

ANALYTICAL REPORT

Wealth Management Book-of-Business Audit

Self-directed portfolio analysis on a fully synthetic holdings dataset

Data Cleaning

Rule-Based Parsing

Descriptive Audit

Prepared for Arbutus Visual Analytics

Prepared by Marc Charbonneau, Principal Analyst, Arbutus Visual Analytics

Date July 17, 2026

Confidentiality Self-directed portfolio piece: fully synthetic dataset, no real client, advisor, or firm

EXECUTIVE SUMMARY*Written for a non-technical audience.*

This report audits a wealth-management advisory practice's book of business across three structural questions a custodian's raw export cannot answer on its own: how many *distinct true funds* are actually held once fund "series" are consolidated, how much is held in direct preferred shares once fund-wrapped preferred-yield products are correctly excluded, and how much of the book sits in thinly-held, potentially concentrated positions. All figures in this report are generated from a fully synthetic dataset built to match the structure of a real advisory book of business; no real client, advisor, or firm data appears anywhere in this analysis. See Conclusions, right after the infographic overview, for the full findings.

POSITIONS AUDITED**331**

Synthetic holdings-export rows across 12 asset types, generated to match the structure of a real book of business

[↴ jump to finding](#)**RAW FUND SERIES****36**

Distinct fund/ETF series names in the raw export, before consolidation

[↴ jump to finding](#)**TRUE FUND COUNT (Q1)****24**

Distinct underlying funds after rule-based series consolidation — a 33% reduction from the raw series count, none of it visible in a standard custodian report

[↴ jump to finding](#)**DIRECT PREFERRED-SHARE AUA (Q2)****\$2.61M**

Across 17 positions and 10 issuers, after excluding 3 fund-wrapped "preferred yield" products (\$667K) the custodian's own tagging conflates with direct holdings

[↴ jump to finding](#)**AUA HELD BY 1-2 HOUSEHOLDS (Q3)****\$30.1M**

Across 143 non-GIC positions — 78% of all non-GIC positions are held by fewer than 4 households, a practice-concentration signal invisible in any single-position report

[↴ jump to finding](#)

Wealth Management Book-of-Business Audit

A self-directed portfolio analysis for Arbutus Visual Analytics

1 PROBLEM

The Challenge

- 🎯 How many distinct funds does a book actually hold, once the same fund's no-load and front-end-load "series" are counted as one?
- 🔍 How much is genuinely in direct preferred shares, once mutual funds that merely invest in preferred shares are excluded?
- ⚠️ How much of the book sits in positions held by only 1 or 2 households — a concentration signal no single-position report surfaces?



2 WORKFLOW

How We Built It

- 📄 Audited 331 holdings-export rows across 12 asset types, logging every cleaning and scoping decision before writing analysis code
- 🔗 Built a deterministic, auditable rule-based parser to strip load-code and series-class suffixes and recover each position's true underlying fund
- 📊 Extracted issuer tickers from preferred-share symbols and separated direct holdings from fund-wrapped preferred-yield products
- 🔗 Filtered and cross-tabbed thinly-held positions (fewer than 4 households, ex-GICs) against risk rating to surface a suitability/concentration lens



3 SOLUTION

What We Delivered

- ↪ A true fund count (24) that is 33% lower than the raw series count (36) — a discrepancy invisible in any standard custodian report
- 💡 A clean direct-preferred-share total (\$2.61M) separated from \$667K of fund-wrapped preferred-yield products the custodian's own asset-type tag conflates with it
- ⚠️ A household-concentration breakdown showing \$30.1M held by only 1-2 households, cross-tabbed by risk rating to flag where thin holding and elevated risk overlap
- A fully auditable input-to-output mapping (every raw security name traced to its consolidated true fund) for line-by-line review

Fully synthetic dataset, generated to match the structure of a real advisory book of business. No real client, advisor, or firm data appears in this analysis.

Conclusions

Written for a non-technical audience.

WHAT WE SET OUT TO ANSWER

Can a rule-based audit surface structural facts about a wealth-management book of business (true fund diversification, direct preferred-share exposure, household-concentration risk) that a standard custodian export leaves hidden?

WHAT WE FOUND

- Fund positions split across load/series codes overstate a book's true fund count by a third (36 raw series vs. 24 true funds) unless explicitly consolidated — a discrepancy invisible in a standard custodian report.
- A custodian's own asset-type tagging conflates direct preferred shares with mutual funds that merely invest in preferred shares, overstating genuine direct exposure by 25% (\$3.27M combined vs. \$2.61M direct) unless the two are separated.
- 78% of non-GIC positions are held by fewer than 4 households, and cross-tabbing this against risk rating shows the sharpest practice-management concern concentrated in thinly-held equity positions rated MEDIUM risk or higher, not spread evenly across the book.
- All three findings come from deterministic, fully auditable rule-based logic (string parsing and filter-and-aggregate operations), not a black-box model — every input-to-output mapping can be checked line by line.

WHAT WE RECOMMEND

1. Build fund-series consolidation into any recurring book-of-business report, rather than relying on the custodian's raw per-series line count.
2. Flag and separately report fund-wrapped preferred-yield products whenever a practice reports its preferred-share exposure.
3. Run the household-concentration × risk-rating cross-tab as a recurring practice-management check, prioritizing thinly-held, elevated-risk positions for suitability review first.

Abbreviations

AUA	Assets Under Administration, in Canadian dollars
CUSIP	Committee on Uniform Securities Identification Procedures number, a unique security identifier
ETF	Exchange-Traded Fund
GIC	Guaranteed Investment Certificate
HH	Household

1. Data Source, Cleaning, and Variable Roles

This section covers the synthetic dataset's structure and the reason a fabricated dataset was used instead of the real client data this methodology was originally built for, followed by the cleaning steps applied before any analysis.

1.1. A Synthetic Dataset, Not Anonymized Real Data

This piece generalizes an analytical methodology originally built for a real wealth-management advisory practice's book-of-business audit. The dataset used throughout this report is **fully synthetic**: every CUSIP, security name, fund family, issuer ticker, AUA figure, and household count is fabricated, generated to match the structural shape of a real holdings export (fund positions split across multiple "series," preferred-share tickers in a recognizable pattern, a long-tailed household-concentration distribution) without deriving any value from a real client's data. This is a deliberate choice, not a simplification: replacing labels on real figures while keeping the real AUA totals, real household counts, or real security-level detail would not remove the

re-identification risk for a real advisor's real client holdings — a specific dollar total or a specific household-concentration figure is itself a fingerprint. A fully fabricated dataset removes this risk entirely, since no value in it is derived from any real file.

1.2. Dataset Structure

The synthetic export contains 331 positions across 12 asset types (common stock, GICs, corporate bonds, mutual funds and ETFs across three sub-categories, direct and fund-wrapped preferred shares, alternative securities, limited partnership units, and cash balances), each carrying the same eleven fields a real custodian export would: symbol, CUSIP, last trade price, security name, AUA, household count, asset type, security class, risk rating, sector, and industry. Household counts were generated with a long-tailed distribution

(heavily weighted toward 1-2 households, smoothly decaying, with pooled fund and GIC positions receiving a multiplier to reflect that more clients are typically pooled into a fund than into an individual stock pick), matching the shape that motivates the household-concentration audit in Section 4.3 below.

1.3. Cleaning and Variable Roles

The synthetic export loads with clean headers and no placeholder values, so no header/footer cleaning or type coercion was required (unlike a real custodian export, which typically needs

both). Column names were normalized to snake_case for the analysis script. `CUSIP` and `Security Name` serve as identifiers; `Asset Type`, `Security Class`, `Risk Rating`, `Sector`, and `Industry` are descriptive/grouping attributes used directly in pandas groupby/filter operations; `AUA (CAD\$)` and `# HHs` are the two measures summed or counted throughout. No response variable exists in this analysis — all three questions below are descriptive/structural audits, not predictive models, matching the branch determination used in the real engagement this methodology generalizes from.

2. Exploratory Data Analysis

Distributions of AUA and household count were reviewed before any of the three audit questions were answered.

2.1. AUA and Household-Count Distributions

Figure 1 shows AUA per security on a log scale (mean \$194,317, median \$143,246, heavily right-skewed on the raw scale as is typical for a book with many small positions and a thin tail of larger ones) alongside the household-count

distribution, rebuilt with one bin per integer count and a log-scaled y-axis so the long tail (max 120 households on a single pooled fund position) remains visible rather than collapsing into a single dominant bar. The household-count distribution is heavily weighted toward 1-2 households (mean 17.0, median 3.0), the shape that directly motivates the concentration-risk audit in Section 4.3.

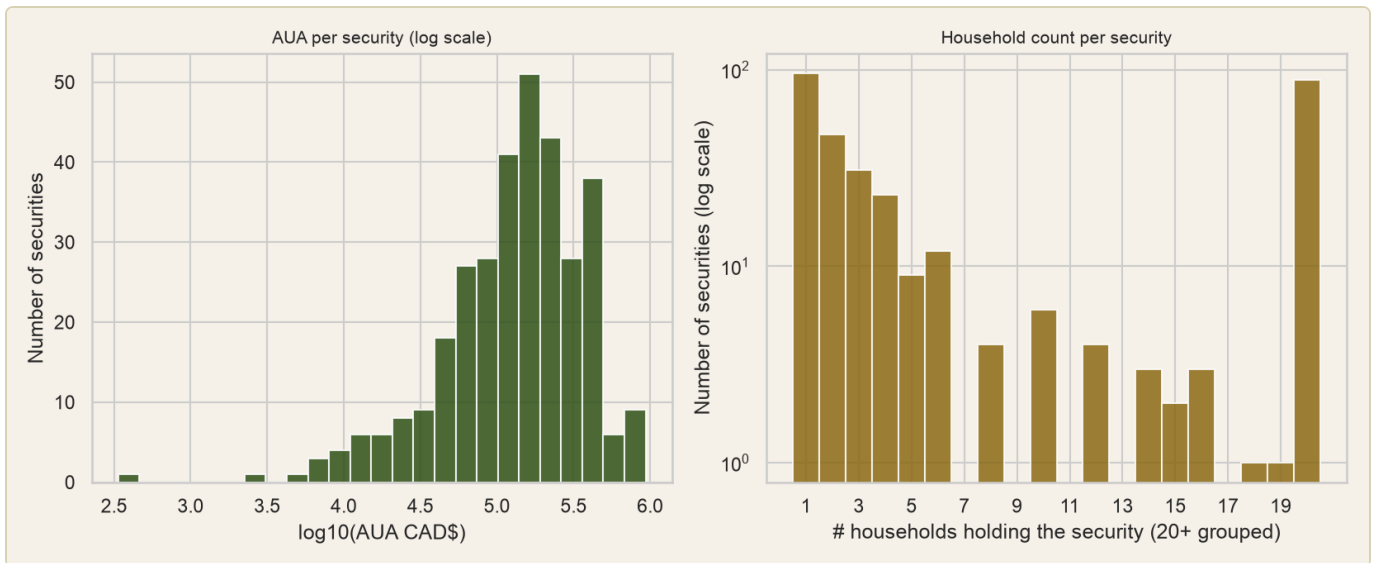


Figure 1. AUA and household-count distributions.

2.2. AUA vs. Household Count

Figure 2 plots AUA against household count directly. The Pearson correlation is -0.296 — a modest negative relationship, consistent with

larger pooled fund/GIC positions sometimes carrying moderate AUA-per-household while some of the single-household positions carry meaningfully large individual dollar amounts, the exact pattern the household-concentration audit (Section 4.3) is designed to surface.

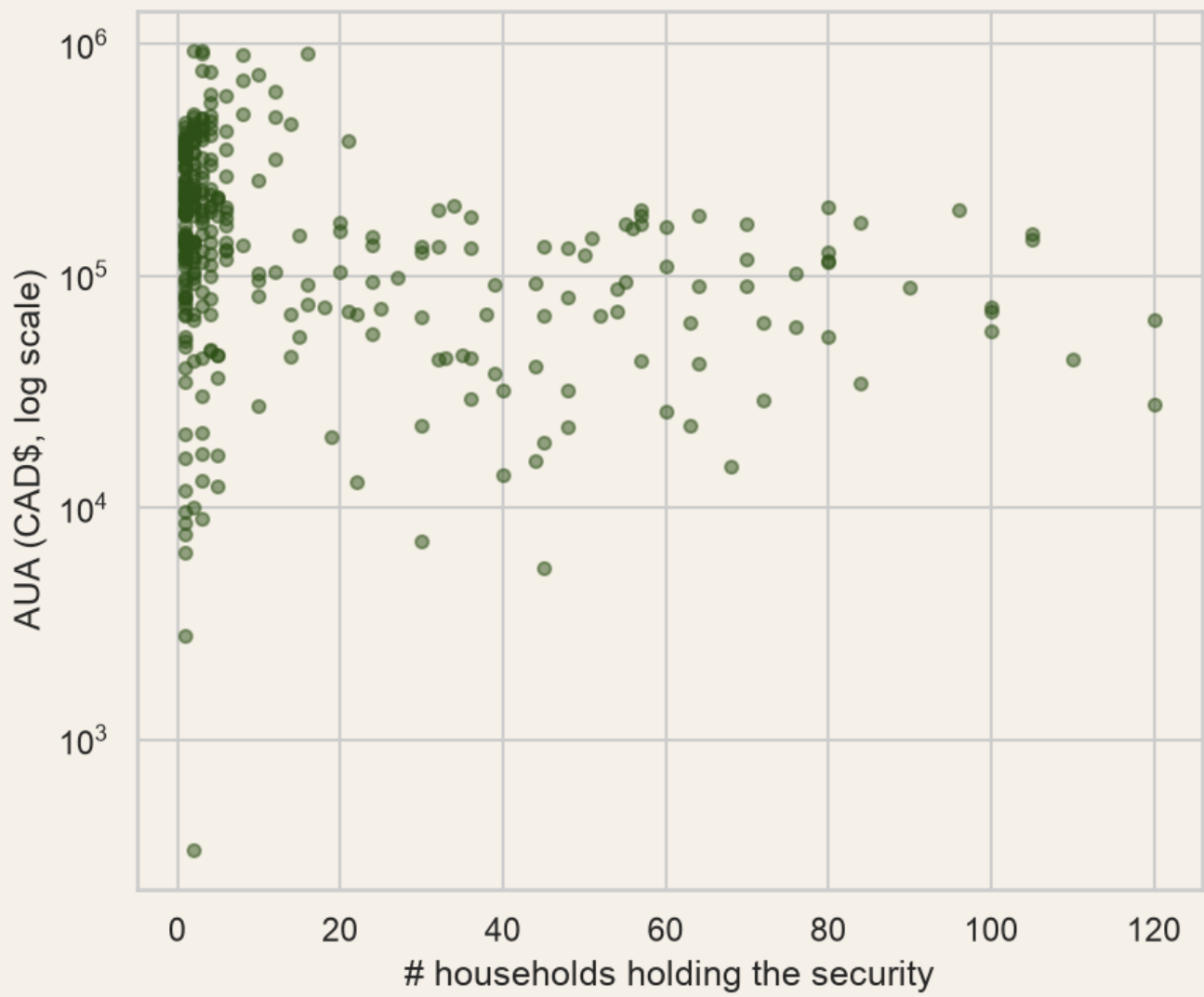


Figure 2. AUA vs. household count.

3. Results: Fund Consolidation, Preferred-Share Exposure, and Household Concentration

The three audit questions, answered in sequence: true fund count, direct preferred-share exposure, and household-concentration risk.

3.1. Q1 — True Fund Count

The synthetic book holds 43 fund/ETF position rows across 36 distinct raw series names. A deterministic, two-stage rule-based parser (strip a trailing load-code token, then a trailing series-class token) consolidates these into **24 distinct true funds** — a 33% reduction from the raw series count. Eight true funds hold 2 or more series (the consolidation target), while 16 hold only a single series each. Figure 3 shows

the distribution of series-per-fund. This is the same class of discrepancy a standard custodian report cannot surface on its own: the raw export counts each load/series combination as a separate line, systematically overstating the practice's true fund diversification unless explicitly consolidated. The full input-to-output mapping (every raw security name traced to its consolidated true fund) is saved alongside this report for line-by-line audit.

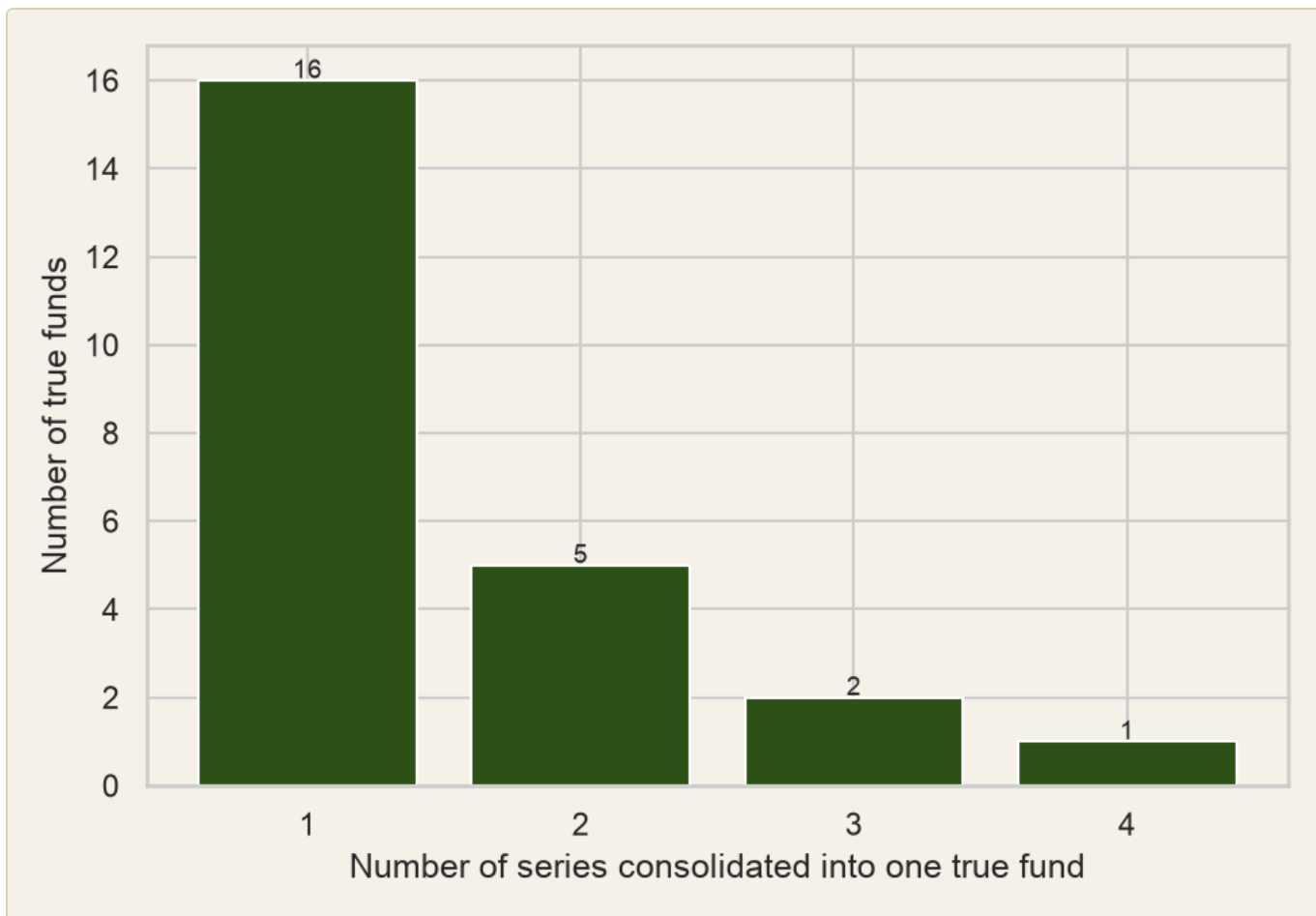


Figure 3. Distribution of series consolidated per true fund (Q1).

3.2. Q2 — Preferred-Share Exposure

20 positions carry an "Preferred Shares" asset-type tag, totaling \$3.27M in AUA. Of these, 3 positions (\$667K) are mutual funds that invest in a basket of preferred shares, not individually-traded preferred securities — a distinction the custodian's own asset-type tagging does not make, since it tags both alike. Separating these out leaves **17 direct preferred-share positions totaling \$2.61M**

across 10 issuers, the figure that actually answers "how much is held in preferred shares" as a knowledgeable advisor would mean it. Figure 4 breaks this direct total down by issuer (ranging from \$11,942 to \$577,221 per issuer) and by sector (led by Consumer Staples and Health Care at roughly \$590,000-\$647,000 each). Reporting the \$3.27M combined figure instead of the \$2.61M direct figure would overstate genuine single-issuer preferred-share exposure by 25%.

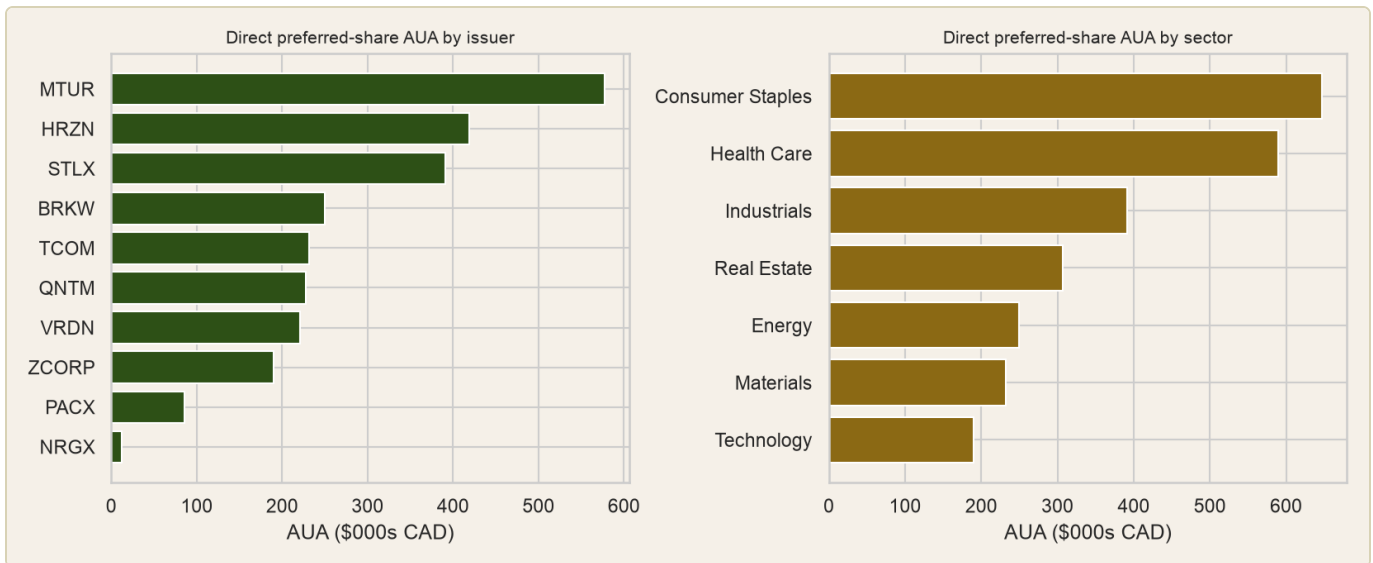


Figure 4. Direct preferred-share AUA by issuer and by sector (Q2).

3.3. Q3 — Household-Concentration Risk

Excluding GICs, 172 of 221 non-GIC positions (77.8%) are held by fewer than 4 households, together totaling \$38.3M in AUA. Narrowing further to positions held by exactly 1 or 2 households — the sharpest concentration signal — 143 positions totaling \$30.1M remain. Figure 5 breaks this down by exact household count and by risk rating. The risk-rating ×

asset-type cross-tab shows the largest thinly-held, elevated-risk concentration sits in Equity positions rated MEDIUM or higher (over \$22M combined across MEDIUM, MEDIUM-HIGH, and HIGH ratings), the profile a suitability/concentration review would typically flag first — as distinct from thinly-held but LOW-risk fixed income or mutual fund positions, which carry materially less practice-management concern despite an identical household count.

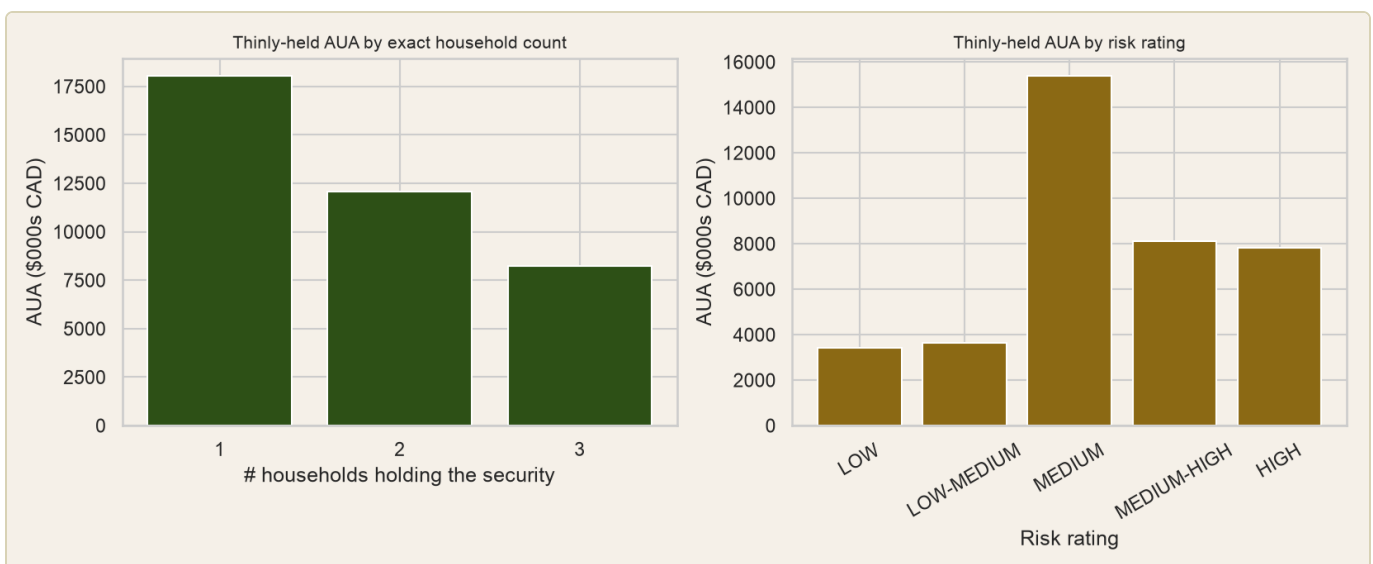


Figure 5. Thinly-held AUA by household count and by risk rating (Q3).

Recommendations

1. Consolidate fund positions by true underlying fund, not by raw series/load code, before reporting fund-count or diversification figures to a practice or its clients.
2. Separate direct preferred-share holdings from fund-wrapped preferred-yield products before reporting preferred-share exposure — the two carry materially different risk and liquidity profiles despite shared asset-type tagging.
3. Prioritize a suitability/concentration review of thinly-held, elevated-risk-rated equity positions first — the segment where low household counts and higher risk ratings overlap most heavily.

Limitations and Caveats

The following limitations apply to this analysis:

Fully synthetic dataset: every figure in this report derives from a fabricated dataset built to match the structure of a real advisory book of business, not from any real client's actual holdings. Absolute dollar figures, security names, and issuer tickers are illustrative only and should not be read as a claim about any specific practice, advisor, or firm.

Simplified relative to the original methodology: the real engagement this piece generalizes from also combined positions across two internal dealer/book codes for the same practice; that step is not needed here since the synthetic dataset represents a single book by construction.

Rule-based parsing, not exhaustive: the fund-series consolidation logic (Section 3.1) covers the load-code and series-class patterns present in this dataset; a real custodian export may use additional suffix conventions requiring the parser to be extended.

No external benchmark: this is a standalone structural audit of one (synthetic) book, not benchmarked against industry norms for fund count, preferred-share allocation, or concentration risk.

AI Usage Disclosure

This report's production and formatting were completed with AI-assisted tools. All methodology, architecture, and analytical review were completed by the author.