shares} \times (\text{Trade_Price} - \text{NBBOMidpoint}) / \max(1, \text{NBBOSpread}))}{\text{sum}(\text{Trade_Shares})}$$

where midpoint and spread values are calculated based on the last NBBO

$$\text{NBBOMidpoint} = (\text{NBBOBid_InPennies} + \text{NBBOAsk_InPennies}) / 2$$

$$\text{NBBOSpread} = \text{NBBOAsk_InPennies} - \text{NBBOBid_InPennies}$$

If Bid == Ask, then it is assumed the midpoint of the Bid/Ask is that price. If the market is crossed (NBBO Bid > NBBO Ask), then it is not possible to know what the correct price is so the last good NBBO Bid and Ask (including the Bid == Ask case) will be used.

TimeWeightBid, **TimeWeightAsk**: time-weighted bid and ask are calculated with

$$\text{sum}(\text{Price}_{\{n\}} \times (\text{Price}_{\{n+1\}_Time} - \text{Price}_{\{n\}_Time})) / \text{Bar_Duration}$$

where Price_0 is the bar open price.

TimeWeightSpread: For spread-weighted calculation, only include spread when it makes sense. Exclude when the spread is Crossed, Locked, or Unrealistic, e.g., 0.01 to 99999.

RelativeSpreadAverage: The relative spread size is the Bid/Ask spread relative to the midpoint price at time t for a trade. It shows how wide the spread is compared to the midpoint price. For each second, the average of the relative spread sizes for each trade is calculated.

For each trade in the bar:

$$\begin{aligned} t &= \text{time of a trade} \\ b &= \text{National Best Bid (NBBO) at } t \text{ (use last NBBO before time } t) \\ a &= \text{National Best Ask (NBBO) at } t \text{ (use last NBBO before time } t) \\ m &= \text{midpoint price at } t, m = (b + a) / 2 \end{aligned}$$



US Equity Trade and Quote Second Bar Extended Guide

`rs = relative spread at t is max(a - b, 0) / m` (max function is used as NBBO spread may be 0 or inverted at times)

`RelativeSpreadAverage = sum(rs) / count(rs)`

TradeCumulDistributionToBid: A distribution of trades relative to the bid and offer. Calculate the distance of each trade to the Bid with 0 executed at the Bid and 1 executed at the Ask. Then calculate the cumulative distribution of the distance from the Bid using the below probabilities.

0, 0.05, 0.1, 0.20, 0.40, 0.60, 0.80, 0.90, 0.95, 1

This shows if there was pressure on either side of the bid offer and if 'retail trade' moves away from the paper – there will be a potential way to recalibrate it without recalculating everything in history again. As an example: say 1000 shares are traded in the given bin. 100 at a bid, 400 at mid, and 500 at the offer. The distribution above will look as follows

0	100
0.05	100
0.1	100
0.2	100
0.4	100
0.6	500
0.8	500
0.9	500
0.95	500
1	1000

RetailTRFBuySize, RetailTRFSellSize: Identify Retail Buy and Sell trades executed internally by a Broker-Dealer or wholesale to dark-pool based on their sub-penny pricing as a TRF reported trade. These indicators are based on the research paper "Tracking Retail Investor Activity" by E. Boehmer, Ch. M. Jones, X. Zhang, and X. Zhang published in the Journal of Finance (https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2822105).

The field calculations are based on the following description from the paper:

*Based on these institutional arrangements, identifying transactions initiated by retail customers is fairly straightforward. Transactions with a retail seller tend to be reported on a TRF at prices that are just above a round penny due to the small price improvement, while transactions with a retail buyer tend to be reported on a TRF at prices just below a round penny. To be precise, for all trades reported to a FINRA TRF (exchange code "D" in TAQ), let P_{it} be the transaction price in stock i at time t , and let $Z_{it} \equiv 100 * \text{mod}(P_{it}, 0.01)$ be the fraction of a penny associated with that transaction price. Z_{it} can take any value in the unit interval $[0,1)$. If Z_{it} is in the interval $(0,0.4)$, we identify it as a retail sell transaction. If Z_{it} is in the interval $(0.6,1)$, then the transaction is coded as a retail buy transaction. To be conservative, transactions at a round penny ($Z_{it} = 0$) or near the half-penny ($0.4 \leq Z_{it} \leq 0.6$) are not assigned to the retail category.*



APPENDIX A. FREQUENTLY ASKED QUESTIONS

Why are Trade Prices often inside the Bid Price to Ask Price range?

The Low/High Bid/Ask is the low and high NBBO price for the bar range. Very often, a trade may not occur at these prices, as the price may only last a few seconds or executions are being crossed at mid-point due to hidden order types that execute at mid-point or as price improvement over current Bid/Ask.

Why are time-based columns not properly recognized when I try importing data to Excel?

Older versions of Excel will automatically convert the TimeBarStart field into an Excel format timestamp, but this fails when TimeBarStart is HHMMSSmmm (millisecond) or HHMMSSmmmmiiinnn (nanosecond). For timestamp with the nanosecond (millisecond) format, import the data using the Excel “From Text” option and set the data type for column “TimeBarStart” to “Text”, so Excel does not automatically try to convert it.



APPENDIX B. BAR CALCULATIONS FROM TRADE AND QUOTE EVENTS

This section describes logic for second bar calculations based on events from the Trade and Quote dataset. Please also refer to the Equity Trade and Quote Guide for more details on data fields and condition flags used.

There is a separate logic for the Standard Bars dataset and Bars with FINRA/TRF and Odd Lots Excluded.

Standard Trade and Quote Second Bar

Excluded data

- Exclude any trade with Price = 0 (not applied to Trade Cancel)
- Exclude any trade with Quantity = 0
- Exclude any quote with Price = 0
- Exclude any quote with Quantity = 0

You should also exclude any event with one or more flags listed in Table 2.

Table 2: Flags for Trade and Quote Events to be Excluded During Bar Calculations

Trade Events		Quote Events	
Bit Mask Position	Flags	Bit Mask Position	Flags
14	tOutOfSequence	3	qClosing
20	tAveragePrice	4	qNewsDissemination
22	tPriceVariation	5	qNewsPending
23	tRule155	6	qTradingRangeIndication
24	tOfficialClose	7	qOrderImbalance
25	tPriorReferencePrice	13	qResume
26	tOfficialOpen		



Included data

- Include only NBBO quotes
- Include both NBBO and exchange trades
- Include Trade Cancel events only for the CancelSize field (no price or flag filter applied)

You should only include events with one or more flags listed in Table 3. If the event has any of the exclude flags enabled, it is not included. If the event does not contain any flags from the include list, it is not included in bar calculations.

Table 3: Flags for Trade and Quote Events to be Included During Bar Calculations

Trade Events		Quote Events	
Bit Mask Position	Flags	Bit Mask Position	Flags
0	tRegular	0	qRegular
1	tCash	1	qSlow
2	tNextDay	2	qGap
5	tIntermarketSweep	11	qOpeningQuote
6	tOpeningPrints	21	qFastTrading
7	tClosingPrints		
10	tFormT		
13	tExtendedHours		
21	tCross		
29	tTradeThroughExempt		
31	tOddLot		

Price validation

Additionally, you should filter out test quote events using the following approach:

- 1) For symbols with price history (last 10 trading days):

$$\text{MinPrice} = 0.05 * \text{AveragePrice}$$

$$\text{MaxPrice} = 10 * \text{AveragePrice}$$



2) For new symbols (no price history):

MinPrice = 0.03

MaxPrice = 19998

If a Quote price is lower than MinPrice or higher than MaxPrice - the event is excluded.

Note: We do not recommend applying price filtering for Trade events.

No-FINRA/TRF and Odd Lots Trade and Quote Second Bar

Excluded data

- Exclude any trade with Price = 0 (not applied to Trade Cancel)
- Exclude any trade with Quantity = 0
- Exclude any quote with Price = 0
- Exclude any quote with Quantity = 0
- Exclude any trade or quote with Exchange = FINRA

You should also exclude any event with one or more flags listed in Table 4.

**Table 4: Flags for Trade and Quote Events to be Excluded During Bar Calculations
(No-FINRA/TRF Dataset)**

Trade Events		Quote Events	
Bit Mask Position	Flags	Bit Mask Position	Flags
14	tOutOfSequence	3	qClosing
20	tAveragePrice	4	qNewsDissemination
22	tPriceVariation	5	qNewsPending
23	tRule155	6	qTradingRangeIndication
24	tOfficialClose	7	qOrderImbalance
25	tPriorReferencePrice	13	qResume
26	tOfficialOpen		
31	tOddLot		

Included data

- Include only NBBO quotes
- Include both NBBO and exchange trades
- Include Trade Cancel events only for the CancelSize field (no price or flag filter applied)



You should only include events with one or more flags listed in Table 5. If the event has any of the exclude flags enabled, it is not included. If the event does not contain any flags from the include list, it is not included in bar calculations.

Table 5: Flags for Trade and Quote Events to be Included During Bar Calculations (No-FINRA/TRF Dataset)

Trade Events		Quote Events	
Bit Mask Position	Flags	Bit Mask Position	Flags
0	tRegular	0	qRegular
1	tCash	1	qSlow
2	tNextDay	2	qGap
5	tIntermarketSweep	11	qOpeningQuote
6	tOpeningPrints	21	qFastTrading
7	tClosingPrints		
10	tFormT		
13	tExtendedHours		
21	tCross		
29	tTradeThroughExempt		

Price validation

Additionally, you should filter out test quote events using the following approach:

- 1) For symbols with price history (last 10 trading days):

$$\text{MinPrice} = 0.05 * \text{AveragePrice}$$

$$\text{MaxPrice} = 10 * \text{AveragePrice}$$

- 2) For new symbols (no price history):

$$\text{MinPrice} = 0.03$$

$$\text{MaxPrice} = 19998$$

If a Quote price is lower than MinPrice or higher than MaxPrice - the event is excluded.

Note: We do not recommend applying price filtering for Trade events.