Initial Margin Models for Energy Markets: A WTI Case Study

1 Introduction and Related Literature

Margin posted to a central counterparty (CCP) aims to make losses over a short liquidation horizon unlikely to exceed collateral. In commodities, risk is episodic and state-dependent (storage constraints, seasonal flows, expiry rolls), which complicates the design of risk-sensitive yet operationally stable initial margin (IM). Reviews of the 2020–2022 stress document large, rapid swings in IM for energy products and an ongoing transition from legacy scenario systems (e.g., SPAN) toward portfolio VaR engines with explicit anti-procyclicality (APC) features [1, 2]. Major clearers have implemented or announced modern frameworks (ICE IRM 2, CME SPAN 2) with floors/buffers to temper reactivity, supported by public communications [3–8]. While the classic SPAN methodology remains instructive and widely referenced [9], operational capabilities now include real-time engines and what-if analysis (e.g., OCC STANS implementations) [10]. The common tension is clear: backward-looking risk measures remain reactive; APC tooling mitigates—but does not eliminate—procyclicality [1, 2].

To address this, we add a forward-looking layer that preserves a transparent VaR-style baseline while allowing traded-option information to steer IM smoothly within existing CCP controls. Let the portfolio-agnostic baseline for horizon h at confidence α be

$$M_t^{\text{base}} \;=\; \kappa\, \widehat{\text{VaR}}_{\alpha,h}(\Delta P(t{:}t{+}h))\,, \tag{1.1}$$

with governance buffer $\kappa \geq 1$. From each trading day, we construct a compact feature vector x_t from the option-implied surface and market state: parameters from an SVI–SABR calibration [11, 12], implied risk-neutral moments (variance, skewness, kurtosis), calibration diagnostics, hidden-Markov regime probabilities, and a futures term-structure index [1, 13, 14]. A gradient-boosted tree (XGBoost) f_θ with a custom objective maps x_t to a short-horizon risk-scaling target: the next-day ATM-volatility ratio (*gamma*),

$$\widehat{\gamma}_t \;=\; f_\theta(x_t) \approx \frac{\sigma_{t+1}^{\text{ATM}}}{\sigma_t^{\text{ATM}}}. \tag{1.2}$$

which we convert into a bounded multiplicative overlay

$$\phi_t \;=\; \text{clip}(\widehat{\gamma}_t,\; F_{\min},\; F_{\max})\,, \tag{1.3}$$

so the production margin is

$$M_t^{\text{prod}} \;=\; \phi_t\, M_t^{\text{base}}. \tag{1.4}$$

This design keeps attribution and governance simple (a single multiplier), remains engine-agnostic (compatible with filtered historical simulation and scenario-based implementations used in IRM 2/SPAN 2 families), and supports intraday recalculation without full portfolio revaluation [3, 4, 8]. When the baseline uses an ATM-proportional proxy, this corresponds to taking $\phi_t \approx \widehat{\gamma}_t$, preserving interpretability while leveraging the richer forward-looking signals in x_t. Related research explores deep-learning proxies for IM/MVA [14, 15] and ML replication of SPAN-style requirements [13]; industry work links changes in market structure to margin outcomes [1, 16].

The remainder of the paper is organised as follows. §2 details the feature construction, forecasting model, and overlay governance; §3 reports coverage, average level, and procyclicality; the Discussion interprets operational implications and limitations; and §5 concludes with implementation guidance for CCPs and clearing members.

2 Methodology

We superimpose a forward-looking, option-implied signal on a transparent VaR-style baseline for WTI futures and options. For a liquidation horizon h and confidence α, the baseline margin is

$$M_t^{\text{base}} \;=\; \kappa\,\widehat{\text{VaR}}_{\alpha,h}(\Delta P(t{:}t{+}h))\,, \qquad \kappa \geq 1, \tag{2.1}$$

and we form the production margin by applying a bounded multiplicative overlay. Unless stated otherwise, $h = 1$ trading day, $\alpha = 0.99$, $\kappa = 1$.

$$M_t^{\text{prod}} \;=\; \phi_t\, M_t^{\text{base}}, \qquad \phi_t \in [\,F_{\min}, F_{\max}\,] = [0.70, 1.50], \tag{2.2}$$

so that coverage is preserved while day-to-day reactivity is damped within governance limits [2–4].

We first describe the data and feature construction. For each business day t we assemble the SOFR overnight rate r_t; settlement prices for a strip of consecutive WTI futures (front through deferred); full option chains; and the OVX index for context. Series advance on business days (Mon–Fri); r_t is carried forward on non-business days. Option chains are included whenever valid prices are available (no hard OI/volume filters and no bid–ask screening). Missing values are handled by as-of joins; features are used in raw units. Our sample spans **2024-07-05–2025-08-22** and covers CL futures and their listed options; futures returns are computed on the *front* contract (no volume/OI-based rolling). This mirrors supervisory and industry documentation on margin dynamics in energy [1].

Next we summarize the futures curve with a curve term-structure index (CTSI). The CTSI aggregates normalised calendar spreads across near-, mid- and far-dated contracts and adds a simple persistence signal reflecting the ordering of prices along the curve. Spreads are stabilised by scale-normalisation and a bounded transform, and the resulting index is clipped to $[-1, 1]$ so that positive values indicate contango and negative values indicate backwardation. We use the CTSI both as a direct feature and as the driver of a light regime model.

We calibrate the daily option surface and extract an implied risk-neutral distribution (RND). Using Black–76, we invert implied volatilities and fit a smooth SVI–SABR total-variance slice for each expiry. We adopt the following SVI–SABR form for total variance $w_\tau(k)$ at log-moneyness k and time to expiry τ:

$$w_\tau(k) \;=\; \tfrac{1}{2}\,\alpha_\tau^2 \left[\, 1 \;+\; \rho_\tau\,\frac{\nu_\tau}{\alpha_\tau}\,k \;+\; \sqrt{\left(\frac{\nu_\tau}{\alpha_\tau}\,k + \rho_\tau\right)^2 + 1 - \rho_\tau^2}\,\right], \tag{2.3}$$

so that the model-implied volatility is $\sigma_\tau(k) = \sqrt{w_\tau(k)/\tau}$. Calibration solves a weighted least-squares problem in total variance,

$$\min_{\alpha_\tau,\nu_\tau,\rho_\tau} \;\sum_{i\in\mathcal{I}_\tau} w_i\,\big(w_\tau(k_i) - \tau\,\sigma_i^2\big)^2 \;+\; \lambda_{\text{reg}}\,\mathcal{R}(\alpha_\tau,\nu_\tau,\rho_\tau), \tag{2.4}$$

with $w_i \propto (\text{volume}{+}\text{OI})_i$ and a small regulariser $\mathcal{R}$ that penalises unstable wings. We apply a lightweight no-arbitrage guard that penalises slices violating $\alpha_\tau\nu_\tau T_\tau\,(1+|\rho_\tau|) \geq 4$ and record calibration diagnostics such as IV RMSE (Appendix B).

From the fitted slice we compute an implied risk-neutral distribution (RND) by numerical differentiation of call prices (Breeden–Litzenberger). Let $C(K,\tau)$ denote the Black–76 price and define the density $q_\tau(K) = \frac{\partial^2 C}{\partial K^2}(K,\tau)$. Changing variables to log-moneyness $x = \ln(K/F_t)$, with $K = F_t e^x$ and $dK = F_t e^x\,dx$, gives the log-forward density $p_\tau(x) = F_t e^x\,q_\tau(F_t e^x)$. Raw moments are:

$$\begin{aligned} m_r &= \int x^r\,p_\tau(x)\,dx, \quad \mu = m_1, \quad v = m_2 - m_1^2, \\ \text{skew} &= \frac{m_3 - 3m_2\,\mu + 2\mu^3}{v^{3/2}}, \quad \text{kurt} = \frac{m_4 - 4m_3\,\mu + 6m_2\,\mu^2 - 3\mu^4}{v^2}. \end{aligned} \tag{2.5}$$

We evaluate on a dense *strike* grid $K \in [F_t/4,\, 4F_t]$ (symmetric around the forward), which induces a non-uniform grid in x. First derivatives use central differences

$$C_K(K_j) \;\approx\; \frac{C(K_{j+1}) - C(K_{j-1})}{2\Delta K}, \qquad C_{KK}(K_j) \;\approx\; \frac{C(K_{j+1}) - 2C(K_j) + C(K_{j-1})}{\Delta K^2},$$

so that $q_\tau(K_j) \approx C_{KK}(K_j)$ (second-order accurate, $O(\Delta K^2)$ for smooth C). Mapping to $x_j = \ln(K_j/F_t)$ yields $p_j = F_t e^{x_j} q_\tau(K_j)$. We approximate the integral with a trapezoidal rule on the non-uniform x-grid:

$$m_r \;\approx\; \sum_j x_j^r \, p_j \, \Delta x_j, \qquad \Delta x_j \equiv \tfrac{1}{2}\big(\ln \tfrac{K_{j+1}}{K_j} + \ln \tfrac{K_j}{K_{j-1}}\big).$$

Finally we renormalise p by $Z = \sum_j p_j \, \Delta x_j$ so that $\sum_j (p_j/Z)\, \Delta x_j = 1$ before computing m_r. Alongside the central moments and standardized skew/kurtosis, we record $\sigma_{\mathrm{ATM},t} = \sigma_\tau(0)$ and simple shape summaries (level, slope/skew, curvature, and a shift parameter) as features. The wide strike range and explicit renormalisation stabilise the tails without requiring a piecewise wing model.

To capture market state, we specify a two-state Gaussian hidden Markov model on the CTSI,

$$\mathrm{CTSI}_t \,|\, s_t = j \sim \mathcal{N}(\mu_j, \sigma_j^2), \qquad \mathbb{P}(s_t = j \,|\, s_{t-1} = i) = \Pi_{ij}, \quad j \in \{1, 2\},$$

with parameters (μ_j, σ_j^2, Π) estimated by EM. We use smoothed state probabilities $\xi_{t,j} = \mathbb{P}(s_t = j \,|\, \mathrm{CTSI}_{1:t})$ and short-run changes $\Delta\xi_{t,j}$ as features. For base series z_t such as σ_{ATM}, the RND moments, and the SVI–SABR parameters, we add one-day lags, short moving averages, and momentum transforms, plus lightweight interactions (for example, skew×kurtosis and $\rho \times \sigma_{\mathrm{ATM}}$) and two OVX summaries (short moving average and momentum). The resulting daily feature vector is denoted x_t.

Returns are computed from daily futures closes. We forecast one-day volatility with an EGARCH(2,2) and Student-t innovations, obtaining volatility σ_t and degrees of freedom ν_t. The parametric 99% VaR is

$$\mathrm{VaR}_{0.99,t} = \sigma_t \, t_{\nu_t}^{-1}(0.01), \qquad M_t^{\mathrm{base}} = \big|\mathrm{VaR}_{0.99,t}\big| \, P_t, \tag{2.6}$$

where P_t is the day-t futures price and one-step P&L is $\mathrm{P\&L}_{t+1} = P_{t+1} - P_t$.

We then learn a short-horizon risk-scaling ("gamma") directly from x_t to the next-day ATM-volatility ratio. We fit a gradient-boosted tree f_θ and set $\hat{\gamma}_t = f_\theta(x_t)$.

$$\gamma_t \;=\; \frac{\sigma_{\mathrm{ATM},t+1}}{\sigma_{\mathrm{ATM},t}}, \qquad \phi_t \;=\; \mathrm{clip}\big(\,\widehat{\gamma}_t\,,\, F_{\min}, F_{\max}\,\big), \tag{2.7}$$

where the in-sample target γ_t is clipped and lightly smoothed to reduce noise. The correction ϕ_t multiplies the baseline margin as in (2.2).

Training is strictly time-ordered. For each test day t we fit f_θ on an expanding window $[1, t-1]$ with a blocked, time-ordered validation split for monitoring. Features are computed without leakage from the training window; models are trained for a fixed number of boosting rounds (no early stopping). We log permutation/GB importances on the validation set to support governance and model monitoring; hyperparameter ranges are listed in Appendix B.

We train an XGBoost regressor on the mapping $x_t \mapsto \gamma_t$ with the following per-observation objective (a piecewise-smooth surrogate):

$$\mathcal{L}(\hat{\gamma}; \gamma) \;=\; \tfrac{1}{2}(\hat{\gamma} - \gamma)^2 \;+ \underbrace{\mathbb{1}\{\hat{\gamma} > \, a\}(\hat{\gamma} - a)^2 + \mathbb{1}\{\hat{\gamma} < \, b\}(\hat{\gamma} - b)^2}_{\text{extreme-clip penalty}} + \lambda_M \, \hat{\gamma} \tag{2.8}$$

whose gradients/Hessians match a mean-squared term, squared-hinge penalties beyond $[b, a]$, and a small linear term discouraging inflated $\hat{\gamma}$. Predictions are bounded to $[F_{\min}, F_{\max}] = [0.70, 1.50]$ at inference.

Finally, we evaluate performance in a walk-forward backtest. We join ϕ_t with the VaR baseline to form M_t^{prod} and compute, over the test path: (i) 99% breach frequency $\hat{p} = \frac{1}{T}\sum_t \mathbb{1}\{|\mathrm{P\&L}_{t+1}| > M_t\}$ for the baseline and the dynamic series; (ii) the average margin ratio $\overline{M/P}$; and (iii) a procyclicality proxy $\mathrm{sd}(\Delta(M/P))$. For $\hat{p}$ we report a normal-approximate 95% interval $\hat{p} \pm 1.96\sqrt{\hat{p}(1-\hat{p})/T}$ and the Kupiec unconditional-coverage test (Christoffersen diagnostics summarised in Appendix B). We also provide sensitivity to $(F_{\min}, F_{\max})$ and the training-window length, plus ablations that drop moments, regimes or surface parameters. If a daily smile calibration fails quality checks, we fall back to the latest valid slice and set $\phi_t = \min(\max(\widehat{\gamma}_{t-1}, F_{\min}), F_{\max})$; if CTSI or OVX are missing we impute from short moving averages. These fallbacks ensure continuity and avoid spurious margin swings; thresholds and QC rules are documented in Appendix B. The overlay remains portfolio- and engine-agnostic (HS/FHS or scenario-based) and supports intraday recalculation within established CCP controls [2–4, 8].

3 Results

We evaluate the forward-looking overlay against a transparent EGARCH–t VaR baseline on a walk-forward path spanning `2024-07-05` to `2025-08-22` ($T = 286$ trading days). At each day t we compute the production margin $M_t^{\text{prod}} = \phi_t \, M_t^{\text{base}}$ and assess three criteria: (i) 99% coverage (breach frequency), (ii) average margin level $\overline{M/P}$, and (iii) a procyclicality proxy $\text{sd}(\Delta(M/P))$.

Table 1: Coverage (99%), average level, and procyclicality on the test window. Parentheses show % change vs baseline. Confidence intervals are normal-approximate.

	99% breach	$\overline{M/P}$	sd $\Delta(M/P)$
Baseline (EGARCH–t VaR)	0.70% [0.00–1.67]%	4.95%	0.0053
Dynamic (overlay)	1.75% [0.23–3.27]% (+150%)	4.70% (=5.1%)	0.0050 (=5.7%)

Table 1 shows that the overlay lowers both the average level and day-on-day variability while maintaining statistical consistency with the 1% target. Kupiec unconditional-coverage p-values are 0.589 (baseline) and 0.250 (dynamic), so we do not reject the 1% hit-rate hypothesis on this window. The dynamic series' higher point estimate (1.75%) is within expected sampling variation for $T = 286$.

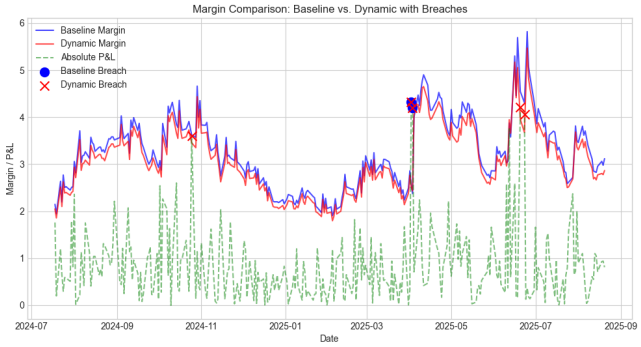


Figure 1: Baseline vs dynamic margins with absolute P&L and breach markers (`2024-07-05`–`2025-08-22`). Dynamic margins (red) track the baseline (blue) while generally remaining lower; markers flag days where $|\text{P\&L}_{t+1}| > M_t$.

3.1 Model signal: what the overlay learns

To understand the mechanism, Appendix A (Figures 3 and 4) reports feature influence through two complementary lenses. In the *usage* view (permutation and model-gain importances), the leading drivers are the option days-to-expiry *DTE* (denoted τ), short-horizon ATM IV dynamics (momentum and lag), and SVI–SABR shape parameters (especially ρ and ν), with market context from F and OVX momentum. In the *marginal-utility* view (single-feature ablations, drop-and-refit), removing `iv_vol_ratio` yields the largest out-of-sample R^2 loss, followed by ATM IV dynamics and τ; coverage and average level remain essentially unchanged across ablations. The divergence between rankings is expected with correlated/interacting predictors. Taken together, both views indicate that the overlay learns forward-looking changes in the surface—level, skew, and curvature—rather than the contemporaneous ATM level. For a single listed tenor (CLV25), the prominence of τ is mechanically sensible: as expiry approaches, the surface re-marks and time decay amplifies the influence of shape parameters, the cues the overlay is designed to capture.

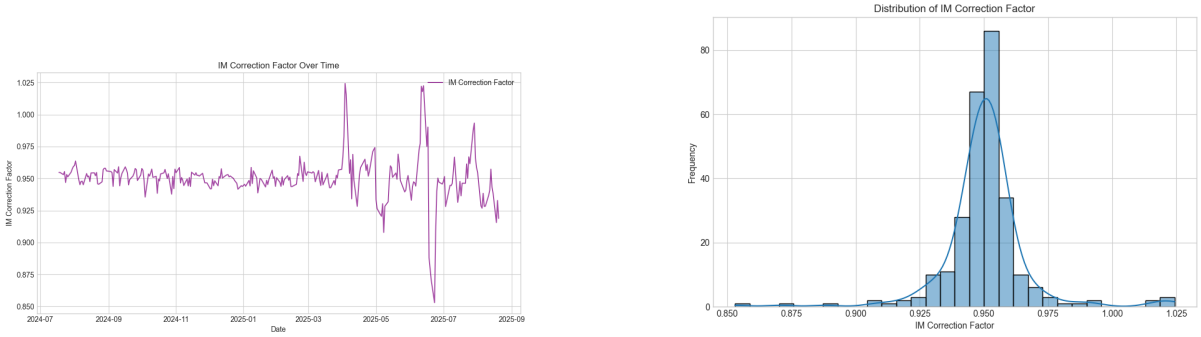
3.2 Overlay behaviour: level and stability

We summarise the behaviour of the multiplier ϕ_t in Table 2. The distribution centres slightly below one with bounded excursions and essentially no clipping at the governance bounds, consistent with anti-procyclicality and with smooth adjustments rather than edge effects.

Table 2: Overlay summary: percentiles and clip activity for ϕ_t.

mean	p1	p5	p25	median	p75	p95	p99	% at $F_{\min}$ / $F_{\max}$
0.9505	0.9028	0.9283	0.9451	0.9507	0.9550	0.9715	1.0191	0.00% / 0.00%

To visualise dynamics and dispersion, Figure 2 shows the time series of ϕ_t alongside its empirical distribution. The series is anchored around 0.95 most days, with interquartile mass in $[0.945, 0.955]$ and a high-percentile tail below the governance ceiling (p95 $\approx$ 0.972, p99 $\approx$ 1.019). Short-lived excursions—brief dips toward 0.90 and spikes just above one—coincide with volatility flare-ups in the margin series but remain bounded and infrequent. Together with the % clipped statistics (essentially zero at both bounds), this indicates the overlay operates as a smooth, information-driven multiplier rather than as a hard cap/stop mechanism.



(a) ϕ_t over time

(b) Distribution of ϕ_t

Figure 2: Overlay correction factor ϕ_t: evolution and cross-sectional shape. The series is centred below one, rarely near the bounds, and adjusts smoothly through stress.

Figure 1 complements these diagnostics: dynamic margins track the baseline envelope through quiet periods, compressing levels by roughly five percent, and remain within the governance band when volatility spikes. Breach markers cluster around a small number of stress days for both series; despite a higher empirical rate for the dynamic variant, both pass unconditional coverage at the 1% target on this sample. In aggregate, the overlay achieves lower and smoother requirements without a statistically detectable loss of coverage, consistent with the intended "forward-looking but bounded" design.

4 Discussion

The results indicate that a light, option-implied overlay can lower and smooth initial margin without a statistically detectable loss of 99% coverage on the test window. Relative to an EGARCH–t VaR baseline, the overlay reduces the average margin-to-price ratio by about 5.1% and the day-on-day variability of M/P by about 5.7%. The empirical breach rate is higher in point estimate (1.75% vs. 0.70%), yet both series pass Kupiec's unconditional-coverage test at the 1% target over $T = 286$ trading days. Taken together, these facts suggest a capital-efficiency gain with modest reactivity costs that remain within standard statistical tolerance for a one-year backtest.

The mechanism is consistent with the model design. Feature importance concentrates on the daily time-to-expiry of the traded option series (*DTE*, denoted T in Appendix A (Figure 3), on ATM implied-volatility momentum and lag, and on SVI–SABR shape parameters—notably ρ and ν—alongside the front futures price and OVX momentum. Because this study fixes a single listed tenor (CLV25, September 2025), DTE varies mechanically and acts as a state variable that co-moves with smile parameters and, through them, with short-horizon risk. It is therefore unsurprising that DTE emerges as the top feature: it captures the time-decay and re-marking effects that drive both the level and the shape of the implied surface. Importantly, the importance profile confirms that the model uses more than ATM alone, leveraging forward-looking skew and convexity information embedded in traded options.

Design choices around governance and robustness behave as intended. The overlay multiplier ϕ_t is centred just below one (mean ≈ 0.951; median ≈ 0.951) with tight interquartile mass and essentially no time spent at the bounds; clipping is therefore not the source of improvement. When smile calibration quality checks fail, the system falls back to the most recent valid slice and applies clip-only updates to ϕ_t, which limits spurious swings. In practice these safeguards are rarely triggered on the sample, and the reduction in level and procyclicality is achieved by the forward-looking forecast rather than by hard caps.

Operationally, the overlay integrates cleanly with CCP-style margin engines. It scales a transparent baseline and requires only portfolio-level summary inputs, so it is compatible with filtered historical simulation and scenario frameworks in use at major clearers. Because ϕ_t is bounded by $[F_{\min}, F_{\max}]$ and computed from public market data, it is easy to communicate to members and supervisors, and it reduces the need for ad hoc parameter changes. Intraday recalculation is straightforward: the baseline updates as usual and the overlay can be refreshed when option surface data are updated, subject to the same governance bounds.

Two caveats are worth noting. First, the backtest covers roughly one year and a finite set of stress episodes; statistical power for tail tests is limited at the 1% level. Second, our baseline is a parametric VaR on front-contract returns; while this choice is transparent and aligns with practice, alternative baselines (e.g., FHS or scenario sets) may change the level and the dispersion of M/P. The overlay is engine-agnostic, so repeating the evaluation on additional baselines and over longer samples with multiple stress regimes would strengthen external validity. In preliminary checks, a curve-regime indicator built from the CTSI added little incremental information on this window; the improvements do not hinge on explicit regime classification.

Overall, the evidence supports a simple conclusion: option-implied, feature-driven signals can steer margin within governance bounds to achieve smoother and lower requirements while preserving coverage tests on the sample. This is attractive for both CCPs and clearing members because it delivers capital relief and reduces procyclicality without sacrificing transparency or requiring full revaluation of portfolios. The next section discusses practical implications, limitations, and extensions, and Section 5 concludes with implementation guidance.

5 Conclusion

This paper proposes a simple, governance-friendly way to make initial margin for energy derivatives more forward-looking without abandoning the transparency of VaR-style engines. We keep a clear baseline—an EGARCH–t 99% VaR on front futures—and learn a bounded multiplicative overlay ϕ_t from traded options and market state. The overlay uses option days-to-expiry (DTE) and surface shape (via SVI–SABR parameters, implied moments, and ATM IV dynamics) to forecast short-horizon risk and scales the baseline as $M_t^{\text{prod}} = \phi_t M_t^{\text{base}}$ within pre-set bounds.

On our WTI case study (`2024-07-05` to `2025-08-22`), the overlay lowers the average margin-to-price ratio by about **5.1%** and reduces day-on-day variability by about **5.7%** relative to the EGARCH–t baseline, while remaining statistically consistent with a 1% coverage target (Kupiec $p = 0.589$ for the baseline and $p = 0.250$ for the dynamic series). The multiplier is centred just below one (mean ≈ 0.951; p95 ≈ 0.972; p99 ≈ 1.019) with essentially zero time at governance bounds, indicating that improvements come from information in the implied surface rather than from hard clipping.

The mechanism is economically intuitive. With a single listed tenor (CLV25), DTE acts as a state variable that co-moves with smile shape and re-marking effects; it therefore emerges as the most important feature, followed by ATM IV momentum/lag and SVI–SABR parameters (ρ, ν). This confirms that the overlay leverages forward-looking smile information beyond the contemporaneous ATM level.

Operationally, the approach is engine-agnostic and portfolio-agnostic: it scales an existing baseline and requires only portfolio-level summaries, so it can sit alongside filtered historical simulation or scenario-based frameworks used by major clearers. Bounded outputs and simple fallbacks (last-good smile; clip-only updates) make the behaviour predictable and easy to communicate to members and supervisors, and intraday recalculation is straightforward under the same governance limits.

Two limitations frame next steps. First, the backtest spans roughly one year with a finite number of stress episodes, so statistical power for 1% tests is limited; extending the sample across multiple stress regimes would strengthen external validity. Second, our baseline is parametric; repeating the evaluation with FHS or scenario engines and across additional commodities would help gauge robustness and portability. Further work includes conditional-coverage/independence diagnostics, liquidity-aware features, intraday extensions, and a read-only parallel run to support change management.

In sum, a light, option-implied overlay can steer margin within transparent bounds to deliver *lower and smoother* requirements while preserving standard coverage tests. This offers a practical path for CCPs and clearing members to reduce procyclicality and funding variability without sacrificing explainability or requiring full revaluation infrastructure.

References

[1] BCBS. Margin dynamics in centrally cleared commodities markets in 2022. Technical Report d550, Basel Committee on Banking Supervision (BCBS), Committee on Payments and Market Infrastructures (CPMI), and IOSCO, May 2023. URL https://www.bis.org/bcbs/publ/d550.pdf. Joint BCBS–CPMI–IOSCO report.

[2] European Central Bank. CCP initial margin models in europe. Occasional Paper Series 314, European Central Bank, April 2023. URL https://www.ecb.europa.eu/pub/pdf/scpops/ecb.op314~afc6d2980c.en.pdf.

[3] Intercontinental Exchange. Ice risk model 2 (irm 2) methodology. https://www.ice.com/clearing/margin-models/irm-2/methodology, 2025. Web page, accessed 2025-08-19.

[4] CME Group. Span 2 methodology and functionality. https://www.cmegroup.com/clearing/risk-management/span-overview/span-2-methodology.html, 2024. Web page, accessed 2025-08-19.

[5] CME Group. Span 2 framework rollout. https://www.cmegroup.com/solutions/risk-management/performance-bonds-margins/span-methodology-overview/launching-span-2.html, 2024. Web page, accessed 2025-08-19.

[6] CME Group. Futures and options margin model. https://www.cmegroup.com/solutions/risk-management/performance-bonds-margins/futures-and-options-margin-model.html, 2025. Web page, accessed 2025-08-19.

[7] Rafik Mrabet. Navigating a new era in derivatives clearing. FIA MarketVoice, January 2024. URL https://www.fia.org/marketvoice/articles/navigating-new-era-derivatives-clearing.

[8] Suzanne Sprague. Cme clearing margining practices. https://www.cftc.gov/media/10226/AAC121423presentation_SuzanneSprague/download, 2024. Presentation to CFTC advisory committee (dated Dec 14, 2023; published 2024).

[9] CME Group. Cme span® methodology. https://www.cmegroup.com/clearing/files/span-methodology.pdf, 2019. Overview of SPAN parameters and scan risk arrays.

[10] Cboe Hanweck. Real-time margin engine for stans® methodology. https://www.cboe.com/services/analytics/hanweck/occ_clearing/, 2025.

[11] Jim Gatheral. Arbitrage-free svi volatility surfaces. Slide deck, 2012. URL https://mfe.baruch.cuny.edu/wp-content/uploads/2013/01/OsakaSVI2012.pdf. Osaka 2012; mirror hosted by Baruch MFE.

[12] Jim Gatheral and Antoine Jacquier. Arbitrage-free svi volatility surfaces. *Quantitative Finance*, 14(1):59–71, 2014. doi: 10.1080/14697688.2013.819986. URL https://arxiv.org/abs/1204.0646.

[13] Clara Branestam and Amanda Sandgren. Exploring the feasibility of replicating span-model's required initial margin calculations using machine learning: A master thesis project for intraday margin call investigation in the commodities market. Master's thesis, Umeå University, 2023. URL https://www.diva-portal.org/smash/get/diva2:1774989/FULLTEXT01.pdf.

[14] Joel P. Villarino and Álvaro Leitão. On deep learning for computing the dynamic initial margin and margin value adjustment. *arXiv preprint arXiv:2407.16435*, 2024. URL https://arxiv.org/abs/2407.16435.

[15] Xun Ma, Sogee Spinner, Alex Venditti, Zhao Li, and Strong Tang. Initial margin simulation with deep learning. *SSRN Electronic Journal*, 2019. doi: 10.2139/ssrn.3357626. URL https://ssrn.com/abstract=3357626.

[16] Jo Burnham. How changing the market you trade in impacts margin. OpenGamma Blog, May 2024. URL https://opengamma.com/how-changing-markets-impacts-margin/.

A Supplementary Tables and Figures

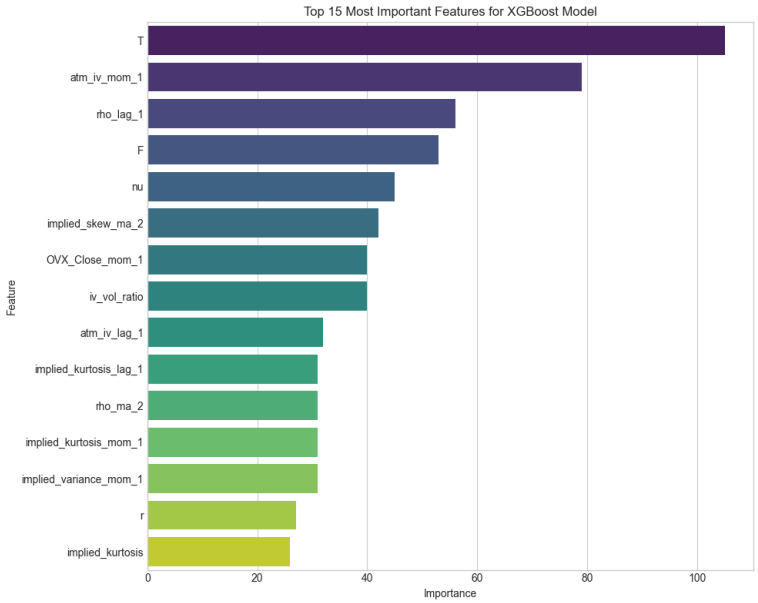


Figure 3: Top-15 features for the XGBoost overlay (permutation/GB importance). DTE (T, also referred to as τ throughout) dominates, followed by ATM IV momentum/lag and SVI–SABR parameters (ρ, ν), with contributions from the front price F and OVX momentum.

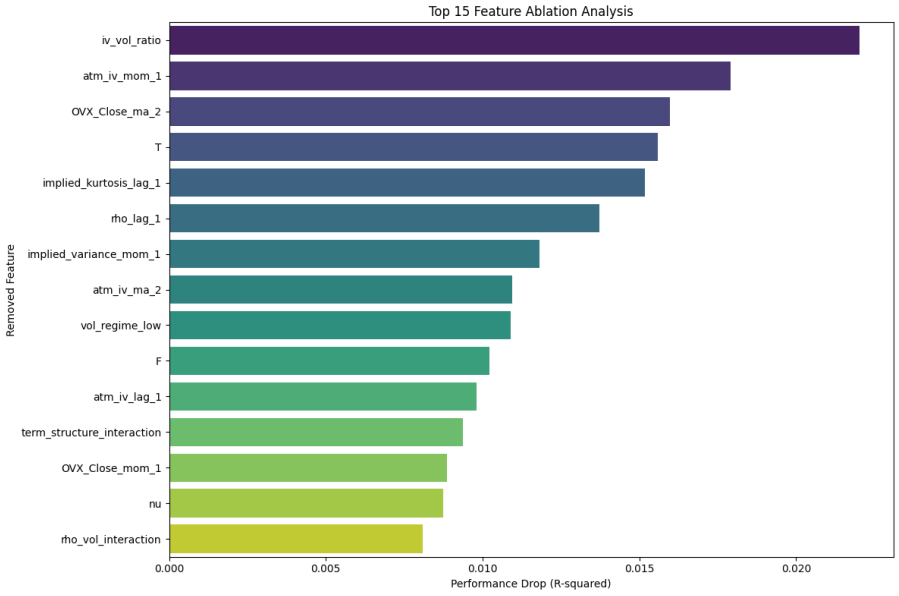


Figure 4: Top-15 features by single-feature ablation (drop and refit). Bars show out-of-sample R^2 loss on the validation split (higher = more important). This *marginal-utility* lens complements the permutation/model-gain *usage* rankings discussed in Section 2.

Table 3: Single-feature ablations with the largest increase in procyclicality.

Removed feature	Breaches	$\overline{M/P}$	sd($\Delta(M/P)$)
atm_iv_mom_1	5	0.047000	0.005040
nu	5	0.047000	0.005037
vol_momentum	5	0.047001	0.005022
atm_iv_lag_1	5	0.047001	0.005022
implied_variance_mom_1	5	0.047000	0.005021

Table 4: Single-feature ablations with the largest decrease in procyclicality.

Removed feature	Breaches	$\overline{M/P}$	sd($\Delta(M/P)$)
τ (days-to-expiry)	5	0.046999	0.005005
OVX_Close_mom_1	5	0.047001	0.005012
OVX_Close_ma_2	5	0.046998	0.005012
skew_kurt_interaction	5	0.047001	0.005013
rho	5	0.047002	0.005014

Table 5: Feature glossary and definitions. Wildcards apply to any base feature x_t.

Feature	**Definition**
F	Front WTI futures settlement price on day t (USD/bbl).
tau	Option time-to-expiry for the listed tenor on day t (in years); monotonically declines as expiry approaches. (Dataset column: T.)
r	SOFR overnight rate used as risk-free rate in Black–76.
atm_iv	At-the-money implied volatility from the calibrated surface at k=0 for the selected expiry.
OVX_Close	CBOE Crude Oil Volatility Index close (proxy for 1m WTI option IV).
rho	SABR/'SVI–SABR' correlation parameter $\rho \in [-0.999, 0.999]$ between forward and volatility; controls skew.
nu	SABR vol-of-vol parameter $\nu > 0$; controls smile curvature and wing steepness.
implied_variance	Risk-neutral variance of log-forward, computed from the calibrated surface via Breeden–Litzenberger on $K \in [F/4, 4F]$.
implied_skew	Risk-neutral standardised third central moment (skewness) of log-forward implied by the surface.
implied_kurtosis	Risk-neutral standardised fourth central moment (non-excess kurtosis) of log-forward implied by the surface.
iv_rmse	Root-mean-squared error of the IV fit across observed option quotes on the day.
hmm_regime	Viterbi regime index from a 2-state Gaussian HMM on the CTSI (0=backwardation, 1=contango).
hmm_prob	Smoothed probability of the backwardation state from the HMM fitted to the CTSI.
CTSI	Curve Term-Structure Index aggregating normalised calendar spreads with a bounded persistence term; clipped to $[-1, 1]$ (contango positive).
vol_regime_high	Indicator: 1 if $\texttt{atm\_iv}_t$ is above its rolling 22-day 75th percentile; else 0.
vol_regime_low	Indicator: 1 if $\texttt{atm\_iv}_t$ is below its rolling 22-day 25th percentile; else 0.
skew_kurt_interaction	Product implied_skew × implied_kurtosis.
vol_momentum	Interaction atm_iv_mom_1 × implied_variance_mom_1.

Feature	Definition
rho_vol_interaction	Interaction $\rho \times$ atm_iv.
iv_vol_ratio	Ratio atm_iv/ stdev_{10}(atm_iv); higher values indicate elevated IV relative to its recent variability.
skew_level	Absolute skew level |implied_skew|.
month1_spread	Calendar spread between the first two listed futures (e.g., $F_2 - F_1$).
month3_spread	Calendar spread between the third and first listed futures (e.g., $F_3 - F_1$).
year1_spread	Calendar spread between the 12th and first listed futures (e.g., $F_{12} - F_1$) when available.
norm_month1	Normalised month-1 spread (e.g., spread scaled by a robust dispersion such as rolling MAD).
norm_month3	Normalised month-3 spread.
norm_year1	Normalised year-1 spread.
persistence	CTSI persistence term reflecting whether the curve ordering (contango/backwardation) persisted from $t-1$ to t (bounded).
persistence_impact	Contribution of the persistence term to the CTSI after scaling; diagnostic feature.
*_lag_1	One-day lag x_{t-1}.
*_ma_2	Two-day simple moving average $\frac{1}{2}(x_t + x_{t-1})$.
*_mom_1	One-day momentum (first difference) $x_t - x_{t-1}$.
target	Training target $\gamma_t = \sigma_{\text{ATM},t+1}/\sigma_{\text{ATM},t}$ (clipped), used to compute the overlay $\phi_t = \text{clip}(\widehat{\gamma}_t, F_{\min}, F_{\max})$.

B Training and Governance

Data and joins. Option chains are included whenever valid option prices are available; no hard filters are applied on open interest or volume, and no bid–ask or BA/mid screening is used. Futures returns are computed from the front contract time series (no volume/OI-based rolling). The timeline advances on business days (Mon–Fri) without an exchange-holiday calendar. Missing macro/market series are merged by as-of alignment without a hard maximum lag. Observed lags are logged and reviewed; no anomalous multi-day gaps were detected in-sample.

SVI–SABR calibration. For each date τ, we fit total variance $w_\tau(k)$ by minimizing a weighted least-squares objective with weights proportional to option activity (volume+OI when present). Parameters are bounded as $\alpha_\tau \in [10^{-5}, 10]$, $\nu_\tau \in [10^{-5}, 10]$, $\rho_\tau \in [-0.999, 0.999]$. No explicit slope cap $|\partial_k w_\tau|$ is enforced; instead, a no-arbitrage guard applies a heavy penalty (and flags the slice) if $\alpha_\tau \nu_\tau T_\tau (1+|\rho_\tau|) \geq 4$. If the guard is tripped or the optimizer fails to converge, the daily slice is marked as a calibration warning. In our sample, no slices were rejected by the guard, and no optimizer failures occurred (fallback unused).

RND grid and tails. Risk-neutral densities are obtained by central differences on calls across a dense *strike* grid $K \in [F/4, 4F]$ (symmetric around the forward). Because $x = \ln(K/F)$, this induces a non-uniform grid in log-moneyness x; no piecewise wing model of the form $a + b|k|$ with a fixed cutoff $k^\star$ is used. The numerical PDF is re-normalized to integrate to one before computing moments.

Feature processing. Features are used in raw scale (no mandatory global standardisation or winsorisation); this preserves interpretability of coefficients/feature importances and avoids data leakage across windows. If a scaler artifact is present, it is applied; otherwise, raw features are passed through. The set includes smile/surface descriptors (e.g., ATM level and shape parameters) and auxiliary market state variables. Regime information is added via a 2-state Gaussian HMM (EM estimation), using the posterior state probabilities.

Model target and bounds. The custom training loss is tolerant inside a band $[b, a] = [0.70, 1.30]$ and applies extra linear penalty weight $\lambda_M = 0.05$ outside the band. At inference, the predicted correction factor is clipped to $[F_{\min}, F_{\max}] = [0.70, 1.50]$.

Training protocol. We use an expanding window with a blocked (time-ordered) validation split for monitoring. Early stopping is not employed; models are trained for a fixed number of boosting rounds ($n_{\text{estimators}} = 200$) with learning rate $\eta = 0.05$. Hyperparameter grids and final selections are persisted for reproducibility.

Backtesting diagnostics. We report the realized breach rate with its normal-approximate 95% confidence interval, the Kupiec unconditional-coverage test, and the Christoffersen independence test. The current implementation does not automatically sweep $(F_{\min}, F_{\max})$ or the training-window length; any such sensitivity checks are run offline and reported separately. Ablations remove moments, regime variables, or surface parameters.

Fallbacks and QC. If a daily smile fails the no-arbitrage guard or the optimizer fails, we carry forward the most recent valid surface and set $\phi_t = \min\big(\max(\widehat{\gamma}_{t-1}, F_{\min}), F_{\max}\big)$. Missing CTSI/OVX observations are imputed from short moving averages. All QC flags and thresholds are timestamped in the logs.