WORKING PAPER · JUNE 2026 · VERSION 1.0

Beyond the Point Estimate

Making Uncertainty Operational in Equity Valuation

Martin Russmann

Independent Researcher · ORCID 0009-0006-4079-1216

mrussmann@proton.me · martinrussmann.com

ABSTRACT

Equity valuation is often implemented as a deterministic pipeline: select point inputs, compute a discounted cash flow (DCF) number, and compare it to market price. This design hides uncertainty, makes model risk difficult to quantify, and encourages overconfident decisions under regime shifts. The point estimate is the problem; the distribution is the deliverable. This working paper specifies an alternative from first principles: a valuation pipeline that produces *probability distributions* over enterprise value and connects those distributions to *explicit decision policies*. Cash flows are modeled as strictly positive stochastic processes on their natural support, the perpetuity constraint $r > g_{\text{term}}$ is enforced without distorting the dependence structure, growth and discount-rate components are coupled through a heavy-tailed copula, and a two-level Monte Carlo design separates path-level variability from parameter (model) risk. The evaluation target is neither a single accuracy number nor the unobservable intrinsic value itself: a *convergence model* maps the valuation gap to realized excess returns, and the framework is falsified if the estimated convergence speed is indistinguishable from zero.

KEYWORDS probabilistic valuation; discounted cash flow; Monte Carlo; copula dependence; uncertainty quantification; model risk; decision policy; backtest overfitting

Disclaimer. This document describes a research and engineering framework. It does not constitute investment advice, a recommendation, or an offer to buy or sell securities. Any empirical performance claims require a fully specified dataset, protocol, and out-of-sample evaluation.

1 Executive Summary

Equity valuation is often implemented as a deterministic pipeline: select point inputs, compute a discounted cash flow (DCF) number, and compare it to market price. This design hides uncertainty, makes model risk difficult to quantify, and encourages overconfident decisions under regime shifts. The point estimate is the problem; the distribution is the deliverable.

This white paper specifies an alternative from first principles: a valuation pipeline that produces *probability distributions* over enterprise value and connects those distributions to *explicit decision policies*. The framework combines five components:

1. **Distributional cash-flow model on natural support:** cash flows are modeled as strictly positive stochastic processes using log-increments, with a revenue–margin decomposition extending coverage to loss-making firms.
2. **Economically constrained terminal value:** the perpetuity constraint $r > g_{\text{term}}$ is enforced in the sampling scheme without distorting the dependence structure, and a value-driver formulation ties terminal growth to returns on invested capital.
3. **Dependence modeling:** growth and discount-rate components are coupled via a heavy-tailed copula to represent regime co-movements and tail dependence.
4. **Explicit uncertainty decomposition:** path-level variability and parameter (model) risk are separated by a two-level Monte Carlo design, so that interval statements are honest about their epistemic component.
5. **Nowcasting layers with auditability:** a robust multivariate anomaly signal and a calibrated sentiment signal update short-horizon priors without duplicating price-based information.

The evaluation target is not a single accuracy number, and it is not the unobservable intrinsic value itself. A *convergence model* maps the valuation gap to realized excess returns; the framework is falsified if the estimated convergence speed is indistinguishable from zero. Evaluation therefore comprises (i) probabilistic forecast quality of the implied return distribution (proper scoring rules and calibration diagnostics) and (ii) the realized behavior of a *fixed* decision policy applied out-of-sample under explicit constraints.

2 Problem Statement and Design Principles

2.1 Why deterministic DCF is structurally brittle

A point-estimate DCF compresses uncertainty into a single number, implicitly assuming that (a) parameter uncertainty is negligible, (b) model form is correct, and (c) the decision-maker can ignore tail outcomes. These assumptions fail precisely in the settings where valuation matters: macro shifts, financing constraints, industry disruption, and information shocks.

2.2 Principles

From a first-principles perspective, a valuation system should:

- represent uncertainty as distributions on the correct support (e.g., $CF_i > 0$, bounded margins);
- enforce economically necessary constraints (e.g., $r > g_{\text{term}}$) without biasing the joint distribution;
- separate *forecasting* (what distributions are plausible) from *policy* (what actions follow);
- separate *path uncertainty* (variability given the model) from *parameter uncertainty* (doubt about the model);
- remain auditable: every valuation should be reproducible from stored data snapshots and model versions;
- quantify tail risk explicitly and avoid hidden "repairs" that bias distributions;
- maintain dimensional consistency throughout: growth priors and rate components must be jointly nominal or jointly real.

3 Framework Overview

Let V denote enterprise value and V_{mkt} a market-implied benchmark (e.g., enterprise value from market cap, net debt, and cash). The framework produces a distribution for V and exposes decision statistics such as:

$$\mathbb{P}(V > V_{\text{mkt}}), \qquad \mathbb{E}\left[\frac{V}{V_{\text{mkt}}} - 1\right], \qquad \mathbb{P}\left(\frac{V}{V_{\text{mkt}}} < \tau\right),$$

for a user-specified downside threshold τ. Through the equity bridge of section 4.8, the same statistics are exposed in per-share space, e.g. $\mathbb{P}(p^\star > p_{\text{mkt}})$, which is the statement an end user actually acts on.

4 Model Specification

4.1 Stochastic cash flows on natural support

Let $CF_0 > 0$ be the latest observed period cash flow (or free cash flow). Model cash flows by log-increments:

$$\log CF_i = \log CF_0 + \sum_{k=1}^{i} \ell_k, \qquad \ell_k = \log(1+g_k), \qquad \ell_k \sim \mathcal{N}(\mu_\ell, \sigma_\ell^2), \tag{1}$$

ensuring $CF_i > 0$ by construction. Regime conditioning can be introduced by letting (μ_ℓ, σ_ℓ) depend on a discrete regime variable or macro state. Firms with $CF_0 \leq 0$ are handled by the revenue–margin decomposition of section 4.5.

4.2 Discount rate as a WACC with interpretable components

The valuation functional of eq. (9) discounts *enterprise* cash flows; the appropriate rate is therefore a weighted average cost of capital, not a cost of equity. Construct it from components with operational meaning:

$$r = w_E\, r_E + w_D\, r_D\,(1-\theta), \qquad r_E = r_f + \beta \cdot \mathrm{ERP} + s_{\text{size}} + s_{\text{country}} + \varepsilon_E, \qquad r_D = r_f + s_{\text{credit}}, \tag{2}$$

with market-value weights $w_E + w_D = 1$ and effective tax rate θ. Leverage enters through the weights, the credit spread s_{credit}, and the levered beta — not as an additive equity premium, which would double-count it. Each component can be forecast by a low-dimensional time-series model or treated as scenario-driven with priors anchored to historical regimes. This decomposition supports sensitivity analysis and governance: stakeholders can contest specific components rather than the entire model. The base specification uses a single r per draw across the horizon; if component forecasts imply a material term structure, a per-period r_i replaces the flat rate, and the flattening assumption must otherwise be stated explicitly. All components and growth priors are expressed in nominal terms; mixing real growth with nominal rates is the most common silent DCF error and is treated as a validation failure.

4.3 Terminal value with economically necessary constraint

For a Gordon-growth perpetuity:

$$TV = \frac{CF_{n+1}}{r - g_{\text{term}}}, \qquad CF_{n+1} = CF_n(1+g_{\text{term}}), \qquad r > g_{\text{term}}. \tag{3}$$

The inequality $r > g_{\text{term}}$ is not a stylistic choice; it is required for a finite perpetuity value and for preventing sign inversions in the denominator.

4.4 Terminal value via value-driver fade

The Gordon perpetuity treats g_{term} as a free parameter, but growth is only valuable when the return on incremental invested capital exceeds the discount rate [6]. The value-driver formulation makes this economics explicit:

$$TV = \frac{\mathrm{NOPAT}_{n+1}\left(1 - \dfrac{g_{\mathrm{term}}}{\mathrm{ROIC}_\infty}\right)}{r - g_{\mathrm{term}}}, \qquad 0 \le g_{\mathrm{term}} < r, \qquad g_{\mathrm{term}} \le \mathrm{ROIC}_\infty. \tag{4}$$

Equation (4) encodes two economically necessary properties that the plain perpetuity hides: (i) if $\mathrm{ROIC}_\infty = r$, growth adds no value and TV collapses to the no-growth perpetuity NOPAT_{n+1}/r regardless of g_{term}; (ii) the implied reinvestment rate $g_{\mathrm{term}}/\mathrm{ROIC}_\infty$ is bounded by one, preventing the implicit assumption of value-creating growth financed from nothing.

Rather than sampling ROIC_∞ directly, model competitive erosion over the explicit horizon by mean reversion toward the cost of capital plus a sector-specific persistent spread s_∞:

$$\mathrm{ROIC}_{i+1} = \mathrm{ROIC}_i + \phi\big((r + s_\infty) - \mathrm{ROIC}_i\big) + \eta_i, \qquad \eta_i \sim \mathcal{N}(0, \sigma_\eta^2), \tag{5}$$

with fade speed $\phi \in (0,1]$ and s_∞ pooled hierarchically by sector. This replaces a fragile scalar prior with a structured process whose parameters are contestable component by component, consistent with the governance design of eq. (2).

4.5 Loss-making firms: revenue–margin decomposition

The log-increment model of eq. (1) requires $CF_0 > 0$ and would silently exclude loss-making firms — a large and valuation-relevant slice of any realistic universe. Decompose instead:

$$CF_i = R_i \cdot m_i, \tag{6}$$

where revenue $R_i > 0$ follows the log-increment process of eq. (1) (revenue is on natural positive support even when cash flow is not), and the margin m_i follows a bounded mean-reverting process via a logistic transform of a latent state:

$$m_i = m_{\min} + (m_{\max} - m_{\min})\,\sigma(x_i), \qquad x_{i+1} = x_i + \phi_m\,(x^\star - x_i) + \zeta_i, \qquad \zeta_i \sim \mathcal{N}(0, \sigma_x^2), \tag{7}$$

with $\sigma(u) = 1/(1+e^{-u})$, bounds $m_{\min} < 0 < m_{\max}$ calibrated per sector, and long-run anchor $x^\star$ implied by the pooled sector margin. The current (possibly negative) margin initializes x_0 through the inverse transform. Valuation is then well defined for firms with $CF_0 \le 0$, and the probability of never reaching profitability over the horizon, $\mathbb{P}(\max_i m_i \le 0)$, becomes a reportable risk statistic in its own right.

4.6 Dependence modeling via a heavy-tailed copula

Let the driver vector be $\mathbf{Z} = (\ell_{1:n}, g_{\mathrm{term}}, r)$. Specify marginals as above and bind them through a t-copula to represent tail dependence and regime co-movement:

$$(U_1, \ldots, U_d) \sim C_\nu(\cdot; \Sigma), \qquad Z_j = F_j^{-1}(U_j), \tag{8}$$

where F_j are marginal CDFs and Σ is estimated by sector and regime with shrinkage or hierarchical pooling to reduce finite-sample noise.

4.7 Valuation functional

For horizon n, enterprise value is

$$V = \sum_{i=1}^{n} \frac{CF_i}{(1+r)^i} + \frac{TV}{(1+r)^n}. \tag{9}$$

4.8 Equity bridge and per-share distribution

The framework outputs enterprise value, but decisions are taken in equity and per-share space. The bridge must itself be distributional where its components are uncertain:

$$E = V - \mathrm{ND} - \mathrm{MI} - \mathrm{PD} + \mathrm{NOA}, \tag{10}$$

with net debt ND, minority interests MI, pension and lease-adjusted deficits PD, and non-operating assets NOA, each taken from point-in-time fundamentals with explicit adjustment rules recorded in the audit trail. Per-share value uses fully diluted shares n_{fd} via the treasury-stock method evaluated *per draw*; because dilution from options depends on the sampled per-share value, this requires a fixed-point evaluation within each draw:

$$p^{\star(j)} = \frac{E^{(j)}}{n_{\mathrm{fd}}\left(p^{\star(j)}\right)}. \tag{11}$$

The headline decision statistics then carry over to per-share space.

5 Uncertainty Decomposition: Parameter and Path Uncertainty

Sampling paths under a single point estimate of the model parameters $\theta = (\mu_\ell, \sigma_\ell, \Sigma, \nu, \text{marginal parameters})$ conflates two distinct sources of uncertainty: *aleatoric* (path-level) variability conditional on θ, and *epistemic* uncertainty about θ itself. Intervals computed from path variability alone are systematically too narrow, and the calibration diagnostics of section 10 will detect this as under-coverage.

5.1 Two-level Monte Carlo

Let $p(\theta \mid \mathscr{D})$ denote the posterior over parameters given the estimation data $\mathscr{D}$ (hierarchical pooling across sector peers enters here as the prior structure). The engine becomes:

1. **Outer loop:** draw $\theta^{(m)} \sim p(\theta \mid \mathscr{D})$, $m = 1, \ldots, M$.

2. **Inner loop:** for each $\theta^{(m)}$, run the path-level sampler of algorithm 1 to obtain $\{V^{(m,j)}\}_{j=1}^{N_{\text{in}}}$.

The full predictive distribution is the mixture over (m, j). When a full posterior is impractical, a bootstrap-over-estimation-windows approximation of $p(\theta \mid \mathscr{D})$ preserves the decomposition at lower implementation cost; the substitution must be recorded in the model configuration.

5.2 Variance decomposition as a model-risk report

The law of total variance separates the two layers:

$$\operatorname{Var}(V) = \underbrace{\mathbb{E}_\theta\big[\operatorname{Var}(V \mid \theta)\big]}_{\text{path (aleatoric)}} + \underbrace{\operatorname{Var}_\theta\big(\mathbb{E}[V \mid \theta]\big)}_{\text{parameter (epistemic)}} . \tag{12}$$

The second term is the operational definition of *model risk* for this framework and is reported alongside every valuation. A large epistemic share indicates that the dominant uncertainty is estimation-driven and cannot be reduced by increasing the inner sample size; it is reducible only by data, pooling, or structural priors. Both terms are persisted per run (section 9) so that interval statements are reproducible.

6 Monte Carlo Engine: Sampling and Variance Control

The Monte Carlo engine produces samples $\{V^{(j)}\}_{j=1}^{N}$ using variance reduction (e.g., Latin Hypercube Sampling and antithetic variates). Convergence monitoring should be based on standard errors of distributional quantities (selected quantiles, exceedance probabilities), not only the mean.

6.1 Constraint-consistent sampling

Let $\mathbf{Z} = (\ell_{1:n}, g_{\text{term}}, r)$ with joint law P induced by the copula and marginals, and let $\mathscr{A} = \{\mathbf{z} : r > g_{\text{term}}\}$ be the admissible region. The economically meaningful target is the *truncated joint* $P_{\mathscr{A}}(\cdot) = P(\cdot \cap \mathscr{A})/P(\mathscr{A})$.

Why pair-only resampling is biased. A tempting implementation resamples only the pair (g_{term}, r) upon violation, while retaining the already-drawn $\ell_{1:n}$. This does *not* target $P_{\mathscr{A}}$: it conditions the pair on the constraint independently of the growth coordinates, severing precisely the tail dependence between growth and rates that the t-copula was introduced to represent. The distortion is concentrated in the region (high growth, low rates) where valuation outcomes are most extreme; the resulting bias is therefore not benign.

Correct schemes.

- **Full-vector rejection.** Draw the entire vector $\mathbf{U} \sim C_\nu(\cdot;\Sigma)$, map through the marginal quantiles, and accept if and only if $\mathbf{Z} \in \mathscr{A}$. Accepted draws follow $P_{\mathscr{A}}$ exactly. Cost: expected $1/p_{\mathscr{A}}$ proposals per accepted sample, where $p_{\mathscr{A}} = \mathbb{P}(\mathbf{Z} \in \mathscr{A})$.

- **Truncated conditional (Gibbs-type).** Draw $\ell_{1:n}$ and g_{term} from their joint, then draw r from its conditional distribution under the copula restricted to $(g_{\text{term}}, \infty)$ via the conditional quantile transform. Efficient when the violation probability is high; the conditional CDF is available in closed form for the t-copula.

Interaction with variance reduction. Rejection invalidates Latin Hypercube stratification: the accepted subset of an LHS design is no longer marginally stratified. Two remedies: (i) apply LHS on the conditional-quantile parameterization of the truncated joint, so that every stratified point is admissible by construction; or (ii) oversample by a factor $\approx 1/\hat{p}_{\mathscr{A}}$ and accept that stratification holds only approximately, reporting the realized effective sample size.

Acceptance rate as a misspecification alarm. The acceptance rate $\hat{p}_{\mathscr{A}}$ is logged per run. A low value means the prior joint places substantial mass on non-economic states $(r \le g_{\text{term}})$ — this is a statement about model misspecification, not a sampling nuisance, and triggers review of the rate and terminal-growth marginals before any valuation output is released.

Algorithm 1 Constraint-Consistent Copula Monte Carlo Valuation (full-vector rejection)

Inputs: CF_0, horizon n, marginals $F_\ell, F_{g_{\text{term}}}, F_r$, copula $C_\nu(\cdot;\Sigma)$, market benchmark V_{mkt}, sample size N, admissible region $\mathscr{A}$.
$j \leftarrow 0$; $\quad n_{\text{prop}} \leftarrow 0$
while $j < N$ **do**
  Draw full vector $\mathbf{U} \sim C_\nu(\cdot;\Sigma)$; $\quad n_{\text{prop}} \leftarrow n_{\text{prop}} + 1$
  Map $\mathbf{Z} \leftarrow \left(F_\ell^{-1}(U_{1:n}),\, F_{g_{\text{term}}}^{-1}(U_{n+1}),\, F_r^{-1}(U_{n+2})\right)$
  if $\mathbf{Z} \in \mathscr{A}$ **then**
    $j \leftarrow j + 1$
    Compute $CF_i^{(j)}$ via $\log CF_i^{(j)} = \log CF_0 + \sum_{k=1}^{i} \ell_k^{(j)}$
    Compute $TV^{(j)}$ by eq. (3) or eq. (4), and $V^{(j)}$ by eq. (9)
  end if
end while
Persist $\hat{p}_{\mathscr{A}} = N/n_{\text{prop}}$ with the run artifacts; alert if below threshold.
Report: exceedance probability $\mathbb{P}(V > V_{\text{mkt}})$, expected relative value $\mathbb{E}[V/V_{\text{mkt}} - 1]$, tail probabilities, and calibrated prediction intervals.

6.2 Tail probabilities via importance sampling

Decision statistics such as the downside probability $p_\tau = \mathbb{P}(V/V_{\text{mkt}} < \tau)$ concern rare events for which plain Monte Carlo standard errors scale poorly: estimating $p_\tau \approx 10^{-3}$ to $10\,\%$ relative error requires on the order of 10^7 plain samples. Use importance sampling on the Gaussianized copula factors instead.

Write the t-copula draw as $\mathbf{U} = T_\nu\left(\mathbf{Y}/\sqrt{W/\nu}\right)$ componentwise, with $\mathbf{Y} \sim \mathcal{N}(\mathbf{0}, \Sigma)$ and $W \sim \chi^2_\nu$. Tilt the Gaussian factor toward the stress direction $\mathbf{v}$ (low growth, high rates), i.e. sample $\mathbf{Y} \sim \mathcal{N}(c\,\mathbf{v}, \Sigma)$, and weight each draw by the likelihood ratio

$$w^{(j)} = \frac{\varphi_\Sigma(\mathbf{Y}^{(j)})}{\varphi_\Sigma(\mathbf{Y}^{(j)} - c\,\mathbf{v})}, \qquad \hat{p}_\tau = \frac{1}{N}\sum_{j=1}^{N} \mathbb{1}\{V^{(j)}/V_{\text{mkt}} < \tau\}\, w^{(j)}. \tag{13}$$

The tilt magnitude c is selected by pilot runs to roughly center the event; the effective sample size $\text{ESS} = \left(\sum_j w^{(j)}\right)^2 / \sum_j (w^{(j)})^2$ is reported with every tail estimate and persisted alongside the run artifacts. The same machinery serves stress reporting: the tilted runs *are* the scenario engine, with $\mathbf{v}$ drawn from named historical regimes.

7 Risk Monitoring: Robust Multivariate Anomaly Signals

This module is not a replacement for valuation; it is a *nowcasting* layer that detects abnormal market microstructure and risk conditions that should influence short-horizon priors and position sizing.

7.1 Robust standardization under nonstationarity

For an indicator series X_t, two robust standardizations are common:

$$Z_X^{\text{rob}}(t) = \frac{X_t - \text{Med}(X)}{1.4826 \cdot \text{MAD}(X)}, \qquad Z_X^{\text{ewma}}(t) = \frac{X_t - \mu_t}{\sigma_t},$$

where (μ_t, σ_t^2) follow EWMA updates. The choice is empirical: robust location/scale handles outliers; EWMA adapts to volatility clustering.

7.2 Multivariate distance with robust covariance

Stack standardized indicators into $\mathbf{X}_t \in \mathbb{R}^d$. Compute:

$$D_t^2 = (\mathbf{X}_t - \hat{\mu})^\top \hat{\Sigma}_R^{-1} (\mathbf{X}_t - \hat{\mu}), \tag{14}$$

with $\hat{\Sigma}_R$ estimated via robust methods (e.g., Minimum Covariance Determinant). Thresholding by χ^2_d is a *nominal* reference only; for MCD-based distances the Hardin–Rocke F-approximation [5] is the appropriate finite-sample reference, and even that should be empirically calibrated on rolling windows to reflect heavy tails and autocorrelation.

7.3 Composite alert score

For interpretability, map standardized deviations to a bounded score. Define per-indicator activations anchored at zero,

$$a_i(t) = \frac{\sigma\big(\kappa(|Z_i(t)| - z_0)\big) - \sigma(-\kappa z_0)}{1 - \sigma(-\kappa z_0)} \in [0,1), \tag{15}$$

so that $a_i(t) = 0$ when $Z_i(t) = 0$, and aggregate with normalized weights:

$$A(t) = \frac{\sum_{i=1}^{m} w_i \rho_i(t)\, a_i(t)}{\sum_{i=1}^{m} w_i \rho_i(t)} \in [0,1], \tag{16}$$

where $\sigma(u) = 1/(1+e^{-u})$ and $\rho_i(t)$ is a recency function defined by the operational sampling cadence. The normalization guarantees boundedness and a zero baseline by construction, properties an unnormalized weighted sum does not have. Parameters are selected by walk-forward optimization against explicit targets (e.g., next-h-day drawdowns), with nested cross-validation to prevent leakage.

8 Sentiment: Aggregation, Calibration, and Fundamental Linkage

Sentiment is treated as a noisy measurement process, not an oracle. The objective is to incorporate text-derived signals without duplicating price dynamics.

8.1 Aggregation with credibility and recency as estimable quantities

Let sources $i = 1, \ldots, k$ emit raw sentiment $s_i \in [-1, 1]$ at times t_i. Aggregate by:

$$S_{\text{agg}}(t) = \frac{\sum_{i=1}^{k} c_i r_i(t)\, s_i}{\sum_{i=1}^{k} c_i r_i(t)}, \qquad r_i(t) = \exp\left(-\frac{t - t_i}{\tau}\right), \tag{17}$$

where credibility weights c_i and decay τ are learned out-of-sample with timestamp alignment to the trading calendar and deduplication of cross-posted content.

8.2 Calibration to probabilistic targets

Define a prediction target (e.g., h-day excess return vs. sector, or probability of a drawdown event). Calibrate S_{agg} using proper scoring objectives and post-hoc calibration (e.g., isotonic regression) on strictly out-of-sample folds. The calibrated output is a probability statement with measurable reliability.

8.3 Structural linkage to cash-flow priors

To avoid price-signal duplication, sentiment updates cash-flow growth priors — but only at short horizons. A uniform shift across all periods would propagate a transient text signal into the terminal value, contradicting the short-horizon intent. Use a horizon-decaying coefficient:

$$\ell_k \mid S_{\text{agg}}(t) \sim \mathcal{N}\big(\mu_\ell + \gamma_k S_{\text{agg}}(t),\, \sigma_\ell^2\big), \qquad \gamma_k = \gamma e^{-\lambda(k-1)}, \tag{18}$$

where γ and the decay rate $\lambda > 0$ are estimated out-of-sample. The terminal-growth prior receives no sentiment update by construction. This preserves interpretability: text signals move *near-term growth assumptions*, not prices directly and not the perpetuity.

9 System Architecture and Governance

9.1 Modules

A production-ready implementation separates concerns into four modules:

- **Valuation Engine:** distributional valuation (algorithm 1), two-level uncertainty decomposition (section 5), diagnostics, sensitivity analysis.
- **Risk Monitor:** robust multivariate anomaly signals (section 7) and regime labeling.
- **Sentiment Engine:** ingestion, normalization, calibration, and prior updates (section 8).
- **Peer/Pooling Layer:** sector priors, hierarchical pooling for copula parameters, rate components, fade spreads, and margin anchors.

9.2 Auditability and reproducibility

A valuation output is only operationally meaningful if it is reproducible. Each run should persist:

- data snapshots (point-in-time fundamentals, adjusted prices, corporate actions metadata),
- model configuration and priors (including copula Σ and marginal parameters, or the posterior approximation used),
- RNG seed material and sampling method identifiers,
- model/version identifiers (content-hash or semantic version),
- sampling diagnostics: constraint acceptance rate $\hat{p}_{\mathscr{A}}$, importance-sampling effective sample size, and the variance decomposition of eq. (12),
- the sensitivity attribution of section 10.5.

Operational settings (caching, concurrency, refresh cadence) are deployment parameters and must be justified by measurement (latency, cost, data freshness) rather than asserted as fixed numeric constants.

10 Evaluation Plan

10.1 Protocol

Use rolling-origin (walk-forward) evaluation with point-in-time data to avoid look-ahead bias. The universe definition (e.g., top-N by market cap, reconstituted periodically) must be explicit, and survivorship bias must be addressed via historical constituents. Corporate actions and restatements must be handled with point-in-time revision logic.

10.2 From valuation gap to returns: the convergence model

Intrinsic value V is never observed, so a predictive distribution over V cannot be scored directly against a realization: proper scoring rules require an observable target. The bridge is a *convergence model* from the valuation gap to realized returns [3]. Let $\alpha_t = V_t/V_{\mathrm{mkt},t} - 1$ denote the model gap and $y_{t,h}$ the realized h-period excess return versus the sector. Posit:

$$y_{t,h} = a_h + \pi_h\,\alpha_t + \varepsilon_{t,h}, \tag{19}$$

where the convergence coefficient $\pi_h \in [0,1]$ measures the fraction of the gap realized over horizon h, estimated strictly out-of-sample with overlapping-window corrections. π_h is itself a reportable scientific quantity: $\pi_h \approx 0$ falsifies the operational usefulness of the valuation signal regardless of how well calibrated the V-distribution appears internally.

The predictive *return* distribution is obtained by pushing the sampled gap distribution through eq. (19) with parameter uncertainty from section 5. This distribution — not the distribution of V — is the object scored below and the object on which the policy operates.

10.3 Metrics for probabilistic forecasts

For the predictive return CDF $\hat{F}_t$ and realized target $y_{t,h}$, report:

- **Proper scoring rules:** CRPS, log-score where applicable [4].
- **Calibration:** PIT histograms, reliability curves for probability statements.
- **Interval validity:** prediction interval coverage probability (PICP) and mean interval width (MPIW).

$$\mathrm{CRPS}(\hat{F}_t, y_t) = \int_{-\infty}^{\infty} \left(\hat{F}_t(v) - \mathbb{1}\{v \geq y_t\}\right)^2 dv. \tag{20}$$

10.4 Policy evaluation under a fixed rule

Define a *fixed* decision policy before testing and report realized outcomes strictly out-of-sample. Let Y denote the predicted h-period return under eq. (19). The policy is a constrained Kelly-style allocation:

$$\begin{aligned} f^{\star} &= \arg \max_{0 \le f \le f_{\max}} \mathbb{E}\big[\log(1+fY)\big], \\ \text{s.t.} \quad & \mathbb{P}(Y < -L) \le \varepsilon, \qquad f_{\max} < 1/|y_{\min}|, \end{aligned} \tag{21}$$

where $y_{\min}$ bounds the worst-case return in the predictive support, so that $\log(1+fY)$ is defined almost surely — a domain condition that a gap-based formulation omits. The probabilistic constraint expresses tail control directly. Fractional Kelly ($\lambda f^{\star}$, $\lambda \in (0,1]$) is the recommended default given estimation error in both the gap distribution and π_h; the choice of λ is part of the pre-registered policy, not a post-hoc tuning parameter. Portfolio metrics (CAGR, volatility, drawdown, turnover) should be reported with uncertainty via block bootstrap under realistic transaction cost and liquidity assumptions.

10.5 Sensitivity attribution under dependence

Sensitivity reporting answers the governance question "which assumption drives this valuation?" Classical Sobol indices decompose $\mathrm{Var}(V)$ over inputs but assume independent inputs — an assumption the copula deliberately violates. Two consistent options:

- **Sobol on the uniform scale.** Compute indices with respect to the independent uniforms *before* the copula transform. Mathematically valid, but it attributes variance to abstract coordinates rather than to economic quantities, weakening interpretability.
- **Shapley effects.** Allocate $\mathrm{Var}(V)$ across the economic inputs $(\ell_{1:n}, g_{\text{term}}, r, \dots)$ by the Shapley value of the variance-explained game [7]. Shapley effects are well defined under dependent inputs, sum exactly to $\mathrm{Var}(V)$, and are never negative — properties Sobol indices lose under dependence.

The recommended report is a Shapley decomposition of $\mathrm{Var}(V)$ at the economic-input level, computed per run and persisted with the audit trail. The expected qualitative finding — dominance of (r, g_{term}) through the terminal value — then carries a number, and changes in that number across runs become a drift diagnostic for the model itself.

10.6 Backtest overfitting controls

Walk-forward evaluation prevents look-ahead bias but not *selection* bias: a framework with tunable components (copula structure, alert-score parameters, sentiment calibration, policy thresholds) implicitly searches a configuration space, and the best out-of-sample result among many trials is itself an in-sample statistic over trials. Three controls are mandatory:

- **Trial registry.** Every evaluated configuration is hashed and logged before its results are seen, including abandoned trials. The number of effective trials K is a first-class reported quantity.

- **Deflated performance statistics.** Report the deflated Sharpe ratio [1], which adjusts the observed Sharpe for the expected maximum over K correlated trials and for non-normality of returns. An undeflated Sharpe from a multi-trial search is not reported.
- **Probability of backtest overfitting.** Estimate PBO via combinatorially symmetric cross-validation (CSCV) [2]: partition the sample into blocks, evaluate all configurations on each train/test combination, and measure how often the in-sample best ranks below median out-of-sample. A PBO materially above the acceptable threshold blocks promotion of the configuration to production, regardless of headline performance.

These controls extend the auditability principle of section 9 from *runs* to *research process*: the search itself is an object of governance.

11 Limitations and Roadmap

Key limitations:

- **Terminal value dominance:** long-horizon valuation is highly sensitive to (r, g_{term}); the value-driver fade of section 4.4 structures but does not eliminate this sensitivity, which the Shapley report quantifies per run.
- **Convergence nonstationarity:** the coefficient π_h can drift with market regimes and crowding; it must be monitored as a live diagnostic, not estimated once.
- **Structural breaks:** factor and spread processes can change; regime modeling and stress testing are necessary.
- **Copula estimation noise:** dependence estimates are fragile in small samples; hierarchical pooling and shrinkage should be standard.
- **Posterior approximation cost:** the two-level design multiplies compute; bootstrap approximations trade fidelity for tractability and must be flagged in the configuration.
- **Signal leakage:** sentiment and anomaly targets must be aligned to timestamps and trading calendars to avoid inadvertent look-ahead.

Roadmap extensions:

- regime-switching marginals for ℓ_k and rate components, with regime labels supplied by the risk monitor;
- vine copulas for higher-dimensional driver vectors where the single t-copula is too rigid;

- per-period discount rates from component term-structure forecasts;
- explicit model-risk surfaces: scenario stress tests built on the importance-sampling machinery of section 6.2.

12 Conclusion

This white paper specified a probabilistic valuation framework designed to make uncertainty operational. The core contribution is structural: model cash flows on their natural support, enforce economically necessary constraints without distorting the dependence structure, separate path uncertainty from parameter uncertainty, and integrate nowcasting signals through calibrated, auditable updates to short-horizon priors. The framework's falsifiable claim is not that any single valuation is correct, but that the valuation gap predicts returns at a measurable convergence speed π_h — and the evaluation target is the joint behavior of the implied return forecasts and a fixed policy under strict out-of-sample protocols.

References

[1] Bailey, D.H. and López de Prado, M. (2014) 'The deflated Sharpe ratio: correcting for selection bias, backtest overfitting, and non-normality', *Journal of Portfolio Management*, 40(5), pp. 94–107.

[2] Bailey, D.H., Borwein, J., López de Prado, M. and Zhu, Q.J. (2017) 'The probability of backtest overfitting', *Journal of Computational Finance*, 20(4), pp. 39–70.

[3] Frankel, R. and Lee, C.M.C. (1998) 'Accounting valuation, market expectation, and cross-sectional stock returns', *Journal of Accounting and Economics*, 25(3), pp. 283–319.

[4] Gneiting, T. and Raftery, A.E. (2007) 'Strictly proper scoring rules, prediction, and estimation', *Journal of the American Statistical Association*, 102(477), pp. 359–378.

[5] Hardin, J. and Rocke, D.M. (2005) 'The distribution of robust distances', *Journal of Computational and Graphical Statistics*, 14(4), pp. 928–946.

[6] Koller, T., Goedhart, M. and Wessels, D. (2020) *Valuation: Measuring and Managing the Value of Companies*. 7th edn. Hoboken, NJ: Wiley.

[7] Song, E., Nelson, B.L. and Staum, J. (2016) 'Shapley effects for global sensitivity analysis: theory and computation', *SIAM/ASA Journal on Uncertainty Quantification*, 4(1), pp. 1060–1083.

Citation. Russmann, M. (2026) 'Beyond the point estimate: making uncertainty operational in equity valuation'. Working paper, version 1.0, June 2026.

Status. Working paper circulated for discussion. Comments and corrections are welcome.