Base Correlation Model

The model serves the purpose of finding an appropriate correlation to value a collateral debt obligation (CDO) tranche from market information via base correlations. The model includes:

1. Normal Copula Semi-closed Form Solution;
2. New Method of Base Correlation Calibration;
3. Extrapolation of Base Correlation Curve.

In the following of the report, the collateral pool of a CDO trade is defined as a set of N reference names, $\Omega = \{1, 2, \cdots, N\}$, in which each reference name is described by a hazard rate curve $h_i(t)$, recovery rate R_i, and a notional amount $Notl_i$. The collateral pool is tranched into M tranches with the ith tranche defined by an attachment point K_{i-1}, a detachment point K_i, and a fixed time horizon $[0, T_M]$ with fee payments at interval $T_1 < T_2 < \cdots < T_M$.

Within the current GSP credit derivatives framework a reduced form of default probability of a reference name has been implemented and well maintained. Let the default time of ith reference name in the collateral pool be τ_i, a deterministic risk-neutral hazard rate, $h_i(s)$, is defined such that the default probability between s and $s+ds$ $P[s < \tau_i < s + ds \mid \tau_i > s] = h_i(s)ds$. With this definition the default probability functions built upon the hazard rate are

(1) $\qquad S_i(t) = P[t < \tau_i] = e^{-\int_0^t h_i(u)du} =$ Survival probability,

(2) $\qquad F_i(t) = 1 - S_i(t) = 1 - e^{-\int_0^t h_i(u)du} =$ Default probability,

(3) $\qquad f_i(t) = h_i(t)S_i(t) =$ Default probability density function.

- **Normal Copula Semi-Closed Form Model**

Normal copula model has become the market standard to compute the joint default probabilities of a collateral pool, which is essential to value a CDO trade [4,5]. Within the GSP credit derivative modeling framework, a normal copula Monte Carlo (MC) simulation model has already implemented and approved [6].

A copula is a mathematical function that combines marginal probability into a joint distribution. For n uniform random variables, $U_1, U_2 \cdots, U_N$, the joint distribution function is defined as the following.

(4) $C(u_1, u_2, \cdots, u_N) = \Pr[U_1 < u_1, U_2 < u_2, \cdots U_N < u_N]$,

The function C is called the copula function.

The normal copula function is a multivariate cumulative normal distribution with correlation matrix Σ. Applying the normal copula function to the modelling of correlated default events of a collateral asset pool, the uniform random variables are mapped to the default probabilities with standard normal distribution. The default copula, or the cumulative joint default probability for the collateral pool with n assets, can be expressed as:

(5)
$$
\begin{aligned}
F(t_1, t_2, \cdots, t_N) &= \Pr[t_1 > \tau_1, t_2 > \tau_2, \cdots, t_N > \tau_N] \\
&= C(F_1(t_1), F_2(t_2), \cdots, F_N(t_N)) \\
&= \Phi_{N,\Sigma}[\Phi^{-1}(F_1(t_1)), \Phi^{-1}(F_2(t_2)), \cdots \Phi^{-1}(F_N(t_N))],
\end{aligned}
$$

where Φ is the standard cumulative normal distribution function with correlation matrix Σ.

— **One-Factor Normal Copula Model and Analytical Solution**

The normal copula function is actually an n-dimensional integral, which is hard to calculate directly if n is large. We use the one-factor normal copula model to reduce the dimensionality in order to achieve an analytical solution. For the ith obligor in the collateral pool, we assume that its asset process follows

(6) $X_i = \rho_i Z + \sqrt{1 - \rho_i^2} Z_i$,

where Z, Z_i $(i = 1, 2, \cdots N)$ are   independent   standard   normal   random   variables,   that   is $\text{cov}(Z, Z_i) = \text{cov}(Z_i, Z_j) = 0 \ \forall i, j \in \Omega$. The latent variable Z is shared by all reference names in the

collateral pool. From the above equation, the correlation between i^{th} and j^{th} obligor, which could be viewed as the asset correlation $\rho_{i,j}$, can be found to have the following relationship

(7) $$\text{cov}(X_i, X_j) = \rho_{i,j} = \begin{cases} \rho_i \rho_j & i \neq j \\ 1 & i = j \end{cases}.$$

The assumption that $\rho_{i,j} = \rho_i \rho_j$ can be easily achieved for a flat correlation matrix by assuming $\rho_i = \sqrt{\rho}\ (i = 1, 2, \cdots, N)$. With this assumption, the default probability of the ith reference name conditional on the latent variable has the following form

(8) $$F_i^Z(t) = \Pr\left(t > \tau_i \mid Z\right) = \Phi\left[\frac{\Phi_i^{-1}[F_i(t)] - \sqrt{\rho}Z}{\sqrt{1-\rho}}\right].$$

(9) $$S_i^Z(t) = \Pr\left(t > \tau_i \mid Z\right) = \Phi\left[\frac{\sqrt{\rho}Z - \Phi_i^{-1}[F_i(t)]}{\sqrt{1-\rho}}\right]$$

The Normal copula can be proved to have the form

(10) $$C(X_1, X_2, \ldots X_N) = \int \prod_{i=1}^{N} \Phi\left(\frac{X_i - \sqrt{\rho}Z}{\sqrt{1-\rho}}\right) \frac{1}{\sqrt{2\pi}} e^{-\frac{M^2}{2}} dZ.$$

Finally, we can express the one-factor normal default and survival copulas as the following

(11)
$$F(t_1, t_2, \cdots, t_N) = E[\prod_{n=1}^{N} F_n^Z(t_n) \mid Z] = \int \prod_{n=1}^{N} \Phi\left(\frac{\Phi^{-1}(F_n(t_n)) - \sqrt{\rho}M}{\sqrt{1-\rho}}\right) \frac{1}{\sqrt{2\pi}} e^{-\frac{M^2}{2}} dM,$$

$$S(t_1, t_2, \cdots, t_N) = E[\prod_{n=1}^{N} S_n^Z(t_n) \mid Z] = \int \prod_{n=1}^{N} \Phi\left(\frac{\sqrt{\rho}M - \Phi^{-1}(F_n(t_n))}{\sqrt{1-\rho}}\right) \frac{1}{\sqrt{2\pi}} e^{-\frac{M^2}{2}} dM.$$

The n-dimensional copula is computed through a one-dimensional integral.

— **Valuation of a CDO Trade Using One-Factor Semi-Closed Form Solution**

Even if the normal copula can be reduced into a one-dimensional integral shown in Eqs. (10) and (11), it is still quite difficult to be used directly in the valuation of CDO trades. Instead, we proceed with a loss function defined as

$$(12) \qquad L(t) = \sum_{n=1}^{N} LGD_n 1_{t>\tau_n},$$

where $LGD_i = Notl_i \times (1 - R_i)$ is the loss given default (LGD) for the ith reference name.

In order to price a CDO tranche with an attachment point K_{i-1} and a detachment point K_i, the expected loss function of the tranche is introduced, which has the form

$$(13) \qquad E[L^i(t)] = E[(L(t) - K_{i-1})^+ - (L(t) - K_i)^+]$$

The value of the protection leg and the fee leg can then be written as:

Fee Leg:

$$(14) \qquad V_P(K_{i-1}, K_i, \rho) = \int_0^{T_M} D(s)dE[L^i(t)].$$

Annuity of the fee leg:

$$(15) \qquad \begin{aligned} A_f(K_{i-1}, K_i, \rho) &= \sum_{j=1}^{M} a(T_{j-1}, T_j)D(T_j)(K_i - K_{i-1} - E[L^i(T_j)]) \\ &+ \sum_{j=1}^{M} a(T_{j-1}, T_j)\frac{1}{T_j - T_{j-1}}\int_{T_{j-1}}^{T_j}(s - T_{j-1})D(s)dE(L^i(s)) \end{aligned},$$

with the first term being the annuity of the fee payment stream and the second the fee accrual.

The b/e spread for the CDO tranche can be calculated by

$$(16) \qquad \bar{s}_{K_{i-1},K_i} = \frac{V_p(K_{i-1},K_i,\rho)}{A_f(K_{i-1},K_i,\rho)},$$

and, if the traded spread s_{K_{i-1},K_i} is given, the MTM reads

$$
\begin{aligned}
(17) \qquad MTM &= V_p(K_{i-1},K_i,\rho) - s_{K_{i-1},K_i} A_f(K_{i-1},K_i,\rho) \\
&= (\bar{s}_{K_{i-1},K_i} - s_{K_{i-1},K_i}) A_f(K_{i-1},K_i,\rho)
\end{aligned}.
$$

— Iteration Approach To Calculate Expected Loss Distribution Function

In order to calculate the expected loss function denoted in Eq. (13), we need to the find the probability distribution function of the loss function. In general there are two approaches: 1) via the characteristic function and the moment generating function and 2) via iteration.

By conditional independence, we could proceed with the characteristic function as proposed by Laurent and Gregory (2003)[4]:

$$
\begin{aligned}
(18) \qquad \Phi_{L(t)}(u) &= E[e^{iuL(t)}] \\
&= E[\prod_{n=1}^{N} E[e^{iuLGD_n \mathbf{1}_{t>\tau_n}}] \, \big| Z], \\
&= E[\Phi_{L(t)}^{Z}(u) \, \big| Z]
\end{aligned}
$$

within which the characteristic function of the single reference name can be calculated separately with the results shown as

$$(19) \qquad \Phi_{L(t)}^{Z}(u) = \prod_{n=1}^{N} (S_n^{Z}(t) + F_n^{Z}(t)e^{iLGD_n u}).$$

The probability distribution of the loss function can be calculated from the characteristic function.

The GSP toolbox adopts an iteration approach, which has been proved to be a more efficient method. This method was first proposed by L. Anderson et al. [8] and several other versions were later developed by several authors [7~10].

Conditional on the latent variable, we could find that the probability of zero loss is given by

$$(20) \qquad F_N^Z(L(t)=0,t) = \prod_{j=1}^N S_j^Z(t) = S_i^Z(t) \prod_{\substack{j=1 \\ j \neq i}}^N S_j^Z(t) = F_{N-1}^Z(L(t)=0,t)S_i^Z(t),$$

while the probability of the ith reference name default could be written as

$$F_N^Z(L(t)=LGD_i,t) = F_i^Z(t) \prod_{\substack{j=1 \\ j \neq i}}^N S_j^Z(t)$$

$$(21) \qquad = F_N^Z(L(t)=0,t)\frac{F_i^Z(t)}{S_i^Z(t)} = F_{N-1}^Z(L(t)=0,t)F_i^Z(t)$$

From the above two equations, we could find an iteration relationship by assuming that the reference name is added one by one. For example, when the ith reference is added, the following relationship holds

$$(22) \qquad F_N^Z(L(t),t) = F_{N-1}^Z(L(t),t)S_i^Z(t) + F_{N-1}^Z(L(t)-LGD_i,t)F_i^Z(t).$$

Once we got the default probability distribution conditional on the latent variable, the unconditional probability can be found:

$$(23) \qquad F_N(L(t),t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} F_N^Z(L(t),t)e^{-\frac{Z^2}{2}} dZ$$

Then the expected loss function can be written as

$$E[L^i(t)] = E[(L(t)-K_{i-1})^+] - E[(L(t)-K_i)^+]$$

$$(24) \qquad = \int_{K_{i-1}}^{\infty}(x-K_{i-1})f(x,t)dx - \int_{K_i}^{\infty}(x-K_i)f(x,t)dx,$$

with $f(L(t),t)dL(t) = dF_N(L(t),t)$ being the probability density function of expected loss.

With respect to the GSP iteration approach, it is worthy to note that

1. The iteration adopted by the GSP model and denoted in this report is different from those specified by Anderson *et al.* [8] and Hull&White [7]. In the current GSP approach, the limitations on the LGD of each reference name are lifted by building loss function grid points without any approximation. The iteration is then conducted by Eq. (22), which is directly over the reference name instead of the loss function. The advantage of this approach is that no approximation is needed; the disadvantage is that the simulation time will increase quickly as the heterogeneity of the LGD of the reference pool increases.

2. Gauss-Hermite method is used for the integration over the latent variable. As shown in Eqs.(8) and (9), the function is no longer smooth enough if the correlation is very large. Hence 0.9 is set to be the maximum value that the semi-closed form solution can predict a reliable solution.

3. The iteration is carried out quarterly and the integration over time, as denoted by Eqs.(14) and (15), is done by linear interpolation over the quarterly spaced expected loss function.

The pertinent tests have been designed to assess these treatments and approximations in the following section.

- **Base Correlation Curve Calibration**

In the initial version of the GSP toolbox, the base correlation is calculated based on the implied correlation, bearing the assumption that base correlation and implied correlation generates the same expected loss and are internally consistent. However, research shows that they are different and the base correlation should be built in a more self-consistent way [12].

Taking the standard CDX North America index as an example ($K_i = \{0, 3\%, 7\%, 10\%, 15\%, 30\%, 100\%\}$), the base correlation curve is calculated by the following procedures:

1. The 0~3% tranche base correlation is equal to the implied correlation. That is, the value of correlation is find such that the MTM of the 0~3% tranche is zero with

$$(25) \qquad MTM(0,3\%, \rho_{base}^{3\%}) = V_p(0,3\%, \rho_{base}^{3\%}) - s_{0,3\%} \times A_f(0,3\%, \rho_{base}^{3\%}) - U_{0\sim3\%} = 0,$$

where $U_{0\sim3\%}$ is the up-front fee and $s_{0\sim3\%}(=5\%)$ is the quoted spread of the equity tranche. Eq.(25) can be solved using the normal copula model in reverse.

2. For base correlation of 0~7% tranche, we solve for the value of $\rho_{base}^{7\%}$ that solves

(26)
$$
\begin{aligned}
& MTM(3\%,7\%\,\rho_{base}^{7\%}) \\
&= (V_p(0,7\%,\rho_{base}^{7\%}) - V_p(0,3\%,\rho_{base}^{3\%})) \\
&\quad - s_{3\%,7\%} \times (A_f(0,7\%,\rho_{base}^{7\%}) - A_f(0,3\%,\rho_{base}^{3\%})) \\
&= 0
\end{aligned}
$$

where $V_p(0,3\%,\rho_{base}^{3\%})$ and $A_f(0,3\%,\rho_{base}^{3\%})$ are the value of protection and annuity of the fee leg, which are implied from the 0~3% tranche market quote and calculated from the first step.

3. The base correlation of the higher attachment point is calculated by following the same bootstrapping procedure based on the base correlation of the previous base tranche.

The new method, called MTM in the toolbox, is consistent with Refs. [12] and [13].

- **Base Correlation Curve Extrapolation and Interpolation**

The base correlation curve is consisted of the correlation values of several base tranches. When applied to a bespoke tranche, it may encounter the situation that the bespoke tranche has the attachment point and/or detachment point below the first base tranche or higher than the last base tranche. For example, the first base tranche of CDX is 0~3% while the bespoke tranche could be 1.5%~2.5%. In order to find the appropriate correlation in this case, several extrapolation methods on the base correlation curve have been proposed in the submitted template.

In the toolbox, one can pick an extrapolation method by assuming a correlation value for 0~0% and 0~100% base tranches. For example, we could assume 0 point correlation be 1) 0, 2) the same as 0~3% tranche (a flat curve), and 3) an arbitrary value which could be a linear extrapolation of the curve up to 0~7% base tranche.

To the best of our knowledge, there is no generally acceptable methodology of such extrapolation. Moreover, it is even doubtful if we could apply such method, because we actually extending market information to the unknown range.

We argue that, if we need to make a choice, the flat extrapolation is appropriate for two reasons. First, it is more self-consistent. As discussed in Ref. [12], the base correlation is not a proper model of correlation skew. A flat extrapolation assumes no correlation skew within the extrapolated region (0~3% and 30~100% for CDX). As shown in Eqs.(14) and (15), assuming a flat base correlation curve fulfils the arbitrage-free requirements claimed in Ref. [12], hence it is more self-consistent.

Second, a flat extrapolation is robust when applied to a bespoke tranche. If the attachment point and detachment points of a tranche are within the extrapolation range, a positive value of protection is assured by the flat extrapolation while other methods have the possibility of failure.

- **Benchmarking/Comparison to Other Models**

Two models are employed as the benchmark/comparison model, which are:

1. **Normal Copula MC simulation**

The normal copula model was first proposed as an MC simulation approach [5]. The GSP normal copula MC simulation model could serve as the benchmark model to test the closed form solution.

In order to generate correlated defaults times using normal copula, a series of random variables $X_1, X_2, \cdots X_n$ are first generated from an n-dimensional normal distribution with correlation ρ in each MC scenario. The default times for each obligor in the collateral pool $\tau_1, \tau_2, \cdots \tau_n$ are obtained using

(25) $$\tau_i = F^{-1}(\Phi(X_i)), \qquad i = 1, 2, \cdots N.$$

2. **FTD closed form solution**

If we assume a homogeneous collateral pool ($LGD_i = LGD_j = Const \ \forall i, j \in \Omega$) and set the notional of the first tranche equal to LGD, the first tranche becomes a first-to-default (FTD) trade.

Within the one factor normal copula framework, this special FTD actually has a nice closed form solution. Instead of directly solving first default probability, we are interested in the probability that all names survive up to a particular point in time.

The joint distribution for survival time can be defined as follows given the definition of survival copula:

$$(26) \quad \tilde{S}(t) = S(t, t, \ldots, t) = E\left[\prod_{i=1}^{N} S_i(t) \mid Z\right] = \int \prod_{i=1}^{N} \Phi\left[\frac{\sqrt{\rho}M - \Phi_i^{-1}[F_i(t)]}{\sqrt{1-\rho}}\right] \frac{1}{\sqrt{2\pi}} e^{-\frac{M^2}{2}} dM .$$

We can then write the hazard rate and default probability density function for the collateral pool as the following:

$$(27) \quad \tilde{h}(t) = -\frac{1}{\tilde{S}(t)} \frac{d\tilde{S}(t)}{dt}$$

$$(28) \quad \tilde{f}(t) = \tilde{h}(t)\tilde{S}(t)$$

Then we treat this FTD trade just like a single name credit default swap (see https://finpricing.com/product.html). Following this approach, the protection leg and fee leg have the forms.

$$(29) \quad V_{prot}^{FTD} = LGD \int_0^{T_M} D(t)\tilde{f}(t)dt .$$

$$(30) \quad A_{fee}^{FTD} = Notl\sum_{j=1}^{T_M} a(T_{j-1}, T_j)\left(D(T_j)\tilde{S}(T_j) + \frac{1}{T_j - T_{j-1}} \int_{T_{j-1}}^{T_j} (t - T_{j-1})D(t)\tilde{f}(t)dt\right).$$

This closed form solution does not involve iteration hence can be used as a alternative model to check the iteration method.