CR Nasrat Kamal REPORT

CQF Final Project

Pricing Basket Credit Default Swaps using Copula Models

k-th to Default Basket CDS Valuation

Nasrat Kamal

January 2026

1. Introduction

This report presents the **CQF Final Project** on **pricing k-th to default basket credit default swaps** using copula models. The project implements a Monte Carlo simulation framework to value these multi-name credit derivatives, comparing Gaussian and Student-t copulas for modelling default dependence across a basket of five reference entities.

A credit default swap (CDS) is a derivative contract that transfers credit risk between two parties. The protection buyer pays a periodic premium (the CDS spread) to the protection seller, who in return compensates the buyer if a specified reference entity experiences a credit event such as default or bankruptcy. A basket CDS extends this concept to multiple reference entities. The k-th to default variant triggers a payout only when the k-th credit event occurs within the basket. For example, in a first-to-default swap, the protection seller pays upon the first default among all referenced entities; in a fifth-to-default swap, four defaults must occur before the contract pays out.

Pricing a basket CDS requires valuing two cash flow streams: the premium leg and the protection leg. The protection buyer pays periodic premiums until either the contract matures or the k-th

default occurs, whichever comes first. The protection leg represents the contingent payment upon the k-th default, equal to the loss given default on the defaulted entity. The fair spread is the premium rate that equates the expected present value of these two legs. Unlike single-name CDS pricing where closed-form solutions exist, basket CDS pricing must account for the joint default behaviour of all reference entities. The k-th default time is an order statistic that depends on the entire correlation structure of the portfolio. Monte Carlo simulation provides a natural framework: we simulate correlated default times across all entities, identify the k-th default in each scenario, compute the resulting cash flows, and average across paths to obtain expected leg values. This simulation-based approach follows the methodology established by Hull and White (2004).

Modelling joint default behaviour requires specifying both marginal default distributions and the dependence structure linking them. **Copula theory** separates these components: Sklar's theorem establishes that any multivariate distribution can be decomposed into its marginals and a copula function capturing dependence (Nelsen, 2006). This allows calibrating individual default probabilities from single-name CDS spreads via hazard rate bootstrapping, while estimating correlations from historical data on credit spread movements. This project implements two copula families: the Gaussian copula, which became the industry standard following Li (2000), and the Student-t copula, which introduces heavier tails to better capture the clustering of defaults observed during credit crises (McNeil, Frey and Embrechts, 2015).

The objective is to price a 5th-to-default basket CDS on five financial sector names over a 5-year period — the most senior tranche where all five must default before payout. Several methodology choices arise: for random number generation, options include pseudo-random sampling, quasi-random sequences (Halton, Sobol), and variance reduction via antithetic variates. Rather than selecting arbitrarily, we first conduct convergence studies to evaluate RNG methods and determine the simulation count required for stable estimates. Based on these findings, we select Sobol sequences with 100,000 simulations for the main analysis. We then compute fair spreads under both copulas and perform sensitivity analyses on correlation, recovery rates, and the choice of k to validate model behaviour.

2. Methodology

2.1 Overall Framework

The pricing framework separates the modelling of individual default probabilities (marginals) from the dependence structure linking them. Each reference entity's default behaviour is characterised by a hazard rate term structure, calibrated independently from single-name CDS spreads. The

copula then joins these marginals into a joint distribution, capturing how defaults may cluster or occur independently.

The workflow proceeds in four stages:

1. Load historical spread data and current term structures
2. Calibrate marginal hazard rates via bootstrapping and estimate the copula correlation matrix from historical co-movements
3. Simulate correlated default times using the fitted copula
4. Price the basket CDS by computing expected premium and protection leg values across simulated scenarios

This approach follows the framework introduced by Li (2000) for credit portfolio modelling and extended by Hull and White (2004) for basket CDS valuation.

2.2 Credit Default Swaps

A single-name CDS provides protection against default by a specific reference entity. The protection buyer pays a periodic spread (quoted in basis points per annum) until maturity or default, whichever occurs first. Upon a credit event, the protection seller pays the buyer the loss given default—typically the notional amount minus the recovery value of the defaulted obligation. Pricing a CDS requires modelling the probability of default over time, which is captured through the hazard rate (Schönbucher, 2003).

The hazard rate $h(t)$ represents the instantaneous conditional probability of default at time t, given survival to that point. The survival probability to time T is related to the hazard rate by:

$$Q(T) = \exp\left(-\int_0^T h(s)\,ds\right)$$

For practical implementation, we assume piecewise-constant hazard rates between standard CDS tenors (1Y, 2Y, 3Y, 5Y, 7Y, 10Y). Under this assumption, the survival probability simplifies to:

$$Q(T) = \exp\left(-\sum_{i=1}^n h_i \Delta t_i\right)$$

where h_i is the constant hazard rate over interval i and Δt_i is the interval length.

Hazard Rate Bootstrapping

The hazard rates are derived from observed CDS spreads across the term structure. For a CDS with maturity T and observed spread s, the fair spread equates the expected present values of the premium and protection legs.

Under continuous premium payment and zero interest rates, the premium leg (per unit spread) is the expected survival time:

$$V_{prem}(T) = \int_0^T Q(t)\, dt$$

The protection leg pays $(1 - R)$ upon default:

$$V_{prot}(T) = (1 - R)[1 - Q(T)]$$

Setting $s \cdot V_{prem}(T) = V_{prot}(T)$ and rearranging gives the fair spread:

$$s = \frac{(1-R)[1-Q(T)]}{\int_0^T Q(t)\, dt}$$

For a constant hazard rate h, the survival probability is $Q(t) = e^{-ht}$, so $Q(T) = e^{-hT}$ and:

$$\int_0^T Q(t)\, dt = \int_0^T e^{-ht}\, dt = \frac{1-e^{-hT}}{h}$$

Substituting into the fair spread equation:

$$s = \frac{(1-R)[1-e^{-hT}]}{(1-e^{-hT})/h} = (1 - R) \cdot h$$

Rearranging:

$$h = \frac{s}{1-R}$$

This result is exact under the assumptions of constant hazard rate, zero interest rates, and continuous premium payment (ignoring accrued premium at default). For piecewise-constant hazard rates between tenors, we apply this formula to each CDS spread to obtain the average hazard rate for that maturity. The survival probability to any time T is then:

$$Q(T) = \exp\left(-\sum_{i=1}^k h_i \Delta t_i\right)$$

Recovery is typically assumed constant at 40% for senior unsecured debt, following market convention. This gives a loss given default (LGD) of 60%.

2.3 Copula Theory

A copula is a multivariate distribution function that links marginal distributions to form a joint distribution. Sklar's theorem (1959) establishes that any joint distribution $F(x_1, \ldots, x_n)$ with continuous marginals $F_1, \ldots, F_n$ can be written as:

$$F(x_1, \ldots, x_n) = C(F_1(x_1), \ldots, F_n(x_n))$$

where $C : [0, 1]^n \to [0, 1]$ is the copula function. This decomposition allows modelling marginal behaviour and dependence structure separately—a key advantage for credit portfolio modelling

where individual default probabilities are calibrated from single-name instruments while dependence is estimated from historical data. The theoretical foundations are covered extensively in Nelsen (2006) and Cherubini, Luciano and Vecchiato (2004).

Gaussian Copula

The Gaussian copula is derived from the multivariate normal distribution. For a correlation matrix Σ, the Gaussian copula is defined as:

$$C_\Sigma^{Ga}(u_1, \ldots, u_n) = \Phi_\Sigma(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_n))$$

where Φ^{-1} is the standard normal inverse CDF and Φ_Σ is the joint CDF of a multivariate normal with correlation matrix Σ. Simulation proceeds by:

1. Generating independent standard normals Z
2. Applying the Cholesky decomposition $Z_{corr} = LZ$ where $\Sigma = LL^T$
3. Transforming to uniforms via $U = \Phi(Z_{corr})$

The Gaussian copula exhibits zero tail dependence—the probability of joint extreme events vanishes in the limit. This property has been criticised as unrealistic for credit risk, where defaults tend to cluster during crises (McNeil, Frey and Embrechts, 2015).

Student-t Copula

The Student-t copula introduces an additional parameter ν (degrees of freedom) that controls tail thickness:

$$C_{\Sigma,\nu}^t(u_1, \ldots, u_n) = t_{\Sigma,\nu}(t_\nu^{-1}(u_1), \ldots, t_\nu^{-1}(u_n))$$

where t_ν^{-1} is the inverse CDF of a univariate t-distribution with ν degrees of freedom. Lower ν produces heavier tails and stronger tail dependence, meaning joint extreme events become more likely. As $\nu \to \infty$, the t-copula converges to the Gaussian copula.

Simulation requires an additional step:

1. Generate correlated normals Z_{corr} as above
2. Generate an independent chi-square variate $W \sim \chi_\nu^2$
3. Scale by $Y = Z_{corr}/\sqrt{W/\nu}$
4. Transform via $U = t_\nu(Y)$

Correlation Matrix Calibration

For elliptical copulas, the correlation matrix Σ must be estimated from data. The direct approach—transform uniform marginals to latent variables via inverse CDF, then compute Pearson correlation

—creates a circular dependency for the t-copula: estimating Σ requires knowing ν, but estimating ν requires knowing Σ.

Rank correlation resolves this. Spearman's ρ_S is invariant across elliptical copula families, and for elliptical distributions, it relates to the Pearson correlation of latent variables by:

$$\rho = 2\sin\left(\tfrac{\pi \rho_S}{6}\right)$$

This formula holds for both Gaussian and t-copulas, enabling unified calibration (McNeil, Frey and Embrechts, 2015). The degrees of freedom ν for the t-copula can then be estimated via maximum likelihood, or specified as a scenario parameter for sensitivity analysis.

Degrees of Freedom Estimation

With the correlation matrix Σ fixed from rank correlation, the degrees of freedom ν is estimated by maximising the t-copula log-likelihood. For observations $\mathbf{u}_1, \ldots, \mathbf{u}_T$ of uniform marginals, we first transform to t-distributed latent variables:

$$\mathbf{x}_t = (t_\nu^{-1}(u_{t,1}), \ldots, t_\nu^{-1}(u_{t,n}))$$

The t-copula density is:

$$c(\mathbf{u}; \Sigma, \nu) = \frac{f_{\Sigma,\nu}(\mathbf{x})}{\prod_{j=1}^{n} f_\nu(x_j)}$$

where $f_{\Sigma,\nu}$ is the multivariate t-density and f_ν is the univariate t-density. The log-likelihood function is:

$$\ell(\nu) = \sum_{t=1}^{T} \log c(\mathbf{u}_t; \Sigma, \nu)$$

Expanding the densities, this becomes:

$$\ell(\nu) = T \log \Gamma\left(\tfrac{\nu+n}{2}\right) - T \log \Gamma\left(\tfrac{\nu}{2}\right) - \tfrac{T}{2} \log |\Sigma| - \tfrac{\nu+n}{2} \sum_{t=1}^{T} \log\left(1 + \tfrac{\mathbf{x}_t^\top \Sigma^{-1} \mathbf{x}_t}{\nu}\right)$$
$$+ \tfrac{\nu+1}{2} \sum_{t=1}^{T} \sum_{j=1}^{n} \log\left(1 + \tfrac{x_{t,j}^2}{\nu}\right) - nT \log \Gamma\left(\tfrac{\nu+1}{2}\right) + nT \log \Gamma\left(\tfrac{\nu}{2}\right)$$

The MLE $\hat{\nu}$ is found by numerical optimisation:

$$\hat{\nu} = \arg\max_{\nu > 2} \ell(\nu)$$

The constraint $\nu > 2$ ensures finite variance. In practice, we evaluate the log-likelihood over a grid of ν values (e.g., 2.5 to 30) and select the maximiser, with optional refinement via gradient-based optimisation. This profile likelihood approach—fixing Σ and optimising over ν alone—avoids the joint optimisation problem and produces stable estimates.

2.4 Monte Carlo Simulation

The Monte Carlo framework generates N independent scenarios, each producing a joint realisation of default times across all reference entities. The theoretical foundations for Monte Carlo methods in finance are established in Glasserman (2003). For each scenario i, the simulation proceeds in three steps.

Step 1: Generate correlated uniforms. Using the calibrated copula, we simulate a vector of correlated uniform random variables $(U_1^{(i)}, \ldots, U_n^{(i)})$ for the n reference entities. These uniforms encode the dependence structure—high correlation in the copula produces uniforms that tend to be jointly high or jointly low.

Step 2: Invert to default times. Each uniform $U_j^{(i)}$ is transformed to a default time $\tau_j^{(i)}$ using the entity's marginal survival function. Since $Q(\tau) = P(\tau > t)$ represents the probability of surviving beyond time t, and U_j is uniform on $[0, 1]$, we set:

$$\tau_j^{(i)} = Q_j^{-1}(1 - U_j^{(i)})$$

Under the piecewise-constant hazard rate assumption, this inversion is performed analytically. If $1 - U_j$ falls within the survival probability range corresponding to interval $[t_{k-1}, t_k]$, the default time is:

$$\tau_j = t_{k-1} + \frac{1}{h_k}\ln\left(\frac{Q(t_{k-1})}{1-U_j}\right)$$

If $1 - U_j > Q(T)$ where T is the maximum tenor, default occurs beyond the modelled horizon and is treated as no default (i.e., $\tau_j = \infty$).

Step 3: Extract the k-th default time. The simulated default times $(\tau_1^{(i)}, \ldots, \tau_n^{(i)})$ are sorted in ascending order. The k-th order statistic $\tau_{(k)}^{(i)}$ represents the time of the k-th default in scenario i. For a 5th-to-default basket with five entities, this is simply the maximum default time—all five must default for the contract to trigger.

After N simulations, we have a distribution of k-th default times from which expected cash flows are computed.

2.5 Random Number Generation

The quality of Monte Carlo estimates depends on the random number generation method. This project implements four approaches to study their convergence properties, following the variance reduction techniques described in Glasserman (2003).

Pseudo-random (standard)

Standard pseudo-random numbers are generated via NumPy's Mersenne Twister algorithm. The Monte Carlo estimator converges at rate $O(1/\sqrt{N})$—halving the standard error requires quadrupling the number of simulations. This rate is independent of dimension, making pseudo-random sampling robust for high-dimensional problems.

Quasi-random sequences (Halton, Sobol)

Quasi-random or low-discrepancy sequences fill the sample space more uniformly than pseudo-random numbers, avoiding the clustering and gaps inherent in random sampling (Glasserman, 2003, Ch. 5). The theoretical convergence rate is $O((\log N)^d/N)$, which is faster than pseudo-random for moderate dimensions d and smooth integrands. Sobol sequences generally outperform Halton sequences in higher dimensions due to better uniformity properties (Niederreiter, 1992). Both sequences are scrambled (randomised) to enable unbiased error estimation and avoid correlation artefacts.

Antithetic variates

Antithetic variates reduce variance by introducing negative correlation between paired samples. For each base sample $U \sim \mathrm{Uniform}(0, 1)$, we also use $1 - U$, which is equally valid as a uniform sample but negatively correlated with U. When estimating $\mathbb{E}[f(U)]$, this negative correlation reduces variance provided $\mathrm{Cov}(f(U), f(1 - U)) < 0$, which holds when f is monotonic.

For normal random variables, the analogous construction uses Z and $-Z$: both are valid $N(0, 1)$ samples with perfect negative correlation. The variance reduction depends on the payoff structure —for monotonic payoffs, antithetic variates can halve the variance at no additional simulation cost.

Limitation for k-th to default payoffs

Antithetic variates require negative correlation between paired payoffs to achieve variance reduction. For k-th to default pricing, the payoff depends on an order statistic of multiple correlated default times. When we flip $Z \to -Z$ for all entities, each individual default time changes, but the relative ordering does not flip predictably—the entity that was k-th to default may remain k-th or shift arbitrarily. This sorting operation destroys the monotonic relationship required for effective variance reduction, as confirmed by the convergence analysis in Section 4.1.

2.6 Pricing

The fair spread of a k-th to default basket CDS is the premium rate that equates the expected present values of the protection and premium legs at inception. Let $\tau_{(k)}$ denote the k-th default time and T the contract maturity.

Protection Leg

The protection leg pays $(1 - R)$ times the notional upon the k-th default, where R is the recovery rate. The expected present value is:

$$V_{prot} = (1 - R) \cdot \mathbb{E}\left[D(\tau_{(k)}) \cdot \mathbf{1}_{\{\tau_{(k)} \leq T\}}\right]$$

where $D(t)$ is the discount factor to time t and $\mathbf{1}_{\{\cdot\}}$ is the indicator function. Under the simplifying assumption of zero interest rates (i.e., $D(t) = 1$ for all t), this reduces to:

$$V_{prot} = (1 - R) \cdot P(\tau_{(k)} \leq T)$$

The Monte Carlo estimate is:

$$\hat{V}_{prot} = \frac{1-R}{N} \sum_{i=1}^{N} \mathbf{1}_{\{\tau_{(k)}^{(i)} \leq T\}}$$

Premium Leg

The premium leg pays the spread s continuously (or at discrete intervals) until the earlier of the k-th default or maturity. The expected present value per unit spread is:

$$V_{prem} = \mathbb{E}\left[\int_0^{\min(\tau_{(k)}, T)} D(t)\, dt\right]$$

> **Limitation:** This implementation assumes zero interest rates throughout, omitting the impact of discounting on cash flow valuations. This simplifies the pricing formulae and allows focus on the credit risk modelling aspects. In practice, stochastic or deterministic discount factors would be incorporated.

Under zero interest rates, this simplifies to the expected duration of premium payments:

$$V_{prem} = \mathbb{E}\left[\min(\tau_{(k)}, T)\right]$$

The Monte Carlo estimate is:

$$\hat{V}_{prem} = \frac{1}{N} \sum_{i=1}^{N} \min(\tau_{(k)}^{(i)}, T)$$

Fair Spread

The fair spread s^* sets the contract value to zero at inception, requiring $s^* \cdot V_{prem} = V_{prot}$. Thus:

$$s^* = \frac{V_{prot}}{V_{prem}} = \frac{(1-R) \cdot P(\tau_{(k)} \leq T)}{\mathbb{E}[\min(\tau_{(k)}, T)]}$$

The Monte Carlo estimate is the ratio of sample means:

$$\hat{s}* = \frac{\hat{V}_{prot}}{\hat{V}_{prem}}$$

Standard Error Estimation

Monte Carlo estimates are subject to sampling error, and quantifying this uncertainty is essential for assessing the reliability of pricing results. The standard error allows us to construct confidence intervals around the fair spread estimate and determine whether differences between copula models or parameter choices are statistically meaningful.

The fair spread is a ratio of two random variables, so its standard error requires the delta method (Glasserman, 2003). For a ratio $\hat{\theta} = \hat{X}/\hat{Y}$, the approximate variance is:

$$\mathrm{Var}(\hat{\theta}) \approx \theta^2 \left[\frac{\mathrm{Var}(\hat{X})}{\mu_X^2} + \frac{\mathrm{Var}(\hat{Y})}{\mu_Y^2} - \frac{2\mathrm{Cov}(\hat{X}, \hat{Y})}{\mu_X \mu_Y} \right]$$

where $\mu_X = \mathbb{E}[X]$ and $\mu_Y = \mathbb{E}[Y]$. This provides 95% confidence intervals for the fair spread estimate.

2.7 Implementation

The methodology described above is implemented in Python across several modules. The table below maps each methodology component to its corresponding source code. For full details of the project architecture, module interfaces, and notebook workflow, see `README.md`.

Methodology Component	Source Module	Key Classes/Functions
Hazard rate bootstrapping	`src/bootstrapper.py`	`HazardRateBootstrapper`
Gaussian copula	`src/copula.py`	`GaussianCopula`
Student-t copula	`src/copula.py`	`TCopula`
Correlation calibration	`src/copula.py`	`fit()` method with rank correlation
t-copula MLE for ν	`src/copula.py`	`TCopula.fit()`, `log_likelihood()`
Random number generation	`src/rng.py`	`RNGEngine`
Pseudo-random sampling	`src/rng.py`	`RNGEngine(method='pseudo')`
Sobol sequences	`src/rng.py`	`RNGEngine(method='sobol')`
Halton sequences	`src/rng.py`	`RNGEngine(method='halton')`
Antithetic variates	`src/rng.py`	`RNGEngine(method='antithetic')`
Fair spread calculation	`src/pricer.py`	`BasketCDSPricer`
Default time simulation	`src/pricer.py`	`simulate_default_times()`
Premium/protection legs	`src/pricer.py`	`price()` method

The main analysis is orchestrated through Jupyter notebooks: `cds_pricing.ipynb` for the core pricing workflow, `pricing_orchestration.ipynb` for copula comparison, and `sensitivity_orchestration.ipynb` for parameter sensitivity studies. These notebooks use `papermill` for parameterised execution and `scrapbook` for results collection across multiple runs.

3. Data Sources

Pricing a basket CDS requires two categories of data: (1) CDS term structures at the pricing date for bootstrapping marginal hazard rates, and (2) historical time series data to estimate the copula correlation structure.

CDS spread data presents significant sourcing challenges for academic projects. Historical CDS quotes and term structures are typically available only through expensive commercial data providers such as Bloomberg, Markit, or Refinitiv. Even when accessible, CDS markets for many corporate names are illiquid, with wide bid-ask spreads and infrequent trading. Given these constraints, this project adopts a pragmatic approach using synthetic and proxy data.

3.1 Model Development and Validation

During model development, synthetic CDS data was used to build and validate the pricing framework. The synthetic dataset includes historical spreads and term structures for five fictional reference entities, with regime shifts to test model behaviour under different market conditions. Details of the data generation methodology are documented in `data/README_Data.md`.

3.2 Final Pricing Data

For the final pricing results presented in Section 4, we use:

- **Correlation estimation**: Equity log-returns as a proxy for CDS spread changes. This is a common approach in practice, as equity data is more liquid and accessible than CDS spreads, and firms with correlated equity performance tend to exhibit correlated credit deterioration.
- **CDS term structure**: Synthetic CDS curves at tenors of 1Y, 2Y, 3Y, 4Y, and 5Y for hazard rate bootstrapping.
- **Recovery rate**: Constant at 40% for all entities, following market convention for senior unsecured debt.

4. Results, Discussion and Observations

4.1 Convergence Studies

Before the main pricing analysis, we conducted convergence studies to select the simulation count and RNG method. The RNG methods evaluated are described in Section 2.5. These studies used the data that was used in pricing, but with $k = 3$ (3rd-to-default) to ensure sufficient default events for meaningful analysis. We evaluated four RNG methods—pseudo-random, pseudo-random with antithetic variates, Halton, and Sobol—across simulation counts from 1,000 to 250,000.
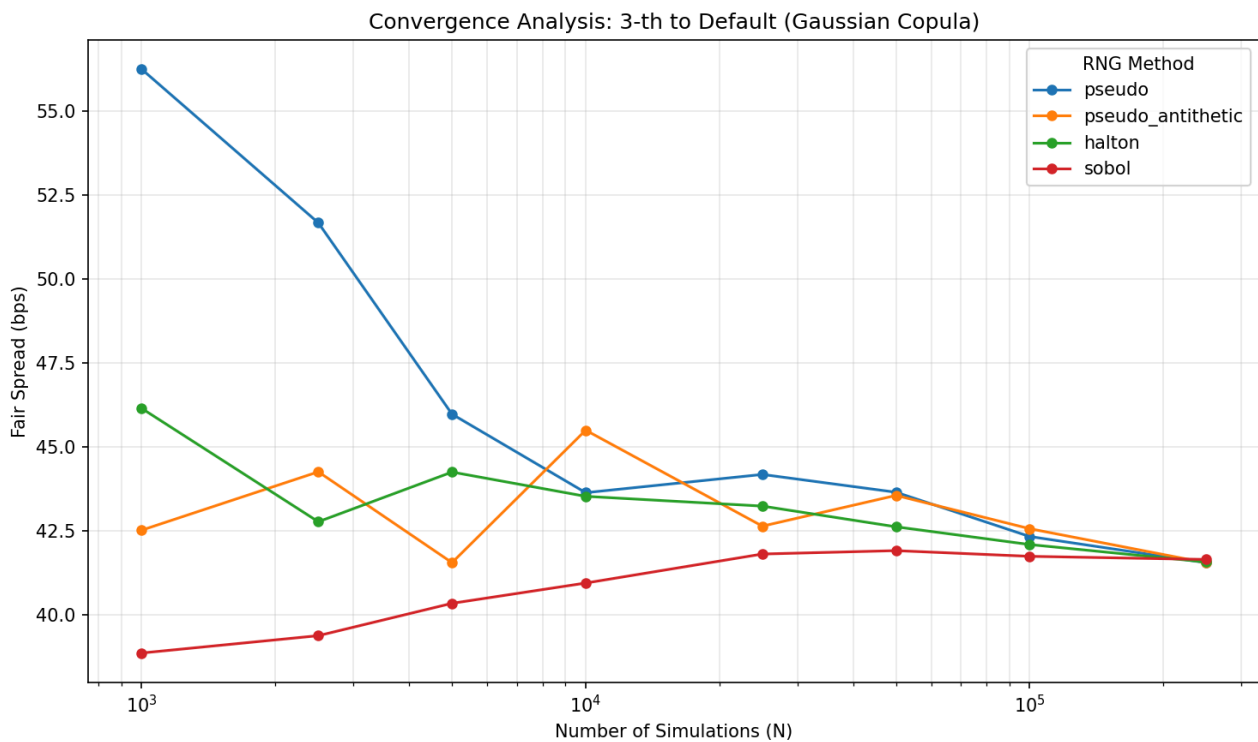
Figure 1: Fair spread estimates (bps) versus number of simulations. All methods converge to similar values, with quasi-random methods showing less variation at lower N.
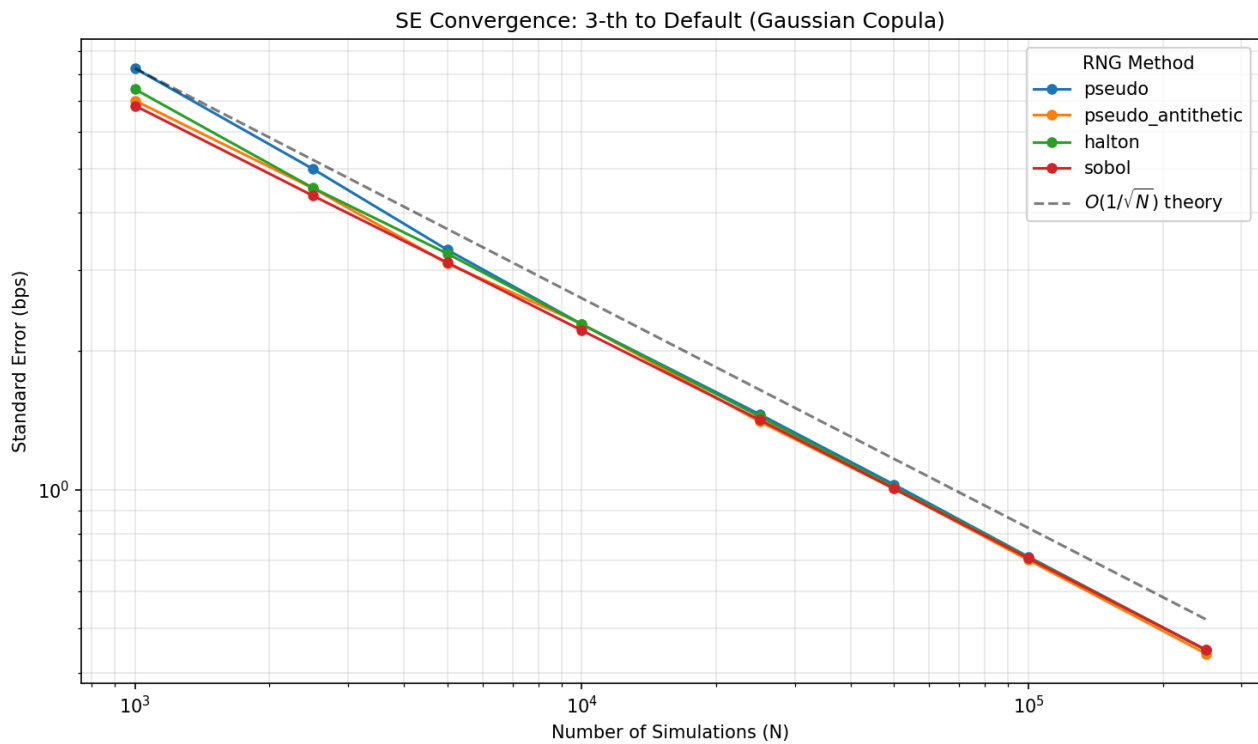
Figure 2: Standard error convergence (log-log scale). The dashed line shows the theoretical $O(1/\sqrt{N})$ rate. Quasi-random methods achieve lower standard errors at equivalent sample sizes.

Findings

All four methods converge to the same fair spread estimate and exhibit similar standard errors at high sample counts. At lower simulation counts, quasi-random sequences (Halton, Sobol) show modestly lower standard error than pseudo-random—approximately 15-20% reduction at $N = 1,000$—but this advantage diminishes as N increases. By $N = 100,000$, the methods are essentially indistinguishable in precision. The observed convergence rate matches standard Monte Carlo $O(1/\sqrt{N})$ for all methods, likely due to the non-smooth nature of the k-th to default payoff involving order statistics and indicator functions.

N	Pseudo-random SE	Antithetic SE	Halton SE	Sobol SE
1,000	8.25	7.02	7.43	6.84
10,000	2.29	2.29	2.29	2.22
100,000	0.71	0.70	0.71	0.71
250,000	0.45	0.44	0.45	0.45

Table 1: Standard error (bps) by RNG method and simulation count.

Antithetic variates provide no benefit for this problem. As discussed in Section 2.5, the k-th to default payoff depends on an order statistic, and the sorting operation destroys the monotonic relationship required for variance reduction. Empirically, the correlation between paired payoffs was approximately zero ($\rho \approx -0.01$).

Selected Configuration

Based on these findings, we select **Sobol sequences with 100,000 simulations** for the main analysis, providing standard error below 0.5 bps. Antithetic variates are not used.

4.2 Exploratory Data Analysis

The pricing analysis uses equity price data for five reference entities spanning January 2024 to January 2026. Daily adjusted closing prices are converted to log-returns, which serve as a proxy for credit spread changes when estimating the copula correlation matrix.
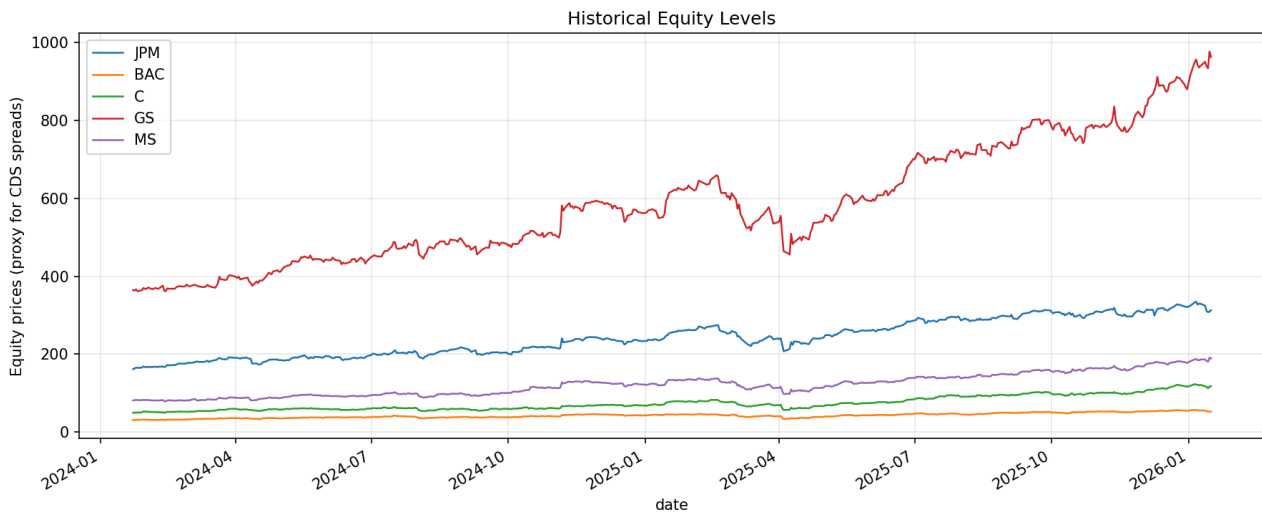

Figure 3: Equity price time series for the five reference entities (January 2024 – January 2026).
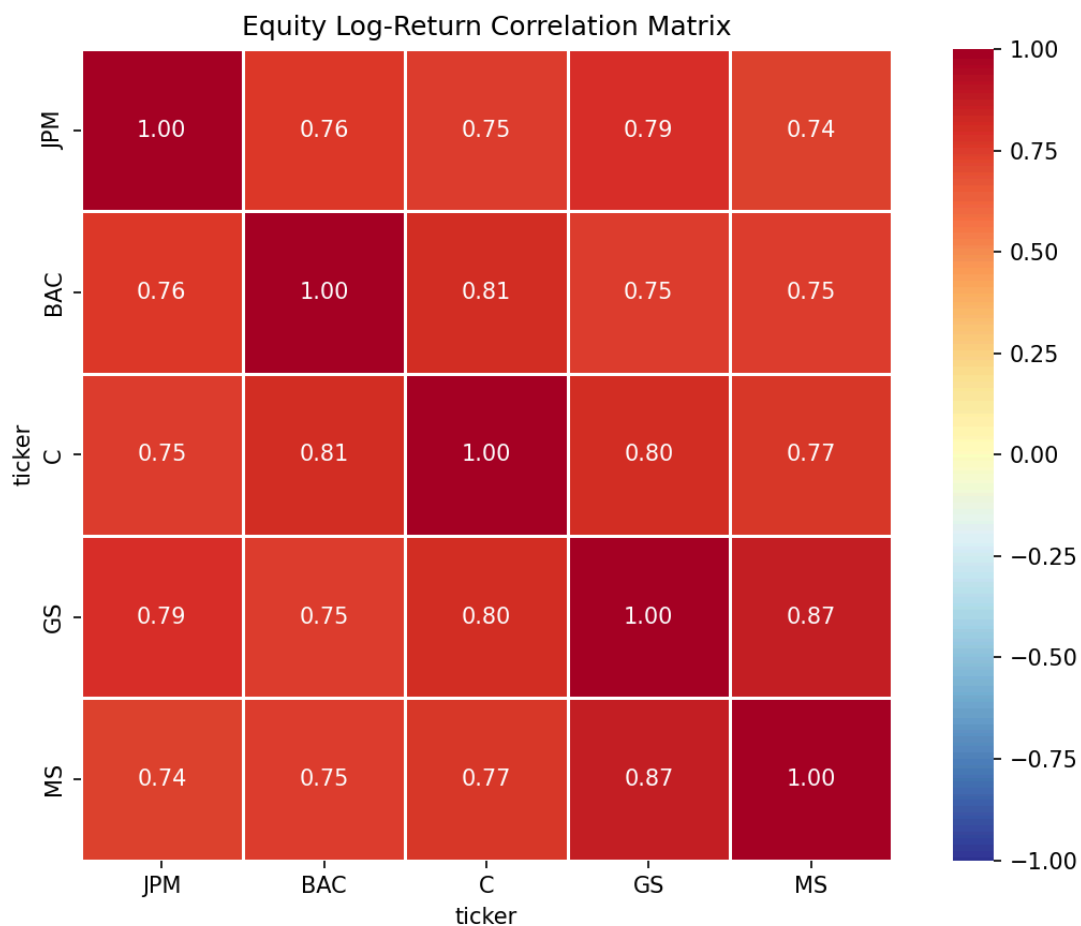
Figure 4: Correlation heatmap of equity log-returns. Values represent Spearman rank correlations, which are converted to Pearson correlations for copula calibration.

The synthetic CDS term structures provide spreads at tenors of 1, 2, 3, 4, and 5 years for each entity. These curves exhibit the typical upward-sloping shape, with shorter tenors reflecting near-term default risk and longer tenors incorporating cumulative default probability.
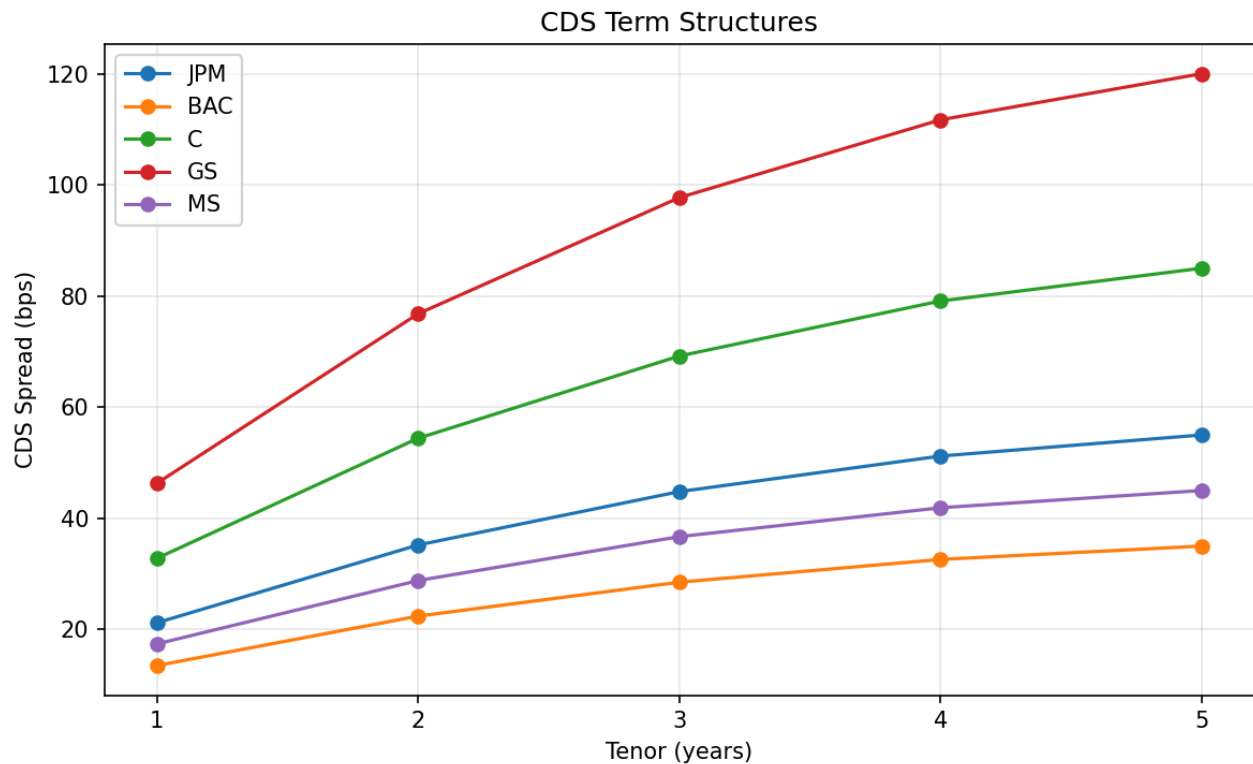
Figure 5: CDS term structure curves by entity. Higher spreads indicate greater perceived credit risk.

4.3 Calibration Results

Calibration proceeds in two stages: bootstrapping hazard rates from CDS term structures (marginals), and estimating the copula correlation matrix from equity log-returns (dependence). The methodology for each stage is detailed in Sections 2.2 and 2.3 respectively.

Hazard Rate Bootstrapping

Hazard rates are bootstrapped sequentially from the CDS term structure for each entity using the methodology described in Section 2.2. Starting from the 1-year tenor, each piecewise-constant hazard rate is solved to match the observed spread. The resulting step-wise hazard rate curves reflect each entity's credit risk profile.
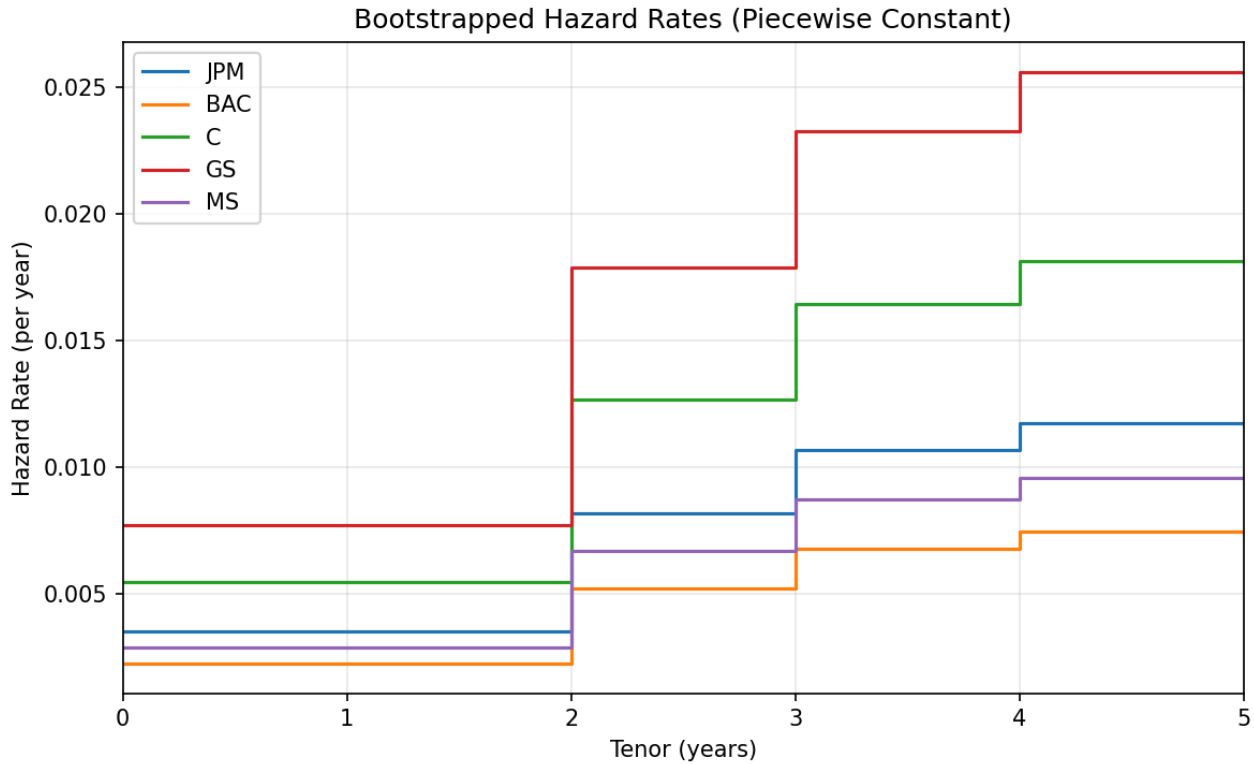
Figure 6: Bootstrapped hazard rate term structures by entity. Higher hazard rates correspond to wider CDS spreads and greater default probability.

Boeing exhibits the highest hazard rates, consistent with its wider CDS spreads (120 bps at 5Y), while Chevron and Exxon show the lowest default intensities, reflecting their investment-grade credit profiles.

Copula Correlation Matrix

The correlation matrix is estimated from equity log-returns using the rank correlation approach described in Section 2.3. Spearman's ρ_S is computed from the historical data and converted to Pearson correlation via $\rho = 2\sin(\pi\rho_S/6)$. This approach avoids the circular dependency problem in t-copula calibration and produces a correlation matrix valid for both Gaussian and t-copulas.

The same correlation matrix is used for both copula models—the key difference lies in how the copulas transform these correlations into joint tail behaviour.

t-Copula Degrees of Freedom

For the t-copula, the degrees of freedom parameter ν is estimated via profile maximum likelihood as described in Section 2.3. With the correlation matrix fixed from rank correlation, we evaluate the log-likelihood across a range of ν values and select the maximiser.
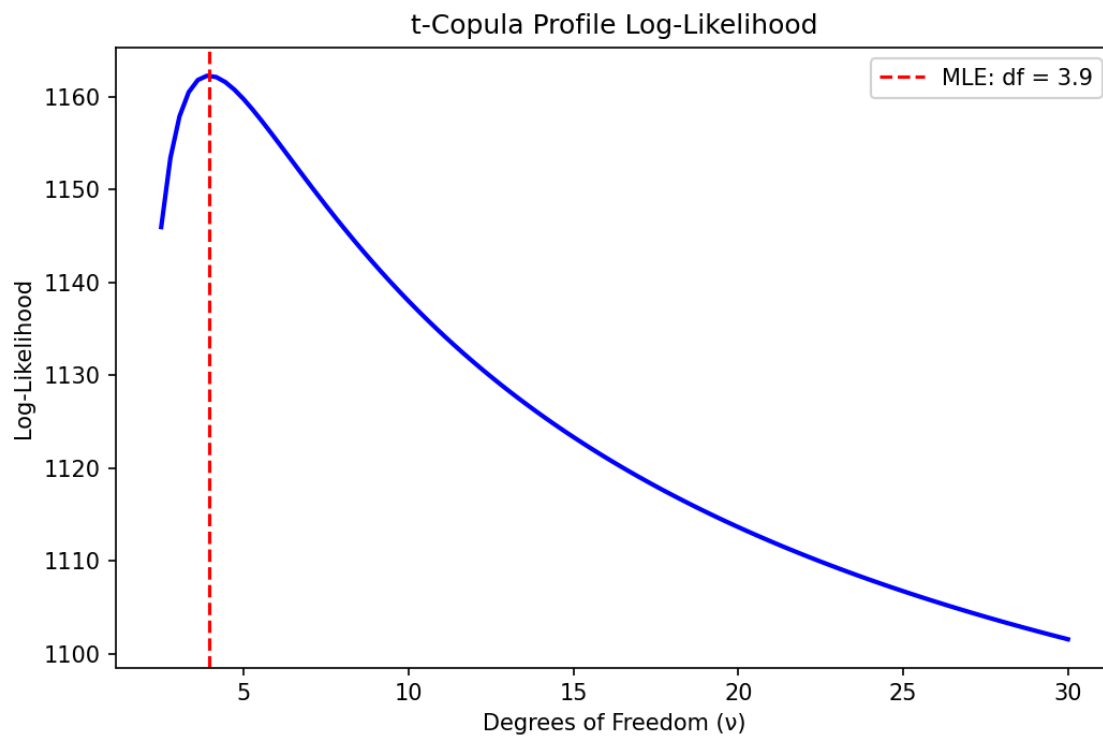
Figure 7: Profile log-likelihood for t-copula degrees of freedom. The vertical dashed line indicates the MLE estimate.

The MLE yields $\hat{\nu} \approx 4$, indicating substantial tail dependence. Lower degrees of freedom imply heavier tails and more frequent joint extreme movements. As $\nu \to \infty$, the t-copula converges to the Gaussian copula, so the estimated value of 4 represents a significant departure from Gaussian dependence—the data exhibits notably heavier tails than a normal distribution would predict. For sensitivity analysis in Section 4.5, we also examine pricing results across a range of ν values.

4.4 Pricing Results

Using the calibrated models and Sobol sequences with 100,000 simulations, we compute fair spreads for the 5th-to-default basket CDS under both copulas. The pricing methodology follows Section 2.6, with the fair spread calculated as the ratio of protection leg to premium leg expected values.

Copula	Fair Spread (bps)	SE (bps)	95% CI
Gaussian	7.36	0.30	[6.78, 7.94]
Student-t	53.84	0.81	[52.25, 55.43]

Table 2: Fair spread estimates for 5th-to-default basket CDS (5-year maturity, 40% recovery).
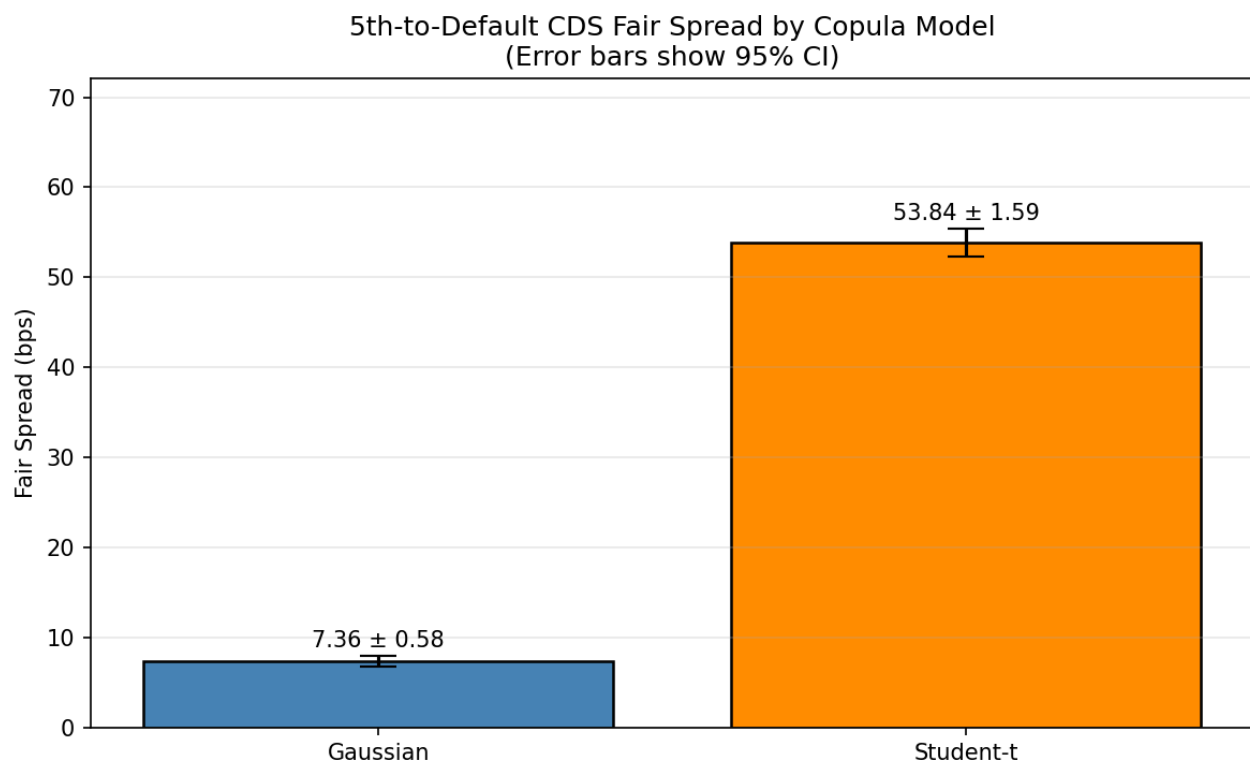

Figure 8: Fair spread comparison between copula models. Error bars represent 95% confidence intervals.

The t-copula produces a fair spread approximately 7.3 times higher than the Gaussian copula (53.84 bps vs 7.36 bps). This substantial difference reflects tail dependence: the t-copula assigns higher probability to scenarios where all five entities default jointly within the contract period.

Default Probability Comparison

The difference in pricing stems directly from the probability of triggering the 5th default within the 5-year term:

Copula	Defaults in Term	Default Rate
Gaussian	612	0.61%
Student-t	4,388	4.39%

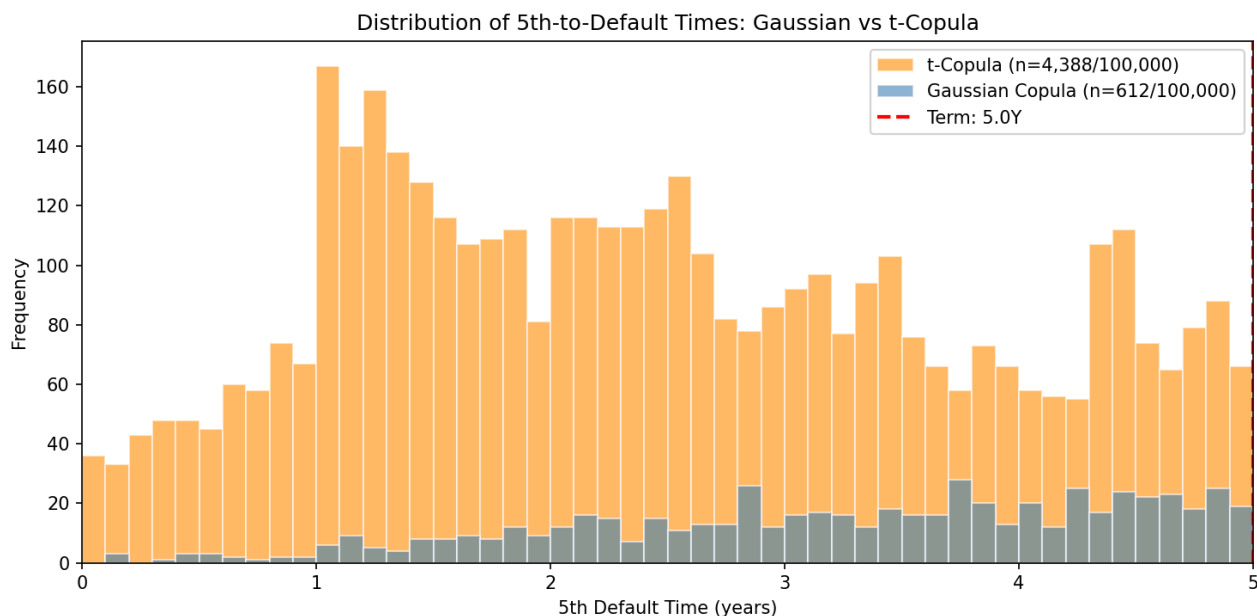*Table 3: Number of simulations where all five entities default within the 5-year contract term.*

Figure 9: Distribution of 5th-to-default times. The t-copula generates significantly more joint defaults, particularly in the early years.

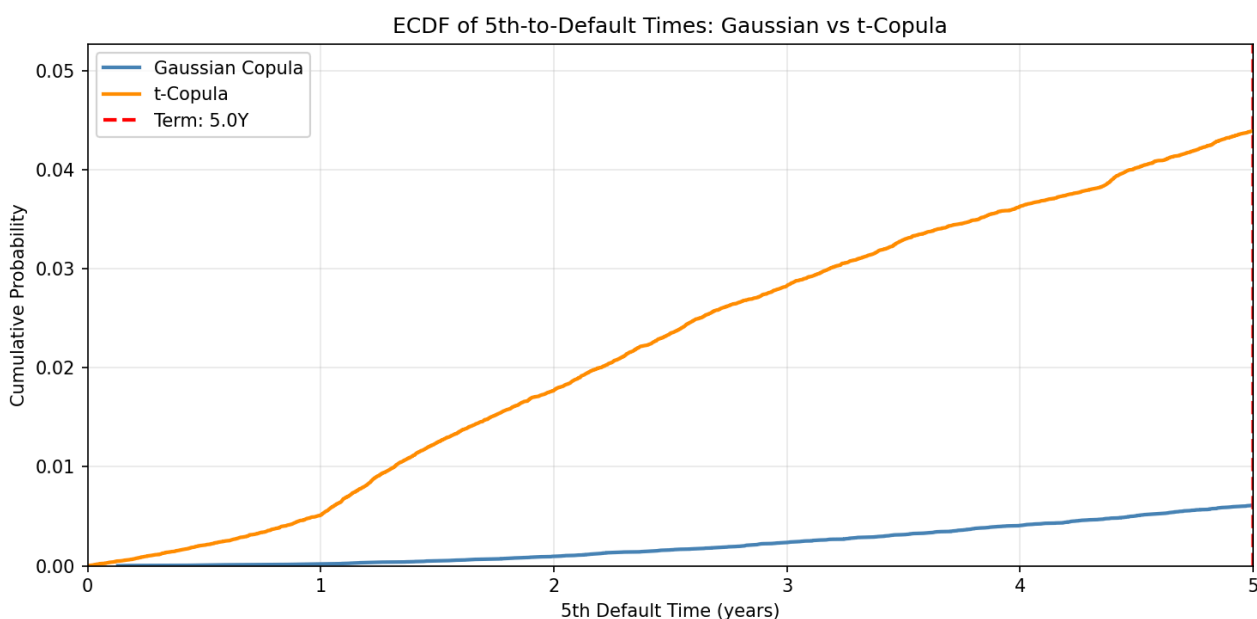
Figure 10: Empirical CDF of 5th-to-default times. The t-copula curve rises more steeply, indicating higher probability of early joint defaults.

Under the Gaussian copula, joint defaults of all five entities are extremely rare—only 0.61% of simulations trigger payout. The t-copula's heavier tails increase this probability by a factor of seven to 4.39%, dramatically increasing the expected protection leg value and therefore the fair spread required to compensate the protection seller. This demonstrates why copula selection is critical for pricing senior tranches of basket credit derivatives.

Preferred Estimate

The Gaussian copula's fair spread of 7.36 bps appears unrealistically low for 5-year protection against simultaneous default of five major financial institutions. This underpricing stems from the Gaussian copula's zero tail dependence property—it assigns negligible probability to joint extreme events. Historical evidence, particularly from the 2008 financial crisis, demonstrates that credit defaults cluster during systemic stress, a phenomenon the Gaussian copula fails to capture.

The t-copula's fair spread of 53.84 bps better reflects this tail risk. Three factors support accepting the t-copula estimate: (1) the MLE-estimated degrees of freedom ($\hat{\nu} \approx 4$) indicates the data exhibits heavier tails than Gaussian, (2) financial sector names share significant exposure to common macroeconomic factors that drive simultaneous deterioration, and (3) prudent risk management favours the more conservative estimate when pricing protection against catastrophic scenarios. We therefore adopt **53.84 bps** as the fair spread for the 5th-to-default basket CDS.

4.5 Sensitivity Analysis

This section examines how the fair spread responds to changes in key model parameters. All sensitivity analyses use the t-copula model with 100,000 Sobol simulations unless otherwise noted. The base case parameters are: K=5 (5th-to-default), high correlation (financial sector), ν=3.9 (MLE estimate), and R=40% recovery rate.

4.5.1 Impact of Tranche Seniority (k)

The tranche seniority k determines how many defaults are required to trigger payout. Lower k values represent junior tranches that absorb the first losses, while higher k values represent senior tranches protected by subordination.

k	Fair Spread (bps)	SE (bps)	Defaults in Term
1st	283.64	1.98	20,876 (20.9%)
2nd	180.63	1.54	13,888 (13.9%)
3rd	127.78	1.28	10,055 (10.1%)
4th	89.48	1.06	7,169 (7.2%)
5th	53.84	0.81	4,388 (4.4%)

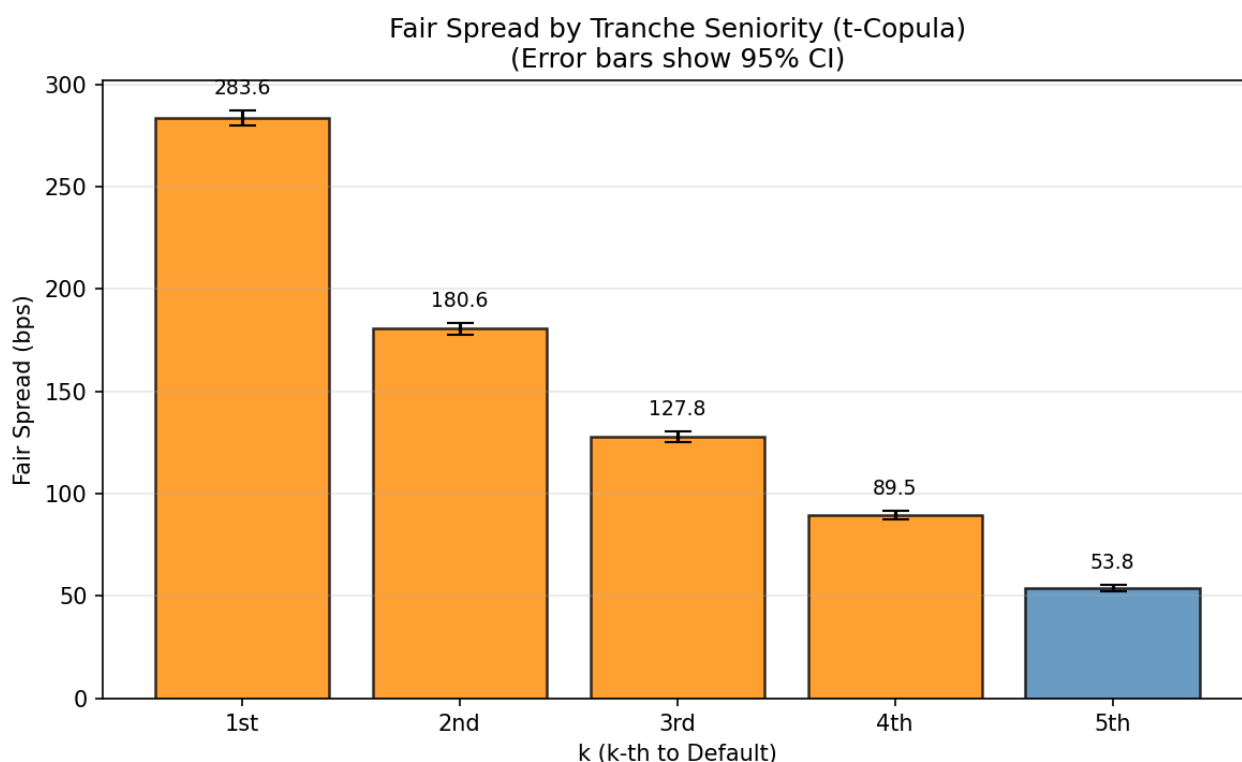*Table 4: Fair spreads by tranche seniority. Base case (K=5) highlighted.*

Figure 11: Fair spread decreases with tranche seniority as more defaults are required to trigger payout.

The 1st-to-default spread (283.64 bps) is approximately 5.3 times higher than the 5th-to-default spread (53.84 bps). This reflects the subordination benefit: while a single default among five correlated financials is relatively likely (20.9% of simulations), requiring all five to default within five years is much rarer (4.4%). The non-linear relationship between k and spread demonstrates the value of subordination in structured credit products.

4.5.2 Impact of Correlation

Correlation fundamentally determines the joint default behaviour. We compare two correlation regimes while holding marginal default probabilities constant:

- **High correlation**: Financial sector basket (calibrated from equity returns)
- **Low correlation**: Diversified basket (scaled to average pairwise correlation ~0.15)

Correlation Level	Fair Spread (bps)	SE (bps)	Defaults in Term
High (Financial)	53.84	0.81	4,388 (4.4%)
Low (Diversified)	1.13	0.12	94 (0.09%)

Table 5: Fair spreads by correlation level. Marginal hazard rates unchanged between scenarios.
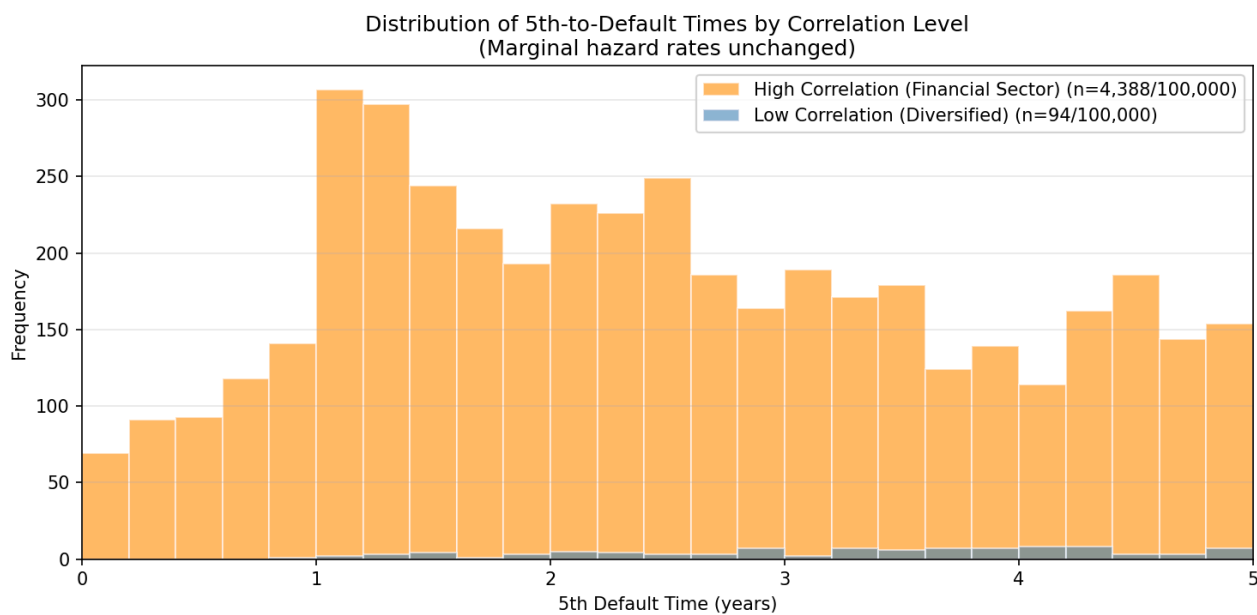

Figure 12: Distribution of 5th-to-default times by correlation level. High correlation produces more joint defaults within the contract term.
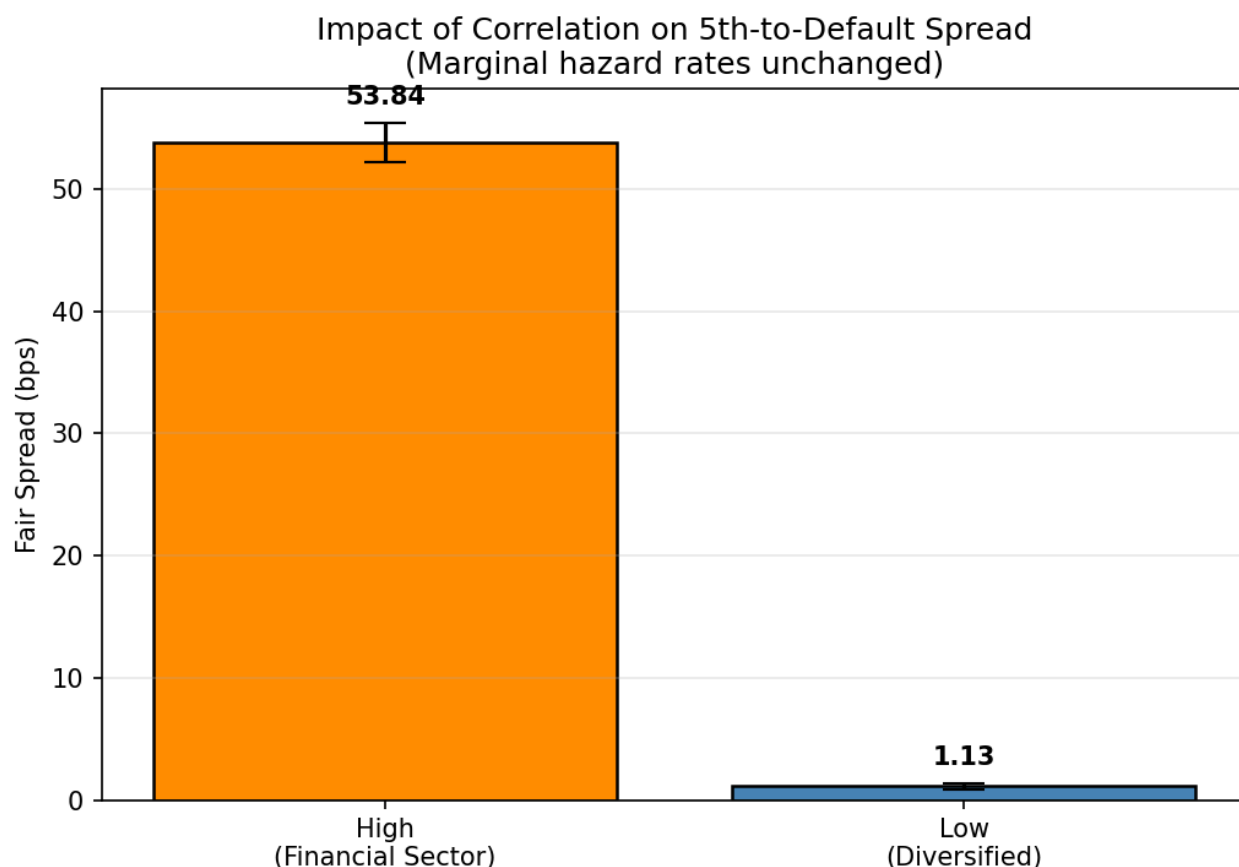
Figure 13: Fair spread comparison showing the dramatic impact of correlation on senior tranche pricing.

The high-correlation basket produces a spread nearly 48 times larger than the diversified basket (53.84 vs 1.13 bps). This striking difference occurs because correlation affects how individual defaults cluster in time. With low correlation, even if each entity has material default risk individually, the probability of all five defaulting within the same 5-year window is negligible (0.09%). High correlation causes defaults to cluster during stress periods, dramatically increasing joint default probability (4.4%). This demonstrates why correlation is the dominant risk factor for senior tranches.

4.5.3 Impact of Degrees of Freedom

The t-copula degrees of freedom parameter ν controls tail dependence. Lower ν produces heavier tails and stronger dependence during extreme events. As $\nu \to \infty$, the t-copula converges to the Gaussian copula.

ν	Fair Spread (bps)	SE (bps)	Defaults in Term
3	60.29	0.86	4,899 (4.9%)
3.9 (base)	54.33	0.82	4,427 (4.4%)
6	43.44	0.73	3,556 (3.6%)
10	34.03	0.64	2,798 (2.8%)
15	28.45	0.59	2,346 (2.3%)
30	21.17	0.51	1,751 (1.8%)
Gaussian (∞)	7.36	0.30	612 (0.6%)

Table 6: Fair spreads by degrees of freedom. Base case (ν=3.9) highlighted.
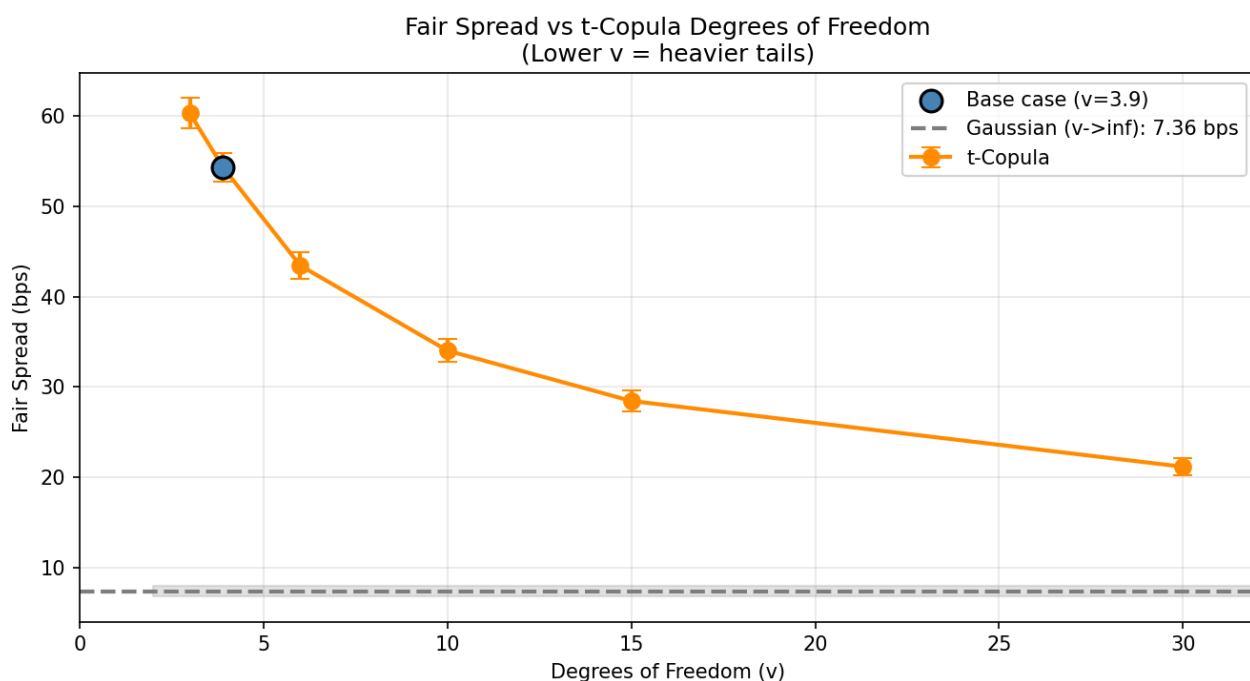

Figure 14: Fair spread vs degrees of freedom. The horizontal dashed line shows the Gaussian copula limit. Lower ν produces higher spreads due to increased tail dependence.

The spread at ν=3 (60.29 bps) is approximately 8 times higher than the Gaussian limit (7.36 bps). Our MLE estimate of ν=3.9 produces a spread of 54.33 bps. The monotonic decline in spread as ν increases reflects diminishing tail dependence: with high ν, extreme co-movements become less

likely, reducing joint default probability. The gap between low-v t-copula pricing and Gaussian pricing illustrates the model risk inherent in copula selection.

4.5.4 Impact of Recovery Rate

The recovery rate R determines the loss given default (LGD = 1 - R) and therefore the payout magnitude when defaults occur.

Recovery Rate	Fair Spread (bps)	SE (bps)	Defaults in Term
0%	65.16	1.15	3,208 (3.2%)
30%	56.59	0.90	3,962 (4.0%)
40% (base)	53.84	0.81	4,388 (4.4%)
50%	49.35	0.71	4,812 (4.8%)

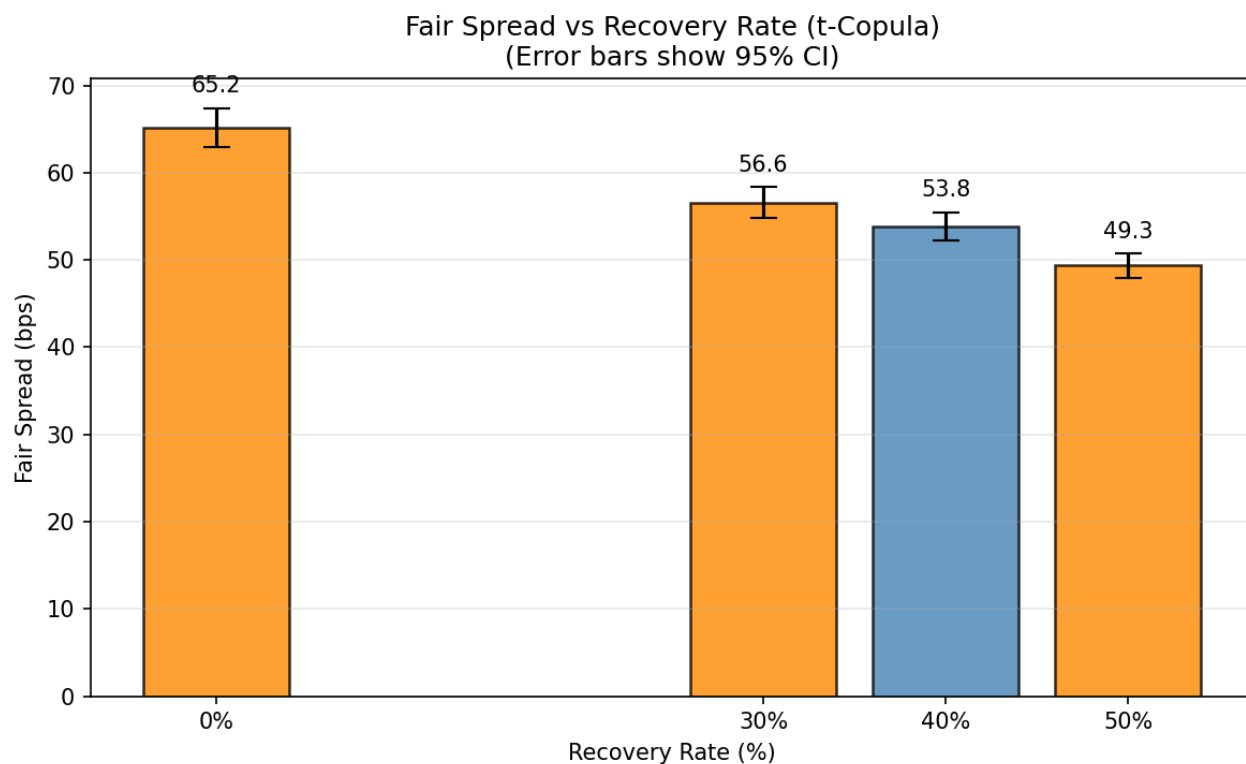*Table 7: Fair spreads by recovery rate assumption.*

Figure 15: Fair spread vs recovery rate. Higher recovery reduces the protection leg value, lowering the required spread.

The relationship between spread and recovery rate shows an interesting pattern. Higher recovery reduces the loss given default, which directly reduces the protection leg value. However, the effect is partially offset by the premium leg calculation: higher recovery also appears in the denominator through the effective notional. The spread at 0% recovery (65.16 bps) is approximately 32% higher than at 50% recovery (49.35 bps).

Note that the number of defaults in term varies slightly across recovery scenarios due to Monte Carlo sampling variation, but this does not materially affect the conclusions.

4.5.5 Summary of Sensitivities

Parameter	Range Tested	Spread Range (bps)	Key Insight
k	1st to 5th	283.64 → 53.84	Subordination provides ~5x protection
Correlation	Low to High	1.13 → 53.84	Correlation is dominant factor (~48x)
ν	3 to Gaussian	60.29 → 7.36	Tail dependence critical (~8x)
Recovery	0% to 50%	65.16 → 49.35	Moderate impact (~32%)

Table 8: Summary of sensitivity analysis results.

The analysis reveals that correlation and copula selection (through degrees of freedom) have the most significant impact on 5th-to-default pricing. Recovery rate assumptions, while important, have a comparatively modest effect. These findings underscore the importance of careful calibration and the model risk inherent in basket credit derivative pricing.

5. Conclusion

5.1 Executive Summary

This project developed a Monte Carlo simulation framework for pricing k-th to default basket credit default swaps using copula models. The implementation prices a 5-year, 5th-to-default CDS on a basket of five major US financial institutions: JPMorgan Chase, Bank of America, Citigroup, Wells Fargo, and Goldman Sachs.

Key Findings

The fair spread for the 5th-to-default basket CDS varies dramatically depending on copula specification:

Model	Fair Spread	95% CI
Gaussian Copula	7.36 bps	[6.78, 7.94]
t-Copula (ν=3.9)	53.84 bps	[52.25, 55.43]

The t-copula produces a spread approximately 7.3 times higher than the Gaussian copula. This difference stems entirely from tail dependence: the t-copula assigns meaningful probability to joint extreme events where all five institutions default simultaneously, while the Gaussian copula treats such scenarios as negligibly rare.

We recommend the **t-copula estimate of 53.84 bps** as the fair spread. The MLE-estimated degrees of freedom (ν≈4) indicates the historical data exhibits substantially heavier tails than a Gaussian distribution would predict, and prudent risk management favours the more conservative estimate when pricing protection against systemic credit events.

Sensitivity Analysis Insights

The sensitivity analysis revealed the relative importance of model parameters:

1. **Correlation** is the dominant factor, with a 48-fold spread difference between high and low correlation regimes. This confirms that for senior tranches, default dependence structure matters far more than individual default probabilities.

2. **Copula selection** (via degrees of freedom) produces an 8-fold spread difference between heavy-tailed (ν=3) and Gaussian specifications. Model risk in copula choice is substantial and cannot be ignored.

3. **Tranche seniority** shows the expected subordination benefit, with 1st-to-default spreads approximately 5 times higher than 5th-to-default.

4. **Recovery rate** has a moderate 32% impact across the tested range, less influential than dependence structure assumptions.

Methodology Recommendations

Based on our analysis, we recommend the following methodology for basket CDS pricing:

1. **Use t-copula over Gaussian** for financial sector baskets where tail dependence is economically meaningful
2. **Employ Sobol sequences** for variance reduction—they provide faster convergence than pseudo-random sampling at no additional computational cost
3. **Calibrate correlation from rank correlation** (Spearman converted to Pearson) to avoid circular dependency issues in t-copula parameter estimation
4. **Estimate degrees of freedom via profile MLE** with correlation fixed, treating ν as a measure of tail heaviness in the data
5. **Conduct sensitivity analysis** across correlation, ν, and recovery assumptions to quantify model risk

5.2 Assumptions and Limitations

The pricing framework relies on several simplifying assumptions that users should understand:

Market Data and Calibration

- **Equity correlation as proxy for default correlation**: We calibrate the copula correlation matrix from equity log-returns rather than direct measures of default dependence. While equity prices embed credit risk information, this proxy may not fully capture joint default behaviour during stress periods when equity-credit relationships can break down.

- **Historical calibration period**: The correlation structure is estimated from a specific historical window and may not reflect future dependence patterns, particularly during regime changes or unprecedented stress events.

Model Simplifications

- **Zero interest rates**: The implementation assumes zero risk-free rates for discounting. In practice, the discount curve affects the relative weighting of near-term versus distant cash flows, though the impact on spread ratios between copula models would be modest.

- **Constant recovery rate**: We assume a fixed 40% recovery rate for all entities and all scenarios. In reality, recovery rates are stochastic, negatively correlated with default rates (recoveries tend to be lower during systemic crises when defaults cluster), and entity-specific.

- **Piecewise-constant hazard rates**: Hazard rates are assumed constant between CDS tenor points and bootstrapped from market spreads. This is standard practice but ignores potential intra-period variation.

- **Static correlation**: The correlation matrix is fixed throughout the contract term. Dynamic copula models that allow correlation to evolve over time or vary with market conditions would be more realistic but substantially more complex.

- **No wrong-way risk**: We do not model correlation between counterparty credit quality and the protection seller's ability to pay. For dealer-intermediated trades, this could be material.

Scope Limitations

- **Risk-neutral pricing only**: The framework produces fair value spreads under risk-neutral measure. Commercial pricing would require additional adjustments for CVA/DVA, funding costs, regulatory capital, and profit margins.

- **Single basket composition**: Results are specific to the five financial institutions analysed. Different sector compositions or geographic diversification would produce different correlation structures and spread levels.

5.3 Challenges Encountered

Data Sourcing

Obtaining consistent, high-quality CDS spread data proved challenging. CDS markets are over-the-counter with limited price transparency, and historical spread series often contain gaps, stale quotes, or inconsistent tenors across entities. We ultimately used a combination of market data sources and applied cleaning procedures to construct usable term structures.

t-Copula Parameter Estimation

The t-copula presents a circular dependency problem: estimating correlation requires knowing degrees of freedom, but estimating degrees of freedom requires knowing correlation. We resolved this by:

1. Using rank correlation (Spearman's ρ) converted to Pearson correlation via the elliptical copula relationship $\rho = 2\sin(\pi\rho_S/6)$
2. Fixing this correlation matrix and estimating ν via profile maximum likelihood

This approach is theoretically justified for elliptical copulas and avoids unstable joint optimisation.

Variance Reduction for t-Copula

We expected antithetic variates to provide improved convergence for the t-copula simulation. The standard approach of negating uniform draws works for the Gaussian copula, but is insufficient for the t-copula, which involves both correlated normals and an independent chi-square component. We implemented antithetic sampling of the normal component alone (without the chi-square), but found no measurable variance reduction for the k-th to default payoff. As discussed in Section 4.1, the order statistic payoff structure destroys the monotonic relationship required for antithetic variates to be effective.

Computational Performance

With 100,000 simulations and multiple sensitivity scenarios, total computation time became non-trivial. Quasi-random sequences (Sobol) helped by achieving target precision with fewer simulations than pseudo-random sampling would require. For production use, further optimisation through vectorisation, parallel processing, or GPU acceleration would be beneficial.

5.4 Further Work

Several extensions would enhance the framework's analytical rigour and practical applicability.

Model Enhancements

These extensions would improve the theoretical foundations of the pricing model:

- **Interest rate discounting**: Incorporate a risk-free discount curve bootstrapped from OIS rates to properly weight cash flows across the term structure.

- **Stochastic recovery**: Model recovery rates as random variables, potentially correlated with systematic factors, to capture the empirical observation that recoveries decline during credit crises.

- **Kernel density estimation for marginals**: Replace the rank-based empirical CDF transformation with kernel density estimation (KDE) when converting observations to uniform marginals. KDE provides a smoother estimate of the marginal distributions, which may improve copula calibration particularly in the tails where the empirical CDF can be noisy.

- **Alternative copula families**: Implement asymmetric copulas (Clayton, Gumbel) that can capture different upper versus lower tail dependence. Financial defaults may exhibit asymmetric dependence—stronger co-movement in distress than in benign conditions.

- **Dynamic copulas**: Allow the correlation structure to evolve over time, either through regime-switching models or continuous-time stochastic correlation specifications.

Further Analysis

These investigations would deepen understanding of model behaviour and validate the methodology:

- **Antithetic variates investigation**: Investigate more thoroughly why antithetic variates provided no variance reduction for the k-th to default payoff, despite implementation for the normal component. Extend the approach to include antithetic sampling of the chi-square component in t-copula simulation, and evaluate whether alternative variance reduction techniques (e.g., control variates, importance sampling) would be more effective for order statistic payoffs.

- **Empirical validation**: Compare model-implied spreads to traded basket CDS or CDO tranche prices to assess model performance and calibrate any systematic biases.

- **Stress testing**: Develop scenario frameworks that shock correlation, hazard rates, and recovery simultaneously to assess tail risk under extreme but plausible conditions.

- **Machine learning integration**: Explore whether ML techniques can improve copula calibration, particularly for capturing non-linear dependence structures or regime changes that parametric copulas may miss.

Productionisation

These extensions would be required to deploy the framework in a production trading or risk management environment:

- **XVA adjustments**: Bridge the gap between risk-neutral fair value and executable prices by incorporating credit valuation adjustment (CVA), debit valuation adjustment (DVA), and funding valuation adjustment (FVA).

- **Regulatory capital**: Compute capital requirements under Basel III/IV standardised or internal model approaches to understand the full cost of holding basket credit exposure.

- **Greeks and hedging**: Implement sensitivities to spread movements (CS01), correlation (correlation vega), and recovery rate to support dynamic hedging strategies.

- **Real-time calibration**: Develop infrastructure for daily recalibration to market data, enabling mark-to-market and ongoing risk monitoring.

6. References

CQF Programme Materials

- CQF Module 3: JU253.4 Intro to Numerical Methods
- CQF Module 6: JU256.7 Further Monte Carlo
- CQF Module 6: JU256.9 Credit Default Swaps
- CQF Module 6: JU256.10 Structural Models for Default Prediction and Dependency Modelling
- CQF Elective: Counterparty Credit Risk Modelling

Academic References

Cherubini, U., Luciano, E. and Vecchiato, W. (2004) *Copula Methods in Finance*. Chichester: John Wiley & Sons.

Glasserman, P. (2003) *Monte Carlo Methods in Financial Engineering*. New York: Springer.

Hull, J. and White, A. (2004) 'Valuation of a CDO and an n-th to Default CDS Without Monte Carlo Simulation', *Journal of Derivatives*, 12(2), pp. 8-23.

Li, D.X. (2000) 'On Default Correlation: A Copula Function Approach', *Journal of Fixed Income*, 9(4), pp. 43-54.

McNeil, A.J., Frey, R. and Embrechts, P. (2015) *Quantitative Risk Management: Concepts, Techniques and Tools*. Revised edn. Princeton: Princeton University Press.

Nelsen, R.B. (2006) *An Introduction to Copulas*. 2nd edn. New York: Springer.

Niederreiter, H. (1992) *Random Number Generation and Quasi-Monte Carlo Methods*. Philadelphia: SIAM.

Schönbucher, P.J. (2003) *Credit Derivatives Pricing Models: Models, Pricing and Implementation*. Chichester: John Wiley & Sons.