Contents

1 BEQI: An Open Methodology for Retail Forex Broker Execution Quality Measurement

Author: Boris Fesenko **Affiliation:** BJF Trading Group Inc., Ontario, Canada **Contact:** boris@fesenko.ca **Version:** BEQI v1.0 (May 2026) **Reference implementation:** https://github.com/bjftradinggroup-inc/forex-broker-audit-toolkit **Methodology page:** https://bjftradinggroup.com/forex-broker-audit-toolkit/

1.1 Abstract

Retail forex broker execution quality has historically been unmeasurable at scale. Brokers do not publish matching latency, slippage symmetry, spread widening factors, last-look hold times, or requote rates. Third-party broker reviews focus on user experience and withdrawal performance — surface variables that do not predict whether an algorithmic strategy will survive in production deployment. This paper presents BEQI (Broker Execution Quality Index), an open methodology for measuring retail forex broker execution quality across five orthogonal dimensions, each independently derivable from a trading statement. We define the per-dimension thresholds, the composite scoring function — a weighted geometric mean with square-root smoothing — and the four-tier broker classification that maps the composite to qualitative bands. A reference implementation in Python is published under MIT license, enabling independent replication, methodology critique, and contribution. A community-driven

public spreadsheet aggregates anonymized audits over time, accumulating toward the first public benchmark dataset of retail forex execution quality. The methodology is intentionally conservative in two ways: we do not publish a leaderboard of named brokers, and we expose all sub-scores alongside the composite so reviewers may recompute under alternative weighting schemes. We discuss limitations, sample-size requirements, and a roadmap for v1.1 and v2.0. This is a free community tool with no support guarantee; bug reports, methodology critiques, and pull requests are welcomed via the GitHub repository.

Keywords: retail forex, broker execution, market microstructure, latency arbitrage, slippage measurement, last-look, open methodology, public benchmark.

1.2 1. Introduction

1.2.1 1.1 The retail forex execution measurement gap

The retail foreign exchange market processes more than five trillion U.S. dollars in average daily turnover, of which a non-trivial fraction is intermediated by retail brokers offering margined contracts-for-difference and similar synthetic instruments to non-institutional clients. Broker execution behavior — how fast orders are matched, how slippage is distributed between client and broker, how spread responds to signal-driven order flow, whether and for how long brokers hold orders before confirmation, and how often orders are rejected with requotes — directly determines whether a given algorithmic trading strategy is profitable in deployment.

Despite this centrality, retail forex broker execution quality is not publicly measured. Brokers compete on advertised spread and headline commission; their actual matching behavior is observable only by individual clients on individual accounts, and the data does not aggregate. Third-party broker review sites (Forex Peace Army, FX Empire, BrokerChooser, and similar) score on user experience, withdrawal speed, regulatory licensing, and other surface dimensions; none publish quantitative measurements of execution behavior. Regulatory disclosures, where they exist, focus on prudential capital and client-money segregation rather than execution mechanics.

The result is an information asymmetry: traders must commit capital to an account before they can observe how that account is actually executed, and there is no reusable framework for translating that observation into a comparison across venues.

1.2.2 1.2 Related work

Three strands of prior work bear on the problem.

The market-microstructure literature (Hasbrouck 2007; O'Hara 1995) establishes the conceptual framework — order flow, asymmetric information, dealer behavior — but operates almost entirely on equity markets where TAQ-style consolidated tape data is available. Retail forex lacks an equivalent reference dataset.

The latency-arbitrage literature (Aquilina, Budish, and O'Neill 2022; Budish, Cramton, and Shim 2015) examines the economics of speed competition on lit equity

venues. The conclusions translate qualitatively to forex — fast feeds dominate slow feeds — but the broker-intermediated retail forex setting introduces opacity that the equity literature does not address.

The fintech-research practitioner community has produced individual broker reviews and ad-hoc execution measurements published in trader forums (ForexFactory, Reddit r/Forex, traden.de) but without a shared methodology, results from different observers cannot be combined into an aggregate signal.

Our contribution is the missing methodological layer: a small set of well-defined measurements that can be computed from any retail trading statement, with a composite score that is honest about what it measures and what it does not.

1.2.3 1.3 Contributions

This paper makes four contributions:

1. We define five orthogonal execution-quality dimensions — matching latency, slippage asymmetry, spread widening factor, last-look hold time, and requote rate — together with per-dimension threshold tables that map a raw measurement to a 0–100 sub-score.
2. We propose a composite scoring function — a weighted geometric mean with square-root smoothing — designed so that no single catastrophic dimension can be hidden by good performance on the others.
3. We publish a reference implementation in Python under the MIT license. The toolkit is single-file, has no third-party dependencies in its v1.0 release, and produces both human-readable and machine-readable output suitable for replication and aggregation.
4. We open a public submission database for anonymized audits, with a deliberate decision against publishing a named-broker leaderboard, on grounds we discuss in Section 6.

The methodology, the toolkit, the submission protocol, and this paper are all version-controlled and openly licensed for replication, critique, and extension.

1.3 2. The five BEQI dimensions

1.3.1 2.1 Matching latency

Definition. The median time, in milliseconds, between order submission (`send_time`) and confirmed fill (`fill_time`) across all filled orders in the audit period.

Source. Trading statement timestamps; or, for FIX-connected accounts, the difference between the `NewOrderSingle (D)` send timestamp and the corresponding `ExecutionReport (8)` fill timestamp.

Why median, not mean. Latency distributions in retail brokers are right-skewed by occasional 1–5 second outliers attributable to network re-routing or broker-side

garbage-collection pauses. The median is robust to these tail events; the toolkit additionally reports the 90th and 99th percentiles so that the reader can inspect the tail behavior independently.

Tier thresholds.

Median latency	Tier	Sub-score band
0–15 ms	1 (Institutional ECN)	85–100
15–80 ms	2 (Fast retail / fast bridge)	65–85
80–200 ms	3 (Standard retail)	40–65
200–1000 ms	4 (Problematic)	0–40
> 1000 ms	4	0

1.3.2 2.2 Slippage asymmetry

Definition. The ratio of negative-slippage filled orders to positive-slippage filled orders.

Sign convention. For a buy order, positive slippage means the order filled below the requested price (favorable to the trader); for a sell order, positive slippage means the order filled above the requested price. A neutral broker, intermediating between symmetric liquidity providers without preferential routing, should produce a ratio close to unity.

Asymmetry metric.

$$A = \frac{N_{\text{neg}}}{N_{\text{pos}}}$$

where N_{neg} and N_{pos} are the counts of orders with negative and positive signed slippage, respectively. In the degenerate case where $N_{\text{pos}} = 0$, the metric is capped at 5.0 for scoring purposes.

Tier thresholds.

Asymmetry	Tier	Sub-score band
1.00–1.05	1	85–100
1.05–1.20	2	65–85
1.20–1.50	3	40–65
1.50–5.00	4	0–40

A reading above 1.20 is statistically loud at modest sample sizes and admits multiple causal interpretations: routing through asymmetric liquidity-provider pools, selective last-look behavior, or a server-side asymmetric slippage engine. The metric measures behavior, not intent.

1.3.3 2.3 Spread widening factor

Definition. The ratio of the median bid-ask spread observed in a 50 ms window centered on each order's `send_time` to the median spread observed in a 5-minute baseline window centered on the same time.

Rationale. Order flow that arrives at moments of price uncertainty correlates with spread widening, because liquidity providers withdraw quotes to protect against being picked off by adverse-selection traders. A broker whose spread expands sharply at signal moments imposes an asymmetric execution cost on the trader that is invisible to a spread-only comparison.

Computation. For each order i in the audit period, the toolkit computes

$$W_i = \frac{\text{median}\{s_t : t \in [t_i - 25\text{ms}, t_i + 25\text{ms}]\}}{\text{median}\{s_t : t \in [t_i - 2.5\text{min}, t_i + 2.5\text{min}]\}}$$

where s_t is the bid-ask spread at tick time t. The reported widening factor is the median of W_i across all orders.

Coverage. This dimension requires tick data covering the audit period. If absent, the dimension is reported as null and the composite score is computed over the remaining four dimensions, with weights renormalized.

Tier thresholds.

Widening factor	Tier	Sub-score band
1.00–1.30	1	85–100
1.30–2.00	2	65–85
2.00–3.50	3	40–65
3.50–10.00	4	0–40

1.3.4 2.4 Last-look hold time

Definition. The median delay between the moment a broker matches an order and the moment the trader receives the fill confirmation.

Rationale. "Last look" is a deliberate hold during which the broker may reject or re-price the order based on price movement during the hold. Long last-look windows allow the broker to systematically avoid trades that move adversely to the broker's book — an asymmetric option whose cost is borne by the trader.

Source. FIX `ExecutionReport (8)` `TransactTime` versus `SendingTime`, or an extended trading statement that exposes both matched and confirmed timestamps. Standard retail HTML statements typically do not expose this granularity, in which case the dimension is reported as null and renormalized as in Section 2.3.

Tier thresholds.

Median last-look	Tier	Sub-score band
0–5 ms	1	85–100
5–150 ms	2	65–85
150–500 ms	3	40–65
500–3000 ms	4	0–40

1.3.5 2.5 Requote rate

Definition. The number of requote events per 100 submitted orders in the audit period. A requote is the broker rejecting a submitted price and offering a worse price, which the trader may accept or cancel.

Source. Explicit requote events in the trading statement. If the statement does not flag requotes, the toolkit infers them by detecting send/cancel/resend triplets within a 200 ms window — a heuristic that admits false positives but is documented and can be disabled by the user.

Tier thresholds.

Requotes / 100 orders	Tier	Sub-score band
0.0–0.5	1	85–100
0.5–2.0	2	65–85
2.0–5.0	3	40–65
5.0–50.0	4	0–40

1.4 3. Composite scoring

1.4.1 3.1 Sub-score interpolation

Within each tier band, the sub-score is interpolated linearly. Formally, given a measurement value x falling in tier band $[L_k, U_k]$ with corresponding score band $[s_k^{\text{low}}, s_k^{\text{high}}]$:

$$\text{score}(x) = s_k^{\text{high}} - \frac{x - L_k}{U_k - L_k} \cdot (s_k^{\text{high}} - s_k^{\text{low}})$$

clipped to $[0, 100]$.

1.4.2 3.2 Composite formula

The composite BEQI is a weighted geometric mean with square-root smoothing over the dimensions for which a sub-score is available:

$$\text{BEQI} = \left(\sum_{i \in D} w_i' \cdot \sqrt{s_i} \right)^2$$

where D is the set of measurable dimensions for the audit, s_i is the sub-score of dimension i (clamped to a minimum of 1 to avoid degeneracies), and w'_i is the weight of dimension i renormalized over the present dimensions:

$$w'_i = \frac{w_i}{\sum_{j \in D} w_j}.$$

The default weights are

$$w_{\text{latency}} = 0.30,\ w_{\text{asymmetry}} = 0.25,\ w_{\text{widening}} = 0.20,\ w_{\text{last-look}} = 0.15,\ w_{\text{requote}} = 0.10$$

summing to unity over the full set of five dimensions.

The square-root smoothing distinguishes the composite from a pure geometric mean (which would multiplicatively annihilate when any s_i approaches zero) and from an arithmetic mean (which would allow a single catastrophic dimension to be averaged out by good performance elsewhere). The chosen functional form preserves a desirable "weakest-link" property — a 1500 ms last-look reading pulls the composite down strongly — while remaining differentiable and well-defined across the parameter space.

1.4.3 3.3 Tier mapping

The composite BEQI is mapped to a four-tier classification:

Composite	Tier	Label
85–100	1	Institutional ECN
65–85	2	Fast retail / fast bridge
40–65	3	Standard retail
0–40	4	Problematic

The tier labels are descriptive rather than categorical: a Tier 3 broker is one whose composite execution quality is consistent with what we observe across the broad middle of the retail forex industry, not a regulatory or commercial categorization.

1.4.4 3.4 Confidence bands

Every audit reports a confidence band based on sample size N (number of fills covered):

N	Confidence	Use
≥ 1000	high	tier classification reliable
200–999	medium	tier indicative; CIs wide
50–199	low	early signal only
< 50	rejected	not accepted in public dataset

1.5 4. Reference implementation

1.5.1 4.1 Architecture

The reference implementation (`beqi.py`) is a single Python file with no third-party dependencies in its v1.0 release. It exposes a command-line interface and a Python API. Internally it is structured as five dimension-analyzers, a composite-scorer, three input parsers (CSV, HTML statement, FIX log — with the latter two stubbed in v1.0 and scheduled for v1.1), and a JSON serializer.

1.5.2 4.2 Input formats

The toolkit accepts three input formats for trading data, listed in order of richness:

- **FIX 4.4 log** — produces measurements for all five dimensions including last-look hold time. Targeted at institutional-style accounts and traders running through FIX bridges.
- **Common terminal HTML statement** — produces measurements for dimensions 1, 2, and 5 directly; dimension 3 requires accompanying tick data; dimension 4 is null.
- **CSV trading log** — the most flexible format; required columns are `order_id, send_time, fill_time, requested_price, filled_price, side, volume`; optional columns are `pair, requote_flag, last_look_ms, account_type`.

For dimension 3 (spread widening), the toolkit additionally requires a tick-data CSV with columns `timestamp, bid, ask`.

1.5.3 4.3 Output schema

The toolkit emits a JSON report whose top-level keys are

```
{
  "methodology_version": "BEQI v1.0",
  "period_start": ISO-8601,
  "period_end": ISO-8601,
  "sample_size": int,
  "confidence": "high" | "medium" | "low",
  "pairs_covered": [str],
  "account_type": "live" | "demo" | null,
  "dimensions": {
    "matching_latency": {score, raw_value, ...},
    "slippage_asymmetry": {...},
    "spread_widening_factor": {...},
    "last_look_hold": {...},
    "requote_rate": {...}
  },
  "composite_beqi": float,
  "tier": int (1-4),
  "tier_label": str,
  "dimensions_measured": [int]
```

}

The full schema including per-dimension `extras` fields (p90, p99, asymmetry counts, widening baseline values, etc.) is documented in `METHODOLOGY.md` in the repository.

1.5.4 4.4 Validation

The v1.0 release ships with synthetic sample data exhibiting all five dimensions — `samples/sample_trades.csv` and `samples/sample_ticks.csv` — and produces a deterministic BEQI report when run against them. We have additionally validated the toolkit on internally collected statement data spanning multiple broker dialects, used during development of the dialect support inside the SharpTrader software suite. Independent replications and bug reports are explicitly invited via the GitHub repository.

1.6 5. The public dataset

1.6.1 5.1 Submission protocol

A companion Google Sheet — the BEQI Public Audit Database — accepts row-level submissions. Submitters paste the JSON output of the toolkit into a public form; an Apps Script trigger validates the JSON against the BEQI schema, generates a submission identifier, and appends the row to the dataset. Validated rows count toward the aggregate distributions presented in a separate `aggregates` tab.

1.6.2 5.2 Anonymity

Submissions are anonymous by default. The toolkit never collects account identifiers, balances, profit-and-loss, or position sizes. The output JSON contains only timestamps, slippage deltas, event counts, and computed scores. The broker name field on the submission form is optional and may be left blank.

1.6.3 5.3 No leaderboard

The maintainer team — BJF Trading Group Inc. — explicitly does not publish a leaderboard of named brokers compiled from the public dataset. We discuss the rationale in Section 6.1.

1.7 6. Discussion

1.7.1 6.1 Why no leaderboard

Two considerations.

First, an authoritative leaderboard published by the methodology maintainer creates a strong incentive for brokers to optimize for the audit signal rather than for execution quality itself — a Goodhart's-law dynamic. The integrity of the dataset depends

on broker behavior remaining unsignaled by the methodology; once leaderboard publication becomes the dominant attractor for broker behavior, the measurements stop generalizing.

Second, public ranking of regulated financial entities by an unregulated third party invites legal exposure that is disproportionate to the research value. By publishing only the methodology, the code, and the anonymized dataset, we transfer the discretionary act of ranking to whoever wishes to undertake it, with the legal accompaniment that follows. The dataset is openly licensed; nothing prevents downstream parties from publishing their own leaderboards using the same data.

1.7.2 6.2 Weight justification

The default weights — latency 0.30, asymmetry 0.25, widening 0.20, last-look 0.15, requote 0.10 — reflect the maintainer team's judgment about which dimensions most strongly determine in-production strategy survival, informed by experience deploying latency-sensitive strategies through retail brokers. They are not derived from first principles, and we do not claim they are optimal.

The methodology explicitly exposes all sub-scores alongside the composite so that reviewers can recompute under alternative weighting schemes. A user who believes that requote rate should be weighted higher than last-look hold time can apply that weighting trivially; the raw scores remain the authoritative measurement, with the composite a derived view.

1.7.3 6.3 Sample-size requirements

Audits with fewer than 200 fills are tagged "low confidence" and audits with fewer than 50 are rejected from the public dataset. The reasoning is empirical: at $N = 50$ the slippage-asymmetry estimate has a 95% confidence band wider than the spacing between adjacent tier thresholds, making tier classification statistically meaningless. At $N = 200$ the bands narrow to the point where tier classification is indicative; at $N = 1000$ the classification is reliable. Submitters who wish to contribute usefully to the public dataset are encouraged to accumulate at least 200 fills before submitting.

1.7.4 6.4 Methodology critique

The methodology is open to critique, and we encourage it. Specifically:

- The 50 ms signal-window in the spread-widening computation is fixed; some strategies operate on substantially different signal-detection latencies, and the window should be configurable.
- The square-root smoothing in the composite is a design choice rather than a derivation; alternative smoothing functions (logarithmic, power-law) may produce qualitatively different rankings on edge cases.
- The five dimensions cover a substantial fraction of execution-quality variation but are not exhaustive; an explicit "broker-side rejection rate" dimension and an "adaptive throttle" dimension are candidates for inclusion in v1.2.

Proposed methodological revisions should be opened as Discussion threads in the GitHub repository. Methodology version 1.1 will incorporate the strongest community critiques.

1.8 7. Limitations and threats to validity

We identify five material limitations.

1. **Sample-size dependency.** Beyond the rejection threshold above, even high-confidence audits remain point estimates with non-trivial variance. Aggregation across submissions partially mitigates this but does not eliminate it.
2. **Time-of-day variance.** A given broker may post measurably different scores at different sessions, due to liquidity-provider rotation, bridge-routing changes, or staff-shift effects on manual order handling. v1.0 audits do not partition by session; v1.1 will introduce per-session breakdowns.
3. **Account-level differences.** Brokers can apply different routing logic per-account based on profile, balance, prior trading pattern, or jurisdiction. One trader's audit may not generalize to another's on the same nominal broker. The public dataset partially mitigates this through volume of submissions but does not directly address it.
4. **Causal interpretation.** Asymmetric slippage admits multiple causal explanations — routing through asymmetric LP pools is statistically indistinguishable from selective last-look or active manipulation in the absence of additional data. The methodology measures behavior, not intent.
5. **Coverage of dimensions 3 and 4.** Spread widening (dimension 3) requires a tick stream; last-look hold time (dimension 4) requires either a FIX log or an extended statement. Standard retail HTML statements typically support only dimensions 1, 2, and 5, in which case the composite is computed over three of five dimensions and the resulting tier classification is a less reliable estimate of true overall execution quality.

1.9 8. Future work

1.9.1 8.1 v1.1: parser improvements

The most-requested feature based on early community feedback is broader statement-format support. v1.1 will ship parsers for common terminal HTML statements and FIX 4.4 logs, eliminating the need for users to manually convert to CSV.

1.9.2 8.2 v1.2: sixth dimension

We propose a sixth dimension — adaptive throttle behavior — to capture brokers that progressively slow execution in response to detected high-frequency order patterns.

Detection requires a longer audit period (1 month minimum) and a comparison between the latency distribution at the start of the period and at the end.

1.9.3 8.3 v2.0: live dashboard

The companion public dataset will be exposed via a live dashboard at `beqi.bjftradinggroup.com` showing aggregate distributions across all submissions, partitioned by confidence band, account type, and time period. Methodology weights and threshold tables will be revisitable via configurable controls.

1.10 9. Conclusion

We have presented BEQI, an open methodology for measuring retail forex broker execution quality across five dimensions, together with a reference implementation in Python and a public dataset for community contributions. The methodology is conservative in its claims, transparent in its construction, and open to critique. We do not publish a leaderboard of named brokers, on grounds of data integrity and legal exposure, but the dataset is openly licensed and may be used by downstream parties to construct rankings as they see fit.

We invite three forms of community engagement: replications against brokers we have not covered, methodology critiques and weight refinements via GitHub Discussions, and pull requests for additional input-format parsers. The toolkit is a free community resource maintained on a best-effort basis; high-touch support is not guaranteed. Bug reports and feature requests should be opened as GitHub Issues.

The dataset starts empty in May 2026. Its long-term value depends on community contributions. We hope that, over the subsequent twelve to twenty-four months, the accumulated submissions form the first public benchmark of retail forex execution quality, independent of broker advertising and review-site spreads-only scoring.

1.11 Data and code availability

The reference implementation, sample data, methodology document, and public dataset structure are available at:

- **Repository:** https://github.com/bjftradinggroup-inc/forex-broker-audit-toolkit
- **Methodology page (with FAQ):** https://bjftradinggroup.com/forex-broker-audit-toolkit/
- **Public submission database:** https://docs.google.com/spreadsheets/d/10TOJd1NVzAyqwAd

The toolkit is licensed under the MIT License. This paper is licensed under CC-BY 4.0; the methodology specification (`METHODOLOGY.md` in the repository) is licensed under MIT.

1.12 Conflict of interest

The author is the founder and operator of BJF Trading Group Inc., which sells commercial trading software including SharpTrader Pro, SharpTrader Optimizer, and NewsAutoTraderPro. These commercial products are independent of the BEQI methodology and toolkit, which are open-source and broker-agnostic. The author has no commercial interest in any specific broker scoring well or badly under the BEQI methodology.

1.13 Acknowledgements

The methodology benefited from years of internal-tooling development at BJF Trading Group on broker-dialect support for the SharpTrader execution engine. We thank the early pre-release testers from the BJF Trading Group community for feedback on the threshold tables and the composite weighting; remaining errors are entirely the author's responsibility.

References

Aquilina, M., Budish, E., and O'Neill, P. (2022). Quantifying the high-frequency trading "arms race." *Quarterly Journal of Economics*, 137(1), 493–564.

Budish, E., Cramton, P., and Shim, J. (2015). The high-frequency trading arms race: frequent batch auctions as a market design response. *Quarterly Journal of Economics*, 130(4), 1547–1621.

Hasbrouck, J. (2007). *Empirical Market Microstructure: The Institutions, Economics, and Econometrics of Securities Trading*. Oxford University Press.

O'Hara, M. (1995). *Market Microstructure Theory*. Blackwell.

Goodhart, C. (1975). Problems of monetary management: the U.K. experience. *Papers in Monetary Economics*, vol. I, Reserve Bank of Australia.

Fesenko, B. (2026). Why latency arbitrage backtests don't survive in production: the execution-time gap. *BJF Trading Group research*. https://bjftradinggroup.com/latency-arbitrage-backtest-execution-time-gap/

1.14 Appendix A: Sample BEQI report (JSON, abbreviated)


```
{
  "methodology_version": "BEQI v1.0",
  "period_start": "2026-04-01T00:00:00Z",
  "period_end": "2026-04-29T00:00:00Z",
  "sample_size": 1247,
  "confidence": "high",
  "pairs_covered": ["EURUSD", "XAUUSD", "GBPUSD"],
```

```
  "account_type": "live",
  "dimensions": {
    "matching_latency": {
      "score": 58, "raw_value": 94.0,
      "p90_ms": 187.0, "p99_ms": 412.0, "n": 1247
    },
    "slippage_asymmetry": {
      "score": 47, "raw_value": 1.34,
      "negative_count": 712, "positive_count": 532, "n": 1247
    },
    "spread_widening_factor": {
      "score": 51, "raw_value": 2.4,
      "baseline_spread_med": 0.000080, "signal_spread_med": 0.000192
    },
    "last_look_hold": {
      "score": 39, "raw_value": 218.0,
      "p90_ms": 487.0, "p99_ms": 894.0, "n": 1247
    },
    "requote_rate": {
      "score": 44, "raw_value": 3.2,
      "requote_count": 40, "total_orders": 1247
    }
  },
  "composite_beqi": 48.6,
  "tier": 3,
  "tier_label": "Standard retail",
  "dimensions_measured": [1, 2, 3, 4, 5]
}
```


1.15 Appendix B: Composite formula derivation

The composite functional form

$$\text{BEQI} = \left(\sum_i w_i' \sqrt{s_i} \right)^2$$

is selected from a family of generalized f-mean composites

$$M_f(s) = f^{-1} \left(\sum_i w_i' f(s_i) \right)$$

with $f(x) = \sqrt{x}$ and $f^{-1}(y) = y^2$.

The choice of f controls the substitution behavior between dimensions:

- $f(x) = x$ (identity) yields the arithmetic mean: dimensions are perfectly substitutable, a catastrophic dimension can be hidden by good performance elsewhere.

- $f(x) = \log x$ yields the geometric mean: dimensions are partially complementary; a zero score on any dimension drives the composite to zero, which is undesirable for the "all-but-one-missing" cases the methodology routinely encounters.
- $f(x) = \sqrt{x}$ yields the present formulation: between arithmetic and geometric, with the desirable "weakest-link" property but without the degenerate behavior at zero.

We selected $\sqrt{x}$ as the smallest power function that produces visible weakest-link behavior across the empirical range we observed during development. Other smoothing functions remain reasonable; the methodology document explicitly invites alternative-smoothing proposals in repository Discussions.

1.16 Appendix C: CLI usage

```
$ git clone https://github.com/bjftradinggroup-inc/forex-broker-audit-toolkit
$ cd forex-broker-audit-toolkit
$ python beqi.py --input statement.csv --tick-data ticks.csv --output report.json

============================================================
BEQI Audit Report
============================================================
Period:         2026-04-01 to 2026-04-29 (4 weeks)
Pairs:          EURUSD, XAUUSD, GBPUSD
Sample size:    1,247 fills
Confidence:     High (>1000 fills)
------------------------------------------------------------
1. Matching latency (median):     94 ms     [score: 58]
   p90: 187 ms · p99: 412 ms
2. Slippage asymmetry:            1.34      [score: 47]
   1,247 fills · 712 negative · 532 positive · 3 zero
3. Spread widening factor:        2.4x      [score: 51]
   baseline 0.8 pip · at-signal 1.92 pip
4. Last-look hold time:           218 ms    [score: 39]
   measured from FIX log timestamps
5. Requote rate:                  3.2 / 100 [score: 44]
   40 requotes detected in 1,247 orders
------------------------------------------------------------
Composite BEQI:                   48.6
Tier:                             3 (Standard retail)
------------------------------------------------------------
JSON report written to: report.json
```

Manuscript version 1.0 — May 2026. Comments welcome at boris@fesenko.ca or via GitHub Discussions at https://github.com/bjftradinggroup-inc/forex-broker-audit-toolkit/discussions.