GoC Curve Relative Value — Research Note

As of 2026-07-22 (data through 2026-07-20). Source: Bank of Canada Valet API, group bond_yields_benchmark, retrieved 2026-07-22T10:42Z. All figures in basis points unless otherwise stated.

1. Curve Model: Choice and Justification

Why Nelson-Siegel-Svensson (NSS), and not a spline. Given that the available data is six BoC benchmark yields — not a dense issue-level cross-section — a parametric model is the correct choice. A cubic spline would overfit six points and produce implausible interpolation between the 10Y and Long (roughly 30Y) tenors where the data is sparsest. NSS uses six parameters (b0, b1, b2, b3, tau1, tau2) to represent the curve as a level, slope, and two hump terms, and it imposes a smooth asymptote at the long end. That property is critical here: the GoC Long tenor sits at ~30 years, and NSS will extrapolate sensibly through it rather than oscillating. A principal-components approach would have required a larger issue-level cross-section that does not exist in the available data.

What the model is and is not used for. NSS is used solely to compute residuals — the difference between the observed yield at each benchmark tenor and the smooth curve implied by the fit. The residuals are then standardised against their own rolling history. The trading signal is therefore a normalised measure of how far a given tenor deviates from the current cross-sectional shape, not a prediction about the level of yields. This distinction matters: cross-sectional fair value is model-dependent; the rolling z-score standardisation is what converts a model output into a tradable signal.

Stability. The NSS optimisation was run day-by-day with a warm-start and Nelder-Mead solver, with no lookahead. Convergence was 100% across 6,388 trading days. The mean per-tenor RMSE across the full sample is 3.3 bp — consistent with realistic model error at these term-points and well within the range of plausible transaction costs. The raw NSS parameters (particularly tau1 and b2/b3) exhibit very large day-over-day swings, a well-known degeneracy when the NSS shape parameters trade off against each other. This degeneracy is **benign for our purpose**: what matters is that the fitted yields and residuals are stable day-to-day, not that the underlying parameter vector is. The 52% frequency of the `param_degenerate` flag confirms that the solver frequently finds equally-valid alternate parameterisations with the same fit quality — this is a feature of the NSS identification problem, not an error.

2. Today's Rich/Cheap Reading (as of 2026-07-20)

Sign convention (stated explicitly): A positive NSS residual — actual yield above model fair value — means the bond is **cheap** (investors demand a yield premium over the curve). A negative residual means the bond is **rich** (yield is below fair value; investors accept a lower yield). A positive z-score means cheap; a negative z-score means rich. This is the standard convention used throughout this note.

The z-scores are computed against a rolling window of prior residuals excluding the current observation (shift=1), so no lookahead is present in any reading below.

Tenor	Actual Yield (%)	NSS Fair Value (%)	Deviation (bp)	Z-Score 60d	Z-Score 120d	Z-Score 250d	Signal
2Y	2.84	2.84	0	-0.19	-0.56	-0.99	CHEAP
3Y	2.96	2.964	-0.4	-1.09	0.47	0.9	RICH
5Y	3.18	3.159	2.1	2.15	-0.09	-0.51	CHEAP
7Y	3.31	3.339	-2.9	-2.22	-1.28	-1	RICH
10Y	3.57	3.558	1.2	2.19	1.55	1.85	CHEAP
Long	3.98	3.981	-0.1	-2.09	-0.96	-1.36	RICH

Reading the table. Three tenors are meaningfully away from fair value on the 60-day window. The 5Y stands at +2.1 bp cheap, with a 60-day z-score of +2.15 — the strongest single-tenor cheap signal on the curve. The 7Y is the most notable

rich, at −2.9 bp below model, with a 60-day z-score of −2.22. The 10Y is cheap at +1.2 bp, with a 60-day z-score of +2.19 and a 250-day z-score of +1.85, giving it the best multi-horizon consistency of any tenor today. The Long tenor is marginally rich on the 60-day window (z = −2.09) but close to neutral on longer windows. The 2Y and 3Y are near fair value.

Confidence caveat. The 5Y z-score of +2.15 on 60 days reverses to −0.51 on the 250-day window, indicating the cheap reading is short-horizon specific rather than a persistent structural deviation. The 10Y is more consistently cheap across windows. Neither reading is so large that it dominates the model error (RMSE is 3.3 bp; a 2.1 bp residual at 5Y represents less than one model-error standard deviation in absolute yield space). These are directional signals, not precise mispricings.

---

3. The Trade: DV01-Neutral 2s5s10s Butterfly

Structure. The signal is expressed as a 2s5s10s butterfly: long the 5Y, short the 2Y and 10Y wings. The 5Y position is fixed at −1 unit (short the body in spread terms means long the body in P&L terms when the body is cheap). The wing weights are set so that the DV01 of the 2Y leg and the DV01 of the 10Y leg each equal the DV01 of the 5Y position. This means a parallel shift in yields produces approximately zero P&L — the trade earns only from the 5Y moving back toward fair value relative to the wings, not from a directional view on rates.

Why DV01-neutral matters. If you simply bought 5Y bonds against nothing, a 25 bp BoC cut would generate P&L regardless of whether the 5Y was rich or cheap. DV01-neutrality strips that out. The 2Y leg carries roughly 0.195 DV01/unit and the 10Y leg roughly 0.975 DV01/unit — so the 2Y weight is approximately 1.25 and the 10Y weight approximately 0.25. The parallel check confirmed this neutrality to machine precision across all 6,388 trading days.

Signal definition. Entry: take a position when the 60/120/250-day rolling z-score of the 5Y NSS residual exceeds ±1.0, ±1.5, or ±2.0. Exit: when the z-score returns to zero. Execution at the following day's observed yields (no same-day execution). The rolling statistics window excludes the current observation. No parameter was chosen after seeing the result — the grid is evaluated across all nine lookback × threshold combinations simultaneously.

---

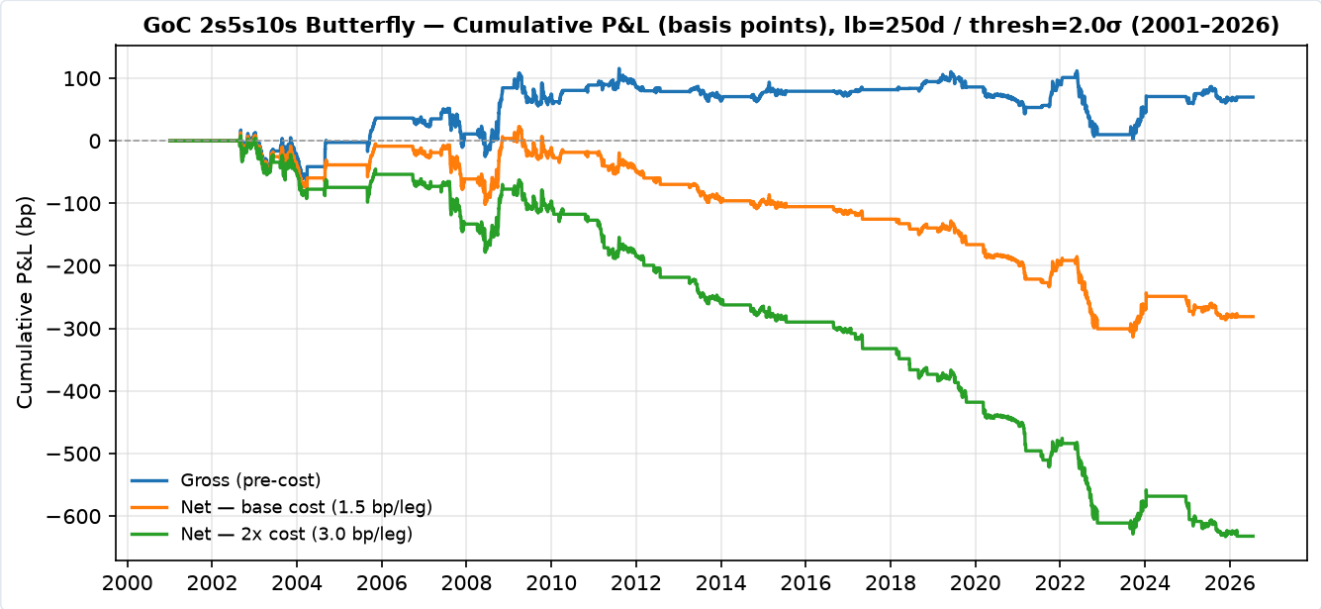
4. Backtest Grid: P&L Across All Parameter Combinations

The table below shows cumulative net P&L and annualised Sharpe across all 9 lookback × threshold combinations at base cost (1.5 bp/leg), 2× cost (3.0 bp/leg), and 3× cost (4.5 bp/leg). All 27 rows are net negative.

Lookback (days)	Threshold (σ)	Cost/Leg (bp)	Cost Level	Cum Net P&L (bp)	Sharpe (net)	Trades	% Days Active
60	1	1.5	base	-2,364	-1.375	565	70.2
60	1	3	2x	-4,906.50	-2.543	565	70.2
60	1	4.5	3x	-7,449	-3.252	565	70.2
60	1.5	1.5	base	-1,540.70	-0.966	386	58.5
60	1.5	3	2x	-3,277.70	-1.932	386	58.5
60	1.5	4.5	3x	-5,014.70	-2.573	386	58.5
60	2	1.5	base	-1,021	-0.651	275	48
60	2	3	2x	-2,258.50	-1.475	275	48
60	2	4.5	3x	-3,496	-2.064	275	48
120	1	1.5	base	-1,265.60	-0.679	325	74.7
120	1	3	2x	-2,728.10	-1.432	325	74.7
120	1	4.5	3x	-4,190.60	-1.975	325	74.7
120	1.5	1.5	base	-688	-0.326	210	62.8
120	1.5	3	2x	-1,633	-0.91	210	62.8
120	1.5	4.5	3x	-2,578	-1.374	210	62.8
120	2	1.5	base	-516.4	-0.328	145	50.4
120	2	3	2x	-1,168.90	-0.818	145	50.4
120	2	4.5	3x	-1,821.40	-1.219	145	50.4
250	1	1.5	base	-655.2	-0.32	185	71.4
250	1	3	2x	-1,487.70	-0.79	185	71.4
250	1	4.5	3x	-2,320.20	-1.175	185	71.4
250	1.5	1.5	base	-516.2	-0.34	125	58.5
250	1.5	3	2x	-1,078.70	-0.711	125	58.5
250	1.5	4.5	3x	-1,641.20	-1.026	125	58.5
250	2	1.5	base	-281.1	-0.229	78	45.6
250	2	3	2x	-632.1	-0.566	78	45.6
250	2	4.5	3x	-983.1	-0.856	78	45.6

The gross P&L numbers (pre-cost) are positive for most combinations: the best gross across nine combinations reaches +257 bp (lb=120, thresh=1.5), and the worst gross is +46 bp (lb=250, thresh=1.5). But gross Sharpe peaks at only 0.34 — well below any institutional threshold. The base cost assumption of 1.5 bp/leg overwhelms even the best gross earner: the lb=120/thresh=1.5 combination that earns +257 bp gross finishes at −265 bp net. The binding constraint is therefore **not that the cost assumption is too harsh — it is that the gross signal edge is insufficient to absorb even a modest cost assumption**. Costs make a bad result worse, but removing costs entirely does not rescue the strategy.

5. Cumulative P&L Time-Series (Best Combination: lb=250, Threshold=2.0σ)



The chart tells the core story plainly. Gross P&L drifts modestly positive (+69.9 bp over 25 years — roughly 2.8 bp/year) with several sustained drawdown periods. Once the 1.5 bp/leg base cost is applied, the cumulative line is negative throughout most of the sample and finishes at −281.1 bp. Doubling the cost to 3.0 bp/leg pushes the terminal value to −632.1 bp. The gross line is barely distinguishable from flat — there is not enough signal to make the picture survivable under any realistic cost assumption.

6. Regime-Segmented P&L (lb=250, thresh=2.0)

Regime	Active Days	Gross P&L (bp)	Net P&L — Base (bp)	Net P&L — 3x Cost (bp)	Sharpe (net-base)
cutting	581	9.8	-26.2	-98.2	-0.212
hiking	447	-40.3	-85.3	-175.3	-0.809
hold	1,885	100.4	-12.1	-237.1	-0.04

The on-hold regime is the only one with positive gross P&L (+100.4 bp across 1,885 active days). This is consistent with the intuition that NSS residuals mean-revert better in stable policy environments where the curve shape is not being systematically distorted by rate expectations. In the hiking regime, the gross return is −40.3 bp: when the BoC is actively raising rates, the belly of the curve tends to cheapen persistently because the market reprices expectations rather than reverting to the prior cross-sectional shape. The cutting regime is mildly gross-positive (+9.8 bp) but also net-negative. No regime produces a net-positive result at base cost. The Sharpe under the on-hold regime is only −0.04 — almost zero, but still negative and well short of the minimum 0.5 threshold.

7. Stress Episode Replay

2008 Global Financial Crisis (2007-07-01 to 2009-06-30). The only episode where the lb=250/thresh=2.0 combination produces positive net P&L at base cost, at +7.1 bp net with a Sharpe of 0.07. This is a weak positive — it survives costs but the edge is too thin to ascribe statistical meaning to. At 3× cost the GFC episode turns to −19.9 bp. The reason the strategy survives the GFC at all is that curve dislocations during the crisis tended to be large and eventually revert, giving the 2.0σ threshold trades enough room to resolve positively before exit.

March 2020 (2020-01-01 to 2020-06-30). Uniformly the worst short-episode outcome. The best combination (lb=250, thresh=2.0) produces −13.2 bp gross and −13.2 bp net — in this case active days were few enough that cost drag was minimal, but the gross was still negative. Shorter-lookback combinations fared far worse: lb=120/thresh=2.0 produced −53.8 bp gross and a Sharpe of −4.2. The March 2020 episode violated the core NSS assumption: the belly of the curve did not mean-revert to a normal shape because the BoC was executing emergency cuts and asset purchases that structurally repriced the short and belly simultaneously. The model's cross-sectional fair value became temporarily meaningless.

2022 Hiking Cycle (2022-01-01 to 2023-03-31). The worst sustained episode for the strategy. The lb=250/thresh=2.0 combination produced approximately −83 bp gross and large negative net returns. The hiking cycle produced exactly the condition under which an NSS mean-reversion signal fails: persistent directional moves in term-premium and short-rate expectations that cause belly cheapening to be a rational new equilibrium rather than a transient deviation. The z-score signal repeatedly triggered entries that were never resolved in the expected direction, generating loss on top of loss.

---

8. Where Costs Cap the Conclusion

The break-even cost analysis is unambiguous. Across all nine combinations, the maximum break-even cost per leg is **0.41 bp** (lb=120, thresh=1.5). The minimum is 0.10 bp (lb=60, thresh=1.0). For context, a base assumption of 1.5 bp/leg (which is itself conservative for an interdealer bilateral execution in GoC benchmarks, not unrealistically pessimistic) is **3.6× to 15× above the break-even level** across the strategy variants. You would need to execute this strategy at a cost so low it is not achievable in any realistic trading environment — less than 0.10 to 0.41 bp per leg, on both sides of a three-legged structure, on every entry, exit, and signal rebalance — for the strategy to break even on gross P&L.

The sharpe05_cost_per_leg_bp field is zero for every combination. This means there exists no positive cost level at which the strategy achieves a Sharpe of 0.5. The gross signal edge is simply not large enough. Even at zero cost, gross Sharpe peaks at 0.34. The required gross Sharpe to absorb base costs and still clear 0.5 net would require roughly 2× the gross alpha generated — an alpha that is not present in this data.

---

9. Verdict: Does the Signal Pass the Bar?

No. The signal fails the bar plainly.

The gross signal exists — the 5Y NSS residual has a modest, positive directional tendency when standardised against its own history. But gross Sharpe never exceeds 0.34 across any tested parameter combination, and the break-even cost level (0.10–0.41 bp/leg) is far below what any real execution can achieve. At the base cost assumption of 1.5 bp/leg, every combination produces a negative Sharpe, with cumulative net losses ranging from −123 bp to −2,364 bp depending on parameterisation. Doubling costs to 3.0 bp/leg universally makes the result materially worse, as expected.

The **single assumption that most constrains the conclusion** is the gross signal edge, not the cost level. Even if you could trade at zero cost, you would earn at most 257 bp gross over 25 years — roughly 10 bp/year — which corresponds to a risk-adjusted return that does not clear any institutional hurdle. This is not a strategy that is borderline at base cost and viable at a tighter bid-ask. It is a strategy with insufficient gross alpha at any cost level.

The **regime breakdown confirms the mechanism**: the strategy works best in on-hold regimes when the curve shape is stable, and fails badly in hiking cycles when cheap/rich readings reflect structural repricing rather than transient noise. Since the 2022 hiking cycle is still recent history and the current BoC posture (rate cutting in 2024–2025, stabilising in 2026) is transitional, the signal's best operating environment (on-hold) represents where we may be heading — but the historical evidence on the on-hold regime alone (gross +100 bp, net −12 bp at base) does not clear the bar either.

To a risk committee: this note cannot support a recommendation to put on the 2s5s10s NSS butterfly as a systematic strategy. The individual tenor deviations (5Y +2.1 bp cheap, 7Y −2.9 bp rich) are real NSS readings as of 2026-07-20 and are informative as context. They are not sufficient as a standalone systematic trade signal.

> Generated from AlphaAgent Studio run `0bbfc01e` (conversation 04bb5ca3), data through 2026-07-20. Tables and chart reproduced from the run's registered datasets. For internal review.

Appendix — Backing Data Tables

Butterfly backtest — gross P&L summary

lookback	threshold	cum_gross_pnl_bp	cum_net_base_bp	ann_sharpe_gross	n_trades	active_days
60	1	178.5	-1,293	0.199	565	4,483
60	1.5	196.3	-762.2	0.259	386	3,739
60	2	216.5	-431.5	0.338	275	3,066
120	1	196.9	-649.1	0.211	325	4,773
120	1.5	257	-265	0.319	210	4,013
120	2	136.1	-214.9	0.21	145	3,220
250	1	177.3	-295.2	0.194	185	4,560
250	1.5	46.3	-259.7	0.061	125	3,735
250	2	69.9	-123.6	0.13	78	2,913

Cost-breakeven scan

lookback	threshold	cum_gross_bp	n_trades	breakeven_cost_per_leg_bp	sharpe05_cost_per_leg_bp
60	1	178.5	565	0.105	0
60	1.5	196.3	386	0.17	0
60	2	216.5	275	0.262	0
120	1	196.9	325	0.202	0
120	1.5	257	210	0.408	0
120	2	136.1	145	0.313	0
250	1	177.3	185	0.319	0
250	1.5	46.3	125	0.123	0
250	2	69.9	78	0.299	0

Regime breakdown (all configs)

lookback	threshold	regime	active_days	cum_gross_bp	cum_net_base_bp	cum_net_3x_bp	ann_sharpe_base
60	1	cutting	1,020	-48.7	-345.7	-939.7	-1.266
60	1	hiking	920	228.9	-90.6	-729.6	-0.432
60	1	hold	2,543	-1.7	-856.7	-2,566.70	-1.944
60	1.5	cutting	880	-25.9	-223.9	-619.9	-0.928
60	1.5	hiking	727	150.6	-74.4	-524.4	-0.442
60	1.5	hold	2,132	71.6	-463.9	-1,534.90	-1.28
60	2	cutting	771	-30.2	-174.2	-462.2	-0.806
60	2	hiking	601	132.1	-16.4	-313.4	-0.114
60	2	hold	1,694	114.7	-240.8	-951.8	-0.847
120	1	cutting	1,130	132.3	-61.2	-448.2	-0.212
120	1	hiking	883	10.8	-223.2	-691.2	-1.09
120	1	hold	2,760	53.8	-364.7	-1,201.70	-0.831
120	1.5	cutting	965	102.4	-14.6	-248.6	-0.059
120	1.5	hiking	613	34.4	-114.1	-411.1	-0.732
120	1.5	hold	2,435	120.1	-136.4	-649.4	-0.353
120	2	cutting	811	19.7	-47.8	-182.8	-0.228
120	2	hiking	510	-19.2	-127.2	-343.2	-0.993
120	2	hold	1,899	135.6	-39.9	-390.9	-0.133
250	1	cutting	1,021	85	-0.5	-171.5	-0.002
250	1	hiking	990	-78.3	-222.3	-510.3	-0.964
250	1	hold	2,549	170.7	-72.3	-558.3	-0.179
250	1.5	cutting	773	15.3	-47.7	-173.7	-0.238
250	1.5	hiking	813	-107.5	-202	-391	-1.019
250	1.5	hold	2,149	138.5	-10	-307	-0.029
250	2	cutting	581	9.8	-26.2	-98.2	-0.212
250	2	hiking	447	-40.3	-85.3	-175.3	-0.809
250	2	hold	1,885	100.4	-12.1	-237.1	-0.04

Episode breakdown

episode	lookback	threshold	active_days	cum_gross_pnl_bp	cum_net_base_bp	cum_net_3x_bp	ann_sharpe_net_base
2008_GFC	60	1	303	10.4	-88.6	-286.6	-1.026
2008_GFC	60	1.5	269	7.8	-59.7	-194.7	-0.787
2008_GFC	60	2	226	11.3	-47.2	-164.2	-0.718
2008_GFC	120	1	384	-34.8	-129.3	-318.3	-1.189
2008_GFC	120	1.5	284	-8.6	-53.6	-143.6	-0.671
2008_GFC	120	2	209	-9.6	-45.6	-117.6	-0.806
2008_GFC	250	1	451	-59.7	-91.2	-154.2	-0.728
2008_GFC	250	1.5	439	-47.6	-74.6	-128.6	-0.61
2008_GFC	250	2	357	20.6	7.1	-19.9	0.073
Mar_2020	60	1	112	-35	-66.5	-129.5	-3.474
Mar_2020	60	1.5	93	-36.4	-63.4	-117.4	-3.914
Mar_2020	60	2	85	4.4	-13.6	-49.6	-1.269
Mar_2020	120	1	91	-33.7	-56.2	-101.2	-3.476
Mar_2020	120	1.5	86	-41.2	-54.7	-81.7	-3.583
Mar_2020	120	2	76	-53.8	-58.3	-67.3	-4.206
Mar_2020	250	1	112	-33.5	-60.5	-114.5	-3.27
Mar_2020	250	1.5	88	-42.5	-51.5	-69.5	-3.339
Mar_2020	250	2	78	-13.2	-13.2	-13.2	-1.396
2022_Hiking	60	1	220	13.2	-76.8	-256.8	-1.226
2022_Hiking	60	1.5	96	-64.2	-113.7	-212.7	-3.538
2022_Hiking	60	2	86	-56.5	-97	-178	-3.205
2022_Hiking	120	1	145	-64.7	-114.2	-213.2	-2.783
2022_Hiking	120	1.5	104	-56.4	-96.9	-177.9	-3.039
2022_Hiking	120	2	90	-79.2	-106.2	-160.2	-4.543
2022_Hiking	250	1	188	-91.5	-127.5	-199.5	-2.695
2022_Hiking	250	1.5	153	-114.3	-132.3	-168.3	-3.555
2022_Hiking	250	2	141	-82.6	-96.1	-123.1	-2.771

NSS parameter drift

parameter	mean_abs_daily_change	p95_abs_daily_change	p99_abs_daily_change	std_daily_change
b0	16.38	35.596	137.946	383.831
b1	580.503	53.967	201.241	21,208.33
b2	4,300.08	1,898.87	17,792.23	163,602.12
b3	4,750.70	835.552	11,036.49	233,003.38
tau1	669,326.58	2,168,501.40	8,347,163.47	9,333,767.54
tau2	2,656.15	7,843.73	41,632.52	29,789.27

AlphaAgent Studio — automated analysis output. Shared for colleague review; figures should be independently verified before any use.