Analytic Dynamic Factor Copula Model*

Ken Jackson† Alex Kreinin‡ Wanhe Zhang§

November 13, 2011

Abstract

The Gaussian factor copula model is the market standard model for multi-name credit derivatives. Its main drawback is that factor copula models exhibit correlation smiles when calibrating against market tranche quotes. To overcome the calibration deficiency, we introduce a multi-period factor copula model by chaining one-period factor copula models. The correlation coefficients in our model are allowed to be time-dependent, and hence they are allowed to follow certain stochastic processes. Therefore, we can calibrate against market quotes more consistently. Usually, multi-period factor copula models require multi-dimensional integration, typically computed by Monte Carlo simulation, which makes calibration extremely time consuming. In our model, the portfolio loss of a completely homogeneous pool possesses the Markov property, thus we can compute the portfolio loss distribution analytically without multi-dimensional integration. Numerical results demonstrate the efficiency and flexibility of our model to match market quotes.

*This research was supported in part by the Natural Sciences and Engineering Research Council (NSERC) of Canada.

†Department of Computer Science, University of Toronto, 10 King's College Road, Toronto, ON, M5S 3G4, Canada; krj@cs.toronto.edu

‡Algorithmics Inc., 185 Spadina Avenue, Toronto, ON, M5T 2C6, Canada; alex.kreinin@algorithmics.com

§Royal Bank of Canada, 200 Bay Street, Toronto, ON, M5J 2J5, Canada; zhangw@cs.toronto.edu

1 Introduction

Due to their computational efficiency, factor copula models are popular for pricing multi-name credit derivatives. Within this class of models, the Gaussian factor copula model is the market standard model. However, it cannot match market quotes consistently without violating the model assumptions as explained in Hull and White (2006) and Torresetti et al. (2006). For example, it has to use different correlation factor loadings for different tranches based on the same underlying portfolio. To better match the observable spreads, several modifications have been proposed based on the conditional independence framework. See, for example, Andersen and Sidenius (2004), Baxter (2007) and Hull and White (2008). Most of these approaches are static one-period models that generate a portfolio loss distribution at a fixed maturity. They may not be flexible enough to match market quotes or applicable for new products with strong time-dependent features, such as forward-starting tranches, tranche options and leveraged super-senior tranches as pointed out in Andersen (2006). Another popular approach to calibrate factor copula models is base correlation, such as McGinty et al. (2004), which calibrates the correlation for the first loss tranche, i.e., the sum of all tranches up to a detachment point. Although it guarantees the existence of the correlation parameter, it is not arbitrage free. For example, it is easy to construct a tranche with a negative spread using this method as noticed in Torresetti et al. (2006).

Another methodology for multi-name credit derivatives is the top-down approach, which models the portfolio loss directly. For example, Bennani (2005), Schönbucher (2005) and Sidenius et al. (2008) proposed similar frameworks to model the dynamics of the aggregate portfolio losses by modeling the forward loss rates. With these pool loss dynamics, the pricing of credit derivatives becomes straightforward. This approach has been further extended by many researchers, such as Giesecke et al. (2011). However, these models require a large amount of data to calibrate and are currently speculative as explained in Andersen (2006).

It is tempting to see whether we can introduce dynamics into the factor copula model to combine its computational efficiency with the ability to calibrate more consistently against market quotes. The main challenge in developing such a dynamic factor copula model is that the arbitrage free property and computational efficiency become more difficult to achieve, as the number of state variables grows rapidly with the introduction of dynamics. Fingers (2000) proposed several extensions of one-period default models. However, they are not based on the factor copula approach. Therefore, only Monte Carlo simulation is available to implement these extensions. In addition, the correlation coefficients are not allowed to be time-dependent. Andersen (2006) and Sidenius (2007) introduced several "chaining" techniques to build multi-period factor copula models from one-period factor copula models. As these models must integrate over all the common factors, they require a multi-dimensional integration, which is usually computed by Monte Carlo simulation. This makes the model calibration extremely time consuming. Except for some special cases, for example, where the factors are the same for all periods, existing "chaining" methods cannot avoid multi-dimensional integration. Therefore, current multi-period models are hard to generalize to more than two periods.

In this paper, we develop a novel chaining method to build a multi-period factor copula model, which does not allow arbitrage opportunities and avoids multi-dimensional integration. Based on our model, the portfolio loss of a completely homogeneous pool possesses the Markov property, so we can compute its distribution across time by a recursive method instead of by Monte Carlo simulation. Numerical results demonstrate the accuracy, efficiency and flexibility of our model in calibrating against market quotes.

The rest of the paper is organized as follows. Section 2 describes the pricing equations for synthetic CDOs. Section 3 reviews the widely used Gaussian factor copula model as an example of the conditional independence framework. Section 4 reviews existing "chaining" methods before introducing our new multi-period model. Section 5 discusses calibration. Section 6 presents the numerical results. Section 7 concludes the paper and discusses future work.

2 Pricing equations

In a synthetic CDO, the protection seller absorbs the pool loss specified by the tranche structure. That is, if the pool loss over $(0, T]$ is less than the tranche attachment point a, the seller does not suffer any loss; otherwise, the seller absorbs the loss up to the tranche size $S = b - a$. In return for the protection, the buyer pays periodic premia at specified times $t_1 < t_2 < \ldots < t_n = T$.

We consider a synthetic CDO containing K names with loss-given-default N_k for name k in the original pool. Assume that the recovery rates are constant. Let D_i denote the risk-free discount factors at time t_i, and d_i denote the expected value of D_i in a risk-neutral measure. Denote the pool loss up to time t_i by L_i. Then, the loss absorbed by the specified tranche is

$$\mathcal{L}_i = \min(S, (L_i - a)^+), \quad \text{where } x^+ = \max(x, 0) \tag{1}$$

We make the standard assumption that the discount factors D_i's and the pool losses L_i's are independent, whence D_i's and $\mathcal{L}_i$'s are also independent.

In general, valuation of a synthetic CDO tranche balances the expectation of the present values of the premium payments (premium leg) against the effective tranche losses (default leg), such that

$$\mathbb{E}\Big[\sum_{i=1}^{n} s(S - \mathcal{L}_i)(T_i - T_{i-1})D_i\Big] = \mathbb{E}\Big[\sum_{i=1}^{n}(\mathcal{L}_i - \mathcal{L}_{i-1})D_i\Big] \tag{2}$$

The fair spread s is therefore given by

$$s = \frac{\mathbb{E}\big[\sum_{i=1}^{n}(\mathcal{L}_i - \mathcal{L}_{i-1})D_i\big]}{\mathbb{E}\big[\sum_{i=1}^{n}(S - \mathcal{L}_i)(T_i - T_{i-1})D_i\big]} = \frac{\sum_{i=1}^{n}(\mathbb{E}\mathcal{L}_i - \mathbb{E}\mathcal{L}_{i-1})d_i}{\sum_{i=1}^{n}(S - \mathbb{E}\mathcal{L}_i)(T_i - T_{i-1})d_i} \tag{3}$$

In the last equality of (3), we use the fact that D_i and $\mathcal{L}_i$ ($\mathcal{L}_{i-1}$) are independent. Alternatively, if the spread is set, the value of the synthetic CDO is the difference between the two legs:

$$\sum_{i=1}^{n} s(S-\mathbb{E}\mathcal{L}_i)(T_i-T_{i-1})d_i - \sum_{i=1}^{n}(\mathbb{E}\mathcal{L}_i-\mathbb{E}\mathcal{L}_{i-1})d_i$$

Therefore, the problem is reduced to the computation of the mean tranche losses, $\mathbb{E}\mathcal{L}_i$. To compute this expectation, we have to compute the portfolio loss L_i's distribution. Therefore, we need to specify the correlation structure of the portfolio defaults.

3 One factor copula model

Due to their tractability, factor copula models are widely used to specify a joint distribution for default times consistent with their marginal distribution. A one factor model was first introduced by Vasicek (1987) to evaluate the loan loss distribution, and the Gaussian copula was first applied to multi-name credit derivatives by Li (2000). After that, the model was generalized by Andersen et al. (2003), Hull and White (2004) and Laurent and Gregory (2005), to name just a few. In this section, we review the one factor Gaussian copula model to illustrate the conditional independence framework.

Let τ_k be the default time of name k, where $\tau_k = \infty$ if name k never defaults. Assume the risk-neutral default probabilities

$$\pi_k(t) = \mathbb{P}(\tau_k \le t), \quad k = 1, 2, \ldots, K$$

are known. In order to generate the dependence structure of default times, we introduce random variables U_k, such that

$$U_k = \beta_k X + \sqrt{1-\beta_k^2}\varepsilon_k, \quad \text{for } k = 1, 2, \ldots, K \tag{4}$$

where X is the systematic risk factor; ε_k are idiosyncratic risk factors, which are independent of each other and also independent of X; and the constants $\beta_k \in [-1, 1]$.

The default times τ_k and the random variables U_k are connected by a percentile-to-percentile transformation, such that $\mathbb{P}(\tau_k \leq t) = \mathbb{P}(U_k \leq b_k(t))$, where each $b_k(t)$ can be viewed as a default barrier.

Models satisfying the assumptions above are said to be based on the conditional independence framework. If, in addition, we assume X and ε_k follow standard normal distributions, then we get a Gaussian factor copula model. In this case, each U_k also follows a standard normal distribution. Hence we have

$$b_k(t) = \Phi^{-1}(\pi_k(t)) \tag{5}$$

where Φ is the standard normal cumulative distribution function. Conditional on a particular value x of X, the risk-neutral default probabilities are defined as

$$\pi_k(t,x) \equiv \mathbb{P}(\tau_k \leq t \mid X = x) = \mathbb{P}(U_k \leq b_k(t) \mid X = x) = \Phi\left[\frac{\Phi^{-1}(\pi_k(t)) - \beta_k x}{\sqrt{1-\beta_k^2}}\right] \tag{6}$$

In this framework, the default events of the names are assumed to be conditionally independent. Thus, the problem of correlated names is reduced to the problem of independent names. The pool losses L_i satisfy

$$\mathbb{P}(L_i = l) = \int_{-\infty}^{\infty} \mathbb{P}_x(L_i = l) d\Phi(x) \tag{7}$$

where $L_i = \sum_{k=1}^{K} N_k \mathbf{1}_{\{U_k \leq b_k(t_i)\}}$, and $\mathbf{1}_{\{U_k \leq b_k(t_i)\}}$ are mutually independent, conditional on $X = x$.[1] Therefore, if we know the conditional distributions of $\mathbf{1}_{\{U_k \leq b_k(t_i)\}}$, the conditional distributions of L_i can be computed easily, as can $\mathbb{E}[\mathcal{L}_i]$. To approximate the integral (7), we use a quadrature rule. Thus, the integral (7) reduces to

$$\mathbb{P}(L_i = l) \approx \sum_{m=1}^{M} w_m \mathbb{P}_{x_m}(L_i = l)$$

where the w_m and x_m are the quadrature weights and nodes, respectively.

[1] As we assume constant recovery rates, the pool losses L_i are discrete random variables here. This approach can be extended to the continuous case with stochastic recovery rates.

A significant drawback of this model is that it does not allow the β_k's to be time dependent, which is often required to calibrate the model effectively. If β_k is a function of time, $\pi_k(t,x)$ may be a decreasing function of time, which may lead to an arbitrage opportunity, as explained in the next section. More specifically, for $0 < t_1 < t_2$, to guarantee $\pi_k(t_1,x) \le \pi_k(t_2,x)$, or equivalently,

$$\Phi\left(\frac{b_k(t_1) - \beta_k(t_1)x}{\sqrt{1-\beta_k(t_1)^2}}\right) \le \Phi\left(\frac{b_k(t_2) - \beta_k(t_2)x}{\sqrt{1-\beta_k(t_2)^2}}\right)$$

we need

$$\frac{b_k(t_1) - \beta_k(t_1)x}{\sqrt{1-\beta_k(t_1)^2}} \le \frac{b_k(t_2) - \beta_k(t_2)x}{\sqrt{1-\beta_k(t_2)^2}}$$

As x may be any real value, for any fixed $\beta_k(t_1) \neq \beta_k(t_2)$, it is easy to find an x to violate this inequality. For example, if $b_k(t_1) = -2$, $b_k(t_2) = -1.4$, $\beta_k(t_1) = 0.6$ and $\beta_k(t_2) = 0.8$, then

$$\pi_k(t_1, 2) = \mathbb{P}(\tau_k \le t_1 \mid X = 2) = \Phi(-4)$$
$$\pi_k(t_2, 2) = \mathbb{P}(\tau_k \le t_2 \mid X = 2) = \Phi(-5)$$

4 Multi-period factor copula models

To overcome this drawback, Andersen (2006) and Sidenius (2007) pioneered the technique of "chaining" a series of one-period factor copula models to produce a multi-period factor copula model. However, their approaches must integrate over the multi-dimensional common factors to evaluate the portfolio loss distribution over time, requiring the evaluation of a high-dimensional integral, usually computed by Monte Carlo simulation. Therefore, their models are hard to generalize to more than two periods, except for some special, but possibly unrealistic, cases, such as, the common factors are the same for all periods. In this section, we first review the approaches of Andersen (2006) and Sidenius (2007). Then, we present our new model, which avoids multi-dimensional integration.

In general, the conditional independence framework, including one-period and multi-period factor copula models, has to satisfy two properties: consistency and no arbitrage. By consistency, we mean that the model has to match the marginal default probabilities of the underlyings, i.e.,

$$\mathbb{P}(\tau_k \leq t) = \int_{\mathcal{D}} \mathbb{P}(\tau_k \leq t \mid X^{(t)} = x) dF(x) \tag{8}$$

Here, $X^{(t)}$ represents the common factors up to time t (it may be a multiple dimensional random variable in the discrete case or a stochastic process in the continuous case); $\mathcal{D}$ is the domain of $X^{(t)}$; and $F(\cdot)$ is the cumulative distribution function of $X^{(t)}$. By no arbitrage, we mean that the pool loss distribution is a non-decreasing function of time, i.e.,

$$\mathbb{P}(L_i = l) \leq \mathbb{P}(L_j = l), \text{ for } t_i \leq t_j \tag{9}$$

To satisfy this constraint in practice, we usually require a stronger condition: the conditional default probability of a single name is non-decreasing over time, i.e.,

$$\mathbb{P}(\tau_k \leq t_1 \mid X^{(t_1)} = x) \leq \mathbb{P}(\tau_k \leq t_2 \mid X^{(t_2)} = y), \text{ for } t_1 \leq t_2 \text{ and } x(t) = y(t), \text{ for } t \leq t_1 \tag{10}$$

where $x(t)$ means the value of x at time t. Obviously, if we satisfy condition (10), then the pool loss (9) is non-decreasing, which implies no arbitrage. Generally, the consistency property is easy to satisfy, but the no arbitrage property is not, as shown in the previous section.

In the rest of the paper, we extend the factor copula model to a discrete-time dynamic model. For each period $(t_{i-1}, t_i]$ and each name k, we associate a latent random variable

$$Y_{k,i} = \beta_{k,i} X_i + \sqrt{1 - \beta_{k,i}^2}\epsilon_{k,i} \tag{11}$$

where X_i is a random variable associated with the common factors for period $(t_{i-1}, t_i]$ and $\epsilon_{k,i}$ are mutually independent random variables associated with idiosyncratic factors for name

k and period $(t_{i-1}, t_i]$. To guarantee the no arbitrage property, Andersen (2006) employed a discrete version of the first hitting time model to construct the conditional default probabilities. More specifically, he connected the default time τ_k and the latent random variables by

$$\begin{aligned} \mathbb{P}(\tau_k < t) &= \mathbb{P}(Y_{k,1} \le b_k(t_1)), & t \le t_1 \\ \mathbb{P}(t_{i-1} < \tau_k \le t) &= \mathbb{P}(Y_{k,1} > b_k(t_1), \ldots, Y_{k,i-1} > b_k(t_{i-1}), Y_{k,i} \le b_k(t_i)), & t \in (t_{i-1}, t_i] \end{aligned}$$

Then the conditional default probability for $t \le t_1$ is the same as that in the one-factor copula model. For $t \in (t_{i-1}, t_i]$, the conditional default probability satisfies

$$\mathbb{P}(t_{i-1} < \tau_k \le t \mid X^{(i)} = x^{(i)}) = \mathbb{P}(Y_{k,1} > b_k(t_1), \ldots, Y_{k,i-1} > b_k(t_{i-1}), Y_{k,i} \le b_k(t_i) \mid X^{(i)} = x^{(i)})$$

Here, $X^{(i)}$ is associated with the common factors for the periods up to t_i, or equivalently, $X^{(i)} = \{X_1, X_2, \ldots, X_i\}$.

Similar to the one-factor copula model, we must compute the boundary $b_k(t_i)$ satisfying the consistency property (8). For $t \le t_1$, the computation is the same as that for the one factor copula model. However, for $t \in (t_{i-1}, t_i]$, it appears that we must integrate the common factors up to t_i. The complexity of this multi-dimensional integration depends on the assumptions associated with the X_i's. Andersen (2006) showed two special cases: (1) X_i are the same and (2) a two-period model, where X are two dimensional random variables. Besides the computation of the default boundary, the multi-dimensional integration also arises when computing the unconditional portfolio loss distribution from the conditional loss distributions.

Sidenius (2007) attacked the no arbitrage problem by introducing conditional forward survival probabilities

$$\mathbb{P}(\tau_k > t \mid \tau_k > t_{i-1}, X^{(i)} = x^{(i)}) = \frac{\mathbb{P}(\tau_k > t \mid X^{(i)} = x^{(i)})}{\mathbb{P}(\tau_k > t_{i-1} \mid X^{(i)} = x^{(i)})}, \quad t \in (t_{i-1}, t_i]$$

Using this, he expressed the conditional survival probability for $t \in (t_{i-1}, t_i]$ as

$$\mathbb{P}(\tau_k > t \mid X^{(i)} = x^{(i)}) = \mathbb{P}(\tau_k > t \mid \tau_k > t_{i-1}, X^{(i)} = x^{(i)}) P(\tau_k > t_{i-1} \mid X^{(i-1)} = x^{(i-1)})$$

For $t \leq t_1$, the conditional survival probability is the same as that in the one factor copula model.

The model allows a conditional forward survival probability for each time period $(t_{i-1}, t_i]$ to be associated with each correlation factor, i.e., $\mathbb{P}(\tau_k > t \mid \tau_k > t_{i-1}, X^{(i)} = x^{(i)}) = \mathbb{P}(\tau_k > t \mid \tau_k > t_{i-1}, X_i = x_i)$. For example, if the X_i's associated with the latent random variables $Y_{k,i}$ in (11) are independent, then the conditional forward survival probability can be computed by

$$\mathbb{P}(\tau_k > t \mid \tau_k > t_{i-1}, X^{(i)} = x^{(i)}) = \frac{\mathbb{P}\Big(\beta_{k,i} X_i + \sqrt{1 - \beta_{k,i}^2} \epsilon_{k,i} > b_k(t_i) \mid X_i = x_i\Big)}{\mathbb{P}\Big(\beta_{k,i} X_i + \sqrt{1 - \beta_{k,i}^2} \epsilon_{k,i} > b_k(t_{i-1}) \mid X_i = x_i\Big)}$$

Using the consistency property (8), we can calibrate the $b_k(t_i)$ recursively. However, it is impossible to preserve any tractability for general cases. Similarly, the multi-dimensional integration problem cannot be avoided, except in some special cases, such as all X_i are the same.

To overcome the high-dimensional integration problem, we use a similar approach based on the same latent random variables (11), but we connect $Y_{k,i}$ and τ_k by the forward default probability

$$\mathbb{P}(Y_{k,i} \leq b_k(t_i)) = \mathbb{P}(\tau_k \in (t_{i-1}, t_i] \mid \tau_k > t_{i-1}) = \frac{\mathbb{P}(\tau_k \leq t_i) - \mathbb{P}(\tau_k \leq t_{i-1})}{1 - \mathbb{P}(\tau_k \leq t_{i-1})}$$

If X_i and $\epsilon_{k,i}$ follow standard normal distributions, then each $Y_{k,i}$ also follows a standard normal distribution. Therefore, we can compute the conditional default boundary $b_k(t_i)$ by

$$b_k(t_i) = \Phi^{-1}\big(\mathbb{P}(\tau_k \in (t_{i-1}, t_i] \mid \tau_k > t_{i-1})\big)$$

We can also compute each conditional forward default probability by

$$\mathbb{P}\big(\tau_k \in (t_{i-1}, t_i] \mid \tau_k > t_{i-1}, X_i = x_i\big) = \Phi\left(\frac{b_k(t_i) - \beta_{k,i}x_i}{\sqrt{1-\beta_{k,i}^2}}\right)$$

The idea of using forward default probabilities has been used for CDO analysis by Morokoff (2003). However, in Morokoff's model the correlation coefficients are constant and no analytical methods are available.

To compute the conditional pool loss distribution, we need to construct $\mathbb{P}(\tau_k \le t_i \mid X_1 = x_1, \ldots, X_i = x_i)$ from $\mathbb{P}\big(\tau_k \in (t_{i-1}, t_i] \mid \tau_k > t_{i-1}, X_i = x_i\big)$. Based on the definitions of these terms, we have

$$\begin{aligned}
&\mathbb{P}(\tau_k \le t_i \mid X_1 = x_1, \ldots, X_i = x_i) \\
&\quad = \mathbb{P}(\tau_k \le t_{i-1} \mid X_1 = x_1, \ldots, X_{i-1} = x_{i-1}) + \mathbb{P}(\tau_k \in (t_{i-1}, t_i] \mid X_1 = x_1, \ldots, X_i = x_i) \\
&\quad = \mathbb{P}(\tau_k \le t_{i-1} \mid X_1 = x_1, \ldots, X_{i-1} = x_{i-1}) \\
&\qquad\quad + \mathbb{P}(\tau_k > t_{i-1} \mid X_1 = x_1, \ldots, X_{i-1} = x_{i-1}) \cdot \mathbb{P}\big(\tau_k \in (t_{i-1}, t_i] \mid \tau_k > t_{i-1}, X_i = x_i\big)
\end{aligned}$$

For the rest of the paper, we denote $P(\tau_k \le t_{i-1} \mid X_1 = x_1, \ldots, X_{i-1} = x_{i-1})$ by $q_{k,i-1}$ and $\mathbb{P}\big(\tau_k \in (t_{i-1}, t_i] \mid \tau_k > t_{i-1}, X_i = x_i\big)$ by $p_{k,i}$ for simplicity. If $q_{k,i}$ and $p_{k,i}$ are the same for all $k = 1, \ldots, K$, we denote them by q_i and p_i, respectively.

Using the conditional default probabilities $q_{k,i}$, we can compute efficiently the conditional distribution of the pool loss for a completely homogeneous pool, where $\beta_{k,i}$, $\pi_k(t)$ and N_k are the same for $k = 1, \ldots, K$. In this special, but important, case, the distribution of L_i can be computed by the distribution of number of defaults l_i, as $L_i = N_1 \sum_{k=1}^{K} \mathbf{1}_{\{\tau_k \le t_i\}} = N_1 l_i$.

Therefore, the conditional pool loss distribution of a completely homogeneous pool satisfies

$$
\begin{aligned}
\mathbb{P}\big(L_i = rN_1 \mid X_1 = x_1, \ldots, X_i = x_i\big) &= \mathbb{P}\big(l_i = r \mid X_1 = x_1, \ldots, X_i = x_i\big) \\
&= \binom{K}{r}\big(q_{i-1} + (1 - q_{i-1})p_i\big)^r \big((1 - q_{i-1})(1 - p_i)\big)^{K-r} \\
&= \binom{K}{r}\left(\sum_{m=0}^{r}\binom{r}{m} q_{i-1}^m (1 - q_{i-1})^{r-m} p_i^{r-m}\right)(1 - q_{i-1})^{K-r}(1 - p_i)^{K-r} \\
&= \sum_{m=0}^{r}\binom{K}{m} q_{i-1}^m (1 - q_{i-1})^{K-m} \cdot \binom{K-m}{r-m} p_i^{r-m}(1 - p_i)^{K-m-(r-m)} \\
&= \sum_{m=0}^{r}\mathbb{P}\big(l_{i-1} = m \mid X_1 = x_1, \ldots, X_{i-1} = X_{i-1}\big)\mathbb{P}\big(\hat{l}_{(i-1,i]}^{K-m} = r - m \mid X_i = x_i\big) \qquad (12)
\end{aligned}
$$

where $\hat{l}_{(i-1,i]}^{K-m}$ is the number of defaults during $(t_{i-1}, t_i]$ with the pool size $K - m$, and its distribution is computed using the conditional forward default probability p_i.

To compute the tranche loss, we need to compute the unconditional pool loss distribution from the conditional ones, i.e., we need to integrate over the common factors X_i. Generally, this requires a multi-dimensional integration, for which Monte Carlo simulation is usually used. However, we can avoid the multi-dimensional integration in this special case by exploiting the independence of the X_i's:

$$
\begin{aligned}
\mathbb{P}\big(l_i = r\big) &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \sum_{m=0}^{r} \mathbb{P}\big(l_{i-1} = m \mid X_1 = x_1, \ldots, X_{i-1} = X_{i-1}\big) \\
&\qquad \cdot \mathbb{P}\big(\hat{l}_{(i-1,i]}^{K-m} = r - m \mid X_i = x_i\big) d\Phi(X_1) \ldots d\Phi(X_i) \\
&= \sum_{m=0}^{r} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \mathbb{P}\big(l_{i-1} = m \mid X_1 = x_1, \ldots, X_{i-1} = X_{i-1}\big) d\Phi(X_1) \ldots d\Phi(X_{i-1}) \\
&\qquad \cdot \int_{-\infty}^{\infty} \mathbb{P}\big(\hat{l}_{(i-1,i]}^{K-m} = r - m \mid X_i = x_i\big) d\Phi(X_i) \\
&= \sum_{m=0}^{r} \mathbb{P}\big(l_{i-1} = m\big)\mathbb{P}\big(\hat{l}_{(i-1,i]}^{K-m} = r - m\big) \qquad (13)
\end{aligned}
$$

Therefore, the unconditional pool loss distribution possesses the Markovian property and can be computed recursively. Iscoe (2003) derived a similar Markovian property for the two name case of Morokoff's model in Morokoff (2003), where the correlation coefficients are constant. Our derivation is more general, and our model fixes the correlation decay of Morokoff's model as described in Iscoe (2003).

Remark: We need to assume that the X_i's are independent to derive the key formula (13), which enables us to avoid the costly multi-dimensional integration in computing the unconditional pool loss distribution from the conditional one. However, it is worth noting that this is the only place in the paper where we need to assume that X_i and X_j are independent for all $i \neq j$. Therefore, we could use more general processes for the X_i's if we do not need to compute the unconditional pool loss distribution from the conditional one or if we could replace (13) by another efficient formula to compute the unconditional pool loss distribution from the conditional one.

The difference between our approach and Andersen's approach in Andersen (2006) can be understood intuitively as follows. In Andersen's approach, the latent random variables Y_k, which reflect the healthiness of name k, are reset back to zero at the beginning of each period. Therefore, the process forgets its previous position. The latent process of our model is also reset to zero at the beginning of each period. However, in our model it describes the healthiness of the forward default probability. The process for the default probability actually remembers its position at the end of the previous period: how the process evolves for the new period depends on the latent process of the forward default probability. In addition, as noted above, it appears that Andersen's approach requires a costly multi-dimensional integration to compute the unconditional pool loss distribution from the conditional one, except in some simple special cases. For a completely homogeneous pool, assuming that the X_i and X_j are independent for all $i \neq j$, our approach uses the much less costly recurrence (13) to compute the unconditional pool loss distribution from the conditional one.

For a more general pool[2], it still holds that the event that r defaults occur before t_i is equivalent to the event that m defaults occur before t_{i-1} and $r - m$ defaults occur during

[2]There are two other types of underlying pools: (1) homogeneous pools, where all N_k are the same, for all $k = 1, \dots, K$, and either $\beta_{k,i}$ or $\pi_k(t)$ are different for some k; (2) inhomogeneous pools, where N_k, $\beta_{k,i}$ and $\pi_k(t)$ are different for some k.

$(t_{i-1}, t_i]$, for $m = 0, \dots r$. That is,

$$\mathbb{P}(l_i = r) = \sum_{m=0}^{r} \mathbb{P}(l_{i-1} = m, l_{(i-1,i]} = r - m) = \sum_{m=0}^{r} P(l_{i-1} = m) \cdot \mathbb{P}(l_{(i-1,i]} = r - m \mid l_{i-1} = m)$$

Moreover, this relationship extends to the conditional probabilities:

$$\mathbb{P}(l_i = r \mid X_1 = x_1, \dots, X_i = x_i) = \sum_{m=0}^{r} \mathbb{P}(l_{i-1} = m \mid X_1 = x_1, \dots, X_{i-1} = x_{i-1}) \cdot \mathbb{P}(l_{(i-1,i]} = r - m \mid l_{i-1} = m, X_1 = x_1, \dots, X_i = x_i)$$

Under the assumptions of our model, we can simplify the expression above using

$$\mathbb{P}(l_{(i-1,i]} = r - m \mid l_{i-1} = m, X_1 = x_1, \dots, X_i = x_i) = \mathbb{P}(l_{(i-1,i]} = r - m \mid l_{i-1} = m, X_i = x_i)$$

Therefore,

$$\mathbb{P}(l_i = r \mid X_1 = x_1, \dots, X_i = x_i) = \sum_{m=0}^{r} \mathbb{P}(l_{i-1} = m \mid X_1 = x_1, \dots, X_{i-1} = x_{i-1}) \cdot \mathbb{P}(l_{(i-1,i]} = r - m \mid l_{i-1} = m, X_i = x_i)$$

To obtain the unconditional pool loss distribution, we need to integrate over the common factors, as we did in (13). Therefore, in our model, the Markov property holds for a general pool:

$$\mathbb{P}\big(l_i = r\big) = \sum_{m=0}^{r} \mathbb{P}\big(l_{i-1} = m\big) \cdot \mathbb{P}\big(l_{(i-1,i]}^{K-m} = r - m \mid l_{i-1} = m\big)$$

However, as the default probability of each name may be different in a general pool, we end up with another combinatorial problem: we need to consider all possible combinations of $l_{i-1} = m$.

Obviously, the completely homogeneous pool is a special case. However, it is of considerable practical importance, since such pools often arise in practice. Moreover, the pool loss of a general pool is generally approximated by the pool loss of a completely homogeneous one for computational efficiency in calibration and the valuation of bespoke contracts.

Remark: For simplicity, we used the Gaussian factor copula model to illustrate our new discrete dynamical multi-period factor copula model. However, it is important to note that our approach can be applied to construct a multi-period factor copula model from any one factor copula model based on the conditional independence framework.

5 Calibration

Our goal is to calibrate our model against the market tranche quotes on the same underlying pool. To illustrate our approach, we use the tranche quotes of the credit indexes, CDX and ITRAXX. As our model allows the correlation factor loadings to be time-dependent, we can introduce dynamics into the model by letting the correlation factor loadings follow particular dynamic processes. This added flexibility gives our dynamic model enough degrees of freedom to calibrate against market quotes.

We obtain the spread quotes for the indexes and tranches on CDX and ITRAXX from the Thomson Datastream. We approximate the default probabilities of a single name using the index spreads, which are the average spreads of the 125 names in CDX or ITRAXX. Due to the data availability and popularity, we calibrate our model against the four mezzanine tranches with maturities 5 years, 7 years and 10 years. Therefore, we have to fit 12 market tranche quotes on the same underlying pool.

To fit these 12 tranche quotes, we must incorporate sufficient degrees of freedom into our model. As the correlation factor loadings are time-dependent in our model, they can be any dynamic process within the range $[0, 1]$. Therefore, we can obtain sufficient degrees of freedom by constructing a suitable dynamic process for the correlation factor loadings. To illustrate our approach, we employ a binomial tree structure for the correlation factor loadings in our numerical examples. We assume that the correlation factor loading process is a piecewise constant function over time and each branch of the tree describes one possible path of the factor loading process. To compute the tranche prices, we only need to take the

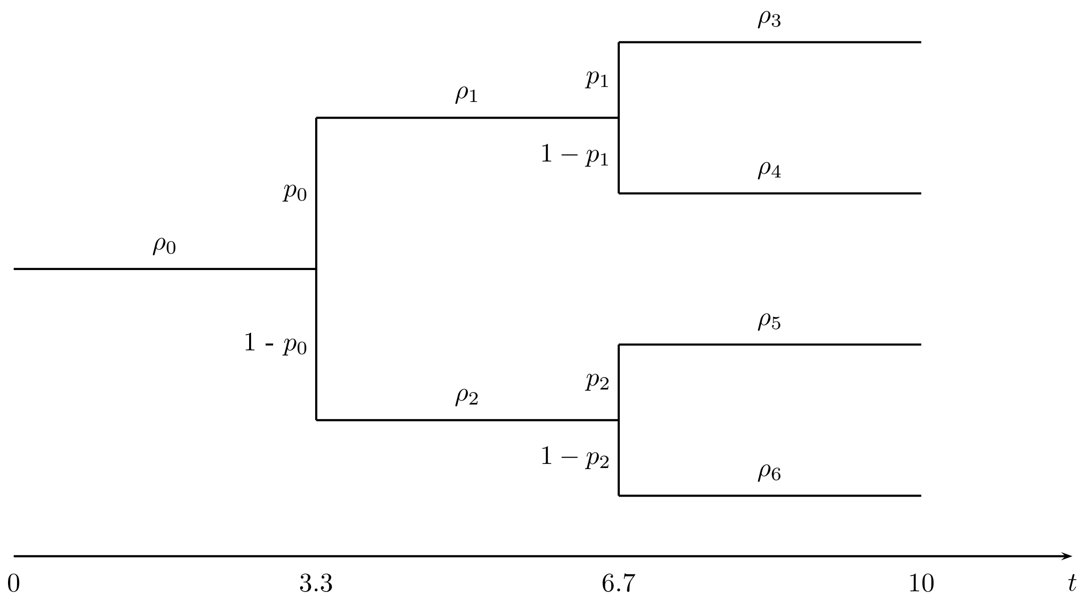

Figure 1: A dynamic tree structure

expectation of the tranche prices on each branch. Figure 1 illustrates an equally-spaced three-period[3] tree, where ρ_j is the value of the correlation factor loading and p_j is the probability of the process taking the upper branch. With this tree structure, the correlation factor loading process has four possible paths for a 10-year maturity contract. For example, for an annual payment tranche contract, one possible path for the $\beta_{k,i}$'s is $(\rho_0, \rho_0, \rho_0, \rho_1, \rho_1, \rho_1, \rho_3, \rho_3, \rho_3, \rho_3)$ with probability $p_0 p_1$. We can increase or decrease the degrees of freedom of the tree by adjusting the number of periods or the tree structure, e.g., constraining the general tree to be a binomial tree. Recently, Kaznady (2011) proposed an improved alternative multi-path parameterization of the correlation coefficients dynamics.

6 Numerical examples

We begin by comparing the results generated by the Monte Carlo method to those obtained by the recursion (13) on an example with arbitrarily chosen parameters. The numerical experiments are based on 5-year CDOs with 100 underlying names and annual premium payments. The tranche structure is the same as those of CDX, i.e., six tranches with attachment

[3]As illustrated in this example, the number of periods for the tree may be different from the number of periods of our model, which equals the number of premium payments.

Time	1Y	2Y	3Y	4Y	5Y
Probability	0.0041	0.0052	0.0069	0.0217	0.0288

Table 1: Risk-neutral cumulative default probabilities

Tranche	Monte Carlo	95% CI	Recursion
0% – 3%	953.40	[946.71, 960.62]	951.60
3% – 7%	182.09	[179.51, 184.81]	181.59
7% – 10%	58.95	[57.26, 60.33]	58.77
10% – 15%	22.21	[21.01, 23.39]	22.09
15% – 30%	3.47	[3.03, 3.78]	3.44
30% – 100%	0.07	[0.03, 0.09]	0.07

Table 2: Tranche premia (bps)

and detachment points, 0%–3%, 3%–7%, 7%–10%, 10%–15%, 15%–30% and 30%–100%. We assume a constant interest rate of 4% and a constant recovery rate of 40%. For simplicity, we assume that all $\beta_{k,i} = 0.6$. The risk-neutral cumulative default probabilities are listed in Table 1.

Each Monte Carlo simulation consists of 100,000 trials, and 100 runs (with different seeds) for each experiment are made. Based on the results of these 100 experiments, we calculate the mean and the 95% non-parametric confidence interval. Table 2 presents the risk premia for the CDOs. For our example, the running time of one Monte Carlo experiment with 100,000 trials is about 14 times that used by our recursive method. These results demonstrate that the recursive relationship (13) is accurate and efficient.

To calibrate against the market quotes, we employ the tree structure for the correlation factor loadings discussed in the previous section. In particular, we use an equally-spaced four-period tree. However, we add constraints by using the same growth rate μ_j and probability p_j for period j, as shown in the tree in Figure 2. Therefore, we have 7 parameters in total to calibrate against 12 tranche quotes. We compute the parameters by solving an associated optimization problem. For the objective function of the optimization problem, we could use either the absolute error in the spreads

$$f_{\text{abs}} = \sum (m_i - s_i)^2, \text{ for } i = 1, \ldots, 12$$

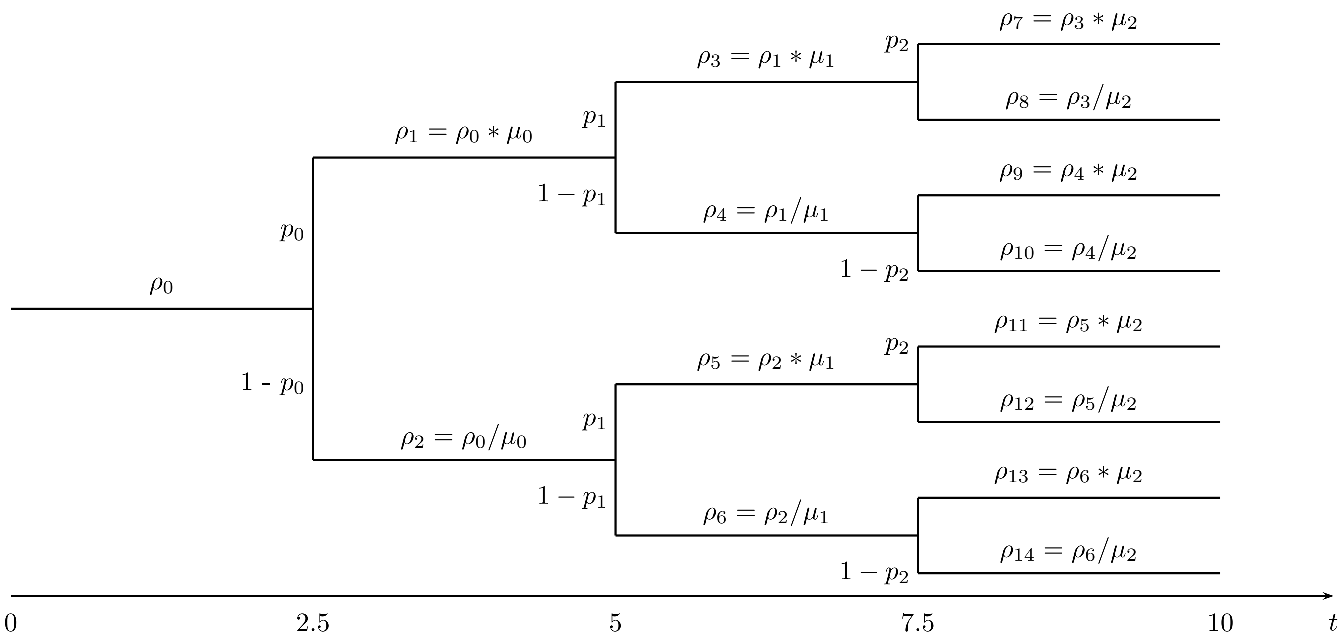

Figure 2: A particular dynamic tree example

or the relative error in the spreads

$$f_{\text{rel}} = \sum (m_i - s_i)^2/m_i^2, \text{ for } i = 1, \ldots, 12$$

where m_i is the market spread quote for tranche i and s_i is the model spread for tranche i.

Table 3 lists the calibration result for the tranche quotes of CDX series 8 on April 4, 2007. The upper half of the table uses the absolute spread error as the objective function, while the lower half of the table uses the relative spread error as the objective function. In both cases, the rows "Parameter" display the values of the parameters in our model, in the order $\rho_0, \mu_0, p_0, \mu_1, p_1, \mu_2, p_2$.

Table 4 lists the calibration results for the same data using the Gaussian factor copula model and the normal inverse Gaussian factor copula model in Kalemanova et al. (2007). In the table, "NIG(1)" means the normal inverse Gaussian factor copula model with one extra parameter for fat-tailness, and "NIG(2)" means the normal inverse Gaussian factor copula model with two extra parameters for skewness and fat-tailness. Our results in Table 3 are far superior to the results of the three models in Table 4.

Maturity	5 yr			7 yr			10 yr		
Tranche	Market	Model	Abs Err	Market	Model	Abs Err	Market	Model	Abs Err
3 – 7	111.81	110.13	1.68	251.44	254.65	3.21	528.31	528.37	0.06
7 – 10	22.31	20.90	1.41	54.69	59.51	4.82	134.00	134.21	0.21
10 – 15	10.42	7.99	2.43	26.47	28.45	1.98	63.30	61.38	1.92
15 – 30	4.34	1.97	2.37	9.50	12.08	2.58	20.46	23.36	2.90
Parameter	0.73	0.43	0.98	0.32	0.57	0.11	0.63	$f_{\text{abs}} = 8.52$	
Tranche	Market	Model	Rel Err	Market	Model	Rel Err	Market	Model	Rel Err
3 – 7	111.81	109.88	1.73%	251.44	300.00	19.31%	528.31	560.57	6.11%
7 – 10	22.31	21.37	4.23%	54.69	54.00	1.26%	134.00	141.36	5.49%
10 – 15	10.42	10.79	3.57%	26.47	25.01	5.52%	63.30	60.20	4.90%
15 – 30	4.34	4.36	0.37%	9.50	9.86	3.79%	20.46	22.30	8.99%
Parameter	0.55	0.65	0.80	0.42	0.71	0.15	0.57	$f_{\text{rel}} = 25.01\%$	

Table 3: Calibration result of CDX 8 on April 4, 2007

In additional to the market data on a single day, we calibrate our model against market spreads of CDX series 8 on each Wednesday from March 23, 2007 to July 4, 2007. Figure 3 plots the absolute errors and relative errors of the 12 tranches using the four-period tree structure with 7 parameters. The unit of the absolute error is basis points and the unit of the relative error is percentage. For market data before the credit crunch (July, 2007), our model is able to match the data quite well with 7 parameters. For market data after the credit crunch, the calibration error increases dramatically. We believe this is because the market quotes exhibit arbitrage due to the large demand and supply gap. As the financial crisis developed, traders tried to sell the credit derivatives they were holding, but no one wanted to buy them. For more numerical results about calibration of our model, refer to Kaznady (2011).

7 Conclusions

In this paper, we introduce a dynamic multi-period factor copula model, which can be calibrated fairly easily and matches the market quotes quite well. Using the independence of the common factors and the forward default probability, we show that the loss of a completely homogeneous pool possesses the Markov property. Therefore, we can avoid the

Maturity	5 yr				7 yr				10 yr			
Tranche	Market	Gaussian	NIG(1)	NIG(2)	Market	Gaussian	NIG(1)	NIG(2)	Market	Gaussian	NIG(1)	NIG(2)
3 − 7	111.81	149.77	84.48	92.12	251.44	379.65	240.59	240.36	528.31	653.48	537.32	536.43
7 − 10	22.31	14.61	32.42	33.21	54.69	80.52	62.03	64.61	134.00	248.90	154.68	148.07
10 − 15	10.42	1.51	21.42	19.71	26.47	14.80	36.18	35.30	63.30	77.84	66.95	65.44
15 − 30	4.34	0.02	12.28	9.36	9.50	0.49	19.02	16.18	20.46	5.49	29.00	26.38
Abs err		39.98	32.14	24.86		131.62	18.88	18.54		171.19	24.39	17.42
Parameter	Gaussian: 0.30				NIG(1): 0.46, 0.37				NIG(2): 0.44, 0.99, -0.61			
Tranche	Market	Gaussian	NIG(1)	NIG(2)	Market	Gaussian	NIG(1)	NIG(2)	Market	Gaussian	NIG(1)	NIG(2)
3 − 7	111.81	164.22	89.70	86.76	251.44	383.20	289.34	265.15	528.31	635.06	642.09	616.51
7 − 10	22.31	21.07	23.40	24.17	54.69	94.04	53.01	53.28	134.00	255.83	173.08	151.31
10 − 15	10.42	2.88	12.52	12.50	26.47	20.92	24.02	25.14	63.30	89.10	54.25	53.37
15 − 30	4.34	0.07	4.96	4.46	9.50	0.98	8.77	9.23	20.46	8.06	15.40	16.95
Rel err		130.96%	31.95%	31.26%		128.06%	19.53%	8.35%		118.37%	46.17%	31.39%
Parameter	Gaussian: 0.33				NIG(1): 0.34, 0.44				NIG(2): 0.35, 0.99, -0.63			

Table 4: Calibration result of CDX 8 on April 4, 2007 by different models

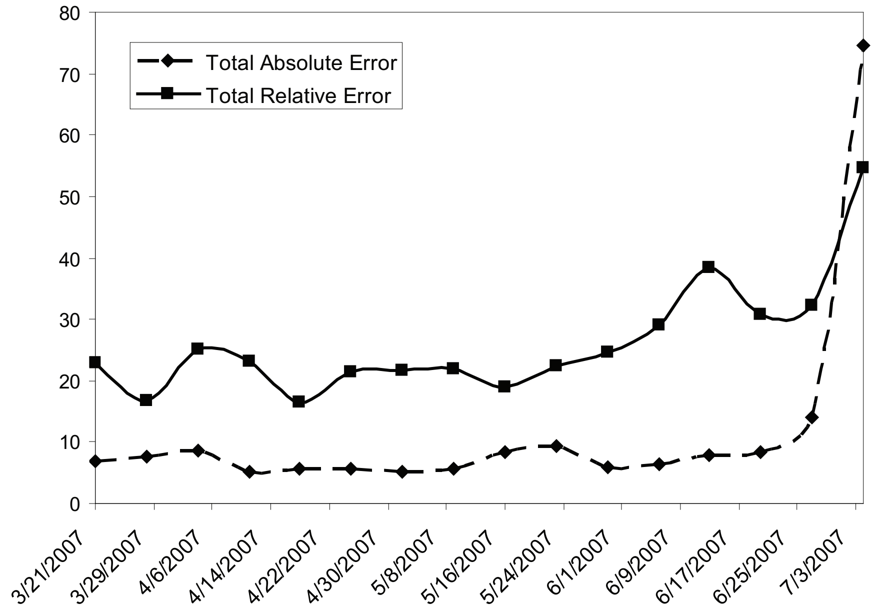

Figure 3: Weekly calibration result of CDX 8

multi-dimensional integration that must be computed in the multi-period factor copula models. The calibration results demonstrate the flexibility of our model in fitting the market quotes. Most importantly, the method is a generic one: it can be applied to construct a multi-period factor copula model from any one-period factor copula model based on the conditional independence framework.

Our numerical results demonstrate that our multi-period factor copula model is able to calibrate consistently against market data. However, we have developed an efficient method for completely homogenous pools only using an independent latent process across time. Therefore, key open questions are how to extend the model to a general pool and a general latent process.

References

Leif Andersen. Portfolio losses in factor models: Term structures and intertemporal loss dependence. *Journal of Credit Risk*, 2(4), 2006.

Leif Andersen and Jakob Sidenius. Extensions to the Gaussian copula: Random recovery and random factor loadings. *Journal of Credit Risk*, 1(1):29–70, 2004.

Leif Andersen, Jakob Sidenius, and Susanta Basu. All your hedges in one basket. *Risk*, 16 (11):67–72, 2003.

Martin Baxter. Gamma process dynamic modelling of credit. *Risk*, 20(10):98–101, 2007.

Norddine Bennani. The forward loss model: A dynamic term structure approach for the pricing of portfolio credit derivatives. Working paper, available at http://www.defaultrisk.com, 2005.

Christopher Fingers. A comparison of stochastic default rate models. *RiskMetrices Journal*, 1(2):49–73, 2000.

Kay Giesecke, Lisa Goldberg, and Xiaowei Ding. A top down approach to multi-name credit. *Operations Research*, 59(2):283–300, 2011.

John Hull and Alan White. Valuation of a CDO and an n^{th} to default CDS without Monte Carlo simulation. *Journal of Derivatives*, 12(2):8–23, 2004.

John Hull and Alan White. Valuing credit derivatives using an implied copula approach. *Journal of Derivatives*, 14(2):8–28, 2006.

John Hull and Alan White. Dynamic models of portfolio credit risk: A simplified approach. *Journal of Derivatives*, 15(4):9–28, 2008.

Ian Iscoe. Losing default correlation, step by step. Technical report, Algorithmics Inc., 2003.

Anna Kalemanova, Bernd Schmid, and Ralf Werner. The normal inverse Gaussian distribution for synthetic CDO pricing. *Journal of Derivatives*, 14(3):80–93, 2007.

Max Kaznady. Calibration of multi-period single-factor Gaussian copula models for CDO pricing. Master's thesis, University of Toronto, 2011. Available at http://www.cs.toronto.edu/NA/reports.html.

Jean-Paul Laurent and Jon Gregory. Basket default swaps, CDOs and factor copulas. *Journal of Risk*, 7(4):103–122, 2005.

David Li. On default correlation: A copula approach. *Journal of Fixed Income*, 9(4):43–54, 2000.

Lee McGinty, Eric Beinstein, Rishad Ahluwalia, and Martin Watts. Credit correlation: A guide. Technical report, JP Morgan, 2004.

William Morokoff. Simulation methods for risk analysis of collateralized debt obligations. Technical report, Moody's KMV, 2003.

Philipp Schönbucher. Portfolio losses and the term structure of loss transition rates: a new methodology for the pricing of portfolio credit derivatives. Working paper, available at http://www.defaultrisk.com, 2005.

Jakob Sidenius. On the term structure of loss distributions — a forward model approach. *International Journal of Theoretical and Applied Finance*, 10(4):749–761, 2007.

Jakob Sidenius, Vladimir Piterbarg, and Leif Andersen. A new framework for dynamic credit portfolio loss modeling. *International Journal of Theoretical and Applied Finance*, 11(2): 163–197, 2008.

Roberto Torresetti, Damiano Brigo, and Andrea Pallavicini. Implied correlation in CDO tranches: A paradigm to be handled with care. Working paper, available at www.defaultrisk.com, 2006.

Oldrich Vasicek. Probability of loss distribution. Technical report, KMV Corporation, 1987.