Shift-Aware Gaussian-Supremum Validation for Wasserstein-DRO CVaR Portfolios

Derek Long
Department of Industrial Engineering and Operations Research
Columbia University
New York, NY 10027
d.long@columbia.edu

Abstract

We study portfolio selection with a Conditional Value-at-Risk (CVaR) constraint under distribution shift and serial dependence. While Wasserstein distributionally robust optimization (DRO) offers tractable protection via an ambiguity ball around empirical data, choosing the ball radius is delicate: large radii are conservative, small radii risk violation under regime change. We propose a shift-aware Gaussian-supremum (GS) validation framework for Wasserstein-DRO CVaR portfolios, building on the work by Lam and Qian (2019). Phase I of the framework generates a candidate path by solving the exact reformulation of the robust CVaR constraint over a grid of Wasserstein radii. Phase II of the framework learns a target deployment law Q by density-ratio reweighting of a time-ordered validation fold, computes weighted CVaR estimates, and calibrates a simultaneous upper confidence band via a block multiplier bootstrap to account for dependence. We select the least conservative feasible portfolio (or abstain if the effective sample size collapses). Theoretically, we extend the normalized GS validator to non-i.i.d. financial data: under weak dependence and regularity of the weighted scores, any portfolio passing our validator satisfies the CVaR limit under Q with probability at least $1 - \beta$; the Wasserstein term contributes a deterministic margin $(\delta/\alpha)\|x\|_*$. Empirical results indicate improved return–risk trade-offs versus the naive baseline.

1 Introduction

Portfolio optimization under uncertainty requires robust risk management, especially for extreme losses. A common risk metric is the Conditional Value-at-Risk (CVaR), which measures the average loss in the worst α fraction of cases. CVaR constraints are used to cap tail risk and prevent systemic crises by ensuring portfolio losses remain below a threshold with high confidence. In practice, data-driven portfolio selection with a CVaR constraint faces significant challenges: the true distribution of returns is unknown and may shift over time (regime shifts), and asset returns are often dependent (serially correlated and cross-correlated), violating i.i.d. assumptions. Classical robust optimization or distributionally robust optimization (DRO) approaches account for some distributional uncertainty by considering worst-case distributions in an ambiguity set (e.g. a Wasserstein ball around the empirical distribution). However, these methods can be sensitive to unanticipated distribution shifts: if market conditions change (e.g. a volatility regime change), a previously feasible portfolio may violate the CVaR limit. Moreover, dependent data reduces effective sample size and can lead to mis-estimated risk. This motivates the need for a shift-aware validation procedure that guarantees CVaR constraints under distribution shifts and dependence.

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: Generative AI in Finance.

We focus on the CVaR-constrained portfolio problem with d assets. Let $x \in \mathbb{R}^d$ be allocations, $\xi \in \mathbb{R}^d$ asset returns, F the (unknown) data-generating law, and $c = -\mathbb{E}_F[\xi] \in \mathbb{R}^d$ (expected losses). The classical formulation is

$$\begin{aligned} \min_x \quad & c^\top x \\ \text{s.t.} \quad & \inf_{t\in\mathbb{R}} \left\{ t + \tfrac{1}{\alpha}\mathbb{E}_F[\max\{-\xi^\top x - t,\, 0\}] \right\} \le \gamma, \\ & \sum_{l=1}^{d} x_l = 1, \qquad x_l \ge 0 \; \forall l \in \{1,\dots,d\}. \end{aligned} \tag{1}$$

To guard against distributional uncertainty, we adopt a Wasserstein DRO version centered at an empirical law $\hat{F}$ with radius δ, so that the robust CVaR constraint becomes

$$\sup_{Q\in\mathcal{B}(\hat{F},\delta)} \inf_{t\in\mathbb{R}} \left\{ t + \tfrac{1}{\alpha}\mathbb{E}_Q[\max\{-\xi^\top x - t,\, 0\}] \right\} \le \gamma. \tag{2}$$

For W_1 ambiguity balls, the robust CVaR constraint admits an exact, tractable reformulation (cf. Cor. 5.1 in [9]):

$$Z = \left\{ x \in \mathbb{R}^d : \begin{array}{l} \delta v + \frac{1}{n}\sum_{i=1}^n z_i \le \alpha r, \\ r \le z_i + \gamma + \xi_i^\top x \;\text{ for } i = 1,\dots,n, \\ \|x\|_* \le v, \\ v \ge 0,\;\; r \ge 0,\;\; z_i \ge 0 \;\text{ for } i = 1,\dots,n, \end{array} \right\}, \tag{3}$$

where $\|\cdot\|_*$ is the dual norm tied to the ground metric. While (3) is computationally convenient, selecting δ is delicate: too small jeopardizes feasibility under shift; too large kills performance. Our contribution is a data-driven selection via a *shift-aware* GS validator that yields feasibility guarantees under serial dependence.

2 Related Work and Contributions

Wasserstein DRO delivers tractable, statistically meaningful robustness [9, 6], with applications to finance and CVaR portfolios [2, 9, 11]. However, calibrating the radius can be challenging and often leads to overly conservative solutions. To avoid these overly conservative solutions, Lam and Qian [8] develop a normalized Gaussian-supremum validator (among other validators) for data-driven optimization with uncertain constraints. Instead of fixing a very conservative reformulation (large safety margins), their approach generates a solution path of candidate portfolios indexed by a conservativeness parameter (e.g., different values of ρ in the DRO formulation). They then evaluate these candidates on a holdout dataset to find the least conservative solution that still meets the constraint with high confidence $(1-\beta)$. The GS validator computes a simultaneous confidence bound across the entire solution path by estimating the quantile of the supremum of Gaussian deviations.

In machine learning, covariate shift refers to the scenario where the input distribution changes between training and testing phases, while the conditional relationship (e.g., returns given factors) remains stable. A rich body of work addresses this via importance weighting techniques: direct ratio estimation (KLIEP, uLSIF) [13, 5], kernel mean matching (KMM) [4], and classifier-based ratios [1]; these ideas enable targeting a *shifted* law in validation. Furthermore, standard statistical inference assumes independent data, but financial time series exhibit autocorrelation, volatility clustering, and other dependencies. To overcome this, time-series inference under dependence uses block bootstraps [7, 10] and multiplier/wild bootstraps for dependent data [12, 3]. Our validation builds on the *Gaussian-supremum validator* of Lam & Qian [8]; we extend it to non-i.i.d. financial data with shift-aware weighting and BMB calibration.

Contributions. *(i)* A **shift-aware GS validator**: we reweight validation data by estimated density ratios to evaluate CVaR under a learned center Q, then perform Wasserstein DRO around Q with a smaller radius. *(ii)* A **dependence-robust calibration**: we replace i.i.d. Gaussian calibration by a block multiplier bootstrap for the GS band, yielding correct joint coverage under serial dependence. *(iii)* **Finite-sample feasibility**: we prove that any selected portfolio meets the CVaR limit under Q with probability $\ge 1-\beta$; proofs follow GS arguments with extensions for weighting and dependence.

3 Methodology and Main Results

We adopt a two-phase procedure. Phase I generates a candidate path $\{x^*(\delta_j)\}_{j=1}^p$ by solving (3) over a grid of radii $\delta_1 < \cdots < \delta_p$. Phase II validates feasibility on a holdout, time-ordered sample $\{\xi_t\}_{t=1}^{n_2}$ potentially from a shifted regime. The full algorithm is detailed in Appendix A.

Shift-aware center via density ratios. Let P denote the holdout law and Q the target law for deployment (current regime). We estimate $w(\xi) = q(\xi)/p(\xi)$ using a probabilistic classifier that distinguishes recent (proxy for Q) vs. holdout (proxy for P) samples, convert scores to odds, *clip*, and normalize: $\tilde{w}_t = w_t / \sum_i w_i$. For any candidate x, we estimate the CVaR functional's expectation under Q by the weighted mean:

$$\widehat{\Psi}_Q(x) \;=\; \min_{t\in\mathbb{R}} \left\{ t + \tfrac{1}{\alpha} \sum_{i=1}^{n_2} \tilde{w}_i \, \max\{-\xi_i^\top x - t,\, 0\} \right\}.$$

Effective sample size is $n_{\text{eff}} = 1/\sum_i \tilde{w}_i^2$; we abstain if n_{eff} collapses.

Dependence-robust GS calibration. For each candidate $x_j = x^*(\delta_j)$ with empirical minimizer $\hat{t}_j$, define scores $\phi_{ij} = \hat{t}_j + \frac{1}{\alpha}\max\{-\xi_i^\top x_j - \hat{t}_j, 0\}$ and weighted means $\widehat{H}_j = \sum_i \tilde{w}_i \, \phi_{ij}$. We form centered *block* sums $R_{kj} = \sum_{i\in\mathcal{B}_k} \tilde{w}_i(\phi_{ij} - \widehat{H}_j)$ over contiguous $\{B_k\}_{k=1}^K$ of length b (with $b \to \infty$, $b/n_2 \to 0$). Drawing i.i.d. Gaussian multipliers ε_k, the bootstrap max statistic is

$$T^* = \max_{1\le j\le p} \frac{1}{\sqrt{n_{\text{eff}}}} \frac{\sum_{k=1}^K \varepsilon_k \, R_{kj}}{\widehat{\sigma}_j}, \quad \widehat{\sigma}_j^2 = \sum_i \tilde{w}_i \, (\phi_{ij} - \widehat{H}_j)^2.$$

Let $\widehat{q}_{1-\beta}$ be the $(1-\beta)$-quantile of $\{T^{*(b)}\}_{b=1}^M$. We build a *simultaneous upper band* for the robust constraint:

$$U_j \;=\; \underbrace{\widehat{H}_j}_{\text{empirical (weighted)}} \;+\; \underbrace{\tfrac{\delta_j}{\alpha}\,\|x_j\|_*}_{\text{DRO margin}} + \widehat{q}_{1-\beta}\,\frac{\widehat{\sigma}_j}{\sqrt{n_{\text{eff}}}},$$

and declare x_j feasible if $U_j \le \gamma$. We select the *least conservative feasible* radius $\delta^* = \min\{\delta_j : U_j \le \gamma\}$; ties break by objective $c^\top x_j$. Proofs for the following theorems can be found in Appendix B — LLMs were used to assist in constructing these proofs.

Theorem 1 (Feasibility under shift- and dependence-aware GS)**.** *Suppose (A1)–(A4) (see Appendix B) and a fixed finite candidate set $\{x_j\}_{j=1}^p$. Let $\hat{q}_{1-\beta}$ be computed by Step 4 with block size $b \asymp n_2^{1/3}$ and $B \to \infty$. Then, with probability at least $1-\beta-o(1)$, any pair $(x_j, \delta^*(x_j))$ produced by Step 5 satisfies*

$$\sup_{F\in\mathcal{B}(Q,\delta^*(x_j))} \mathrm{CVaR}_\alpha(-\xi^\top x_j) \;\le\; \gamma.$$

In particular, the selected $(x^\star, \delta^\star)$ is feasible w.p. $\ge 1-\beta-o(1)$.

Theorem 2 (Validation-driven suboptimality bound)**.** *Let $(x_Q^\dagger, \delta_Q^\dagger)$ be any feasible solution to the Q-robust problem $\min\{c^\top x : \sup_{F\in\mathcal{B}(Q,\delta)} \mathrm{CVaR}_\alpha(-\xi^\top x) \le \gamma,\ \delta \in [\delta_{\min}, \delta_{\max}],\ x \in \mathcal{X}\}$, with $\mathcal{X} = \{x : \mathbf{1}^\top x = 1,\ x \ge 0\}$. Under (A1)–(A4) and convexity of $c^\top x$, the output $(x^\star, \delta^\star)$ of Algorithm 1 satisfies, with probability $1-\beta-o(1)$,*

$$c^\top x^\star - c^\top x_Q^\dagger \;\le\; \underbrace{\inf_j \left(c^\top x_j - c^\top x_Q^\dagger\right)}_{\textit{approximation (menu) error}} \;+\; \underbrace{L_c\,\|x^\star - x_j\|_2}_{\textit{optimization gap for chosen candidate}}$$

for any candidate x_j; in particular, if the menu contains $x_Q^\dagger$ (or approximates it to ε in ℓ_2), the excess cost is $O_p(\varepsilon)$.

4 Numerical Experiments

We evaluate three validators on the CVaR-constrained portfolio problem: NEW (our shift- & dependence-aware GS with block multipliers and analytical radius), OLD-NGS (Lam–Qian normalized GS assuming i.i.d.), and IW-CV (importance-weighted cross-validation baseline). In all

Table 1: Scenario 1 (no shift) results.

Method	Feas. (↑)	Obj. (↓)	CVaR	LHS	δ	Runtime (s)
NEW	0.91	-1.05×10^{-3}	0.031	0.059	0.008	11.1
IW-CV	0.90	-1.07×10^{-3}	0.030	0.060	0.009	**186.0**
OLD-NGS	0.86	-1.27×10^{-3}	0.023	0.086	0.015	10.7

Table 2: Scenario 2 (shift) results.

Method	Feas. (↑)	Obj. (↓)	CVaR	LHS	δ	Runtime (s)
NEW	0.90	-0.95×10^{-3}	0.046	0.092	0.010	11.2
IW-CV	0.90	-0.96×10^{-3}	0.047	0.093	0.010	**190.1**
OLD-NGS	0.28	-0.72×10^{-3}	0.041	0.104	0.016	10.8

our experiments, we set our confidence level $1-\beta = 0.9$. Consider the problem with $d = 8$ assets with AR(1) returns and two scenarios: (1) No shift: training/validation/test all follow P with autoregression $\phi = 0.3$ and covariance Σ, and (2) Shift: test follows Q with mean deterioration ($\mu_Q = \mu_P - \Delta$), volatility inflation ($\Sigma_Q = 1.7^2\Sigma$), and stronger persistence $\phi_Q = 0.45$. Sample sizes are $(n_{\text{train}}, n_{\text{val}}, n_{\text{test}}) = (1000, 1200, 15000)$. We set $\alpha = 0.05$ and fix the risk budget γ *ex ante* from the equal-weight portfolio's empirical CVaR_α on Scenario 1 training, multiplied by a 10% margin; the same γ is used in both scenarios for all methods. NEW uses logistic density-ratio weights on validation (clipped odds), block length $b \asymp n_{\text{val}}^{1/3}$, $B = 800$ multipliers, and the analytical radius $\delta^*(x) = \alpha\,[\gamma - \widehat{H}_w(x) - \hat{q}\,\hat{\sigma}_w(x)/\sqrt{n_{\text{eff}}}]_+/\|x\|_2$ (clipped to $[10^{-3}, 2\times10^{-2}]$). Candidate portfolios $x(\delta)$ are generated by solving the weighted Wasserstein–CVaR reformulation with $\|\cdot\|_2$ (Clarabel), using blended training weights; we also include a small random Dirichlet menu for diversity. OLD-NGS is identical except it sets $w_i \equiv 1/n_{\text{val}}$ and $b = 1$. IW-CV uses the same weights as NEW but selects δ by K-fold cross-validation over a grid (here K=5), making it notably slower. We repeat each scenario for $R = 1000$ Monte Carlo replications and report: *Feas.* = fraction of replications where the test robust constraint (worst-case CVaR with selected δ) is $\leq \gamma$; *Cost* = $c^\top x$ (lower is better); test CVaR_α (*CVaR*), the robust left-hand side *LHS* $= \mathrm{CVaR}_\alpha + \frac{\delta}{\alpha}\|x\|_2$; selected δ; and median wall-clock *runtime* per replication (s).

In Scenario 1, NEW attains the target feasibility $1-\beta$ with cost comparable to IW-CV, while running an order of magnitude faster (no K-fold search). OLD-NGS under-covers due to serial dependence, often selecting larger radii that inflate the robust LHS yet still miss the target. In Scenario 2, the i.i.d. band becomes substantially miscalibrated and feasibility collapses, whereas NEW maintains coverage by combining density-ratio weighting (shift-aware centering/variance) with block multipliers (dependence). Across both scenarios, IW-CV performs similarly to NEW in feasibility/cost but slower from repeated model re-fits across folds; it could also over-regularize when folds experience different volatility levels, whereas NEW uses a single weighted, dependence-aware calibration that aligns more closely with the one-step deployment regime.

5 Conclusion

We developed a validation framework for CVaR-constrained portfolio selection that is simultaneously *shift-aware* and *dependence-robust*. By re-centering Wasserstein-DRO around a learned target distribution Q via density-ratio reweighting and calibrating feasibility with a Gaussian-supremum band using block multiplier bootstrap, the method selects the least conservative portfolio that still satisfies the CVaR limit with confidence $(1-\beta)$. Our guarantees hinge on absolute-regularity dependence, accurate density-ratio weights, and a finite candidate menu; in practice the procedure can be sensitive to norm scaling, block length, and heavy tails, and abstention/clipping choices may reduce coverage or power under severe shift. Future work includes broader numerical studies on real financial data (e.g., multi-asset equity/rates/FX portfolios across calm and crisis regimes) with transaction costs and liquidity/short-sale constraints, head-to-head comparisons against strong DRO baselines, and methodological extensions to heavy-tailed models, alternative ambiguity sets, and adaptive radius/portfolio selection.

References

[1] S. Bickel, M. Brückner, and T. Scheffer. Discriminative learning under covariate shift. *Journal of Machine Learning Research*, 10(75):2137–2155, 2009. URL `http://jmlr.org/papers/v10/bickel09a.html`.

[2] J. Blanchet, L. Chen, and X. Y. Zhou. Distributionally robust mean–variance portfolio selection with wasserstein distances. *Management Science*, 68(9):6382–6410, 2022. doi: 10.1287/mnsc.2021.4155.

[3] A. Bücher and I. Kojadinovic. Dependent multiplier bootstrap for non-degenerate u-statistics under mixing conditions. *Bernoulli*, 22(2):927–968, 2016. doi: 10.3150/14-BEJ675.

[4] J. Huang, A. Gretton, K. M. Borgwardt, B. Schölkopf, and A. J. Smola. Correcting sample selection bias by unlabeled data. In *Advances in Neural Information Processing Systems*, volume 19, pages 601–608, 2007.

[5] T. Kanamori, S. Hido, and M. Sugiyama. A least-squares approach to direct importance estimation. *Journal of Machine Learning Research*, 10:1391–1445, 2009.

[6] D. Kuhn, P. Mohajerin Esfahani, V. A. Nguyen, and S. Shafieezadeh-Abadeh. Wasserstein distributionally robust optimization: Theory and applications in machine learning. arXiv:1908.08729, 2019. URL `https://arxiv.org/abs/1908.08729`.

[7] H. R. Künsch. The jackknife and the bootstrap for general stationary observations. *The Annals of Statistics*, 17(3):1217–1241, 1989. doi: 10.1214/aos/1176347265.

[8] H. Lam and H. Qian. Combating conservativeness in data-driven optimization under uncertainty: A solution path approach. *arXiv preprint arXiv:1909.06477*, 2019. URL `https://arxiv.org/abs/1909.06477`.

[9] P. Mohajerin Esfahani and D. Kuhn. Data-driven distributionally robust optimization using the wasserstein metric: Performance guarantees and tractable reformulations. *Mathematical Programming*, 171(1–2):115–166, 2018. doi: 10.1007/s10107-017-1172-1.

[10] D. N. Politis and J. P. Romano. The stationary bootstrap. *Journal of the American Statistical Association*, 89(428):1303–1313, 1994. doi: 10.1080/01621459.1994.10476870.

[11] C. S. Pun, T. Wang, and Z. Yan. Data-driven distributionally robust CVaR portfolio optimization under a regime-switching ambiguity set. *Manufacturing & Service Operations Management*, 25 (5):1779–1795, 2023. doi: 10.1287/msom.2023.1229. URL `https://pubsonline.informs.org/doi/10.1287/msom.2023.1229`.

[12] X. Shao. The dependent wild bootstrap. *Journal of the American Statistical Association*, 105 (489):218–235, 2010. doi: 10.1198/jasa.2009.tm08651.

[13] M. Sugiyama, T. Suzuki, and T. Kanamori. *Density Ratio Estimation in Machine Learning*. Cambridge University Press, 2012. doi: 10.1017/CBO9781139035613.

Appendix A

Algorithm 1 Shift-aware GS validation with block multipliers for Wasserstein–CVaR portfolios

Require: Training $\{\xi_t\}_{t=1}^{n_1}$, validation $\{\xi_{n_1+1}, \dots, \xi_{n_1+n_2}\}$, level α, miscoverage β, budget γ, norm dual $\|\cdot\|_*$, block length b, bootstrap reps B, recent fraction m/n_2.

1: **Shift-aware validation weights.** Split validation into early/late: $D_{\text{early}} = \{\xi_{1:n_2-m}\}$, $D_{\text{late}} = \{\xi_{n_2-m+1:n_2}\}$. Fit a probabilistic classifier, set $w_i \propto \frac{p(\text{late}|\xi_i)}{1-p(\text{late}|\xi_i)}$, normalize $\sum_i w_i = 1$, and compute $n_{\text{eff}} = (\sum_i w_i^2)^{-1}$.

2: **Candidate generation.** For each δ on a grid (or skip if using the analytical $\delta^\star$), solve the weighted Wasserstein–CVaR reformulation to get candidates $\{x_\ell\}_{\ell=1}^L$; let $c = -\sum_i \tilde{w}_i \xi_i$ on training.

3: **Validation CVaR map.** For each candidate x_j, compute $t_w(x_j) \in \arg\min_t \sum_i w_i \phi_{x_j}(\xi_i; t)$, $\hat{H}_w(x_j) = \sum_i w_i \phi_{x_j}(\xi_i; t_w(x_j))$, and $\hat{\sigma}_w^2(x_j) = \sum_i w_i \{\phi_{x_j}(\xi_i; t_w(x_j)) - \hat{H}_w(x_j)\}^2$.

4: **Block multiplier GS (with block sums).** Partition validation indices into $K = \lfloor n_2/b \rfloor$ contiguous blocks $B_1, \dots, B_K$ (discard remainder). For each (k, j) define the *block sum*

$$S_{kj} := \sum_{i \in B_k} w_i \big\{\phi_{x_j}(\xi_i; t_w(x_j)) - \hat{H}_w(x_j)\big\}.$$

Draw i.i.d. multipliers $\varepsilon^{(r)} \sim \mathcal{N}(0, I_K)$, $r = 1, \dots, B$, and compute

$$T^{(r)} := \max_{1 \le j \le L} \frac{\sum_{k=1}^K \varepsilon_k^{(r)} S_{kj}}{\hat{\sigma}_w(x_j)\sqrt{n_{\text{eff}}}}, \qquad \hat{q}_{1-\beta} := (1-\beta)\text{-quantile of } \{T^{(r)}\}_{r=1}^B.$$

5: **Analytical radius & feasibility filter.** For each j, set

$$\delta^\star(x_j) := \text{clip}\left(\frac{\alpha\,[\,\gamma - \hat{H}_w(x_j) - \hat{q}_{1-\beta}\,\hat{\sigma}_w(x_j)/\sqrt{n_{\text{eff}}}\,]_+}{\|x_j\|_*}\ ;\ \delta_{\min}, \delta_{\max} \right),$$

and the validated upper bound

$$U(x_j) := \hat{H}_w(x_j) + \frac{\hat{q}_{1-\beta}\,\hat{\sigma}_w(x_j)}{\sqrt{n_{\text{eff}}}} + \frac{\delta^\star(x_j)}{\alpha}\,\|x_j\|_*.$$

Define the validated set $\mathcal{F} := \{\, j :\ U(x_j) \le \gamma \,\}$. **If** $\mathcal{F} = \emptyset$, abstain. Note that,

$$U(x) = \max\left\{\gamma,\ \hat{H}_w(x) + \frac{\hat{q}_{1-\beta}\,\hat{\sigma}_w(x)}{\sqrt{n_{\text{eff}}}}\right\}.$$

Hence $U(x) \le \gamma$ iff $\hat{H}_w(x) + \hat{q}_{1-\beta}\hat{\sigma}_w(x)/\sqrt{n_{\text{eff}}} \le \gamma$, and among validated candidates $U(x) \equiv \gamma$.

6: **Selection among validated candidates.** Return the least-conservative feasible solution $x^\star \in \arg\min_{j \in \mathcal{F}} \delta^\star(x_j)$.

Notation. For $x \in \mathbb{R}^d$ and return $\xi \in \mathbb{R}^d$ define the CVaR map

$$\phi_x(\xi; t) \ :=\ t + \frac{1}{\alpha}\big(-\xi^\top x - t\big)_+, \qquad \text{and} \quad \Phi_x(\xi) \ :=\ \inf_{t \in \mathbb{R}} \phi_x(\xi; t).$$

On a weighted sample $\{(\xi_i, w_i)\}_{i=1}^{n_2}$ with $w_i \ge 0$, $\sum_i w_i = 1$, let

$$t_w(x) \in \arg\min_{t \in \mathbb{R}} \sum_{i=1}^{n_2} w_i\, \phi_x(\xi_i; t), \qquad \widehat{H}_w(x) \ :=\ \sum_{i=1}^{n_2} w_i\, \phi_x(\xi_i; t_w(x)).$$

Let $\hat{\sigma}_w^2(x) := \sum_i w_i \big(\phi_x(\xi_i; t_w(x)) - \widehat{H}_w(x)\big)^2$ and $n_{\text{eff}} := 1/\sum_i w_i^2$. We take the dual norm to be $\|\cdot\|_* = \|\cdot\|_2$.

Baselines. (i) *OLD NGS (i.i.d.).* Replace $w_i \equiv 1/n_2$, set $b = 1$, and compute $\hat{q}_{1-\beta}$ from the (unnormalized) Gaussian supremum of the empirical covariance as in [8].
(ii) *IW plug-in.* Use w_i as in Step 1, skip Step 4 (set $\hat{q}_{1-\beta} = 0$), and select $\delta^*(x)$ as in Step 5 with $\hat{\sigma}_w \equiv 0$.

Remark. When $w_i \equiv 1/n_2$ and $b = 1$, Algorithm 1 reduces to the normalized GS validator of [8]; when $b > 1$ but $w_i \equiv 1/n_2$ we obtain a serial-dependence robust variant; with general w and $b > 1$ we obtain the proposed shift- and dependence-aware validator.

Appendix B

We sketch the key ingredients and then give full proofs.

B.1 Assumptions

(A1) Dependence. We assume the stationary sequence (ξ_t) is *absolutely regular* (also known as beta-mixing). Let $b(k)$ denote the absolute-regularity mixing coefficients, defined by

$$b(k) := \sup_{n\in\mathbb{Z}} \mathbb{E}\Big[\sup_{B\in\mathcal{F}_{n+k}^{\infty}} \big|\mathbb{P}(B \mid \mathcal{F}_{-\infty}^{n}) - \mathbb{P}(B)\big|\Big].$$

We require the summability condition $\sum_{k\geq 1} b(k)^{\delta/(2+\delta)} < \infty$ for some $\delta > 0$.

(A2) Moments. $\sup_{\|x\|_2\leq 1} \mathbb{E}\big[|\xi^\top x|^{2+\rho}\big] < \infty$.

(A3) Density ratio regularity. The shift-aware weights w_i satisfy $w_i \geq 0$, $\sum_i w_i = 1$, and there exist constants $0 < c_w \leq C_w < \infty$ such that $c_w/n_2 \leq w_i \leq C_w/n_2$ for all i (this holds if the logistic odds are clipped).

(A4) Lipschitz loss. For any $x, x' \in \mathbb{R}^d$ and any $t \in \mathbb{R}$, $\big|\phi_x(\xi;t) - \phi_{x'}(\xi;t)\big| \leq \frac{\|\xi\|_2}{\alpha}\,\|x - x'\|_2$.

B.2 Technical lemmas

Lemma 1 (Weighted CVaR Lipschitzness)**.** *Let $t_w(x) \in \arg\min_t \sum_i w_i\,\phi_x(\xi_i;t)$. Under (A4), $x \mapsto \widehat{H}_w(x) = \sum_i w_i\,\phi_x(\xi_i;t_w(x))$ is $(\mathbb{E}_w\|\xi\|_2)/\alpha$-Lipschitz, where $\mathbb{E}_w$ denotes expectation w.r.t. the empirical measure with weights w.*

Proof. For any x, x', by optimality of $t_w(x)$ and $t_w(x')$ and the Lipschitz property of $\phi_x(\cdot;t)$ in x we have $\widehat{H}_w(x) - \widehat{H}_w(x') \leq \sum_i w_i\big(\phi_x(\xi_i;t_w(x')) - \phi_{x'}(\xi_i;t_w(x'))\big) \leq \frac{\|x-x'\|_2}{\alpha}\sum_i w_i\|\xi_i\|_2$. Symmetrizing yields the claim. □

Block statistic. For block k with index set B_k, define

$$S_{kj} := \sum_{i\in B_k} w_i\big\{\phi_{x_j}(\xi_i;t_w(x_j)) - \hat{H}_w(x_j)\big\}.$$

Lemma 2 (Weighted CLT with block multipliers)**.** *Under* (A1)–(A3)*, for any fixed finite candidate set $\{x_j\}_{j=1}^{p}$,*

$$\max_{1\leq j\leq p} \frac{\sqrt{n_{\text{eff}}}\,\big\{\hat{H}_w(x_j) - \mathbb{E}_w[\Phi_{x_j}(\xi)]\big\}}{\hat{\sigma}_w(x_j)} \;\Rightarrow\; \max_{1\leq j\leq p} Z_j,$$

where $(Z_1,\ldots,Z_p)$ is mean-zero Gaussian with covariance induced by the long-run variance of $\phi_{x_j}(\xi;t_w(x_j))$. Moreover, the block multiplier statistic

$$T^{(r)} := \max_{1\leq j\leq p} \frac{\sum_{k=1}^{K} \varepsilon_k^{(r)}\,S_{kj}}{\hat{\sigma}_w(x_j)\sqrt{n_{\text{eff}}}}, \qquad \varepsilon^{(r)} \sim \mathcal{N}(0, I_K),$$

with block size $b \asymp n_2^{1/3}$ consistently estimates the law of $\max_j Z_j$, i.e., $\hat{q}_{1-\beta} \to q_{1-\beta}$ in probability.

Proof. This follows from empirical process theory for beta-mixing arrays and multiplier bootstrap consistency (e.g., Theorem 3.1 in block-multiplier CLTs) applied to the finite class $\{\phi_{x_j}(\cdot; t_w(x_j))\}_j$; (A3) ensures Lindeberg conditions under weighting and provides n_{eff}; (A2) yields finite long-run variance. Normalization by $\hat{\sigma}_w(x_j)$ yields a self-normalized process; consistency of the block multiplier quantile for the normalized supremum follows by standard arguments (continuous mapping plus anti-concentration for Gaussian maxima). □

Remark. The bootstrap calibrates the *self-normalized* supremum using the same denominator $\widehat{\sigma}_w(x_j)$ in both the original and bootstrap worlds; under beta-mixing this yields consistent critical values for the normalized process.

Lemma 3 (Wasserstein Lipschitz bound for CVaR)**.** *Let $\|\cdot\|_2$ be the ground norm and $B(Q,\delta)$ the W_1 ball of radius δ around Q. For any $x \in \mathbb{R}^d$ and $\alpha \in (0,1)$,*

$$\sup_{F \in B(Q,\delta)} \mathrm{CVaR}_\alpha(-\xi^\top x) \;\le\; \mathrm{CVaR}_\alpha^Q(-\xi^\top x) + \frac{\delta}{\alpha}\,\|x\|_2.$$

Proof. Fix $t \in \mathbb{R}$ and set $f_t(\xi) = (-\xi^\top x - t)_+$. Then f_t is $\|x\|_2$-Lipschitz w.r.t. $\|\cdot\|_2$. By Kantorovich–Rubinstein duality,

$$\sup_{F \in B(Q,\delta)} \mathbb{E}_F[f_t(\xi)] \;\le\; \mathbb{E}_Q[f_t(\xi)] + \delta\,\|x\|_2.$$

Using Rockafellar–Uryasev $\mathrm{CVaR}_\alpha(L) = \inf_t\{t + \frac{1}{\alpha}\mathbb{E}(L-t)_+\}$, with $L = -\xi^\top x$, we get

$$\sup_{F \in B(Q,\delta)} \mathrm{CVaR}_\alpha(-\xi^\top x) \le \inf_t \left\{t + \tfrac{1}{\alpha}\mathbb{E}_Q[(-\xi^\top x - t)_+]\right\} + \tfrac{\delta}{\alpha}\|x\|_2 = \mathrm{CVaR}_\alpha^Q(-\xi^\top x) + \tfrac{\delta}{\alpha}\|x\|_2.$$

□

B.3 Feasibility guarantee (proof of Theorem 1)

Theorem (Feasibility under shift- and dependence-aware GS)**.** *Suppose (A1)–(A4) and a fixed finite candidate set $\{x_j\}_{j=1}^p$. Let $\hat{q}_{1-\beta}$ be computed by Step 4 with block size $b \asymp n_2^{1/3}$ and $B \to \infty$. Then, with probability at least $1-\beta-o(1)$, any pair $(x_j, \delta^*(x_j))$ produced by Step 5 satisfies*

$$\sup_{F \in \mathcal{B}(Q,\delta^*(x_j))} \mathrm{CVaR}_\alpha(-\xi^\top x_j) \;\le\; \gamma.$$

In particular, the selected $(x^\star, \delta^\star)$ is feasible w.p. $\ge 1-\beta-o(1)$.

Proof. Fix j. By Lemma 2, with probability $\ge 1-\beta-o(1)$ we have $\mathbb{E}_Q[\Phi_{x_j}(\xi)] \le \widehat{H}_w(x_j) + \hat{q}_{1-\beta}\,\hat{\sigma}_w(x_j)/\sqrt{n_{\text{eff}}}$. By Lemma 3,

$$\sup_{F \in \mathcal{B}(Q,\delta)} \mathrm{CVaR}_\alpha(-\xi^\top x_j) \;\le\; \widehat{H}_w(x_j) + \frac{\hat{q}_{1-\beta}\,\hat{\sigma}_w(x_j)}{\sqrt{n_{\text{eff}}}} + \frac{\delta}{\alpha}\|x_j\|_2.$$

Setting $\delta = \delta^*(x_j)$ from Step 5 ensures the right-hand side is $\le \gamma$ (up to the numerical clipping which only reduces δ). This holds simultaneously over $j = 1, \ldots, p$ by the union over a *max* statistic already controlled by $\hat{q}_{1-\beta}$. Selecting $(x^\star, \delta^\star)$ by minimizing the validated upper bound preserves feasibility. □

B.4 Performance guarantee (proof of Theorem 2)

Theorem (Validation-driven suboptimality bound)**.** *Let $(x_Q^\dagger, \delta_Q^\dagger)$ be any feasible solution to the Q-robust problem $\min\{c^\top x : \sup_{F \in \mathcal{B}(Q,\delta)} \mathrm{CVaR}_\alpha(-\xi^\top x) \le \gamma,\ \delta \in [\delta_{\min}, \delta_{\max}],\ x \in \mathcal{X}\}$, with $\mathcal{X} = \{x : \mathbf{1}^\top x = 1,\ x \ge 0\}$. Under (A1)–(A4) and convexity of $c^\top x$, the output $(x^\star, \delta^\star)$ of Algorithm 1 satisfies, with probability $1-\beta-o(1)$,*

$$c^\top x^\star - c^\top x_Q^\dagger \;\le\; \underbrace{\inf_j \left(c^\top x_j - c^\top x_Q^\dagger\right)}_{\textit{approximation (menu) error}} \;+\; \underbrace{L_c\,\|x^\star - x_j\|_2}_{\textit{optimization gap for chosen candidate}}$$

for any candidate x_j; in particular, if the menu contains $x_Q^\dagger$ (or approximates it to ε in ℓ_2), the excess cost is $O_p(\varepsilon)$.

Proof. Feasibility holds by Theorem 1. Among all feasible candidates, Algorithm 1 selects the one minimizing the validated upper bound, which is an upper envelope for the (unknown) robust CVaR. This is a conservative filtering and does not increase $c^\top x$ relative to the best feasible candidate present in the menu. The remainder is a standard approximation argument: $c^\top x$ is L_c-Lipschitz on $\mathcal{X}$ (e.g., $L_c = \|c\|_2$), so if the menu contains an ε-approximation to $x_Q^\dagger$, the selected $x^\star$ is within $O(\varepsilon)$ in cost. □

B.5 Specializations and reductions

1. **Reduction to Lam & Qian (2019).** If $w_i \equiv 1/n_2$ (no shift) and $b = 1$ (i.i.d.), then $n_{\text{eff}} = n_2$, the block sums reduce to centered sample means, and Step 4 coincides with the normalized Gaussian supremum used in [8]; Step 5 reduces to their feasibility test with the same calibration of $q_{1-\beta}$.
2. **Only dependence (no shift).** If $w_i \equiv 1/n_2$ but $b \asymp n_2^{1/3} > 1$, Lemma 2 still applies and we obtain a dependence-robust GS validator.
3. **Only shift (no dependence).** If $b = 1$ but w is non-uniform (clipped density ratios), we recover an importance-weighted GS band with effective sample size n_{eff}.

Implementation note. In practice we use the *analytical radius* in Step 5, $\delta^*(x) = \alpha\big(\gamma - \widehat{H}_w(x) - \hat{q}\,\hat{\sigma}_w(x)/\sqrt{n_{\text{eff}}}\big)_+/\|x\|_2$ clipped to $[\delta_{\min}, \delta_{\max}]$, which avoids coupling δ to any pre-specified grid and prevents pathological selections (e.g., always taking the largest δ when no grid point is feasible).