

Technical Appendix for MuniTrack

24 September 2026

1. Secondary Market Metrics

1.1 Purpose and scope

The pipeline turns MSRB trade reports [1] into weekly (Section 1) and quarterly (Section 2) statistics on U.S. municipal bond secondary trading: dealer markups, par traded, electronic trading, dealer concentration and the share of top dealers. Markups are measured on round-trip *chains* built with an algorithm adapted from Li and Schürhoff [2].

1.2 Data sources

- **MSRB Surveillance Database** [1], the only transaction-level source. The same database supplies the dealer registry, used to roll dealer codes up to parent firms.
- **Mergent Municipal Bond Securities Database** [3]: security reference data (maturity, dated date, coupon type, denomination, put and floating-rate flags) for trades before May 2011.
- **After May 2011**: the same fields come from the MSRB’s security reference data [1]. It has no offering date, so offering date, bond type and amount are taken from the Bond Buyer [4], matched by CUSIP through the primary-market pipeline.
- **Method**: chain construction and sample screens are adapted from Li and Schürhoff [2, Appendix A]; the dealer diversity measure follows Ahern et al. [5].

1.3 Pipeline

Stages run in the order below; each writes Parquet (zstd) and is rerun only with `--overwrite`. The matching and weekly stages are Python packages (`python -m chains.run`, `python -m weekly.run`). All paths are relative to `data/`.

Stage	Script → output	Operations
Stage	0_0 → CSV beside each raw file in 0_raw/msrb/	Unzips and converts the raw trade, capacity, dealer-registry and reference files, applies the column layouts of headerless files, repairs malformed delimiters and writes a manifest.
Assemble	1_MSMB_transactions → 1_processed/msrb/ <stage>.parquet	Appends files from different periods.
Clean	1_clean_msrb_sample (_mc: sharded, same output) → 3_clean/ secondary/trades_whole_ sample.parquet	Cleans the data (see Quality controls), sets trade type CD (customer sells to dealer), DD (inter-dealer) or DC (dealer sells to customer), appends the eras (about 2.3×10^8 trades) and assigns <code>trade_id</code> = CUSIP-sequence.
Match	chains → work/chains/ legs/part-*.parquet, 4_matched/secondary/ matched_chains_all_ years.parquet	Builds round-trip chains in one pass over the whole sample (see Chains and metric definitions), so chains cross year ends; parallel over CUSIPs.

Stage	Script → output	Operations
Weekly	<code>weekly</code> (markups, volume, electronic, spreads, concentration, dealers, merge) → <code>5_final/secondary/</code>	Computes the weekly and quarterly series (see Chains and metric definitions) and merges them into <code>weekly_series</code> and <code>quarterly_series</code> .
Publish	<code>weekly</code> (figures, export, numbered, web) → <code>charts/weekly/</code> , <code>6_web_csvs/YYYY/MM/DD/</code>	Writes per-year CSVs from 2005, PNGs (matplotlib) with one CSV per figure, and the dated web tree mirrored to <code>latest/</code> .

1.4 Chains and metric definitions

Chains. Within each CUSIP, trades are ordered by time, then $CD < DD < DC$, price and par. Each CD trade starts a chain. The next leg is an unused DD or DC trade of the same par whose seller is the previous leg’s buyer, within -10 to $+30$ days and ± 10 trades of the previous leg. A DC leg closes the chain ahead of any DD; otherwise the earliest DD adds a dealer hop (at most 7). If no same-par leg exists, the dealer’s in-window DC trades totalling at most the chain’s par close it as a split. Each trade is used once. The chain markup is $100(P_{\text{end}} - P_{\text{CD}})/P_{\text{CD}}$, where P_{end} is the DC price or the par-weighted split price.

Weeks run Sunday–Saturday and are labelled by their Friday; quarters are calendar quarters. Size buckets: retail par $< \$100\text{k}$, midsize $\$100\text{k}–\500k , institutional $> \$500\text{k}$.

- **Markups** (Figs. 1.1, 1.3, 1.5, 1.7, 1.8): completed chains that open and close in the same week, winsorized at 0.1/99.9% over the whole sample and averaged (unweighted) by the week of the CD leg; by size bucket, electronic status, top-dealer status, compensation type (transaction vs. non-transaction based, from Aug 2018) and chain time (pre-arranged if the chain completes within 15 minutes, otherwise inventory).
- **Par traded** (Fig. 1.2): weekly sum of trade par by size bucket, in \$ billions.
- **Electronic trading** (Figs. 1.3–1.4, from 26 Apr 2018): a customer trade is electronic if its dealer portion is classified as an alternative trading system (ATS) or broker’s broker (BB) based on its EBS code; an inter-dealer trade is electronic if it is flagged ATS or BB. Shares are $100 \times \text{electronic/all trades}$.
- **Top dealers** (Figs. 1.5–1.6): dealer codes are mapped to parent firms with the dealer registry; the 25 parents with the most trades over the whole sample (venues excluded) form the top set. A chain is top-dealer if any leg involves one. Fig. 1.6 is the percent of trades with a top dealer, by trade type.
- **Concentration** (Figs. 2.1–2.2): customer trades by quarter and dealer code, ATS venues excluded. With s_j each dealer’s par share: $\text{HHI} = \sum s_j^2$, balanced diversity $\exp(-\sum s_j \ln s_j)$ [5], and active dealers = dealers with at least 100 trades.

1.5 Quality controls and caveats

- **Report cleaning:** cancelled trades are removed and the latest version of each trade kept; inter-dealer trades are kept once (the sell side). Drop trades with par below \$5,000 or not a multiple of \$5,000, a dealer code that is not a valid four-character code, or no price; drop self-trades.
- **Bond screens** [2, Appendix A]: drop floating-rate, variable-rate, puttable, warrant and derivative issues and non-\$5,000 denominations; require original maturity ≥ 1 year, seasoning ≥ 30 days and remaining maturity ≥ 180 days; truncate prices at 0.5/99.5% separately for zero-coupon and other bonds.
- **Deviations from Li and Schürhoff [2]:** seasoning is 30 days (not 90); the \$5,000-multiple screen is added; chains with agency trades are kept; ties are broken by input order for reproducibility.
- **Caveats:** ATS and non-transaction-based fields exist only from 26 Apr 2018, so those series start then. Concentration uses dealer EBS codes, not parent firms.

2. Primary Market Metrics

2.1 Purpose and scope

The pipeline turns the daily *Bond Buyer* sales-results pages [4] into monthly (Section 3) and quarterly (Section 4) statistics on U.S. municipal bond issuance. The statistics cover volume, offering yields, competitive true interest cost (TIC), bidding intensity and auction-winner concentration, and are published as figure CSVs, PNG charts and a PDF chartpack. The raw archive (`bondbuyer.tar.zst`) holds 8,877 daily pages (4,367 competitive, 4,510 negotiated) from 25 Feb 2008 to 31 Aug 2026. Monthly series start in March 2008 and quarterly series in 2008Q2. The unit of observation throughout is the *maturity row*: one serial or term maturity of one bond issue.

2.2 Data sources

- **The Bond Buyer, Sales Results database** [4], the only transaction-level source. It is downloaded from `data.bondbuyer.com/salesResults`. It supplies issuer, state, dates, bond type, the maturity schedule (par, coupon, yield) and deal terms (winning bid and other bids for competitive sales; reoffered yield for negotiated sales), purchaser, financial advisor and legal opinion.
- **Official-statement (OS) reference data**. These files are used only to fill blank offering yields for Fig. 3.2. These data are extracted from the offering statement in the MSRB PM data product from the Municipal Securities Rulemaking Board [6].
- **MSRB Rule G-33** [7] supplies the calculation conventions, not data: 30/360 day count, semiannual compounding, and the price–yield and accrued-interest formulas.
- **Reference constants**: recession shading bands (Dec 2007–Jun 2009, Feb–Apr 2020), which match NBER dates [8]; ten date-effective dealer merger rules, each citing a Wikipedia article [9]; and about 170 hand-coded dealer name aliases. There are no other external files.

2.3 Pipeline

`primary_runner.py` runs steps 1_1a–1_4c in order. Steps 1_2a–1_3 run once per quarter, the runner stops at the first failure, and it renames existing outputs to `*_updatedYYYYMMDD_HHMMSS` before overwriting them. `graphing_runner.py` then runs `2_1` → `3_0` → `3_1` → `3_2`. A monthly update rebuilds Section 3 and carries Section 4 forward; a quarterly update rebuilds both. All paths below are relative to `data/`.

Stage	Script → output	Operations
Acquire	<code>1_0a, 1_0b (bondbuyer_html_extract) → 0_raw/bondbuyer/htmls/YYYY/Q#{competitive,negotiated}/YYYY.MM.DD.html</code>	One page per sale date. Pages already on disk are skipped unless <code>--overwrite</code> is given.

Stage	Script → output	Operations
Scrape	1_1a, 1_1b → 1_processed/.../1_{type}_uncleaned_YYYYQ#.csv	Parses the state-grouped listings into maturity rows: 24 fields for competitive sales, 20 for negotiated. Tags taxable and bank-qualified sections, and splits text where a par amount runs into a year. IDs are built from content (row_uid, a SHA-1 hash). Files are assigned to quarters by their HTML folder.
Clean	1_2a, 1_2b → 2_intermediate/.../2_{type}_YYYYQ#.csv (+ bidder, purchaser and manager tables)	Maps states to postal codes (rows with unmapped states are dropped) and parses dates. Sets deal par to the largest \$ figure in the deal header. Flags GO, revenue, taxable, AMT, bank-qualified and book-entry bonds. Parses the winning bid into price, rate, rate type (TIC, NIC or effective) and winner. Fills deal-level fields within each deal. Drops rows with no dated date, bond type or issuer.
Combine	1_3 → 3_clean/.../combined_bondbuyer_YYYYQ#.csv; full range 4_matched/primary/TAG/	Joins competitive rows to the purchaser table, then appends negotiated rows. Extracts the bond series and removes logical duplicates. Derives price as 100(payment + premium)/payment. Applies 34 hardcoded fixes for pars inflated 1,000× and merges duplicate auction IDs.
Monthly	1_4a → 5_final/.../monthly_bondbuyer_stats_TAG.csv	Matches OS data, fills yields, screens amounts, standardizes competitive TIC and computes the metrics (see Metric definitions). Months with no data are added as empty rows back to 2008-03, and the run fails if any month is still missing. Writes about 20 audit CSVs.
Quarterly	1_4b → quarterly_bondbuyer_stats_for_graphing, industry_*	Computes the same metrics by offering quarter, keeping complete quarters only, and adds issue-size bins.
Concent.	1_4c → primary_quarterly_concentration_measures	Standardizes purchaser names and computes the concentration indices (see Metric definitions).
Figures	2_1 → 6_web_csvs/YYYY/MM/DD/{figure,appendix}_csvs/	Writes one CSV per figure plus section{3,4}_consolidated.csv, and mirrors them to latest/.
Render	3_0 (LaTeX style), 3_1 (RPPE style), 3_2	Draws the PNGs with matplotlib into charts/YYYY/MM/DD/. 3_2 assembles the PDF with pdflatex, falling back to PdfPages; each figure gets a description, source and methodology line.

2.4 Metric definitions

A month is assigned by offering date (falling back to sale, dated, then recorded date); a quarter by offering_quarter. Weighted (_w) means are par-weighted over rows with positive par. C and N denote competitive and negotiated sales.

- **Issuance** (Figs. 3.1, 4.1–4.2): total par, count, and mean, median and 90th-percentile size; each by sale type and share of par. By sector (issuer-name keywords: utility, healthcare, transportation, education, housing, other) and size bin (<\$1M, \$1–5M, \$5–25M, ≥\$25M).
- **Characteristics**: mean coupon; maturity = (due – offering date)/365.25; shares that are GO, revenue, taxable, bank-qualified, book-entry, callable and competitive.
- **Offering yield** (Fig. 3.2, from June 2009): the Bond Buyer yield (reoffering yield for C, reoffered

yield for N) is used where it is reported. If it is blank or “NRO”, it is filled from the OS yield-to-maturity or yield-to-call. Failing that, the yield is solved from the OS price under G-33 (settlement on the dated date, bisection, kept only if in $(0, 20]$). Each issue takes whichever basis (yield to maturity or to call) is closer to Bond Buyer. A Bond Buyer yield more than 5 bp from both is replaced. Values above 20% are dropped, and the rest are capped at the 99.5th percentile within each sale type. Bond Buyer prices are never used.

- **Competitive TIC** (Fig. 3.3): taken in order of preference from the reported TIC, the reported effective rate, or a TIC converted from a reported NIC. The converted TIC solves one yield y for the issue: $\sum_i \text{par}_i [\text{PV}_i(y) - \text{AI}_i] = P \cdot \sum_i \text{par}_i$. Here P is the purchase price per unit of par, PV_i is the G-33 dirty price and AI_i the accrued interest. The conversion requires a complete maturity schedule whose total par is within 0.5% of deal par, and the result must be in $(0, 20]$. Negotiated TIC is not computed: it needs underwriter-fee inputs.
- **Bidding** (Fig. 3.4, competitive only): depth is the number of bids parsed per auction; dispersion is the coefficient of variation $100 \cdot \text{sd}/\text{mean}$ of the bid rates. Both are computed overall and by rate type.
- **Concentration** (Figs. 4.4–4.5, App. 3): Fig. 4.4, the diversity of auction winners, follows the framework of Ahern et al. [5]. Purchaser names are standardized using the alias table, splitting of syndicate lists, fuzzy clustering (`rapidfuzz token_sort_ratio` ≥ 95) and merger roll-ups [9]. Par is split equally among co-purchasers. The indices include only purchasers with at least 25 wins in the quarter. With s_j each included purchaser’s par share: $\text{HHI} = \sum s_j^2$ on a 0–1 scale, inverse $\text{HHI} = 1/\text{HHI}$, Shannon entropy $H = -\sum s_j \ln s_j$, and effective number e^H .

2.5 Quality controls and caveats

- **Duplicates**: rows are deduplicated on `row_uid`, falling back to a 21-column content key. A warning is logged when a month has more than $1.25\times$ the rows of the same month a year earlier.
- **Amounts** are never winsorized. Maturities exactly $1,000\times$ too large are rescaled. A row is dropped if one maturity exceeds $1.5\times$ deal par, or if a one-series deal’s maturities sum to more than $1.5\times$ deal par. Zero-coupon (capital appreciation) deals with implied accretion $\leq 8\%$ are exempt. Winsorization of coupon and maturity is implemented but switched off.
- **Caveats**: The HHI covers only purchasers with 25 or more wins, so it is not a full-market index on the DOJ 0–10,000 scale. The recession bands are hardcoded in two places (`1_4b` and `graphing_helpers`).

References

- [1] Municipal Securities Rulemaking Board. Surveillance Database, 2026. Accessed: September 18, 2026.
- [2] Dan Li and Norman Schürhoff. Dealer Networks. *Journal of Finance*, 74(1):91–144, 2019. doi: 10.1111/jofi.12728.
- [3] Mergent. Mergent Municipal Bond Securities Database, 2026. Accessed: March 1, 2026.
- [4] The Bond Buyer. Bond Buyer Data Archives: Sales Results. <https://data.bondbuyer.com/salesResults>, 2026. Subscription database (Arizent). Competitive and negotiated sales results, 25 February 2008–31 August 2026; local archive `bondbuyer.tar.zst` (8,877 daily HTML pages). Accessed: September 18, 2026.
- [5] Kenneth R. Ahern, Lei Kong, and Xinyan Yan. The Anatomy of Concentration: New Evidence from a Unified Framework. Working Paper 4510938, SSRN, 2024. URL https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4510938. Accessed: September 24, 2026.
- [6] Municipal Securities Rulemaking Board. MSRB Primary Market Historical Data Product. <https://www.msrb.org/Market-Data-and-Research/Academics>, 2026. Distributed by the Electronic Municipal Market Access (EMMA) system. Accessed: September 18, 2026.
- [7] Municipal Securities Rulemaking Board. Rule G-33: Calculations. <https://www.msrb.org/Rules-and-Interpretations/MSRB-Rules/General/Rule-G-33>, n.d. Accessed: September 24, 2026.
- [8] National Bureau of Economic Research. US Business Cycle Expansions and Contractions. <https://www.nber.org/research/data/us-business-cycle-expansions-and-contractions>, n.d. Business Cycle Dating Committee. The pipeline hardcodes the recession bands without citing a source; the dates match these NBER peaks and troughs. Accessed: September 24, 2026.
- [9] Wikipedia. Articles cited as sources for the dealer merger rules in `primary_purchaser_names.py`, n.d. Rule, effective quarter: Bear Stearns→J.P. Morgan, 2008Q2, https://en.wikipedia.org/wiki/JPMorgan_Chase#Bear_Stearns; Wachovia→Wells Fargo, 2008Q4, https://en.wikipedia.org/wiki/Wachovia#Acquisition_by_Wells_Fargo; Merrill Lynch→Bank of America, 2009Q1, https://en.wikipedia.org/wiki/Bank_of_America#Acquisition_of_Merrill_Lynch; Morgan Keegan→Raymond James, 2012Q2, https://en.wikipedia.org/wiki/Raymond_James_Morgan_Keegan; FirstSouthwest/Southwest Securities→Hilltop Securities, 2016Q1, https://en.wikipedia.org/wiki/Hilltop_Holdings_Inc.#Subsidiary:_HilltopSecurities; Siebert Brandford Shank→Siebert Cisneros Shank, 2016Q2, and →Siebert Williams Shank, 2019Q3, https://en.wikipedia.org/wiki/Siebert_Williams_Shank; BB&T/SunTrust→Truist, 2019Q4, https://en.wikipedia.org/wiki/Truist#SunTrust_merger; FTN Financial→FHN Financial, 2019Q4, https://en.wikipedia.org/wiki/First_Horizon_Corporation; Piper Jaffray/Sandler O’Neill→Piper Sandler, 2020Q1, https://en.wikipedia.org/wiki/Piper_Sandler_Companies#History. Accessed: September 24, 2026.