A Hybrid Approach to Risk-Neutral k_{th}-to-Default Basket CDS Pricing

Combining Intensity, Copula, and Quasi-Monte Carlo Methods

Jan Abdullah

Abstract

Ensuring accurate and fair pricing of k_{th}-to-Default Basket CDS is crucial for safeguarding the interests of trading counterparties. This paper presents a hybrid approach to modeling the fair spread of these instruments within a risk-neutral framework. The model leverages the strengths of intensity models, copula techniques under different statistical distributions and correlation methods, and Monte Carlo simulations. Application of this approach is demonstrated through its deployment in pricing a basket of Asian sovereign CDS. This paper also dives into spread pricing sensitivities, analyzing the influence of individual input variables via recalibration while holding other factors constant. Note, the focus here is to explore the mathematical properties of the model and process, rather than conducting an in-depth economic analysis on the basket constituents.

Introduction

What is a kth-to-Default Basket CDS?

A k_{th}-to-Default Basket CDS is a type of structured credit product that bundles multiple single credit default swaps. In contrast to a traditional credit default swap, where a payout is triggered upon a credit event of a single reference entity, the k_{th}-to-Default swap is triggered when the basket experiences its k_{th} credit event. For example, if a trader buys protection on a 10^{th}-to-Default swap spanning 100 entities, they'll receive a payout only if 10 entities in the basket experience a credit event before the swap's maturity. If fewer than 10 entities default, the swap will expire in the protection seller's favor. Note, the payout is based on the notional value of only the 10^{th} defaulting entity, not the cumulative value of all defaulting entities.

Model Overview

Pricing fair spreads requires the consideration of both the credit risk profiles of each individual constituent, as well as the co-dependence of risk profiles between constituents. As such, the fair spread is defined as an expectation over the joint distribution of default times. Default times are simulated by sampling correlated pseudo-samples from a copula with Monte-Carlo and converting samples to default times using individual hazard rate term structures. Fair spread is then calculated by averaging the default legs and premium legs for each simulation separately. Each k_{th}-to-Default tranche is priced distinctly, and the relationships between tranches are explained in this paper.

Methodology and Mathematical Concepts

Process Summary Basic Overview

Data Requirements

1. Historical weekly CDS spreads for each constituent spanning at least 2-3 years.
2. Credit curves for each constituent, covering up to 5Y tenor.
3. Zero-coupon risk-free yield curve, covering up to 5Y tenor.

Procedure Outline

1. Discount Curve Construction: Convert the risk-free yield curve into discount factors. Apply log-linear interpolation to derive a continuous discount factor curve.
2. Bootstrapping: Construct survival probability and hazard rates term structures for each individual constituent using the respective credit curves.
3. Copula Generation: Construct both Gaussian and Student's t copulae.
4. Monte Carlo Sampling: Draw samples from the copulae to produce correlated random uniform vectors.
5. Default Time Conversion: Transform each simulated random uniform variable to default times using the corresponding constituent's hazard rate term structure.
6. Pricing Legs: Use the simulated default times to price the default leg and premium leg independently and for each k_{th}-to-Default instrument.
7. Determining Fair Spread: Use the converged default leg and premium leg pricing to compute the fair spread for each k_{th}-to-Default instrument.

Discount Curve Interpolation

The discount curve is used to determine the present value of future cash flows. The discount factor given an interest rate r and time t is expressed as:

$$D(r,t) = e^{-rt}$$

Discount factors tend to exhibit exponential behaviors more so than linear. Thus, to transform discrete points into a continuous curve, log-linear interpolation is better suited.

For two known discount factors, $D(t_1)$ at time t_1 and $D(t_2)$ at time t_2, the discount factor for $D(t)$ such that $t_1 < t < t_2$ can be derived as:

$$D(t) = D(t_1) \times \left(\frac{D(t_1)}{D(t_2)}\right)^{\frac{t-t_1}{t_2-t_1}}$$

In a computational environment, one approach to log-linearly interpolate discount factors would be to take the natural logarithm of the discrete points, apply linear interpolation, and then exponentiate the linear interpolated curve to revert to the original scale.

Intensity Modeling | Bootstrapping Hazard Rates

For each individual constituent in the basket, the hazard rate term structure is bootstrapped from the credit curve by way of survival probabilities. Given the survival probability function $P(t)$, the hazard rate or intensity, $\lambda(t)$, can be derived as the potential of defaults occurring at time t conditional on no default having occurred before t. It is essentially a calibration to the Inhomogeneous Poisson Process.

The hazard rate term structure is represented as a piecewise-constant function, with jumps at the tenor points. A piecewise constant function offers simplicity, robustness, and computational efficiency over its more complex counterparts such as piecewise linear or cubic spline. It's also a practical method to capture the evolution of credit risk over time without resorting to a continuous function.

For CDS market spreads with increasing maturities $S_1, S_2 \ldots, S_N$, the survival probabilities $P(T_N)$ can be determined by:

$$Let\ L = Loss\ Given\ Default = 1 - Recovery\ Rate$$

$$P(T_N) = \frac{\sum_{n=1}^{N-1} D(0,T_N)[LP(T_{n-1}) - (L + \Delta t_n S_N)P(T_n)]}{D(0,T_N)(L + \Delta t_n S_N)} + \frac{P(T_{N-1})L}{L + \Delta t_N S_N}$$

The intensity, or hazard rate, can then be expressed as the log-ratio of successive survival probabilities:

$$\lambda_m = -\frac{1}{\Delta t} log\ \left(\frac{P(0,t_m)}{P(0,t_{m-1})}\right)$$

Copula Methods

Copula methods require deriving linear correlation matrices from historical spread data. Ideally, spreads should be sampled at weekly intervals. A span of at least 2 years, translating to around 100 observations, ensures a robust dataset. Given that fixed income markets exhibit strong correlation, daily fluctuations will reflect significant co-movement, much of which is just noise. Thus, deriving correlation matrices on daily spreads might lead to exaggerated correlations. The aim isn't to capture correlations of price movements, but rather of intrinsic default risk. Weekly data helps capture the broader default risk trends while filtering out market noise.

This study considers multiple parameteric distributions and correlation methods as the foundation for copula techniques. The focus is on the below 3 styles:

- Gaussian copula with Pearson's correlation.
- Student's t copula with Spearman's rho correlation ("linearized").
- Student's t copula with Kendall's tau correlation ("linearized").

Each of these correlation methods require input data to be structured in a certain way.

Input Datasets

There are 3 distinct input datasets for the 3 correlation methods:

1. Kendall's tau:
   - Dataset: ΔCDS spreads (denoted as X^{Hist}).
   - Description: Weekly changes in CDS spreads.
2. Spearman's rho:
   - Dataset: X^{Hist} converted to uniform distribution, $U^{Hist} \sim U(0,1)$.
   - Description:
     - Generate pseudo-samples/scores by fitting an Empirical Cumulative Distribution Function (ECDF) to X^{Hist}.
     - Steps involved:
       - i. Estimate X^{Hist}'s Empirical Probability Density Function (EPDF) via kernel density estimation.
       - ii. Smoothen the EPDF using an appropriate kernel function.
       - iii. Adjust the bandwidth as needed to achieve uniformity.
   - The "kcde" function in R's "ks" library solves this.
   - Quality Assurance: Check the uniformity of U^{Hist} distributions.
3. Pearson's:
   - Dataset: U^{Hist} converted to a normal distribution, $Z^{Hist} \sim N(0,1)$.
   - Description: Convert the pseudo-samples to conform to a standard normal distribution.

Correlation Methodologies Explained

Kendall's tau

Kendall's tau is a non-parametric rank correlation measure that measures the relationships between large values in two variables. It calculates the difference between the number of concordant and discordant pairs, then normalizes this difference by the total number of pairs.

A pair of observations is considered concordant if the rank order for both elements is the same. Conversely, a pair of observations is considered discordant if the rank order for both elements is not the same.

To illustrate, suppose there are a pair of observations (x_i, y_i) and (x_j, y_j):

- The pair is concordant if $\left(x_i > x_j \,\Lambda\, y_i > y_j\right) \vee (x_i < x_j \,\Lambda\, y_i < y_j)$.
- The pair is discordant if $(x_i > x_j \,\Lambda\, y_i < y_j) \vee (x_i < x_j \,\Lambda\, y_i > y_j)$.

Given N as the total number of pairs, N_c as the number of concordant pairs, and N_d as the number of discordant pairs, the formula for Kendall's tau estimation is:

$$\rho_\tau = \frac{N_c - N_d}{N(N-1)/2}$$

Spearman's rho

Spearman's rho is another non-parametric rank correlation measure that essentially calculates the linear correlation of the associated empirical cumulative distribution functions. The ECDF provides the ranks of the underlying data points. Validating the uniformity of the ECDF is important for the accuracy of this correlation method.

Given N pairs of observations and variables X and Y, where X and Y are ascending-ordered ranked variables and x_i and y_i are the i^{th} observations in each variable, the formula is:

$$\rho_s = 1 - \frac{6\sum_{i=1}^{N}(x_i - y_i)^2}{N(N^2-1)}$$

Pearson

Lastly, the commonly used Pearson correlation coefficient is the linear relationship between two continuous variables, X and Y:

$$\rho = \frac{\mathbb{C}ov(X,Y)}{\sqrt{\mathbb{V}ar(X)\mathbb{V}ar(Y)}}$$

<u>Cholesky Decomposition</u>

Once the correlation matrix is derived, it will need to be factorized using Cholesky Decomposition. This method decomposes the matrix into a lower triangular matrix A, and its transpose A^T. The purpose of this decomposition is to multiply A with random numbers pulled from a normal distribution to generate correlated uniform variables.

In essence, Cholesky Decomposition transforms the correlation matrix into the product of two triangular matrices, represented by:

$$\Sigma = AA^T$$

<u>Student's t Degrees of Freedom Calibration</u>

The degrees of freedom ν in the Student's t copula requires calibration through Maximum Likelihood Estimation. Using the observed pseudo-sample data U^{Hist} and correlation matrix Σ, and given:

- n: number of constituents
- u_t: a 1x*n* row vector representing scaled observations of the n constituents on day t.

The calibration formula for ν is:

$$\nu = argmax_{1<\nu<25}\left\{\sum_{t=1}^{T} log\, c(n, u_t^{Hist}, \nu, \hat{\Sigma})\right\}$$

where the function c is defined as:

$$c(n, u, \nu, \hat{\Sigma},) = \frac{1}{\sqrt{|\hat{\Sigma}|}} \frac{\Gamma\left(\frac{\nu+n}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \left(\frac{\Gamma\left(\frac{\nu}{2}\right)}{\Gamma\left(\frac{\nu+1}{2}\right)}\right)^n \frac{\left(1 + \frac{T_\nu^{-1}(u)\Sigma^{-1}T_\nu^{-1}(u^T)}{\nu}\right)^{-\left(\frac{\nu+n}{2}\right)}}{\prod_{i=1}^n \left(1 + \frac{T_\nu^{-1}(u_i)^2}{\nu}\right)^{-\left(\frac{\nu+1}{2}\right)}}$$

Monte Carlo

Halton Sequence

The Halton sequence offers a refined approach to random number generation. Instead of using pure randomness, it employs a deterministic, low-discrepancy sequence, that covers sample space more uniformly than a pure random sequence. In quasi-Monte Carlo methods, this feature reduces error margins, minimized the number of simulations required, and accelerates convergence compared to true Monte-Carlo methods, which have been criticized for needing a vast number of samples for high accuracy. (For simplicity, this quasi-Monte Carlo method will be referred to as just "Monte-Carlo".)

Application of Monte-Carlo to Copulae

Monte-Carlo methods are integrated with copulae to generate correlated uniform vectors. However, the exact procedure varies between Gaussian vs Student's t copula.

At this point the Cholesky factorized lower triangular matrix, A, is known and the random variable via the Halton sequence, Z, is generated.

Gaussian

There are 2 remaining steps to generate the matrix of correlated uniform variables:

1. Compute a vector of correlated variables:

$$X = AZ$$

2. Use the Normal Cumulative Distribution Function to map X to uniform vectors:

$$U = \Phi(X)$$

Student's t

There are 4 remaining steps to do the same:

1. Draw an independent chi-squared random variable $s \sim \chi_\nu^2$.
2. Compute the Student's t random variables:

$$Y = \frac{Z}{\sqrt{\frac{s}{v}}}$$

3. Compute a vector of correlated variables:
$$X = AY$$
4. Use the Student's t CDF to map X to uniform vectors:
$$U = T_v(X)$$

Pricing

Convert Random Variables to Default Times

For each sovereignty, every element $u \in U_s$ can be converted to default time τ using its own hazard rate term structure.

Deriving the time to default consists of 2 parts. The first is to find the year of default, the second is to find the additional fractional year.

The year of default and hazard rate at the year of default is expressed using the infimum function:

$$inf\left\{(t_{m-1}, \lambda_m) \mid t > 0\text{: } log(1-u) \geq -\sum_{0}^{t} \lambda_m\right\}$$

Where default occurs if the inequality holds. Thus t_{m-1} is the year of default and λ_{m} is the hazard rate on the year of default.

The fractional portion, δt, can then be expressed as:

$$\delta t = -\frac{1}{\lambda_m} log\left(\frac{1-u}{P(0, t_{m-1})}\right)$$

Thus, the exact default time is expressed as:

$$\tau = t_{m-1} + \delta t$$

After converting all $u \rightarrow \tau$, the default time vectors are arranged in ascending order as $(\tau_1, \tau_2, \tau_3 \ldots \tau_N)$ in preparation for pricing.

Spread Pricing

For every k_{th}-to-Default instrument, the spread is evaluated as the quotient of the default leg's present value to the premium leg's present value:

$$S = \frac{DL}{PL}$$

To model the spread, DL and PL are simulated and averaged separately.

Given:

- N as the number of constituents in the basket.
- K as the number of defaults.
- RR as the recovery rate.
- $L = 1 - RR$ as the Loss Given Default.
- $D(0,t)$ as the discount factor at T=0.
- Minimum τ constrained to 0.25 (very small default times can lead to large spreads and interfere with convergence).
- $\tau_0 = 0$.

Default Leg

The present value function for the default leg, given τ, is:

$$DL(\tau) = \begin{cases} 0 & if\ \tau < 0.25\ \vee\ \tau \geq N \\ LD(0,\tau)\left(\frac{1}{N}\right) & if\ \tau \geq 0.25\ \wedge\ \tau < N \end{cases}$$

Premium Leg

For the premium leg, there are two methods, with removal and without removal.

- With removal:

$$PL_R(\tau) = \begin{cases} D(0,T)T & if\ \tau < 0.25\ \vee\ \tau \geq N \\ D(0,\tau)\tau & if\ \tau \geq 0.25\ \wedge\ \tau < N \end{cases}$$

- Without Removal (not applicable to first-to-default):

$$PL_{NR}(\tau) = \begin{cases} D(0,T)T & if\ \tau < 0.25\ \vee\ \tau \geq N \\ \sum_{k=0}^{K-1} D(0,\tau_{k+1})(\tau_{k+1} - \tau_k)\left(\frac{K-1}{K}\right) & if\ \tau \geq 0.25\ \wedge\ \tau < N \end{cases}$$

Given μ simulations, fair spread can be calculated as follows (using either PL_R or PL_{NR}):

$$S = \frac{\sum_{i=1}^{\mu} DL_i \big/ \mu}{\sum_{i=1}^{\mu} PL_i \big/ \mu}$$

Application to Asian Sovereign CDS

Parameters and Configuration

The model is employed to price a k_{th}-to-Default basket of 5 Asian sovereign CDS:

- Japan
- China
- Singapore
- Malaysia
- Thailand

Credit and Discount Curve Snapshot Date: 06/30/2023

Curve Tenors: 1Y-5Y at 1Y increments.

Historical Weekly CDS Spreads Date Range: 06/30/2018 – 06/30/2023

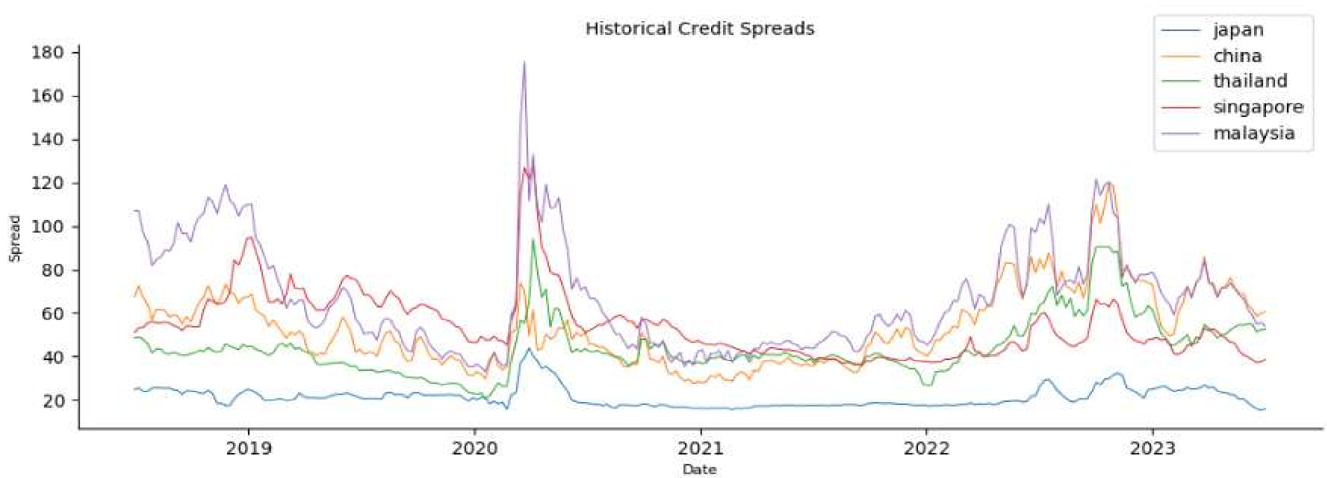
Static Recovery Rate: 40%

*Information on data sources is stated in the appendix.

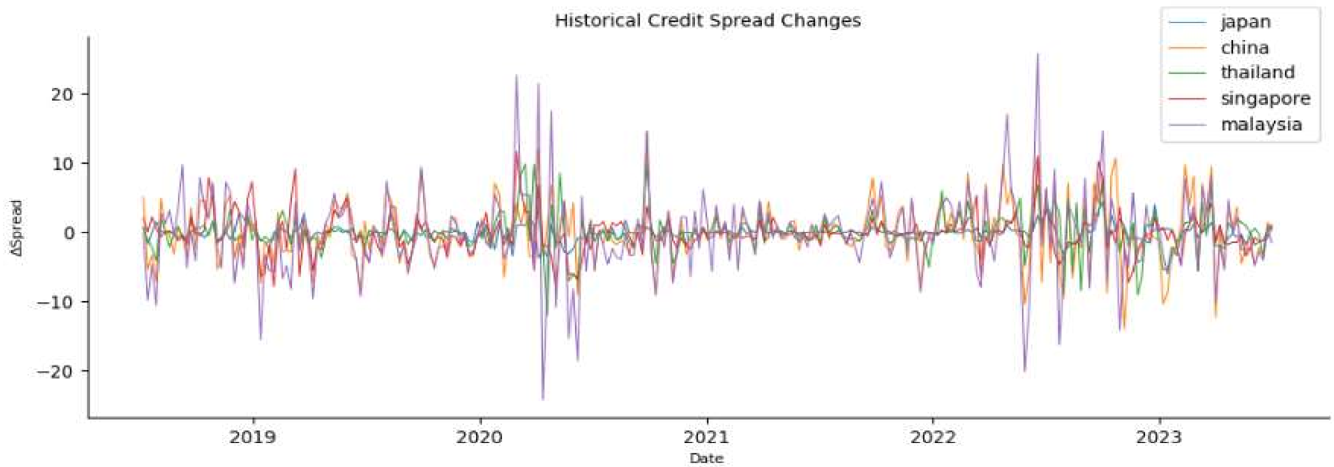
Input Data Exploration

Historical Spreads

The figure below illustrates the historical weekly CDS spreads for the 5 sovereignties. There is a discernible correlation among the data points. The pronounced spread heightening during the COVID-19 period is also visually evident.



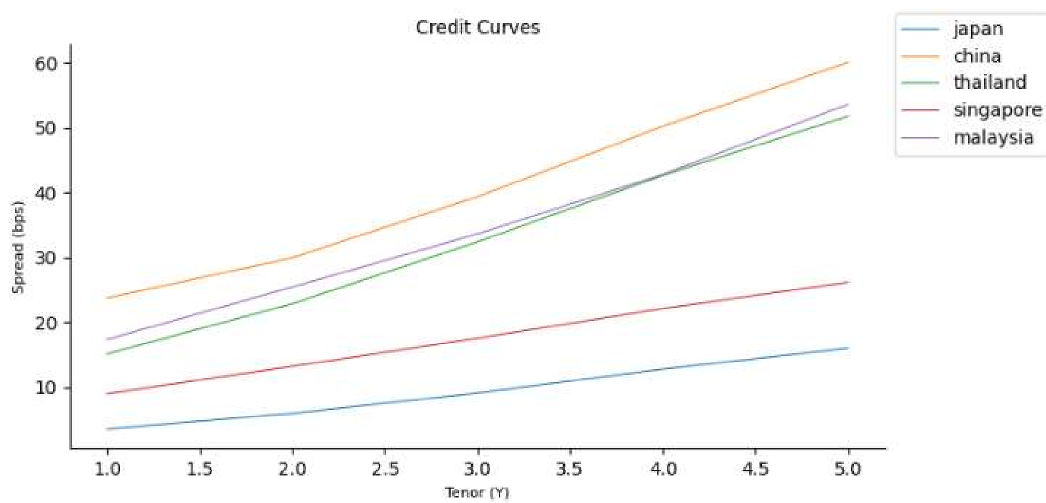
The figure below illustrates the weekly changes in CDS spreads. This provides a direct visual representation of the input dataset into Kendall's tau correlation.



As mentioned, Kendall’s tau emphasizes relationships of pronounced values across variables. As such, this method will likely highlight the significant spread movements during the COVID-19 crash and the macroeconomic shifts in 2022.

Credit Curves

The below credit curve graph illustrates the market-perceived risk of each sovereignty over a 5-year horizon.



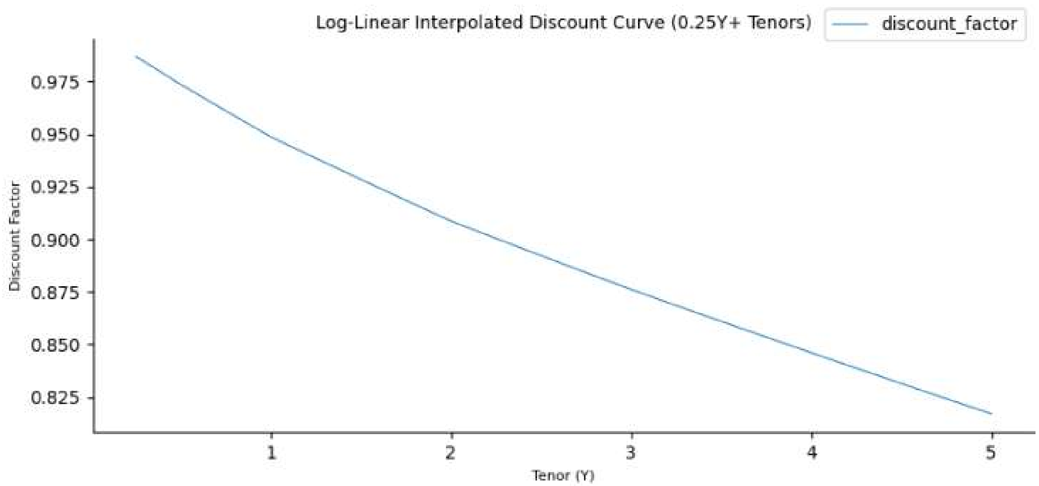
All the sovereignties display normal, upward sloping curves. Credit curves can take on various shapes. The three fundamental shapes are:

- Upward sloping: This standard form suggests that credit risks increase with longer maturities. This pattern is largely due to the inherent challenge of predicting longer-term events. Thus, investors demand higher risk premiums not solely due to evident credit risks, but also the general increased uncertainty with longer durations.
- Flat curve: A consistent spread across all maturities indicates that the market perceives uniform credit risk through varying time horizons.

- ❖ Downward sloping: Or inverted curve. This occurs when short-term credit risks overshadow those in the longer term. The diminishing spread reflects the market's perception of the swap's diminishing relevance. Buyers who purchase longer-tenor CDS on short-term high probability of default credits will likely be paying premium for a longer period than warranted, thus the spread is reduced for further tenors.

Discount Curve

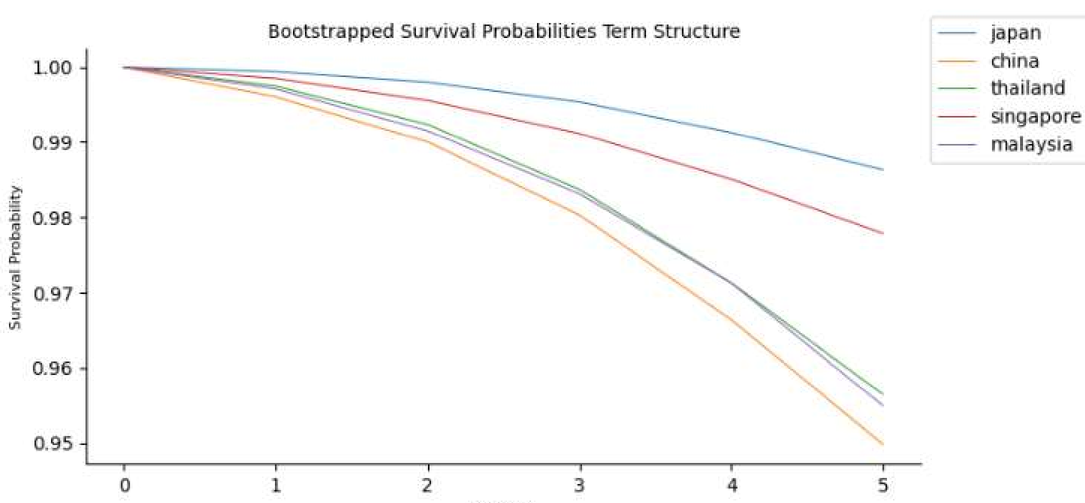
The below graph illustrates the log-linearly interpolated discount curve employed for this pricing exercise. It spans a tenor from 0.25Y to 5Y, with default times below 0.25Y being excluded. Using a continuous discount curve enhances the precision of spread computations, eliminating the need to round to the nearest quarter or half-year interval when discounting.



Bootstrapping

Survival Probabilities

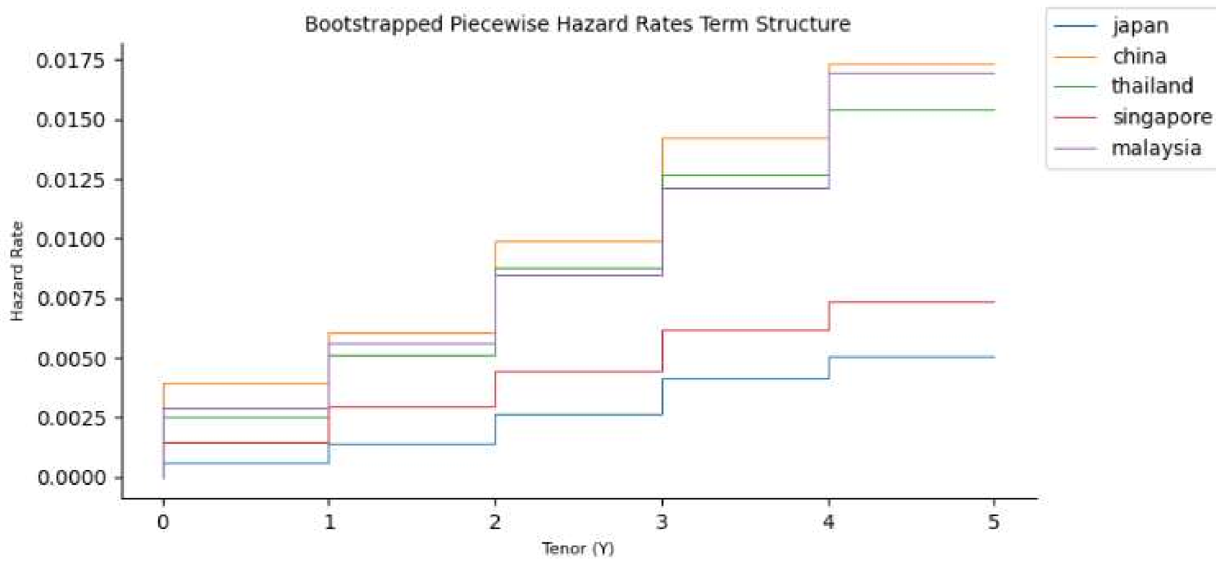
The bootstrapped survival probabilities are depicted in the figure below.



The survival probabilities term structure is bootstrapped from the credit curve. When spreads are higher, the implied survival probability is lower. In contrast, the default probability, which is expressed as $PD(0,t) = 1 - P(0,t)$, moves in tandem with the credit spreads.

Hazard Rates

The hazard rates term structure is bootstrapped from the survival probabilities.



Hazard rates are represented using a piecewise constant function. As mentioned previously, this offers a high level of practicality. The figure above also denotes the empirical basis for generating default times through Monte-Carlo simulations.

The shape of the hazard rates and survival probability term structures for these 5 sovereignties coincide with the standard sloped nature of the underlying credit curves.

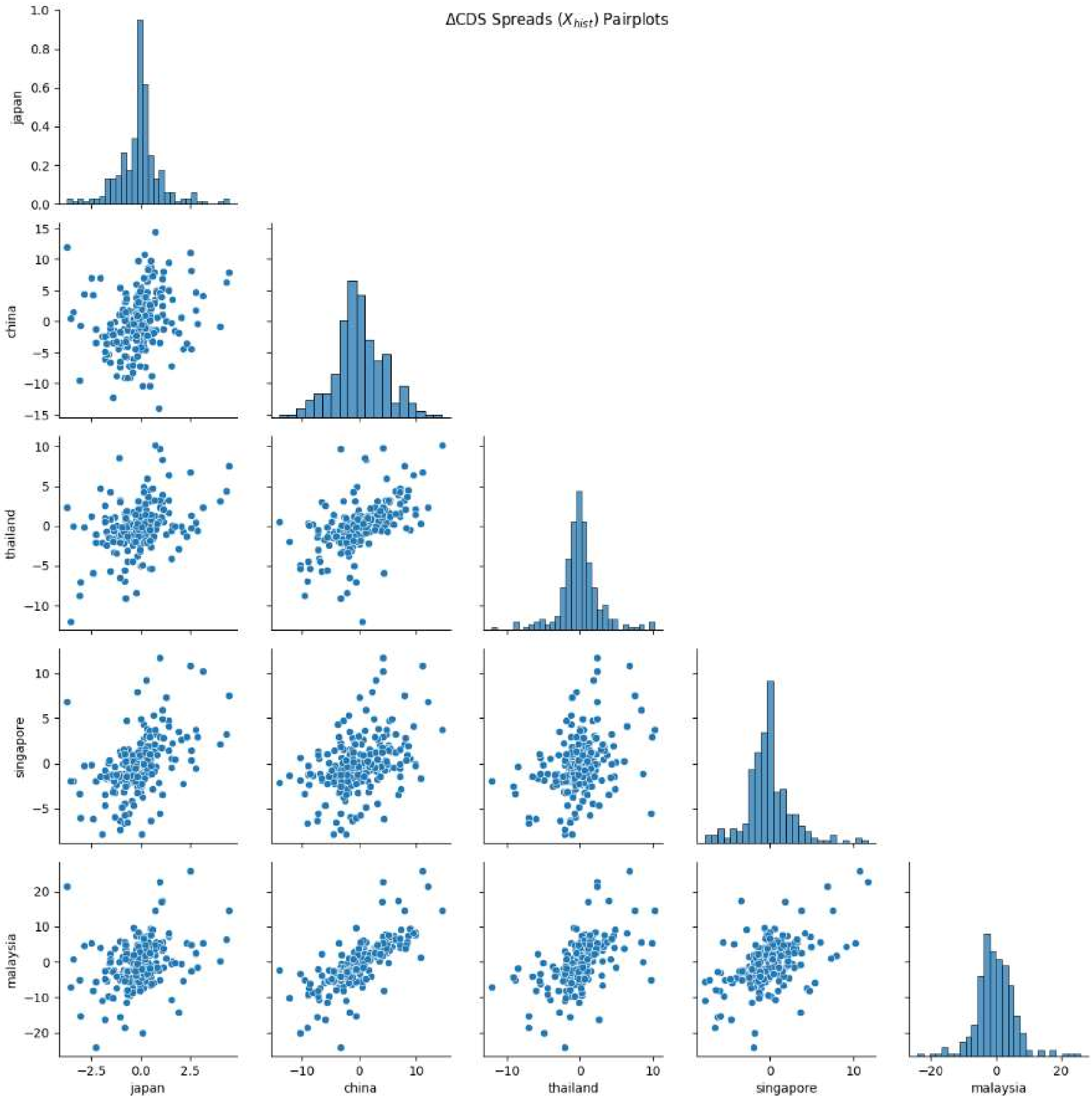
Copula

Correlation Matrices

Three distinct input datasets were devised for the copulae, each tailored for a specific correlation method:

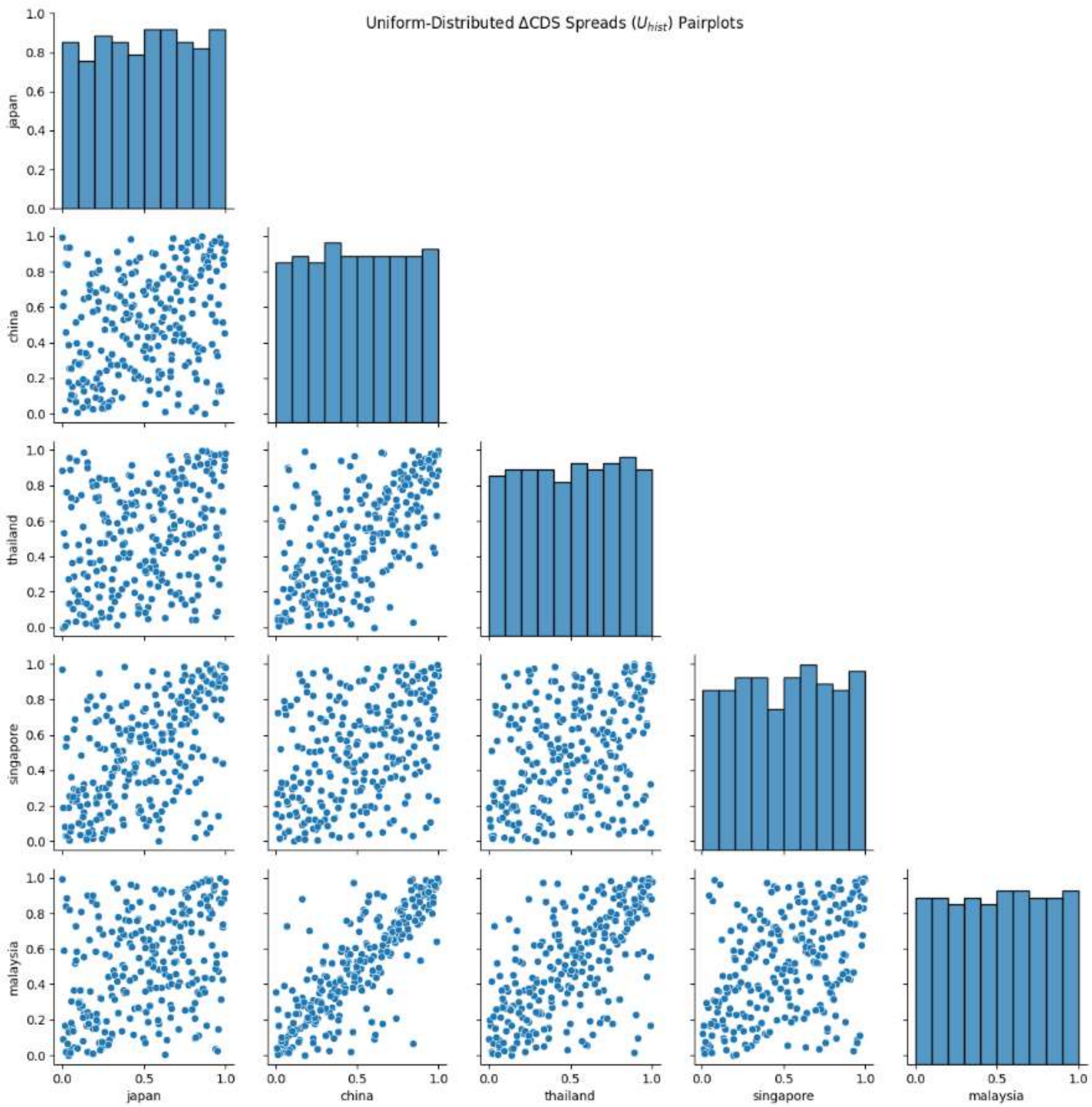
- ΔCDS Spreads: X^{Hist} → Kendall's tau
- Uniform pseudo-samples derived from X^{Hist}: U^{Hist} → Spearman's rho
- Standard normal distribution scaling of U^{Hist}: Z^{Hist} → Pearson

The following pair-plots illustrate the relationships between variables in these three datasets. The lower triangular section showcases scatter plots for pairs of sovereignties, while the diagonals capture the distribution specific to each sovereignty. The accompanying tables display the correlation matrices generated with the respective method.



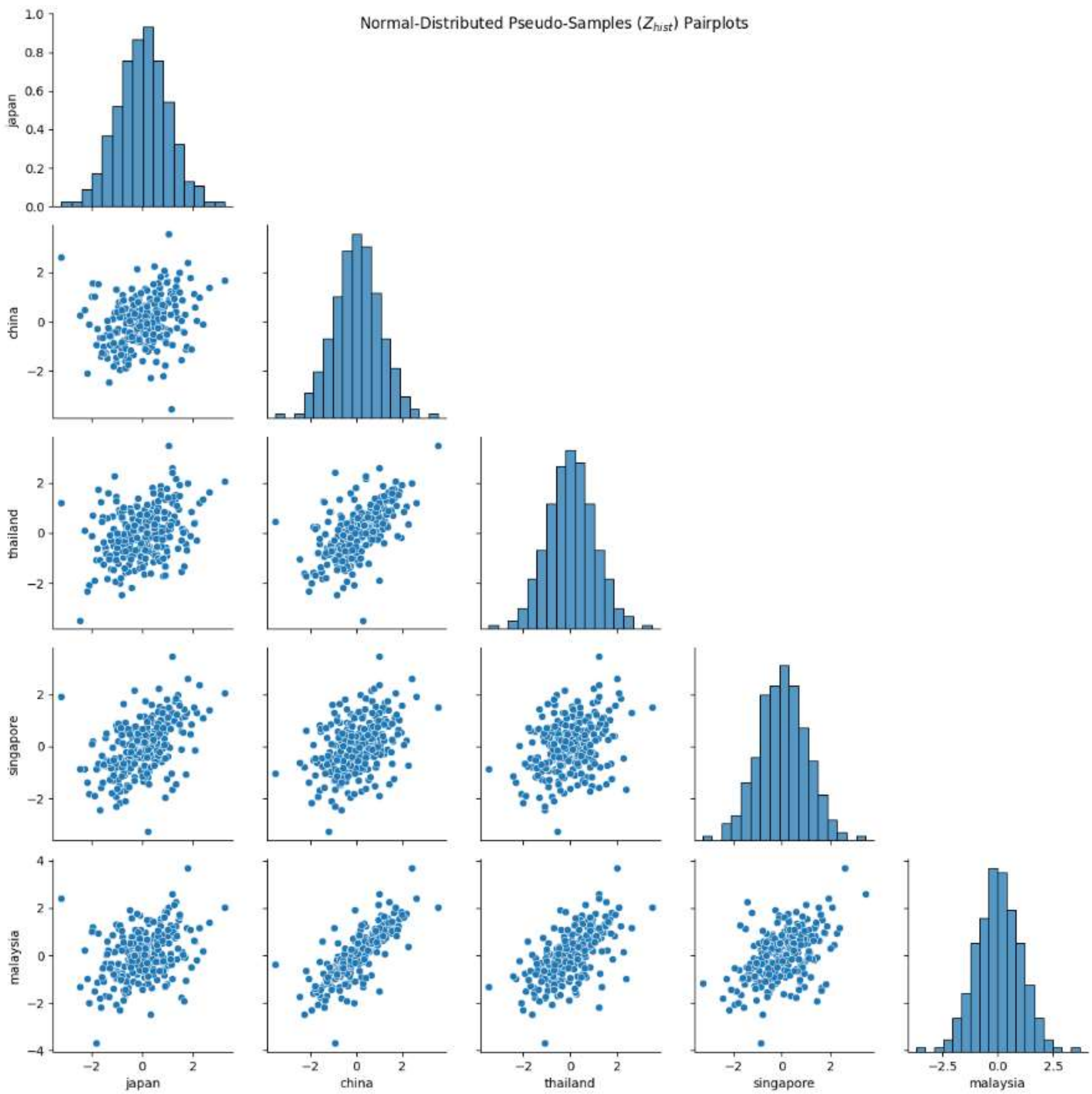
The individual distributions do not conform to any parametric distribution, highlighting the need for an empirical kernel smoothing density estimation function to convert to a uniform distribution.

Student's t (Kendall's tau)	Japan	China	Thailand	Singapore	Malaysia
Japan	1	0.350599	0.32573	0.587066	0.381747
China	0.350599	1	0.659253	0.420798	0.86095
Thailand	0.32573	0.659253	1	0.351383	0.678104
Singapore	0.587066	0.420798	0.351383	1	0.476108
Malaysia	0.381747	0.86095	0.678104	0.476108	1



The transformation of variables to a uniform distribution appears largely effective. The notable even dispersion of points throughout the spaces suggests inherent uniformity. Consequently, the tail dependencies in these copulae are more pronounced compared to the Gaussian copula. This will result in the 2^{nd}, 3^{rd}... n^{th} to-default spreads being higher relative to the Gaussian copula as there will be a higher co-concentration of earlier default times.

Student's t (Spearman's rho)	Japan	China	Thailand	Singapore	Malaysia
Japan	1	0.340008	0.315755	0.567559	0.371533
China	0.340008	1	0.641894	0.416053	0.839401
Thailand	0.315755	0.641894	1	0.344188	0.659404
Singapore	0.567559	0.416053	0.344188	1	0.460743
Malaysia	0.371533	0.839401	0.659404	0.460743	1



The transformation of each variable's scores back to a normal distribution appears clean, and the concentration of points along the diagonal suggests effective normalization. Tail dependency has diminished, thus the 2nd, 3rd... n^{th} to-default spreads are expected to be lower.

Gaussian (Pearson)	Japan	China	Thailand	Singapore	Malaysia
Japan	1	0.281848	0.338171	0.523443	0.359867
China	0.281848	1	0.592775	0.413535	0.792661
Thailand	0.338171	0.592775	1	0.353977	0.616127
Singapore	0.523443	0.413535	0.353977	1	0.476528
Malaysia	0.359867	0.792661	0.616127	0.476528	1

Cholesky Factors

Using Cholesky Decomposition on the 3 correlation matrices yields the below Cholesky factors:

Student's t (Kendall's tau)

1.0000				
0.3506	0.9365			
0.3257	0.5820	0.7451		
0.5871	0.2295	0.0357	0.7755	
0.3817	0.7764	0.1368	0.0889	0.4742

Student's t (Spearman's rho)

1.0000				
0.3400	0.9404			
0.3158	0.5684	0.7598		
0.5676	0.2372	0.0397	0.7874	
0.3715	0.7583	0.1462	0.0815	0.5089

Gaussian (Pearson)

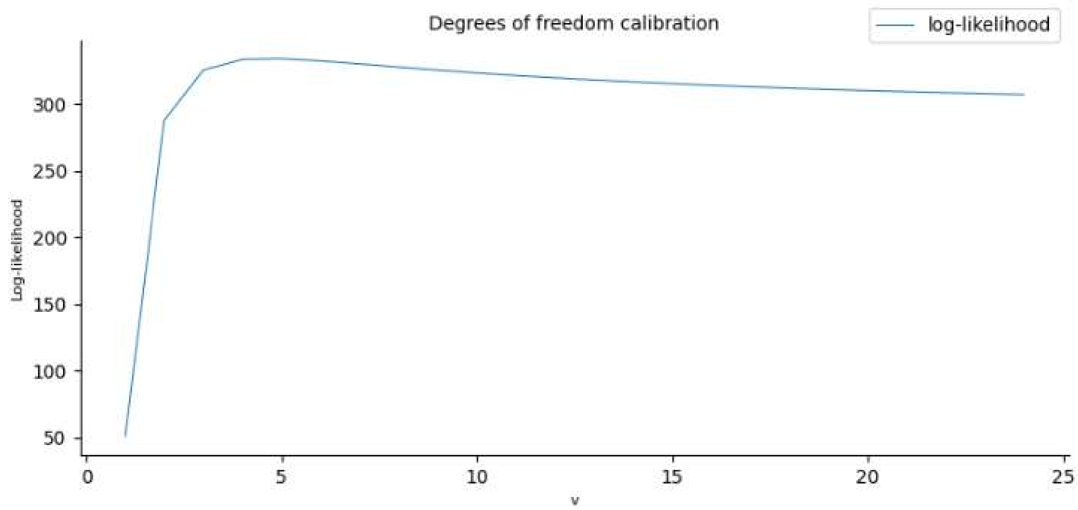
1.0000				
0.2818	0.9595			
0.3382	0.5185	0.7854		
0.5234	0.2772	0.0423	0.8046	
0.3599	0.7204	0.1539	0.1018	0.5634

These will serve as factors for simulated random numbers to generate correlated random variables.

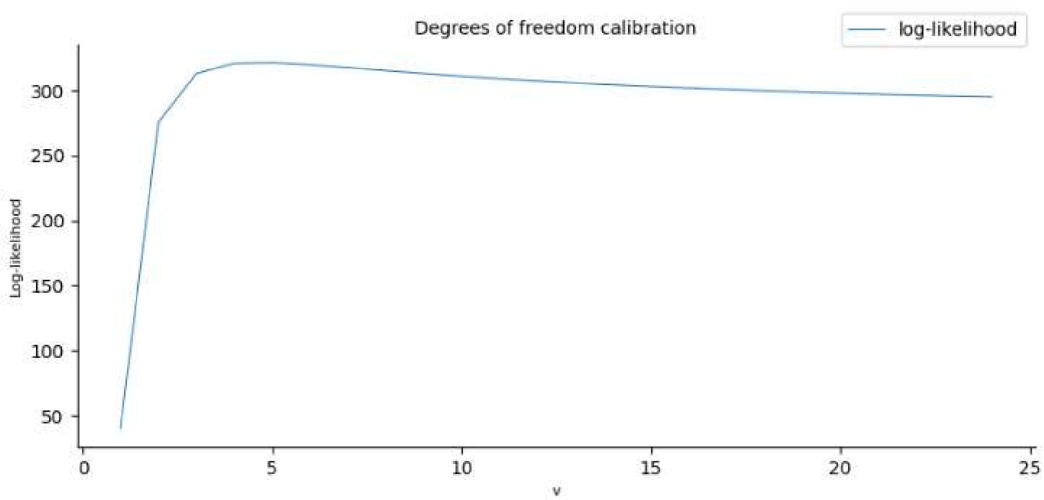
Student's t Degrees of Freedom Calibration

The degrees of freedom for the t-copula were calibrated to 5 using both correlation methodologies. The log-likelihood graphs between the two methods appear near-identical:

❖ Kendall's tau



❖ Spearman's rho



Final Pricing Table

		1	2	3	4	5
Gaussian	**Without Removal**	30.99 bps	9.83 bps	2.98 bps	0.52 bps	0.06 bps
	With Removal		9.78 bps	2.97 bps	0.52 bps	0.06 bps
Students t - Spearman	**Without Removal**	26.68 bps	11.38 bps	4.77 bps	1.27 bps	0.25 bps
	With Removal		11.32 bps	4.76 bps	1.27 bps	0.25 bps
Students t - Kendall	**Without Removal**	26.14 bps	11.58 bps	4.99 bps	1.32 bps	0.28 bps
	With Removal		11.52 bps	4.98 bps	1.32 bps	0.28 bps

The above table, generated with 1,000,000 Monte Carlo simulations, reflects the 1st through 5th-to-default spreads using all relevant combinations of parametric distributions, correlation methods, and premium leg conventions.

Defining the nomenclature for the k_{th}-to-Default spectrum, instruments with lower values of *k* are termed *junior tranches* and instruments with higher values of *k* are termed *senior tranches*. As *k* progresses from 1 to *K*, the spectrum transitions from junior to senior.

Observations below:

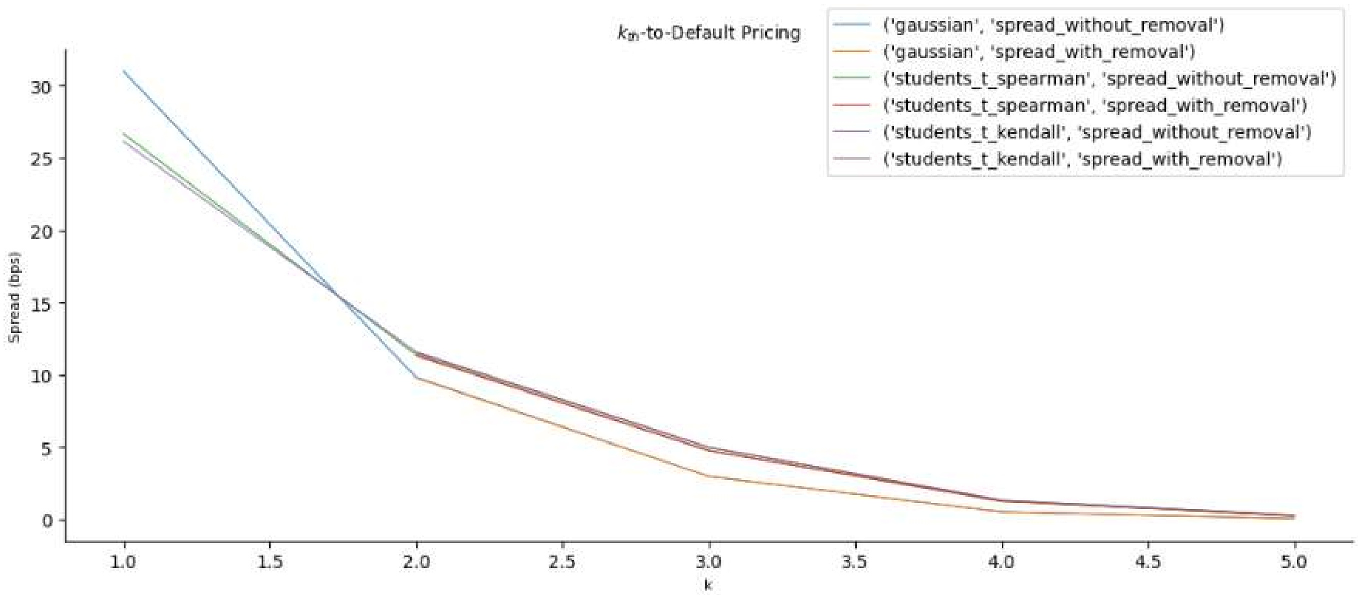
- Comparing premium leg conventions, the "with removal" convention priced cheaper than the "without removal" convention for the junior tranches but converges to the same approaching the senior tranches.
- Across all method combinations, spreads diminish when moving over the spectrum of junior to senior tranches. There is a higher hurdle of activation for the senior tranches as those require more defaults to occur as a prerequisite, thus the likelihood of activation is lower. The magnitude of this trend varies with the underlying correlations, and exceptions may apply if the correlations happen to be uniformly extreme at 1 or -1 across the basket.

  Given a probability of activation $\mathrm{p_k}$ for each tranche in the basket where $0 \leq \mathrm{p_k} \leq 1$, the probability associated with the $\mathrm{K^{th}}$ activation is given by:
  $$\mathrm{p_K} = \prod_{\mathrm{k}=1}^{\mathrm{K}} \mathrm{p_k}$$
  The probability of the K^{th} activation is the cumulative product of activation probabilities of all tranches up to K. Since these probabilities lie between 0 and 1, $\mathrm{p_K}$ tends towards 0 with increasing K. For these reasons, the spread for k should always be less than $\mathrm{k} - 1$.
- For the 1st-to-Default instrument (the "equity tranche"), Gaussian pricing is more expensive than that determined using Student's t. However, the opposite is true for all subsequent tranches. This pattern is due to the higher tail dependencies exhibited by the Student's t copula in comparison to the Gaussian copula. This implies a higher likelihood of joint extreme events, such as defaults, and thus the cumulative product of activation probabilities up to K is higher. Basically, a default in one credit implies a higher likelihood of

default in successive credits, which may chain up to the more senior tranches as default contagion continues.

- Under the Student's t copula, the equity tranche exhibits a lower price with Kendall's tau correlation than with Spearman's rho correlation. However, the opposite holds for subsequent tranches. The spread differences seem to narrow approaching the senior tranches, but this trend's persistence is uncertain since the basket is limited to just 5 constituents. Investigating this would require a larger basket.

The below figure visually represents spread pricing over k.



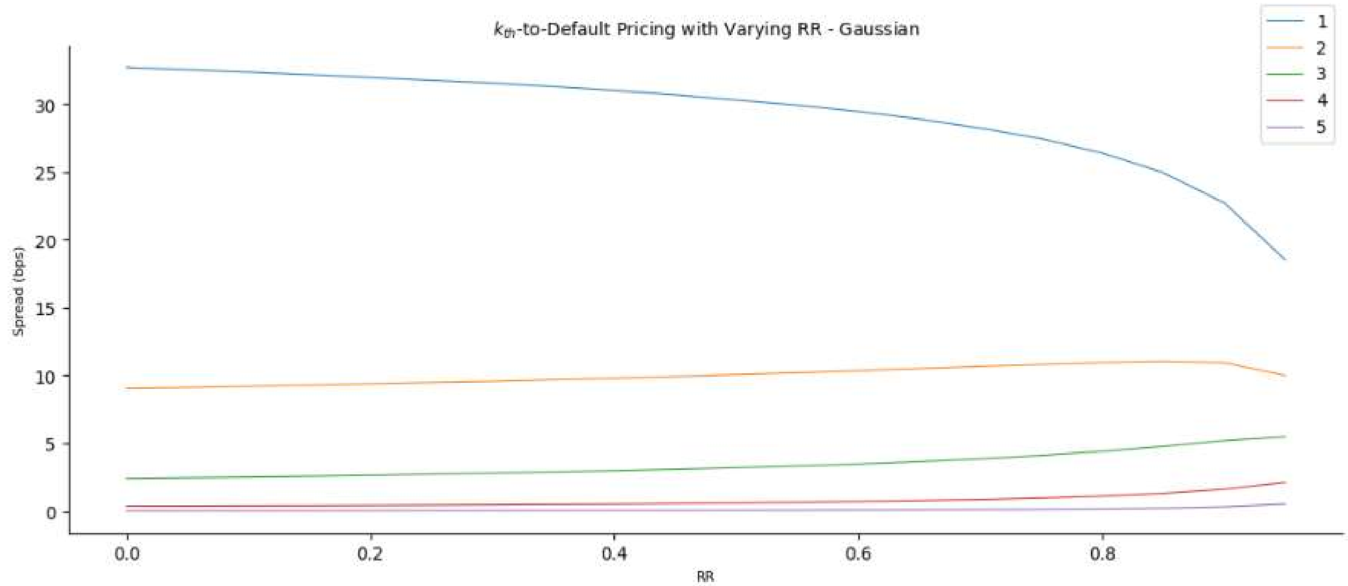
Risk and Sensitivity Analysis

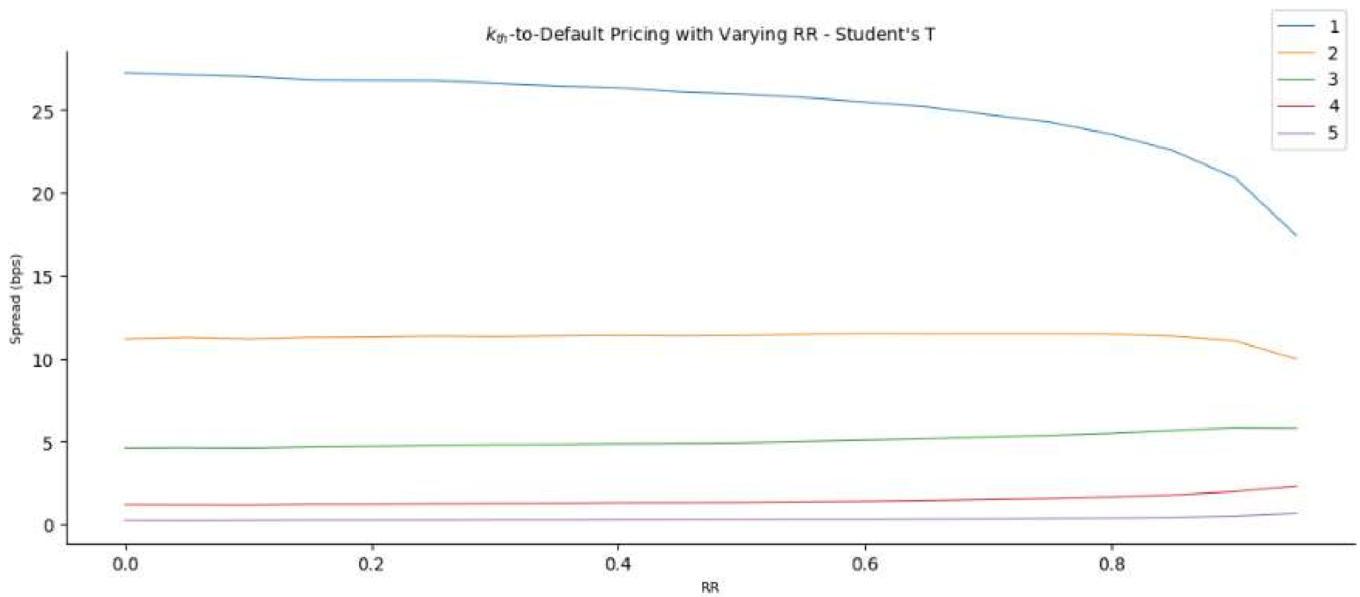
To study the influence of individual factors, the model is recalibrated using a spectrum of values for the specific factor, while keeping all others constant. The individual factors examined are:

- Recovery Rates: Ranging from 0 to 95% with a step size of 5%.
- Correlation Matrices: Ranging from $\pm$35% of actual values with a step size of 1%.
- Credit Curves: Ranging from $\pm$35% of actual values with a step size of 1%.

Each pricing iteration is based on 1,000,000 Monte Carlo simulations. For simplicity, the premium leg convention and correlation method combinations were consolidated, yielding a single average price for each copula type. Thus, an aggregated tranche price is presented for both Gaussian and Student's t copulae.

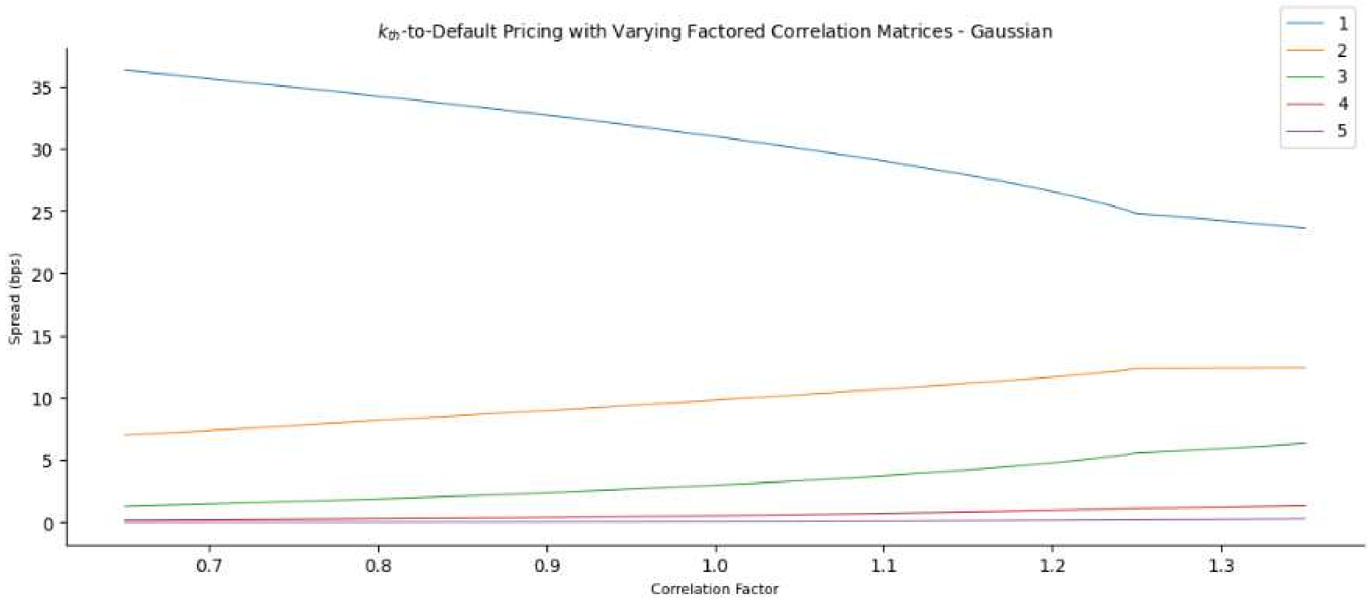
Recovery Rate Sensitivities

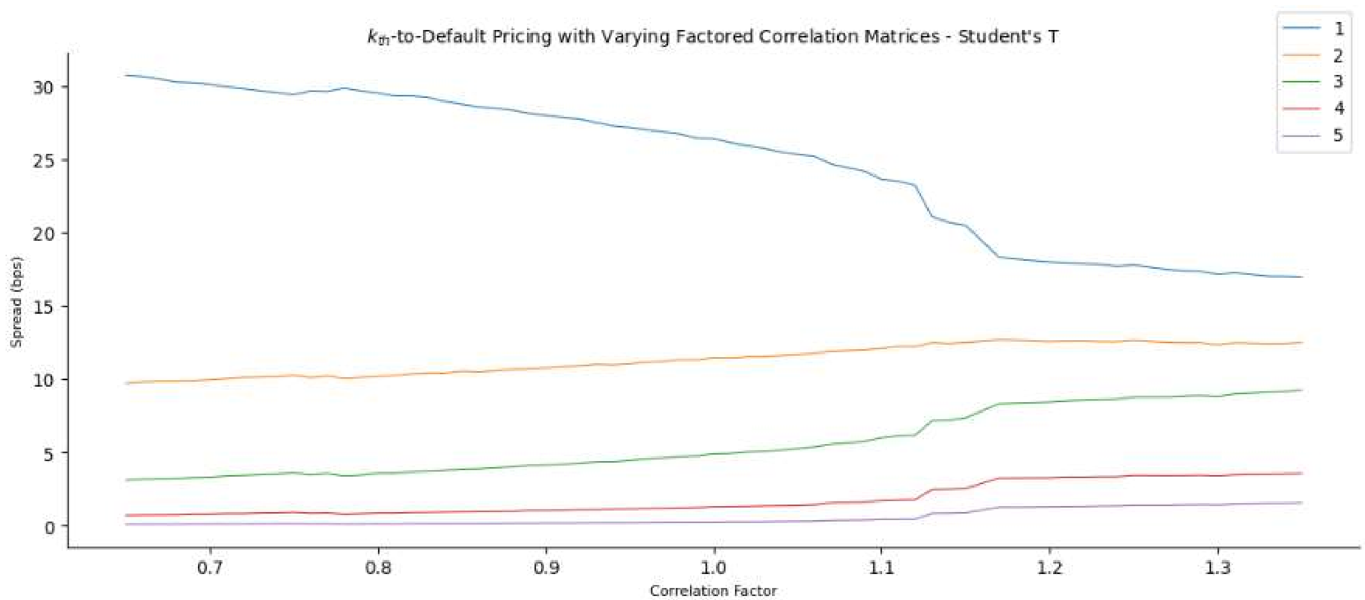




Both copula reflect a consistent pricing trend as the recovery rate approaches 95%. The 1st-to-Default tranche exhibits the highest sensitivity to changes in the recovery rate, while the other tranches remain relatively stable. An increase in recovery rate exerts a downward pressure on the equity tranche. This is attributed to the fact that as recovery rates rise, the potential loss given default declines, reducing the relevance of the CDS. Theoretically, if the recovery rate was known to be 95%, then the swap spread should be close to 0. However, the exact recovery rate cannot be determined ahead of time and the realized rate during a default is dependent on various factors. Thus, even though the equity tranche spread decreases as recovery rate increases, it remains elevated to compensate the protection seller for the inherent uncertainty.

Correlation Factor Sensitivities





The model was recalibrated using correlation matrices factored by $\pm$35% from actual values. Both copulas reflect the same general trend, but the Gaussian copula exhibits a smoother change than Student's t.

The 1st-to-Default tranche spread declines with increasing correlation, whereas the opposite occurs for subsequent tranches. There are a few methods to explain this.

Consider an extreme scenario where all correlations are equal to 1, meaning all constituents have identical historical default risk dynamics. Also, consider hypothetically that hazard rate term structures were similar or identical. With this setup, the constituents in the basket would essentially behave as a single asset. Thus, a default of one translates to the default of all, leading to uniform prices across trances due to this strong interdependence.

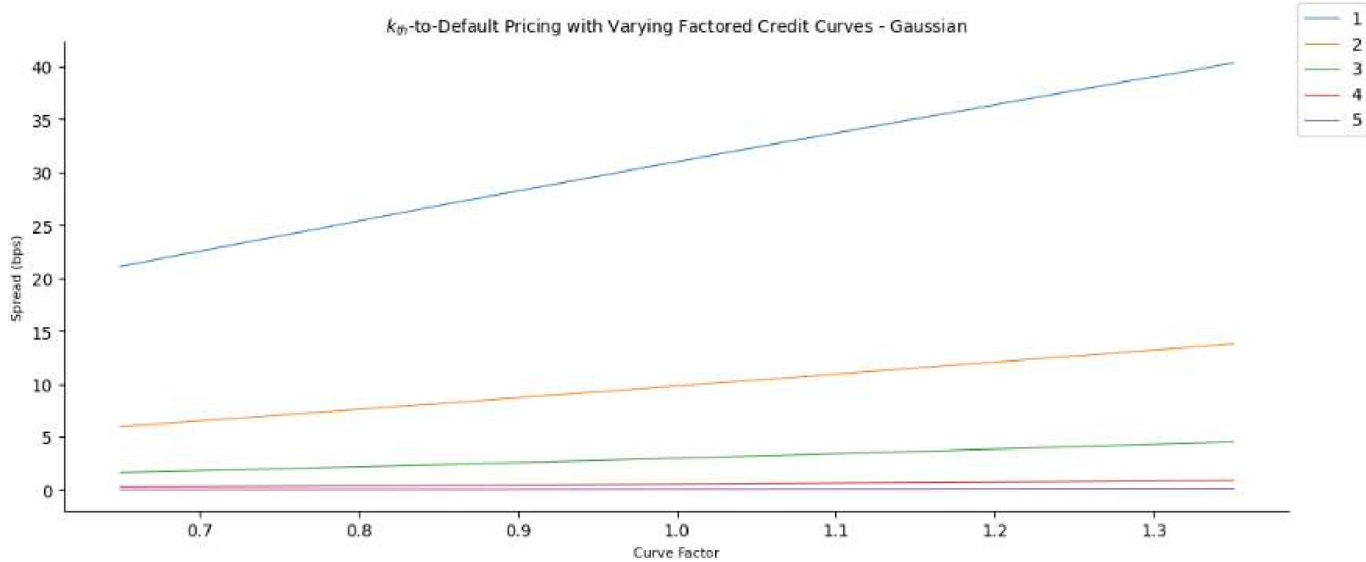
Now consider the opposite scenario where all correlations are equal to 0. Let the hazard rate term structure take any arbitrary form. The basket is now highly diversified, and the default in any one constituent is independent of defaults in another, making multiple defaults less likely. However, the basket experiencing at least one default is more probable given the variety in risk profiles.

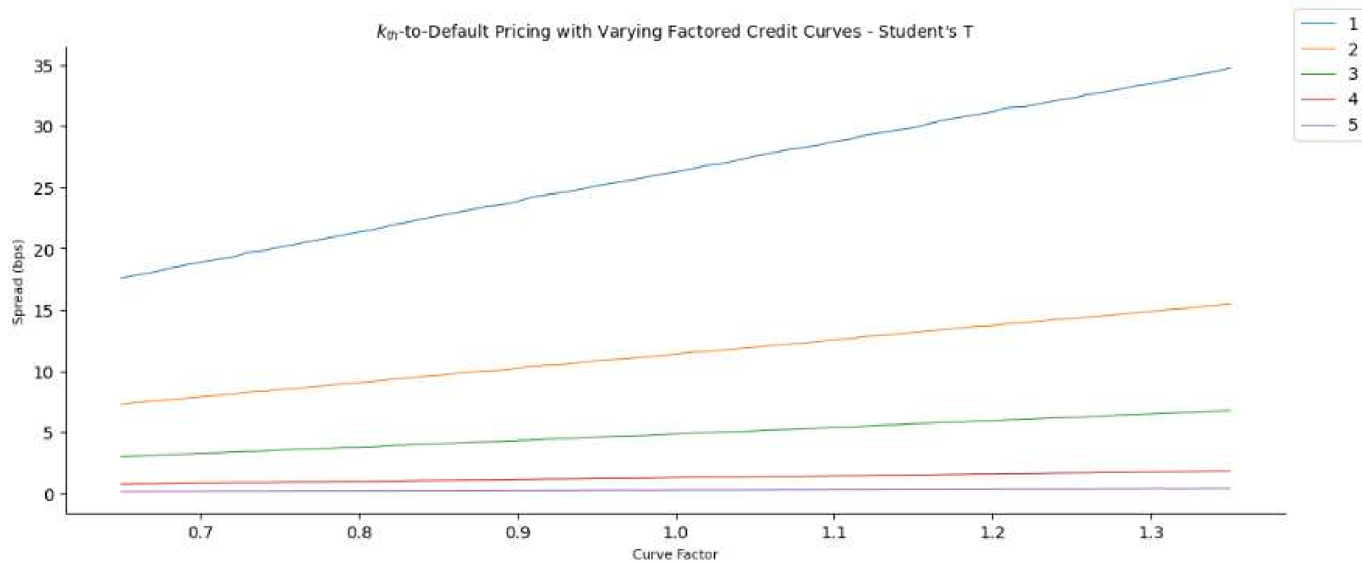
Viewing this through the lens of conditional probabilities where $\mathrm{D_k}$ denotes a realized default for the $\mathrm{k^{th}}$ instrument and P representing probability, then:

- With perfect correlation: $\mathrm{P(D_k|D_{k-1})} \approx 1$, thus price for $\mathrm{k} \approx \mathrm{k} - 1$.
- With zero correlation: $\mathrm{P(D_k|D_{k-1})} = \mathrm{P(D_k)}$, thus price for $\mathrm{k} \perp \mathrm{k} - 1$.

To simplify, high correlation yields an “all or nothing” default dynamic, whereas low correlation yields a “certainly some, but not all” default dynamic.

Credit Curve Factor Sensitivities





The model was recalibrated using credit curves all factored by $\pm$35% from actual values. Both copulas exhibit the same pattern, all spreads increase as the curve factors increase, albeit at different rates. The spread for the 1st-to-Default tranche, which has been established as generally the most sensitive tranche, increases at a faster pace than the subsequent tranches.

This is intuitively obvious, a steeper credit curve indicates greater market-implied default risk, and these heightened risk premiums naturally get embedded into the pricing of basket CDS products.

Concluding Remarks

The k^{th}-to-Default pricing model, underpinned by the Gaussian and Student's t copulae, helps provide insight into the interconnected risks of various assets in a basket. The model emphasizes the roles of credit curves, historical default risk dynamics, and other factors to price tranches distinctly. While this research lays the foundational knowledge of the model through simplified parameters, there are a few caveats to note:

- Pricing is calibrated from a risk-neutral, market-implied standpoint. Actual prices may vary due to live trading dynamics, i.e supply-demand dynamics, liquidity constraints, market dislocations, and trading fees.
- This paper adopts standard market practice by setting a uniform recovery rate of 40% for all constituents. This is a general market assumption when valuing CDS. However, recovery rates differ in a live environment, and modeling such is difficult, and each constituent would have to be modeled individually.
- It is difficult to source historical CDS spread data, it usually incurs a high cost.
- The model relies on historical correlations and current credit curves to project future default scenarios as of a certain date. The model does not account for shifts in credit curves over time, that would need to be modeled separately and integrated.
- The model does not consider counterparty risk, this additional risk exposure would need to be modeled separately and managed.

Appendix

Data Sources

All data were sourced from Bloomberg. Refer to the table below for more information:

Type	Description	BBG Security Description	Pricing Source
CDS Curve	Japan CDS 3M Spread	JGB CDS USD SR 3M D14	CMAN
CDS Curve	Japan CDS 6M Spread	JGB CDS USD SR 6M D14	CMAN
CDS Curve	Japan CDS 1Y Spread	JGB CDS USD SR 1Y D14	CMAN
CDS Curve	Japan CDS 2Y Spread	JGB CDS USD SR 2Y D14	CMAN
CDS Curve	Japan CDS 3Y Spread	JGB CDS USD SR 3Y D14	CMAN
CDS Curve	Japan CDS 4Y Spread	JGB CDS USD SR 4Y D14	CMAN
CDS Curve	Japan CDS 5Y Spread	JGB CDS USD SR 5Y D14	CMAN
CDS Curve	China CDS 3M Spread	CHINAGOV CDS USD SR 3M D14	CMAN
CDS Curve	China CDS 6M Spread	CHINAGOV CDS USD SR 6M D14	CMAN
CDS Curve	China CDS 1Y Spread	CHINAGOV CDS USD SR 1Y D14	CMAN
CDS Curve	China CDS 2Y Spread	CHINAGOV CDS USD SR 2Y D14	CMAN
CDS Curve	China CDS 3Y Spread	CHINAGOV CDS USD SR 3Y D14	CMAN
CDS Curve	China CDS 4Y Spread	CHINAGOV CDS USD SR 4Y D14	CMAN
CDS Curve	China CDS 5Y Spread	CHINAGOV CDS USD SR 5Y D14	CMAN
CDS Curve	Singapore CDS 3M Spread	SINGP CDS USD SR 3M D14	CMAN
CDS Curve	Singapore CDS 6M Spread	SINGP CDS USD SR 6M D14	CMAN
CDS Curve	Singapore CDS 1Y Spread	SINGP CDS USD SR 1Y D14	CMAN
CDS Curve	Singapore CDS 2Y Spread	SINGP CDS USD SR 2Y D14	CMAN
CDS Curve	Singapore CDS 3Y Spread	SINGP CDS USD SR 3Y D14	CMAN
CDS Curve	Singapore CDS 4Y Spread	SINGP CDS USD SR 4Y D14	CMAN
CDS Curve	Singapore CDS 5Y Spread	SINGP CDS USD SR 5Y D14	CMAN
CDS Curve	Thailand CDS 3M Spread	THAI CDS USD SR 3M D14	CMAN
CDS Curve	Thailand CDS 6M Spread	THAI CDS USD SR 6M D14	CMAN
CDS Curve	Thailand CDS 1Y Spread	THAI CDS USD SR 1Y D14	CMAN
CDS Curve	Thailand CDS 2Y Spread	THAI CDS USD SR 2Y D14	CMAN
CDS Curve	Thailand CDS 3Y Spread	THAI CDS USD SR 3Y D14	CMAN
CDS Curve	Thailand CDS 4Y Spread	THAI CDS USD SR 4Y D14	CMAN
CDS Curve	Thailand CDS 5Y Spread	THAI CDS USD SR 5Y D14	CMAN

CDS Curve	Malaysia CDS 3M Spread	MALAYS CDS USD SR 3M D14	CMAN
CDS Curve	Malaysia CDS 6M Spread	MALAYS CDS USD SR 6M D14	CMAN
CDS Curve	Malaysia CDS 1Y Spread	MALAYS CDS USD SR 1Y D14	CMAN
CDS Curve	Malaysia CDS 2Y Spread	MALAYS CDS USD SR 2Y D14	CMAN
CDS Curve	Malaysia CDS 3Y Spread	MALAYS CDS USD SR 3Y D14	CMAN
CDS Curve	Malaysia CDS 4Y Spread	MALAYS CDS USD SR 4Y D14	CMAN
CDS Curve	Malaysia CDS 5Y Spread	MALAYS CDS USD SR 5Y D14	CMAN
CDS Historical Spreads	Japan CDS 5Y Spread	JGB CDS USD SR 5Y D14	CBGN*
CDS Historical Spreads	China CDS 5Y Spread	CHINAGOV CDS USD SR 5Y D14	CBGN*
CDS Historical Spreads	Singapore CDS 5Y Spread	SINGP CDS USD SR 5Y D14	PRXY*
CDS Historical Spreads	Thailand CDS 5Y Spread	THAI CDS USD SR 5Y D14	CBGN*
CDS Historical Spreads	Malaysia CDS 5Y Spread	MALAYS CDS USD SR 5Y D14	CBGN*
Discount Yield Curve	US Treasury 3M Zero-Coupon Yield	USD Treasury Actives (IYC 25) Zero Coupon Yield 3 Month	N/A
Discount Yield Curve	US Treasury 6M Zero-Coupon Yield	USD Treasury Actives (IYC 25) Zero Coupon Yield 6 Month	N/A
Discount Yield Curve	US Treasury 1Y Zero-Coupon Yield	USD Treasury Actives (IYC 25) Zero Coupon Yield 1 Year	N/A
Discount Yield Curve	US Treasury 2Y Zero-Coupon Yield	USD Treasury Actives (IYC 25) Zero Coupon Yield 2 Year	N/A
Discount Yield Curve	US Treasury 3Y Zero-Coupon Yield	USD Treasury Actives (IYC 25) Zero Coupon Yield 3 Year	N/A
Discount Yield Curve	US Treasury 4Y Zero-Coupon Yield	USD Treasury Actives (IYC 25) Zero Coupon Yield 4 Year	N/A
Discount Yield Curve	US Treasury 5Y Zero-Coupon Yield	USD Treasury Actives (IYC 25) Zero Coupon Yield 5 Year	N/A

* Denotes that some values were manually sourced using the CMAN pricing source.

References

- *Credit Default Swaps* by Dr. Jon Gregory
- *Credit Derivatives and Structural Models* by Dr. Jon Gregory
- *Intensity Models* by Dr. Si-Yi Zhou
- *CDO and Correlation Sensitivity* by Dr. Si-Yi Zhou

Running the code

The source code for this model can also be found here: https://github.com/janabdullah96/kth-to-default-basket-cds-pricing

Environment Dependencies

R (4.3.1)

- Please ensure you have R_HOME set as an environment variable. (refer to: https://community.esri.com/t5/spatial-statistics-questions/how-to-call-r-from-python/td-p/565024. Adjust for path and version of R in local system).
- Ensure that the following 2 libraries are installed in the R environment:
  1. ks
  2. randtoolbox

Python (3.7.16)

- The conda environment installation script can be found in this repo at environment/install.bat. This will install the "cqf" conda environment to your local machine.

Run code

The *main.py* script is the entry point into the program.

To run this, open the terminal in the directory that *main.py* is located and activate the conda environment:

```
C:\Windows\System32\cmd.exe
Microsoft Windows [Version 10.0.19044.3086]
(c) Microsoft Corporation. All rights reserved.

U:\fp>activate cqf

(cqf) U:\fp>
```

Make sure that "(cqf)" shows up to the left of the directory in the next line.

Next, run the program by entering the below into the command line:

```
C:\Windows\System32\cmd.exe
Microsoft Windows [Version 10.0.19044.3086]
(c) Microsoft Corporation. All rights reserved.

U:\fp>activate cqf

(cqf) U:\fp>python main.py
```

Note, there are several optional parameters that can be passed in:

Parameter code	Description	Default
-n_sims	Number of MC simulations to run.	10000
-rr	Recovery rate to use.	0.4
-cf	Factor to apply to correlation matrices.	1 (No factor)
-sf	Factor to apply to historical spreads.	1 (No factor)
-crvf	Factor to apply to credit curves.	1 (No factor)

Testing the program setting n_sims = 1000000 will output the below:

```
C:\Windows\System32\cmd.exe
Microsoft Windows [Version 10.0.20348.1787]
(c) Microsoft Corporation. All rights reserved.

U:\fp>activate cqf

(cqf) U:\fp>python main.py -n_sims 1000000
C:\Users\jabdullah\Anaconda3\envs\cqf\lib\site-packages\rpy2\robjects\pandas2ri.py:17: FutureWarning: pandas.core.index is deprecated
and will be removed in a future version.  The public classes are available in the top-level namespace.
  from pandas.core.index import Index as PandasIndex
Running CDS kth-to-default basket pricing.

    Number of Simulations: 1000000
    RR: 0.4
    Correlation Factor: 1
    Spread Factor: 1
    Curve Factor: 1

2023-08-03 20:11:42 - INFO - pymodule.loader - Loading historical CDS spreads.
2023-08-03 20:11:43 - INFO - pymodule.loader - Loading credit curves.
2023-08-03 20:11:43 - INFO - pymodule.loader - Loading discount curve.
2023-08-03 20:11:43 - INFO - pymodule.bootstrap - Constructing survival probabilities term structure table.
2023-08-03 20:11:43 - INFO - pymodule.bootstrap - Constructing hazard rates term structure table.
2023-08-03 20:11:43 - INFO - pymodule.bootstrap - Interpolating discount curve with log-linear interpolation.
2023-08-03 20:11:43 - INFO - pymodule.copula - Setting dataframe of historical CDS spread changes.
2023-08-03 20:11:43 - INFO - pymodule.copula - Converting historical CDS spread changes to pseudo samples with KS Density estimation.
2023-08-03 20:11:45 - INFO - pymodule.copula - Converting pseudo samples of historical CDS spread changes to standard normal.
Loading required package: rngWELL
This is randtoolbox. For an overview, type 'help("randtoolbox")'.
2023-08-03 20:11:47 - INFO - pymodule.simulate - Running MC simulation using Gaussian copula.
2023-08-03 20:11:47 - INFO - pymodule.copula - Running Cholesky decomposition on Gaussian correlation matrix
2023-08-03 20:11:47 - INFO - pymodule.transform - Converting pseudo-samples to default times.
2023-08-03 20:19:29 - INFO - pymodule.price - Constructing Gaussian fair spread table
2023-08-03 20:21:05 - INFO - pymodule.simulate - Running MC simulation using Student's T (Spearman) copula.
2023-08-03 20:21:05 - INFO - pymodule.copula - Running Cholesky decomposition on Student's T Spearman correlation matrix
2023-08-03 20:21:05 - INFO - pymodule.copula - Calibrating degrees of freedom for Student's T copula using MLE.
Calibrated Student's T - Spearman copula to 5 degrees of freedom
2023-08-03 20:21:17 - INFO - pymodule.transform - Converting pseudo-samples to default times.
2023-08-03 20:28:57 - INFO - pymodule.price - Constructing Students T Spearman fair spread table
2023-08-03 20:30:59 - INFO - pymodule.simulate - Running MC simulation using Student's T (Kendall) copula.
2023-08-03 20:30:59 - INFO - pymodule.copula - Running Cholesky decomposition on Student's T Kendall correlation matrix
2023-08-03 20:30:59 - INFO - pymodule.copula - Calibrating degrees of freedom for Student's T copula using MLE.
Calibrated Student's T - Kendall copula to 5 degrees of freedom
2023-08-03 20:31:12 - INFO - pymodule.transform - Converting pseudo-samples to default times.
2023-08-03 20:38:39 - INFO - pymodule.price - Constructing Students T Kendall fair spread table

==================================================
Finalized Spread Pricing Table
                                                        1          2         3         4         5
distribution_type   premium_leg_spread_type
gaussian            spread_without_removal   30.99 bps   9.83 bps  2.98 bps  0.52 bps  0.06 bps
                    spread_with_removal                  9.78 bps  2.97 bps  0.52 bps  0.06 bps
students_t_spearman spread_without_removal   26.53 bps  11.36 bps  4.74 bps  1.24 bps  0.25 bps
                    spread_with_removal                  11.3 bps  4.72 bps  1.23 bps  0.25 bps
students_t_kendall  spread_without_removal   26.06 bps  11.49 bps  4.94 bps  1.34 bps   0.3 bps
                    spread_with_removal                 11.43 bps  4.93 bps  1.33 bps   0.3 bps

(cqf) U:\fp>
```