OPERATIONS RESEARCH
Vol. 56, No. 2, March–April 2008, pp. 286–303
ISSN 0030-364X | EISSN 1526-5463 | 08 | 5602 | 0286

informs

DOI 10.1287/opre.1070.0456

Fast Pricing of Basket Default Swaps

Zhiyong Chen
Financial Analytics and Structured Transactions Group, Bear Stearns and Company Inc., New York, New York 10179, zchen1@bear.com

Paul Glasserman
Graduate School of Business, Columbia University, New York, New York 10027, pg20@columbia.edu

A basket default swap is a derivative security tied to an underlying basket of corporate bonds or other assets subject to credit risk. The value of the contract depends on the joint distribution of the default times of the underlying assets. Valuing a basket default swap often entails Monte Carlo simulation of these default times. For baskets of high-quality credits and for swaps that require multiple defaults to trigger payment, pricing the swap is a rare-event simulation problem. The Joshi-Kainth algorithm is an innovative importance-sampling technique for this problem that forces a predetermined number of defaults to occur on each path. This paper analyzes, extends, and improves the Joshi-Kainth algorithm. We show that, in its original form, the algorithm can actually increase variance; we present an alternative that is guaranteed to reduce variance, even when defaults are not rare. Along the way, we provide a rigorous underpinning in a setting sufficiently general to include both the original method and the version proposed here.

Subject classifications: finance: securities; simulation: efficiency, applications.
Area of review: Financial Engineering.
History: Received March 2006; revision received July 2006; accepted January 2007. Published online in *Articles in Advance* December 17, 2007.

1. Introduction

Basket default swaps are derivative securities tied to an underlying portfolio (or "basket") of corporate bonds or other assets subject to credit risk. A basket default swap provides one party in the swap (the protection buyer) with a type of insurance against the possibility of default in exchange for regular payments made to the other party (the protection seller). Insuring a basket of assets is typically less expensive than insuring each asset separately. For example, in a *first-to-default swap*, the protection buyer is compensated if one asset in the basket defaults but receives no compensation for any subsequent defaults. This is less expensive than buying insurance against all possible defaults and may provide adequate protection if multiple defaults are very unlikely. In an nth-to-default swap, the protection buyer absorbs losses resulting from the first $n-1$ defaults and receives compensation only upon the occurrence of the nth default. See Bruyere et al. (2006) and Schönbucher (2003) for general background on these types of contracts.

Valuing basket default swaps (and, indeed, all portfolio credit derivatives) requires a model of the joint distribution of the default times of the assets in the portfolio. For pricing purposes, we are interested in the distribution under a suitable "risk-neutral" probability measure; we will assume that such a probability exists and is, in effect, chosen by the market. The marginal distribution of the time to default for each asset is typically gleaned from market information related to that asset, such as spreads on corporate bonds or on single-asset credit default swaps (see, e.g., Duffie and Singleton 2003). The main modeling challenge therefore lies in specifying the dependence among the default times, given their marginal distributions. The most widely used mechanism for characterizing this dependence is the Gaussian copula model (Gupton et al. 1997, Li 2000).

Simple cases of this and related models can be evaluated using transform inversion and numerical integration techniques, as in Laurent and Gregory (2003). However, more general cases typically require Monte Carlo simulation. If defaults are rare or if multiple defaults are required to trigger a payment to the protection buyer, then most replications in an ordinary Monte Carlo simulation are uninformative and a very large number of replications may be required to obtain precise price estimates.

To address this difficulty, Joshi and Kainth (2004) introduced an innovative importance-sampling technique that forces all paths to produce at least n defaults in pricing an nth-to-default swap. Their method generates default times sequentially, at each step, increasing the probability that the next asset defaults within the life of the swap contract until n defaults have occurred. Their implementation of this idea cleverly exploits features of the Gaussian copula model.

This procedure is largely unrelated to techniques commonly used in the rare-event simulation literature. It is reminiscent of urn sampling schemes (cf. Feller 1968) used in clinical trials to force a certain percentage of subjects to be assigned to each treatment, but is more complex because of the dependence between default times. Also, unlike most

importance-sampling techniques, the Joshi-Kainth (2004) method violates "absolute continuity," in the sense that their importance-sampling distribution assigns probability zero to certain events that have positive probability under the original sampling distribution.

The purpose of this paper is to analyze, extend, and improve the original method of Joshi and Kainth (2004). We first provide theoretical support for the method through a derivation of the likelihood ratio and verification of unbiasedness, despite the violation of absolute continuity. These results are implicit in Joshi and Kainth (2004), but a rigorous analysis in a more general formulation is useful as a step toward improving the method. We then investigate the variance of the estimator. We show that the order in which the default times are generated affects the variance and that the method does not necessarily reduce variance; indeed, we give conditions under which an increase in variance is guaranteed.

To address these issues, we change the default probabilities used by Joshi and Kainth (2004) to force n defaults on every path. The values used by Joshi and Kainth (2004) are intuitively appealing but somewhat arbitrary; we replace these with values that are, in a certain sense, optimal when the default times are independent. These values guarantee a reduction in variance inversely proportional to the probability of at least n defaults. We apply this method to the Gaussian copula with a factor structure that makes defaults conditionally independent, and we show that this extension also guarantees variance reduction. To obtain still greater variance reduction, we apply stratified sampling to the underlying factors. These ideas apply to other factor copula models as well.

Section 2 provides background on basket default swaps and the Gaussian copula model. Section 3 presents the algorithm of Joshi and Kainth (2004) and discusses some of its theoretical properties. Section 4 presents our alternative, based on conditional default probabilities, in the case of independent default times. Section 5 extends this to the Gaussian copula and §6 combines this method with stratified sampling of underlying factors. Section 7 concludes the paper. Most proofs are contained in the appendix.

2. Problem Description

This section describes the class of credit derivatives we consider and reviews the widely used Gaussian copula model of dependent default times.

2.1. Basket Default Swaps

We consider credit derivatives tied to a basket (or portfolio) of N underlying assets, such as bonds, loans, or credit default swaps. Typical values of N range from 5 to around 100. We let τ_i denote the default time of the ith asset, $i = 1, \ldots, N$, taking $\tau_i = \infty$ if the ith asset never defaults. We consider the pricing of a contract with a discounted payoff $\hat{V}(\tau_1, \ldots, \tau_N)$ depending on the default times.

The specific case of a basket default swap provides protection against the nth default in the basket, with n smaller than N and typically much smaller. This type of nth-to-default swap is less expensive than insuring each asset separately and may provide adequate protection if multiple defaults are unlikely. We use T to denote the life of the contract. Its cash flows are as follows. At dates $0 < T_1, T_2, \ldots, T_m \leqslant T$, the protection *buyer* is scheduled to make fixed payments of $s_1, \ldots, s_m$ to the protection *seller*. However, if the nth default occurs before T, these payments cease and the protection seller makes a payment to the protection buyer. This payment is determined by the identity of the nth asset to default, but is otherwise fixed. If the ith asset is the nth to default, the payment is $1 - r_i$, where r_i is the *recovery rate* and 1 is the (normalized) asset value. (Differences in asset values can be absorbed into differences in recovery rates.) We denote by R the recovery rate for the nth asset to default.

As in Joshi and Kainth (2004), we write the discounted payoff of the swap as the difference between the discounted payoffs of the payments made between the parties, called the *protection* leg and the *value leg*:

$$\hat{V}(\tau_1, \ldots, \tau_N) = V_{\texttt{value}}(\tau_1, \ldots, \tau_N) - V_{\texttt{prot}}(\tau_1, \ldots, \tau_N).$$

We write $D(t)$ for the discount factor for the interval from 0 to t and take this to be deterministic; in the simplest case, $D(t) = \exp(-rt)$ for some fixed rate r. Let τ denote the time of the nth default. Then,

$$V_{\texttt{value}}(\tau_1, \ldots, \tau_N) = (1 - R)D(\tau)I(\tau \leqslant T),$$

with $I(\tau \leqslant T)$ the indicator of the event that the nth default occurs before T; and

$$V_{\texttt{prot}}(\tau_1, \ldots, \tau_N) = \begin{cases} \displaystyle\sum_{i=1}^{j} s_i D(T_i) + s_{j+1} D(\tau) \frac{\tau - T_j}{T_{j+1} - T_j} \\ \qquad\qquad \text{if } T_j \leqslant \tau \leqslant T_{j+1}, \\ \displaystyle\sum_{i=1}^{m} s_i D(T_i) \quad \text{if } \tau > T. \end{cases}$$

The first term reflects the convention that upon the nth default, the protection buyer makes an accrued payment to the protection seller, with these payments accruing linearly between the dates T_j. To price the swap is to compute $\mathbf{E}(\hat{V}(\tau_1, \ldots, \tau_N))$.

It will be convenient to subtract the deterministic component of the swap and define

$$V(\tau_1, \ldots, \tau_N) = \hat{V}(\tau_1, \ldots, \tau_N) + \sum_{i=1}^{m} s_i D(T_i),$$

so it suffices to compute $\mathbf{E}(V(\tau_1, \ldots, \tau_N))$. Moreover, $V(\tau_1, \ldots, \tau_N)$ is zero unless n or more defaults occur before T, so

$$\mathbf{E}(V(\tau_1, \ldots, \tau_N)) = \mathbf{E}(V(\tau_1, \ldots, \tau_N)I(\tau \leqslant T)). \tag{1}$$

Using this formulation, Joshi and Kainth (2004) develop a simulation method that generates at least n defaults on every path.

2.2. The Gaussian Copula Model

The expectation in (1) is over the joint distribution of the default times, which we have not yet specified. The marginal distribution of each τ_i is typically extracted from the market prices of credit default swaps or bonds; these market prices are used to construct a hazard rate function h_i from which we get the distribution

$$F_i(t) = \mathbf{P}(\tau_i \leqslant t) = 1 - \exp\left(-\int_0^t h_i(s)\,ds\right).$$

The Gaussian copula (Gupton et al. 1997, Li 2000) is a widely used mechanism for specifying a joint distribution for the default times consistent with these marginals. The dependence among $\tau_1, \ldots, \tau_N$ is determined by underlying jointly normal random variables $W_1, \ldots, W_N$. Each W_i has a standard normal distribution Φ, so $\Phi(W_i)$ is uniformly distributed on $(0, 1)$ and $\tau_i = F_i^{-1}(\Phi(W_i))$ has distribution F_i. However, $W_1, \ldots, W_N$ are correlated, with correlation matrix $\boldsymbol{\Sigma}$, and this introduces (and, indeed, completely characterizes) dependence among the default times $\tau_1, \ldots, \tau_N$.

For purposes of simulation, it is often convenient to work with the Cholesky decomposition of the correlation matrix $\boldsymbol{\Sigma}$. The Cholesky decomposition produces a lower-triangular matrix $\mathbf{C} = [C_{ij}]$ with $\boldsymbol{\Sigma} = \mathbf{C}\mathbf{C}^\top$. This allows us to write $\mathbf{W} = \mathbf{C}\bar{\mathbf{Z}}$ with $\bar{\mathbf{Z}}$ an $N \times 1$ vector of independent standard normal random variables. We will make the simplifying assumption that $\boldsymbol{\Sigma}$ has full rank so that the diagonal elements of $\mathbf{C}$ are all nonzero.

Using this formulation, each replication of an ordinary Monte Carlo simulation would proceed as follows:

Step 1. Generate N independent uniforms $U_1, \ldots, U_N$.

Step 2. Generate N independent normals $\bar{Z}_1, \ldots, \bar{Z}_N$ by setting $\bar{Z}_i = \Phi^{-1}(U_i)$.

Step 3. Set $\mathbf{W} = \mathbf{C}\bar{\mathbf{Z}}$.

Step 4. Generate $\tau_1, \ldots, \tau_N$ by setting $\tau_i = F_i^{-1}(\Phi(W_i))$.

Step 5. Compute $V(\tau_1, \ldots, \tau_N)$.

For later use, we record the following consequence of the Cholesky decomposition, which is implicit in Joshi and Kainth (2004):

LEMMA 1. *If* $\boldsymbol{\Sigma} > 0$, $\mathbf{W} = \mathbf{C}\bar{\mathbf{Z}}$ *and* $\tau_i = F_i^{-1}(\Phi(W_i))$, $i = 1, \ldots, N$, *then*

$$\mathbf{P}(\tau_i \leqslant t \mid \bar{Z}_1, \ldots, \bar{Z}_{i-1}) = \Phi\left(\frac{\Phi^{-1}(F_i(t)) - \sum_{j=1}^{i-1} C_{ij}\bar{Z}_j}{C_{ii}}\right). \quad (2)$$

PROOF. By construction,

$$\tau_i = F_i^{-1}(\Phi(W_i)) = F_i^{-1}(\Phi(C_{i1}\bar{Z}_1 + \cdots + C_{ii}\bar{Z}_i)),$$

so

$$\begin{aligned}
&\mathbf{P}(\tau_i \leqslant t \mid \bar{Z}_1, \ldots, \bar{Z}_{i-1}) \\
&\quad = \mathbf{P}(W_i \leqslant \Phi^{-1}(F_i(t)) \mid \bar{Z}_1, \ldots, \bar{Z}_{i-1}) \\
&\quad = \mathbf{P}\left(\sum_{j=1}^{i} C_{ij}\bar{Z}_j \leqslant \Phi^{-1}(F_i(t)) \mid \bar{Z}_1, \ldots, \bar{Z}_{i-1}\right) \\
&\quad = \mathbf{P}\left(\bar{Z}_i \leqslant \left[\Phi^{-1}(F_i(t)) - \sum_{j=1}^{i-1} C_{ij}\bar{Z}_j\right] \Big/ C_{ii} \mid \bar{Z}_1, \ldots, \bar{Z}_{i-1}\right),
\end{aligned}$$

from which the result follows. That C_{ii} is nonzero follows from the assumption that $\boldsymbol{\Sigma}$ has full rank. □

Even if $\boldsymbol{\Sigma}$ is only positive semidefinite, so that some C_{ii} may be zero, (2) remains valid provided the expression on the right is interpreted as a step function that jumps from zero to one at the point at which the numerator in the argument of Φ equals zero.

3. The Joshi-Kainth Method

For baskets of high-quality credits or short-maturity swaps, defaults are rare events. Ordinary Monte Carlo simulation therefore produces few paths with n or more defaults; nearly all paths simply return a value of zero for $V(\tau_1, \ldots, \tau_N)$. This makes importance sampling (IS) potentially attractive in this problem.

Joshi and Kainth (2004) introduced an innovative importance-sampling method (henceforth, the JK method) that forces all paths to produce at least n defaults. Their method is largely unrelated to techniques commonly used in the rare-event simulation literature and, in particular, to the techniques developed in the credit risk context by Avranitis and Gregory (2001), Glasserman and Li (2005), Joshi (2005), and Kalkbrener et al. (2004). In this section, we review the JK method, establish some properties of the method, introduce some initial improvements, and present some numerical examples. Our examples motivate an improved method that takes advantage of the key insights of Joshi and Kainth (2004) while producing far greater variance reduction.

3.1. The JK Algorithm

We first give an informal description of the JK method in the case of a first-to-default swap. The JK method first determines whether a particular asset defaults within the life of the swap, and, if a default is to occur, it then determines the time of the default in the interval $[0, T]$. Only the default probabilities are changed. The first asset is assigned a probability $1/N$ of default by time T (regardless of its actual default probability). If it defaults, then the importance sampling is turned off and the simulation proceeds in the usual way. Otherwise, the second asset is assigned a default probability of $1/(N-1)$. The default probabilities for subsequent assets are increased through the sequence $1/(N-2), 1/(N-3), \ldots$ until a default occurs, at which point importance sampling is suspended; if we reach the Nth asset and no default has yet occurred, the Nth asset is given a default probability of $1/(N-(N-1)) = 1$. This ensures that at least one default occurs on every path. The case of an nth-to-default swap works similarly. When the algorithm gets to the ith asset, if $j < n$ defaults have thus far occurred, the ith asset is made to default with probability $(n-j)/(N-i+1)$, ensuring that at least n defaults occur on every path. Once n defaults have occurred, the method reverts to the original sampling procedure. Each

path is then weighted by a likelihood ratio to correct for the change in distribution.

The importance-sampling probabilities used by Joshi and Kainth (2004) may be interpreted as follows. Consider an urn initially containing n black balls and $N-n$ white balls. Balls are drawn from the urn at random, without replacement. If $j<n$ of the first $i-1$ balls are black, then the probability that the ith draw produces a black ball is $(n-j)/(N-i+1)$. All n black balls will eventually be drawn.

We present the algorithm in a somewhat more generic form than that given in Joshi and Kainth (2004); this will facilitate the analysis and extension. For $i=2,\dots,N$, write $F_i(t\mid\tau_1,\dots,\tau_{i-1})$ for the conditional distribution of the default time τ_i given the previous default times $\tau_1,\dots,\tau_{i-1}$. A generic method for simulating the default times generates independent uniforms $U_1,\dots,U_N$ and sets $\tau_1 = F_1^{-1}(U_1)$ and

$$\tau_i = F_i^{-1}(U_i\mid\tau_1,\dots,\tau_{i-1}),\quad i=2,\dots,N,$$

the inverse taken with respect to the first argument.

For the IS algorithm, define default indicator variables $Y_i = I(\tau_i\leqslant T)$, $i=1,\dots,N$. The conditional default probabilities are then given by $p_1 = F_1(T)$ and

$$p_i = \mathbf{P}(Y_i=1\mid\tau_1,\dots,\tau_{i-1}) = F_i(T\mid\tau_1,\dots,\tau_{i-1}),\qquad i=2,\dots,N.$$

These will be replaced by new probabilities $\tilde p_i$, $i=1,\dots,N$, with $\tilde p_1$ fixed and $\tilde p_i$ a function of $Y_1,\dots,Y_{i-1}$. The particular choice used by Joshi and Kainth (2004) can be expressed as

$$\tilde p_1 = n/N,\quad \tilde p_i = \begin{cases} \left(n-\sum_{j=1}^{i-1}Y_j\right)\Big/(N-i+1), & \\ \qquad\qquad \sum_{j=1}^{i-1}Y_j < n, & \\ p_i, \quad \sum_{j=1}^{i-1}Y_j \geqslant n. & \end{cases} \tag{3}$$

Here, the sum over Y_j simply counts the number of assets defaulting by time T.

The JK procedure can be viewed as generating the pairs $(Y_1,\tau_1),\dots,(Y_N,\tau_N)$ recursively using independent $V_1,\dots,V_N$, uniformly distributed between zero and one. Each replication works as follows:

Sampling Procedure. For each $i=1,\dots,N$,

Step 1. Generate V_i uniformly over $(0,1)$.

Step 2. Set $Y_i = I(V_i\leqslant\tilde p_i)$,

$$U_i = \begin{cases} p_iV_i/\tilde p_i & \text{if } Y_i=1,\\ p_i+[(1-p_i)(V_i-\tilde p_i)/(1-\tilde p_i)] & \text{if } Y_i=0,\end{cases} \tag{4}$$

and

$$\tau_i = F_i^{-1}(U_i\mid\tau_1,\dots,\tau_{i-1}). \tag{5}$$

(In the case $i=1$, set $\tau_1 = F_1^{-1}(U_1)$.)

Step 3. Calculate the weight:

$$L_i = \begin{cases} p_i/\tilde p_i & \text{if } Y_i=1,\\ (1-p_i)/(1-\tilde p_i) & \text{if } Y_i=0.\end{cases} \tag{6}$$

Once $\tau_1,\dots,\tau_N$ have been generated, evaluate $V(\tau_1,\dots,\tau_N)$ and return the weighted estimate $V(\tau_1,\dots,\tau_N)L$, with $L = L_1L_2\cdots L_N$ the weight for the path.

Observe that the events $\{\tau_i\leqslant T\}$, $\{U_i\leqslant p_i\}$, and $\{Y_i=1\}$ coincide in this construction, so the Y_i are indeed default indicators, even though we generate Y_i before τ_i. Steps (4) and (5) thus accomplish the following:

$$\text{Generate } \tau_i \text{ conditional on } \{\tau_1,\dots,\tau_{i-1};\tau_i\leqslant T\} \text{ or } \{\tau_1,\dots,\tau_{i-1};\tau_i>T\} \tag{7}$$

depending on whether $Y_i=1$ or $Y_i=0$.

Implementation of this procedure requires the ability to sample from the conditional default time distributions in (5). The specific case in Joshi and Kainth (2004) uses a clever combination of the Gaussian copula and the Cholesky decomposition. Suppose, for simplicity, that each marginal distribution F_i is strictly increasing, so that W_i can be recovered from τ_i as $W_i = \Phi^{-1}(F_i(\tau_i))$, and suppose that $\mathbf{\Sigma}$ has full rank, so that $\bar Z_1,\dots,\bar Z_i$ can be recovered from $W_1,\dots,W_i$. Then, conditioning on $\tau_1,\dots,\tau_i$ is equivalent to conditioning on $\bar Z_1,\dots,\bar Z_i$ and p_i can be evaluated from (2) by setting $t=T$. Moreover, $\tau_i = F_i^{-1}(U_i\mid\tau_1,\dots,\tau_{i-1})$ can be evaluated by using (2) to get

$$\tau_i = F_i^{-1}\left(\Phi\left(\sum_{j=1}^{i-1}C_{ij}\bar Z_j + C_{ii}\Phi^{-1}(U_i)\right)\right).$$

This reduces the problem of sampling from the conditional distributions of the default times to sampling from their marginal distributions. (This mechanism remains valid even if $\mathbf{\Sigma}$ is singular or some F_i fails to be strictly increasing, but in either of those cases we would be sampling τ_i conditional on the slightly larger information set $\{\bar Z_1,\dots,\bar Z_{i-1}\}$, rather than on the more generic information set $\{\tau_1,\dots,\tau_{i-1}\}$.)

3.2. Properties of the JK Method

Write $\tilde{\mathbf{P}}$ for the probability of events defined by $\tau_1,\dots,\tau_N$ constructed using the modified conditional default probabilities $\tilde p_i$, $i=1,\dots,N$, and write $\mathbf{P}$ for the original probability measure. Recall that $\tilde p_i$ may depend on the default indicators $Y_1,\dots,Y_{i-1}$. Extending the original method of Joshi and Kainth (2004), a simple condition ensures that at least n defaults occur on every path:

LEMMA 2. *If* $\tilde p_i = 1$ *whenever* $n-\sum_{j=1}^{i-1}Y_j = N-i+1$, $i=1,\dots,N$, *then* $\tilde{\mathbf{P}}(\tau\leqslant T)=1$.

In referring to the JK method, we will always assume that the condition in Lemma 2 is in force. The particular case (3) used by Joshi and Kainth (2004) satisfies the condition in the lemma, but so do many other choices of probabilities. This raises the question of which probabilities within this class are best, a question we return to in §4.

The next lemma shows that the JK method modifies the distribution of the default times only to the extent that it modifies the probability that default occurs before time T.

LEMMA 3. *For all* $i = 1, \ldots, N$ *and* $t \geqslant 0$, $\tilde{\mathbf{P}}(\tau_i \leqslant t \mid \tau_1, \ldots, \tau_{i-1}, Y_i) = \mathbf{P}(\tau_i \leqslant t \mid \tau_1, \ldots, \tau_{i-1}, Y_i)$.

Using this result, our next result confirms that the weight L calculated in the JK procedure is indeed the appropriate likelihood ratio. Moreover, it confirms that the JK method yields an unbiased estimate, despite the fact (atypical of importance sampling) that $\tilde{\mathbf{P}}$ assigns probability zero to some events that have positive probability under $\mathbf{P}$. Write $\tilde{\mathbf{E}}$ for expectation with respect to $\tilde{\mathbf{P}}$.

THEOREM 1. (i) *For any event* $A \subseteq \{\tau \leqslant T\}$, *we have* $\mathbf{P}(A) = \tilde{\mathbf{E}}[I_A L]$, *where* I_A *is the indicator of the event* A. *In other words,* L *is the likelihood ratio for the restriction of the probability* $\mathbf{P}$ *to* $\{\tau \leqslant T\}$ *relative to the probability* $\tilde{\mathbf{P}}$.

(ii) *For any* $V(\tau_1, \ldots, \tau_N) \geqslant 0$ *with* $V(\tau_1, \ldots, \tau_N) = 0$ *whenever* $\tau > T$, *we have*

$$\mathbf{E}[V(\tau_1, \ldots, \tau_N)] = \tilde{\mathbf{E}}[V(\tau_1, \ldots, \tau_N)L].$$

The JK method goes through the assets sequentially, increasing the default probability at each step until the required number of defaults have occurred. The first issue one faces in implementing the JK algorithm is the order in which to take the assets. Joshi and Kainth (2004) do not comment on this question, and one might wonder if it matters. We will see that it does matter, and the next result gives the optimal sequence for the case of independent defaults with equal recoveries and different constant hazard rates:

PROPOSITION 1. *For a basket of independent assets with identical recovery values* $r_1 = r_2 = \cdots = r_N$ *but different constant hazard rates, variance is minimized in the JK method by taking the assets in ascending order* $p_1 \leqslant p_2 \leqslant \cdots \leqslant p_N$ *of their marginal default probabilities* $p_i = \mathbf{P}(\tau_i \leqslant T)$.

This result remains valid with nonconstant hazard rates as long as we have $h_1(t) \leqslant h_2(t) \leqslant \cdots \leqslant h_N(t)$ for all $t \geqslant 0$. Although this still leaves open the question of the optimal order for more general cases, we will not pursue this question because the alternative we propose in §4 is insensitive to the order in which the default times are generated.

The JK method is an intuitively appealing way to generate informative paths, but its effectiveness depends on its ability to reduce variance. Joshi and Kainth (2004) do not provide conditions ensuring a variance reduction; indeed, the next result shows that a variance increase is guaranteed under certain conditions if one uses the probabilities (3).

PROPOSITION 2. *If the default probabilities of the assets are large enough or the life of the swap long enough, the estimator given by the JK method using* (3) *has larger variance than plain Monte Carlo sampling if* $n \leqslant N - 1$; *i.e., except in the case of a last-to-default swap.*

The shortcoming identified by Proposition 2 results from the possibility that the JK method will actually decrease one or more of the default probabilities as it goes through the assets. As a simple way to address this case, we suggest a *modified JK method*. Whereas the original method sets $\tilde{p}_i = (n - j)/(N - i + 1)$ if j defaults have occurred when we get to the ith asset, the modified method sets

$$\tilde{p}_i = \max\left(\frac{n - j}{N - i + 1}, p_i\right),$$

where, as before, $p_i = \mathbf{P}(\tau_i \leqslant T \mid \tau_1, \ldots, \tau_{i-1})$. An immediate consequence is the following:

PROPOSITION 3. *For all sufficiently large hazard rates or swap maturities, the modified JK estimator coincides with the ordinary Monte Carlo estimator.*

3.3. Numerical Examples

We illustrate the JK method and the modified JK method discussed so far with some numerical results. In these examples, we take a continuously compounded interest rate of $r = 5\%$. For simplicity, we assume just a single protection payment (i.e., $m = 1$) of $s = 0.10$, paid at maturity if fewer than n defaults have occurred. These parameters will be used in subsequent sections as well, and all numerical results are based on 10^6 replications. We use a large number of replications to make accurate variance comparisons; the same relative performance holds even with far fewer replications.

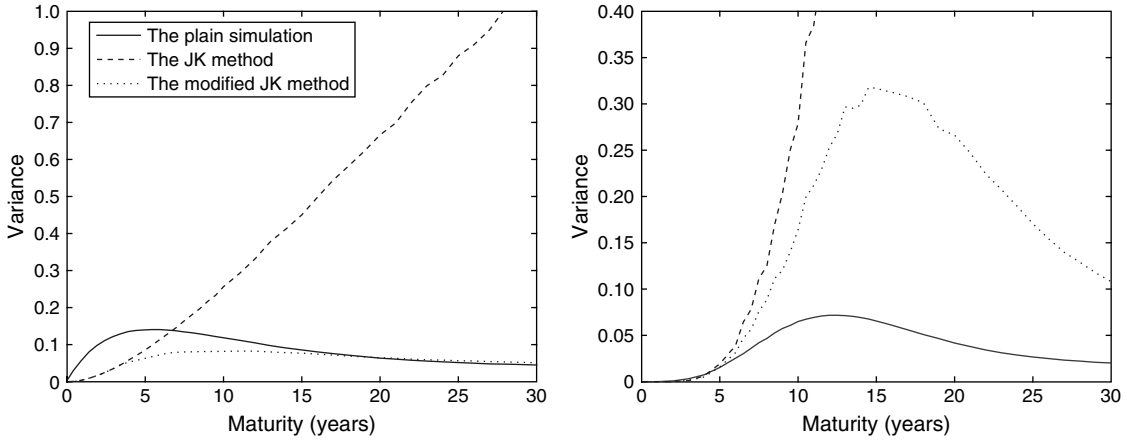
Basket I—Swap A1

As a first illustration, we consider a basket of $N = 10$ independent assets, with constant hazard rates $(0.03, 0.01, 0.02, 0.01, 0.005, 0.001, 0.002, 0.002, 0.017, 0.003)$. The recovery rates are $(0.3, 0.1, 0.2, 0.1, 0.3, 0.1, 0.2, 0.2, 0.1, 0.3)$. Swap A1 is a first-to-default swap in Basket I.

Basket II—Swap A2

Basket II contains $N = 10$ assets with constant hazard rates $(0.05, 0.01, 0.02, 0.02, 0.03, 0.1, 0.03, 0.09, 0.1, 0.05)$. The recovery rates are $(0.3, 0.1, 0.2, 0.1, 0.3, 0.1, 0.2, 0.2, 0.1, 0.3)$. They are correlated and the correlation matrix is $\boldsymbol{\Sigma}$, which is introduced by a multivariate normal vector $\{W_1, \ldots, W_N\}$ in the form of a four-factor model. To generate such a correlation matrix, one can first randomly generate a 10×4 matrix $\mathbf{A}$, compute the supplementary vector $\mathbf{B}$ such that $B_i = \sqrt{1 - \sum_{j=1}^{4} a_{ij}^2}$, then let $\boldsymbol{\Sigma} = \mathbf{A}\mathbf{A}' + \mathbf{B}\mathbf{B}'$. The matrix $\mathbf{A}$ used here can be found

Figure 1. The variance of the estimated price of Swap A1 (left panel) and Swap A2 (right panel).



in the appendix. Swap A2 is a fifth-to-default swap in Basket II.

The results are displayed in Figure 1, which plots variance against swap maturity. The solid line shows the variance of the estimator in the plain Monte Carlo method. The dashed line shows the variance in the JK method and the dotted line shows the variance in the modified JK method. Figure 1 shows that the modified JK method is always better than the JK method, but neither is guaranteed to be better than plain Monte Carlo. As expected, the JK method deteriorates as T increases and defaults become frequent. For the independent assets case, when $T = 7$, where $P(\tau > T) = 0.5034$, the JK method becomes inferior to plain Monte Carlo. For the dependent case, the transition occurs at an even shorter maturity.

4. The Conditional Probability (CP) Method with Independent Assets

The key feature of the JK method is that it replaces the original conditional default probabilities $p_i = \mathbf{P}(\tau_i \leqslant T \mid \tau_1, \ldots, \tau_{i-1})$ with new probabilities having two properties: they guarantee at least n defaults on every replication, and they revert to the original probabilities once n defaults have occurred. Beyond these two properties, the particular values (3) used in the original JK method are somewhat arbitrary. We therefore consider building on their method by searching for probabilities with these two key properties that are, in some sense, optimal. We begin by considering the case of independent default times, a restriction in force throughout this section.

Observe that most of the variability in pricing a basket default swap results from the occurrence or nonoccurrence of the event $\{\tau \leqslant T\}$; the timing and order of defaults produce additional variability, but this is secondary. We therefore consider the related problem of finding effective IS probabilities for the problem of estimating

$$\mathbf{P}(\tau \leqslant T) = \mathbf{E}[I(\tau \leqslant T)] = \mathbf{E}\left[I\left(\sum_{i=1}^{N} Y_i \geqslant n \right) \right]. \tag{8}$$

In fact, it is well known (and easily verified) that the optimal IS distribution for this type of problem samples $Y_1, \ldots, Y_N$ conditional on the event of interest; indeed, this IS technique produces zero variance in estimating (8), a point we return to later. We therefore propose to apply importance sampling with probabilities

$$q_i = \mathbf{P}\left(Y_i = 1 \,\middle|\, Y_1, \ldots, Y_{i-1}, \sum_{j=1}^{N} Y_j \geqslant n \right), \quad i = 1, \ldots, N.$$

It follows immediately from this definition that these probabilities have the key properties identified above:

Property 1. *If there are $k < n$ defaults in the first $N - (n-k)$ assets, then*

$$q_i = \mathbf{P}\left(Y_i = 1 \,\middle|\, Y_1, \ldots, Y_{i-1}, \sum_{j=1}^{i-1} Y_j = k, \sum_{j=1}^{N} Y_j \geqslant n \right) = 1$$

for $i = N - (n-k) + 1, \ldots, N$. Therefore, IS with these probabilities guarantees at least n defaults on every path.

Property 2. *If there are at least n defaults among the first k assets, then*

$$\begin{aligned} q_i &= \mathbf{P}\left(Y_i = 1 \,\middle|\, Y_1, \ldots, Y_{i-1}, \sum_{j=1}^{k} Y_j \geqslant n, \sum_{j=1}^{N} Y_j \geqslant n \right) \\ &= \mathbf{P}(Y_i = 1 \mid Y_1, \ldots, Y_{i-1}) = p_i \end{aligned}$$

for $i = k+1, \ldots, N$.

We will proceed to first show that these probabilities can be evaluated easily, and second show that they produce effective variance reduction in the original pricing problem.

4.1. Calculating the Conditional Probabilities

A simple and fast recursive algorithm allows calculation of the conditional probabilities, which can then be used in

the IS procedure. To explain the algorithm, it is useful to introduce a Markov chain

$$X_i = \sum_{j=1}^{i} Y_j, \quad i = 0, 1, \ldots, N,$$

that counts the number of defaults as we go through the number of underlying assets. (The "time" parameter i of this chain indexes the underlying assets and is unrelated to the evolution of time in the original swap-pricing problem.) That this process is indeed Markov follows from the independence of the default indicators Y_j.

The chain is absorbed at time N into any of the states $\{0, 1, \ldots, N\}$. The required probabilities q_i are simply the transition probabilities of the chain conditional on absorption into the set $G = \{n, n+1, \ldots, N\}$. Define

$$P_i^{(k)} = \mathbf{P}(X_N \geqslant n \mid X_i = k);$$

this is the probability of absorption into set G given that k defaults have been observed as of time i. These probabilities can be computed through backward induction by noting that

$$P_N^{(k)} = \begin{cases} 0, & k = 0, 1, \ldots, n-1, \\ 1, & k = n, n+1, \ldots, N, \end{cases}$$

and

$$P_i^{(k)} = p_{i+1} P_{i+1}^{(k+1)} + (1 - p_{i+1}) P_{i+1}^{(k)}. \tag{9}$$

The backward induction procedure is illustrated in Figure 2 for the case of a third-to-default swap with five underlying assets (i.e., $n = 3$, $N = 5$). The binomial lattice on the left illustrates the evolution of the Markov chain. The chain moves up at time i if the ith asset defaults and moves down if it does not default. The lattice is initializeds with a value of zero at the three lowest terminal nodes (corresponding to zero, one, and two defaults, respectively) and a value of one at the remaining terminal nodes (corresponding to three, four, or five defaults). The right side of the figure illustrates the recursive calculation of (9) at a representative node. (The binomial lattice in Figure 2 should be distinguished from the type of lattice commonly used in pricing options. In option pricing, movement from left to right through the lattice corresponds to the passage of time, and the right-most nodes record payoffs at expiration. The N steps from left to right in Figure 2 correspond to the N underlying assets and are unrelated to the passage of time in the original swap-pricing problem; the terminal nodes record the number of assets that default and indicate if that number exceeds n.)

The required probabilities q_i are easily derived from absorption probabilities $P_i^{(k)}$ and can therefore be calculated together with these probabilities. Set

$$q_i^{(k)} = p_i P_i^{(k+1)} / P_{i-1}^{(k)}. \tag{10}$$

Then,

$$\begin{aligned} q_i^{(k)} &= \mathbf{P}(Y_i = 1) P(X_N \in G \mid X_i = k+1) / \mathbf{P}(X_N \in G \mid X_{i-1} = k) \\ &= \mathbf{P}(Y_i = 1 \mid X_{i-1} = k) P(X_N \in G \mid X_{i-1} = k,\ Y_i = 1) \\ &\quad / \mathbf{P}(X_N \in G \mid X_{i-1} = k) \\ &= \mathbf{P}(Y_i = 1, X_N \in G \mid X_{i-1} = k) / \mathbf{P}(X_N \in G \mid X_{i-1} = k) \\ &= \mathbf{P}(Y_i = 1 \mid X_{i-1} = k,\ X_N \in G). \end{aligned}$$

Thus, if we observe k defaults among $Y_1, \ldots, Y_{i-1}$, we set $q_i = q_i^{(k)}$; this is the conditional default probability for the ith asset, given that at least n of the N assets will default. These probabilities can be used in place of the original JK probabilities; the rest of the algorithm remains the same.

Figure 2. Calculation of absorption probabilities for a third-to-default swap in a basket of five independent assets.

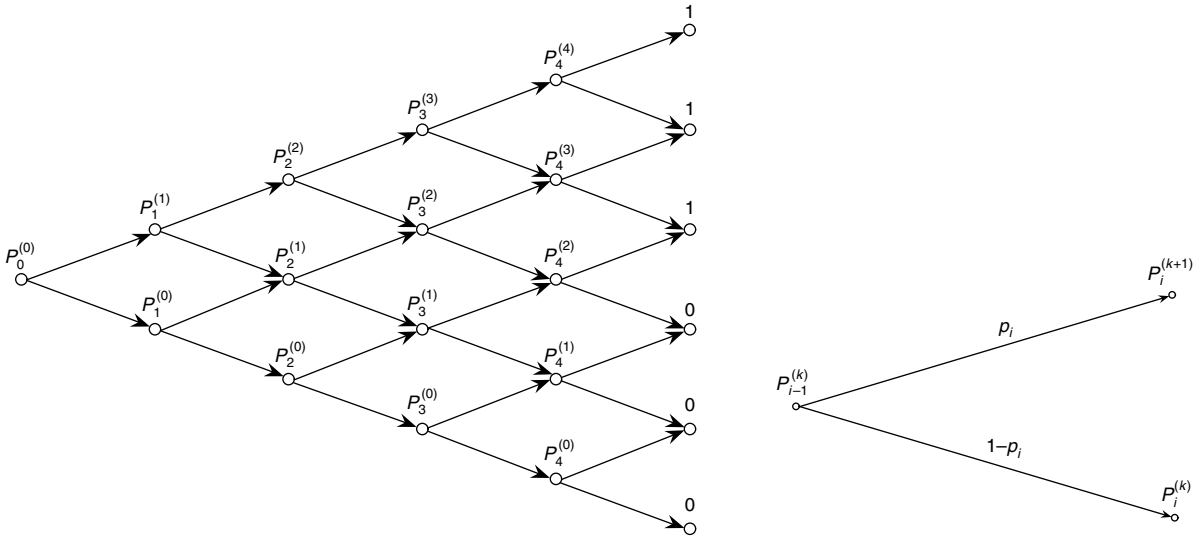


Figure 3. The effect of changing probabilities for a third-to-default swap in a basket of five independent assets.

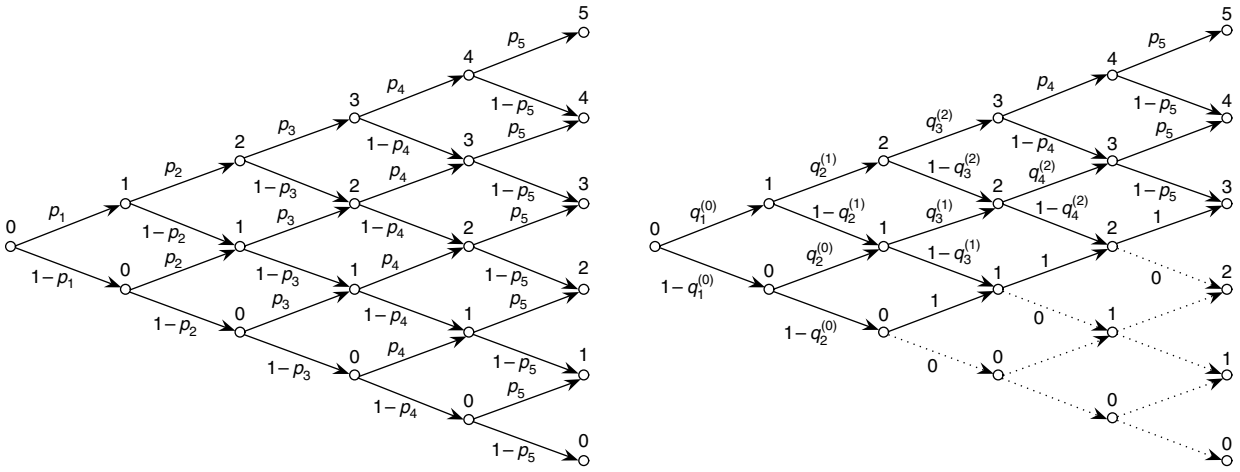


Figure 3 shows how the change of measure forces at least n assets to default in the example of Figure 2. The left panel shows the original transition probabilities and the right panel illustrates the q_i probabilities. The number above each node shows the number of defaults that occur in reaching that node. In changing the default probabilities, we are constraining the Markov chain X to stay within the upper part of the lattice and terminate at a node with at least n defaults.

4.2. Properties of the CP Method

Recall from Theorem 1 (which is sufficiently general to apply to the CP method) that the likelihood ratio is given by $L = L_1 L_2 \cdots L_N$, with L_i given by (6). The following result identifies a special feature of the CP method.

PROPOSITION 4. *In the CP method with independent assets, L is a constant and equals $\mathbf{P}(\tau \leqslant T)$ on every path. Thus, the CP method generates $\tau_1, \ldots, \tau_N$ from their conditional distribution given $\tau \leqslant T$.*

PROOF. It follows from (10) that

$$q_i = p_i P_i^{(X_{i-1}+1)} / P_{i-1}^{(X_{i-1})},$$

so if $Y_i = 1$, then $X_{i-1} + 1 = X_i$ and

$$L_i = p_i / q_i = P_{i-1}^{(X_{i-1})} / P_i^{(X_i)}.$$

Similarly, if $Y_i = 0$, then using (9) and $X_i = X_{i-1}$, we get

$$L_i = (1 - p_i)/(1 - q_i) = P_{i-1}^{(X_{i-1})} / P_i^{(X_i)}.$$

Thus,

$$L = \prod_{i=1}^{N} \frac{P_{i-1}^{(X_{i-1})}}{P_i^{(X_i)}} = \frac{P_0^{(0)}}{P_N^{(X_N)}}.$$

The denominator in this last expression is one, because $X_N \geqslant n$ (by Lemma 2). The numerator is $\mathbf{P}(X_N \in G) = \mathbf{P}(X_N \geqslant n) = \mathbf{P}(\tau \leqslant T)$. The second assertion in the proposition follows from the observation that the likelihood ratio relating the conditional law of the default times to their unconditional law is simply $1/\mathbf{P}(\tau \leqslant T)$. □

Our next result shows that the CP method achieves a guaranteed variance reduction. Moreover, it confirms that the method is particularly effective when the probability of n or more defaults is small.

PROPOSITION 5. *In the case of independent assets, the CP method reduces variance by at least a factor of $1/\mathbf{P}(\tau \leqslant T)$.*

PROOF. By Theorem 1, the estimator in the CP method is unbiased. From Proposition 4, we know that L is identically equal to $\mathbf{P}(\tau \leqslant T)$. Using a tilde to denote expectations computed with respect to the CP importance sampling distribution, we have

$$\begin{aligned}
&\widetilde{\mathbf{Var}}(V(\tau_1, \ldots, \tau_N)L) \\
&\quad = \tilde{\mathbf{E}}(V^2(\tau_1, \ldots, \tau_N)L^2) - (\tilde{\mathbf{E}}(V(\tau_1, \ldots, \tau_N)L))^2 \\
&\quad = L\tilde{\mathbf{E}}(V^2(\tau_1, \ldots, \tau_N)L) - (\mathbf{E}(V(\tau_1, \ldots, \tau_N)))^2 \\
&\quad \leqslant L\mathbf{E}(V^2(\tau_1, \ldots, \tau_N)) - L(\mathbf{E}(V(\tau_1, \ldots, \tau_N)))^2 \\
&\quad = L\mathbf{Var}(V(\tau_1, \ldots, \tau_N)),
\end{aligned}$$

so

$$\frac{\mathbf{Var}(V(\tau_1, \ldots, \tau_N))}{\widetilde{\mathbf{Var}}(V(\tau_1, \ldots, \tau_N)L)} \geqslant \frac{1}{L} = \frac{1}{\mathbf{P}(\tau \leqslant T)}. \quad \square$$

We noted previously that the effectiveness of the JK method depends, in part, on the order in which the default times are generated. In contrast, the CP method is insensitive to the ordering of the assets:

PROPOSITION 6. *In the CP method, the variance reduction achieved does not depend on the order of assets.*

This follows directly from the fact that the CP sampling distribution is always the conditional distribution of $\tau_1, \ldots, \tau_N$ given $\tau \leqslant T$, regardless of the order in which the default times are generated.

5. The CP Method with Dependent Assets

We now return to the case in which the default times are linked through the Gaussian copula model. In this setting, sampling $\tau_1, \ldots, \tau_N$ conditional on $\tau \leqslant T$ does not seem feasible; the algorithm for calculating q_i in §4 does not extend in an obvious way to the case of dependent default times. To apply the CP method, we therefore condition on additional information to make the default times conditionally independent.

5.1. Factor Models

Recall that the dependence in the Gaussian copula model is determined through the correlation matrix $\mathbf{\Sigma}$ of the underlying normal random variables $W_1, \ldots, W_N$. In practice, these correlations are often specified through a factor model of the form

$$W_i = a_{i1}Z_1 + \cdots + a_{id}Z_d + b_i\epsilon_i, \quad i = 1, \ldots, N,$$

or, in matrix-vector form,

$$\mathbf{W} = \mathbf{A}\mathbf{Z} + \mathbf{B}\boldsymbol{\epsilon}, \tag{11}$$

in which

- $Z_1, \ldots, Z_d$, $d < N$, are systematic risk factors, normalized to be independent standard normals;
- ϵ_i, $i = 1, \ldots, N$, are idiosyncratic risks associated with the ith individuals assets, also $N(0, 1)$ distributed and independent of each other and of $Z_1, \ldots, Z_d$;
- $a_{i1}, \ldots, a_{id}$ are the factor loadings for the ith asset, and $\sum_{k=1}^{d} a_{ik}^2 \leqslant 1$;
- $b_i = \sqrt{1 - \sum_{k=1}^{d} a_{ik}^2}$, so that W_i is also $N(0, 1)$ distributed.

The correlation matrix of $\mathbf{W}$ implied by (11) is $\mathbf{A}\mathbf{A}^\top + \mathbf{B}^2$; so, for consistency, we require that this matrix equal $\mathbf{\Sigma}$.

The key property of (11) is that once we condition on $\mathbf{Z}$, the W_i become conditionally independent of each other. This, in turn, makes the default times $\tau_1, \ldots, \tau_N$ conditionally independent of each other, thus allowing us to apply the CP method.

5.2. The Conditional CP Method

Suppose that $\tau_i = F_i^{-1}(\Phi(W_i))$, $i = 1, \ldots, N$, and that the W_i admit the representation in (11). Conditioning on $\mathbf{Z}$ yields

$$\begin{aligned} F_i^Z(t) &\equiv \mathbf{P}(\tau_i \leqslant t \mid \mathbf{Z}) \\ &= \mathbf{P}(F_i^{-1}(\Phi(W_i)) \leqslant t \mid \mathbf{Z}) \\ &= \mathbf{P}\left(b_i\epsilon_i \leqslant \Phi^{-1}(F_i(t)) - \sum_{j=1}^{d} a_{ij}Z_j \right) \\ &= \Phi\left(\left[\Phi^{-1}(F_i(t)) - \sum_{j=1}^{d} a_{ij}Z_j \right] \Big/ b_i \right). \end{aligned} \tag{12}$$

Conditional on $\mathbf{Z}$, we can therefore apply the CP method as in §4, but with F_i replaced by F_i^Z.

More explicitly, the steps are as follows, for each replication:

Step 1. Generate d independent standard normals $Z_1, \ldots, Z_d$.

Step 2. Compute the conditional default probability $p_i^Z = F_i^Z(T)$, for each $i = 1, \ldots, N$.

Step 3. Construct the conditional probabilities q_i^Z using the algorithm of §4.1.

Step 4. Generate $\tau_1, \ldots, \tau_N$ and $L_1, \ldots, L_N$ by following (4)–(6), using p_i^Z for p_i, q_i^Z for $\tilde{p}_i$, and F_i^Z for F_i.

Step 5. Evaluate and return $V(\tau_1, \ldots, \tau_N)L$, with $L = L_1L_2 \cdots L_N$.

For Step 4, we can calculate U_i as in (4) and then set

$$\tau_i = F_i^{-1}(\Phi(W_i)), \qquad W_i = a_{i1}Z_1 + \cdots + a_{id}Z_d + b_i\Phi^{-1}(U_i).$$

Although we do not have a lower bound on the magnitude of variance reduction as we did in the case of independent assets, we nevertheless have the following guarantee:

PROPOSITION 7. *The conditional CP method always reduces variance.*

PROOF. Using the conditional independence of the default times given $Z_1, \ldots, Z_d$, Proposition 4 yields $L = \mathbf{P}(\tau \leqslant T \mid \mathbf{Z})$. Thus, $L \leqslant 1$ and

$$\begin{aligned} &\widetilde{\mathbf{Var}}(V(\tau_1, \ldots, \tau_N)L) \\ &\quad = \tilde{\mathbf{E}}(V^2(\tau_1, \ldots, \tau_N)L^2) - (\tilde{\mathbf{E}}(V(\tau_1, \ldots, \tau_N)L))^2 \\ &\quad = \mathbf{E}(\tilde{\mathbf{E}}(V^2(\tau_1, \ldots, \tau_N)L^2 \mid \mathbf{Z})) - \mathbf{E}(V(\tau_1, \ldots, \tau_N))^2 \\ &\quad = \mathbf{E}(\mathbf{E}(V^2(\tau_1, \ldots, \tau_N)L \mid \mathbf{Z})) - \mathbf{E}(V(\tau_1, \ldots, \tau_N))^2 \\ &\quad \leqslant \mathbf{E}(\mathbf{E}(V^2(\tau_1, \ldots, \tau_N) \mid \mathbf{Z})) - \mathbf{E}(V(\tau_1, \ldots, \tau_N))^2 \\ &\quad = \mathbf{Var}(V(\tau_1, \ldots, \tau_N)). \end{aligned}$$

The second equality uses the fact that $Z_1, \ldots, Z_d$ have the same distribution under the original and CP sampling procedures. □

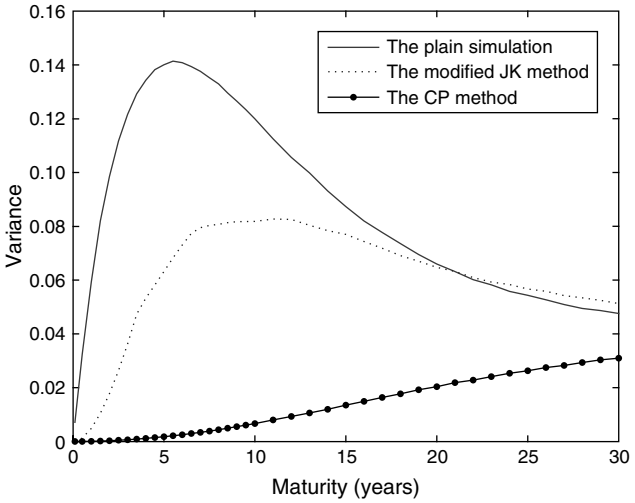
5.3. Numerical Examples

We compare the CP method, the modified JK method, the JK method, and plain Monte Carlo simulation through numerical results for the test cases Swaps A1 and A2. For comparison, we include one more test case—Swap A3, which is very similar to Swap A2, but with different signs in factor loadings. All examples use 10^6 replications.

Basket III—Swap A3

Swap A3 is a fifth-to-default swap in Basket III, where Basket III contains $N = 10$ assets with the same hazard rates and recovery rates as in Basket II. The assets are also correlated through a four-factor Gaussian copula. The

Figure 4. The variance of the estimated price of Swap A1.



only difference between Baskets II and III is the factor-loading matrix **A**. The loadings in Basket III have the same magnitudes as those in Basket II but are all positive.

Figures 4 and 5 plot the variance for each method against the maturity of the swap. The solid line shows the variance in the plain Monte Carlo method, the dashed line shows the variance in the JK method, the dotted line shows the variance in the modified JK method, and the solid line with dots shows the variance in the CP method. As expected, the CP reduces variance in all examples at all maturities. In Figure 4 (which shows an example with independent assets), the CP method works particularly well at short horizons, where the probability of default within the life of the swap is small.

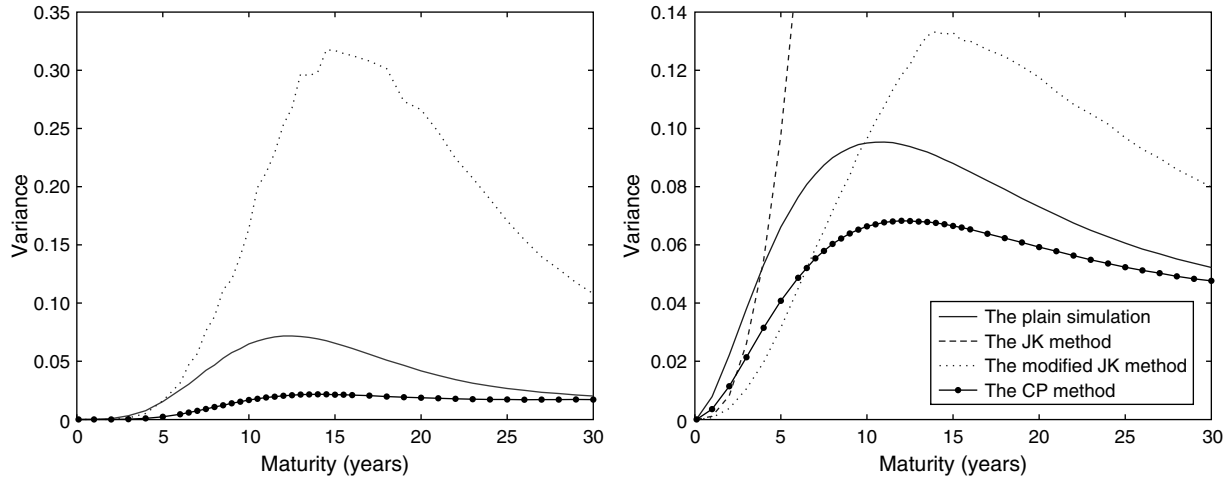
The examples illustrated in Figure 5 are four-factor models, and the CP method continues to work well in these examples. The variance reduction achieved by the CP method is greater in the left panel (Swap A2) than in the right panel (Swap A3). Recall that the only difference between these examples is that Swap A2 has loadings of mixed sign, whereas all the loadings in Swap A3 are positive. As a consequence, an unusually large or small value of one of the factors $Z_1, \ldots, Z_d$ will move all the default times in Swap A3 in the same direction, whereas in Swap A2 the effects on different assets will tend to offset each other. In this respect, Swap A3 is more sensitive to variability in the underlying factors than Swap A2 is; recall that we apply the conditional CP method conditional on the outcomes of the factors.

In the right panel of Figure 5, the JK and modified JK methods actually produce slightly lower variance than the CP method at short horizons. This, too, results from the fact that the conditional CP method applies conditional on the factor outcomes. In contrast, the JK method relies on conditional probabilities calculated through Cholesky decomposition, as in Lemma 1. This can be advantageous when the variability from the factors is large (as in Swap A3). However, because of the conditional probability calculations, the JK method takes about 4–5 times longer than the CP method in Swap A3, so the overall efficiency improvement is greater using the CP method across all maturities. In the next section, we will see that the factor structure can be used to further improve the performance of the CP method.

6. Stratifying the Factors

The CP method, as implemented in §5, reduces variance conditional on the outcome of the factors $Z_1, \ldots, Z_d$, but it does nothing to reduce the variability in the sampling of the factors themselves. This observation leads us to apply stratified sampling to the factors. We refer to this combination as the *CPST method*. We introduce this method in a single-factor setting, then extend it to multiple factors.

Figure 5. The variance of the estimated price of Swap A2 (left panel) and Swap A3 (right panel).



6.1. Single-Factor Models

A single-factor model takes the form

$$W_i = a_i Z + \sqrt{1-a_i^2}\epsilon_i, \quad i = 1, \ldots, N.$$

We apply stratified sampling to the factor Z, which has a standard normal distribution.

To this end, we partition the real line into subintervals $A_1, \ldots, A_K$ of equal probability $1/K$; the interval A_k has endpoints $\Phi^{-1}((k-1)/K)$ and $\Phi^{-1}(k/K)$. To generate one sample of Z in stratum A_k, we generate $\hat{U}$ uniformly over $((k-1)/K, k/K]$, and then set $Z = \Phi^{-1}(\hat{U})$.

The CPST method entails the following steps, for each replication:

Step 1. Generate $U_{(1)}, \ldots, U_{(K)}$ independently and uniformly over $(0, 1)$ and set $\hat{U}_{(k)} = (k - 1 + U_{(k)})/K$. Then, $Z_{(k)} = \Phi^{-1}(\hat{U}_{(k)})$ has the distribution of a standard normal random variable conditioned to fall in A_k.

Step 2. For each $Z_{(k)}$, apply the CP method to obtain a sample $V_{(k)}$ of $V(\tau_1, \ldots, \tau_N)$ and a sample $L_{(k)}$ of the likelihood ratio.

Step 3. Return $\sum_{k=1}^{K} V_{(k)} L_{(k)}/K$ as one sample of the estimator $V(\tau_1, \ldots, \tau_N)L$.

This procedure can be repeated over multiple independent replications to estimate a standard error along with the swap value.

6.2. Multifactor Models

In principle, the same idea can be applied to each of the underlying factors in a multifactor model. However, full stratified sampling becomes impractical when the number of factors d is large. If, for example, we use K_i strata in generating Z_i, $i = 1, \ldots, d$, then the total number of strata is $K_1 \cdots K_d$. Generating at least one point from each stratum therefore requires a sample size at least this large. Unless the K_i are quite small (in which case stratification will provide little benefit), this is likely to be prohibitive for d larger than three.

As an alternative to full stratification in d dimensions, we seek to identify a small number (one or two) of "important" directions along which to stratify. More specifically, we choose an $N \times \tilde{d}$ ($\tilde{d} \leqslant 2$) matrix $\tilde{\mathbf{A}}$ and write

$$\mathbf{AZ} = \tilde{\mathbf{A}}\tilde{\mathbf{Z}} + \hat{\mathbf{A}}\hat{\mathbf{Z}},$$

where $\hat{\mathbf{A}}$ is an $N \times \hat{d}$ ($\hat{d} \leqslant d$) matrix, and $\tilde{\mathbf{Z}}$ ($\tilde{d} \times 1$) and $\hat{\mathbf{Z}}$ ($\hat{d} \times 1$) are independent multivariate normals. The matrix $\hat{\mathbf{A}}$ is constrained to satisfy

$$\mathbf{\Sigma}_Z \equiv \mathbf{A}\mathbf{A}^\top = \tilde{\mathbf{A}}\tilde{\mathbf{A}}^\top + \hat{\mathbf{A}}\hat{\mathbf{A}}^\top.$$

Our objective is to choose $\tilde{\mathbf{A}}$ so that $\tilde{\mathbf{A}}\tilde{\mathbf{Z}}$ explains most of the variability of $\mathbf{AZ}$. We then apply stratified sampling to $\tilde{\mathbf{Z}}$ to eliminate most of this variability.

A convenient way to choose $\tilde{\mathbf{A}}$ is through principal components analysis. Suppose that $\mathbf{\Sigma}_Z = \mathbf{A}\mathbf{A}^\top$ has full rank d; the singular case can be handled similarly by first partitioning the vector $\mathbf{Z}$. Then, $\mathbf{\Sigma}_Z$ has d positive real eigenvalues $\lambda_1 \geqslant \cdots \geqslant \lambda_d > 0$ and an associated set of orthonormal eigenvectors $\{\nu_1, \ldots, \nu_d\}$; i.e., vectors satisfying $\mathbf{\Sigma}_Z \nu_i = \lambda_i \nu_i$ and

$$\nu_i^T \nu_i = 1, \qquad \nu_i^T \nu_j = 0, \quad j \neq i,\ i, j = 1, \ldots, d.$$

It follows that $\mathbf{\Sigma}_Z = \mathscr{V}\Lambda\mathscr{V}^T$, where $\mathscr{V}$ is the orthogonal matrix with columns $\nu_1, \ldots, \nu_d$ and Λ is the diagonal matrix with the diagonal entries $\lambda_1, \ldots, \lambda_d$. Let $\Lambda^{1/2}$ denote the diagonal matrix with the diagonal entries $\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_d}$. Then,

$$\mathbf{\Sigma}_Z = (\mathscr{V}\Lambda^{1/2})(\mathscr{V}\Lambda^{1/2})^\top.$$

Methods for calculating $\mathscr{V}$ and Λ are included in many mathematical software libraries and discussed in detail in, e.g., Golub and Van Loan (1996). As discussed in Glasserman (2004), a good choice for $\tilde{\mathbf{A}}$ is the matrix with columns $\sqrt{\lambda_1}\nu_1, \ldots, \sqrt{\lambda_{\tilde{d}}}\nu_{\tilde{d}}$. The matrix $\hat{\mathbf{A}}$ then has columns $\sqrt{\lambda_{\tilde{d}+1}}\nu_{\tilde{d}+1}, \ldots, \sqrt{\lambda_d}\nu_d$.

We combine the various steps as follows:

Step 1. Compute the eigenvectors and eigenvalues $\mathscr{V}$ and Λ of $\mathbf{\Sigma}_Z$, and find $\tilde{\mathbf{A}}$ and $\hat{\mathbf{A}}$ with $\tilde{d} \leqslant 2$.

Step 2. Stratify the unit hypercube $[0, 1)^{\tilde{d}}$ by partitioning the jth coordinate into K_j intervals to form strata $A_1, \ldots, A_K$, with each A_k of the form $\prod_{j=1}^{\tilde{d}} [i_j - 1, i_j)/K_j$, $k = 1, \ldots, K$, with $K = K_1 \cdots K_{\tilde{d}}$.

Step 3. In stratum A_k, generate a $\tilde{d}$-dimensional uniform $\mathbf{U}_{(k)} = (U_{(k),1}, \ldots, U_{(k),\tilde{d}})$ by setting the jth coordinate of $\hat{\mathbf{U}}_{(k)}$ to be $\hat{U}_{(k),j} = (i_j - 1 + U_{(k),j})/K$. Then, $\tilde{\mathbf{Z}}_{(k)} = \Phi^{-1}(\hat{\mathbf{U}}_{(k)})$ (applied coordinatewise) is a sample of the $\tilde{d}$-dimensional standard normal conditional on $\tilde{\mathbf{Z}} \in \mathscr{A}_k$.

Step 4. Generate a $(d - \tilde{d})$-dimensional normal $\hat{\mathbf{Z}}_{(k)}$, and set $\mathbf{AZ}_{(k)} = \tilde{\mathbf{A}}\tilde{\mathbf{Z}}_{(k)} + \hat{\mathbf{A}}\hat{\mathbf{Z}}_{(k)}$.

Step 5. For each $\mathbf{AZ}_{(k)}$, apply the conditional CP method, and compute the discounted payoff $V_{(k)}$ and likelihood ratio $L_{(k)}$.

Step 6. Take $\sum_{k=1}^{K} V_{(k)} L_{(k)}/K$ as one sample of the estimator $V(\tau_1, \ldots, \tau_N)L$.

As in the single-factor case, these steps can be repeated over multiple independent replications to estimate a standard error. For this method, we have the following observation:

PROPOSITION 8. *The CPST method has lower variance than the CP method and thus has lower variance than ordinary Monte Carlo.*

This follows directly from the fact that stratified sampling with proportional allocation (here meaning that we draw equal numbers of samples from the equiprobable strata) always reduces variance (see, e.g., Glasserman 2004, pp. 215–217).

6.3. Factor-Model Approximation

Although the correlations in the Gaussian copula model are often represented through a factor model of the form (11), this is not always the case. We therefore consider an alternative to the principal components method of the previous section that starts directly from the correlation matrix $\mathbf{\Sigma}$ of $\mathbf{W}$ and approximates it with a low-dimensional factor model. In other words, we seek an approximation of the form

$$\mathbf{W} \approx \tilde{\mathbf{A}}\mathbf{Z} + \tilde{\mathbf{B}}\boldsymbol{\epsilon},$$

with $\mathbf{Z}$ and $\boldsymbol{\epsilon}$ independent standard normal vectors and $\tilde{\mathbf{A}}$ having $\tilde{d}$ columns. Given such a representation with small $\tilde{d}$, we can apply stratified sampling to $\mathbf{Z}$ and combine that with the conditional CP method.

Without a constraint on $\tilde{d}$, such a representation is always possible (e.g., set $\tilde{\mathbf{A}}$ equal to the Cholesky matrix $\mathbf{C}$ and set $\mathbf{B} = 0$), but if we put an upper bound on $\tilde{d}$, this may introduce some approximation error. Thus, unlike the other methods considered in this article, this method involves some bias.

The problem of finding a low-dimensional factor representation is considered in Andersen et al. (2003). They suggest choosing the factor-loading matrix $\tilde{\mathbf{A}}$ $(N \times \tilde{d})$ as the solution to the minimization problem

$$\min_{\tilde{\mathbf{A}}} \operatorname{tr}(\mathbf{\Sigma} - \tilde{\mathbf{A}}\tilde{\mathbf{A}}^\top - \tilde{\mathbf{B}})(\mathbf{\Sigma} - \tilde{\mathbf{A}}\tilde{\mathbf{A}}^\top - \tilde{\mathbf{B}})^\top,$$

where tr is the usual matrix trace operator and $\tilde{\mathbf{B}}$ is a diagonal matrix with

$$(\tilde{\mathbf{B}})_{jj} = 1 - (\tilde{\mathbf{A}}\tilde{\mathbf{A}}^\top)_{jj}, \quad j = 1, \ldots, d. \tag{13}$$

Andersen et al. (2003) propose a recursive algorithm to search for the optimal $\tilde{\mathbf{A}}$; in our tests as well as theirs, the algorithm appears to converge very quickly.

6.4. Numerical Examples

We showed in Propositions 5 and 7 that the CP method guarantees a variance reduction and outperforms plain Monte Carlo. In this section, we examine the additional improvement obtained by applying the CPST method compared with the CP method and the modified JK method. We exclude the JK method from the figures because in all our numerical examples, we have observed that the modified JK method outperforms the JK method. We use 100 strata for one-dimensional stratification and 40 strata in each dimension for two-dimensional stratification. (We have observed quite similar results for one-dimensional stratification using just 10 strata.) All examples use a total of 10^6 samples. In the case of one-dimensional stratification, this means 10^4 replications, with each replication consisting of the average over 100 strata. In the case of two-dimensional stratification, this means 625 replications, each the average over 40^2 strata. This results in an equal number of samples for all methods compared, and thus makes the variance comparisons meaningful. All examples in this section use dependent underlying assets because this is the setting in which CPST is relevant.

Basket IV—Swap A4 and Basket V—Swap A5

Swap A4 is a first-to-default swap in Basket IV, where Basket IV contains $N = 10$ assets with the same hazard rates and recovery rates as in Basket I. The assets are correlated and the correlation matrix is $\mathbf{\Sigma}$, which is introduced by a multivariate normal vector $\{W_1, \ldots, W_N\}$ in the form of a single-factor model. To generate such a correlation matrix, one can first randomly generate a 10×1 matrix $\mathbf{A}$, compute the supplementary vector $\mathbf{B}$, then let $\mathbf{\Sigma} = \mathbf{A}\mathbf{A}^\top + \mathbf{B}\mathbf{B}^\top$. The correlation matrix used here can be found in the appendix. Basket V is the same as Basket IV except that the signs in $\mathbf{A}$ are all positive. Swap A5 is a first-to-default swap in Basket V. We apply stratified sampling to that single factor.

Figure 6 shows the variance of the estimators given by plain Monte Carlo, the modified JK method, the CP

Figure 6. The variance of the estimated price of Swap A4 (left panel) and Swap A5 (right panel).

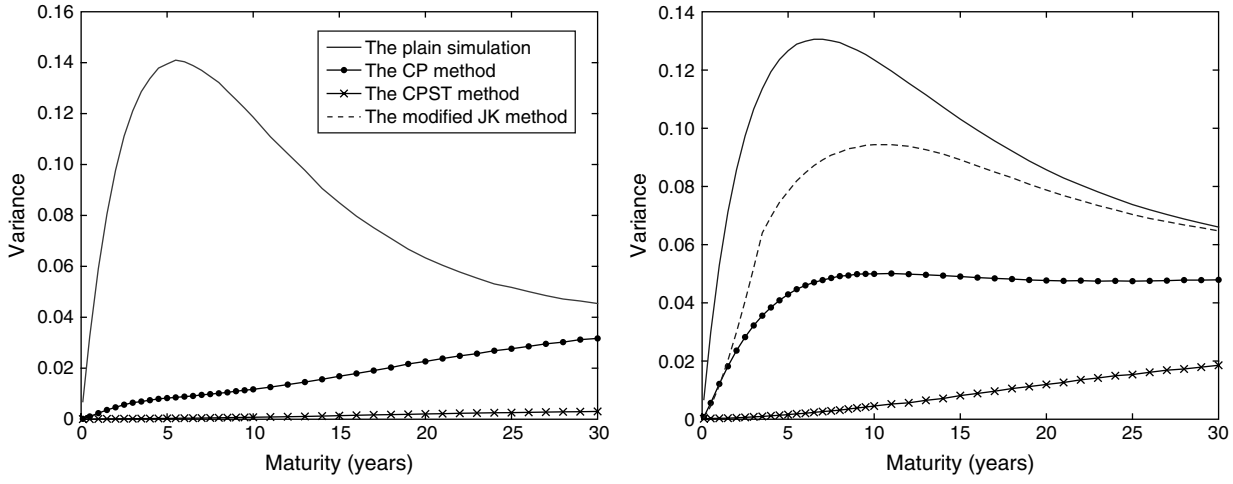
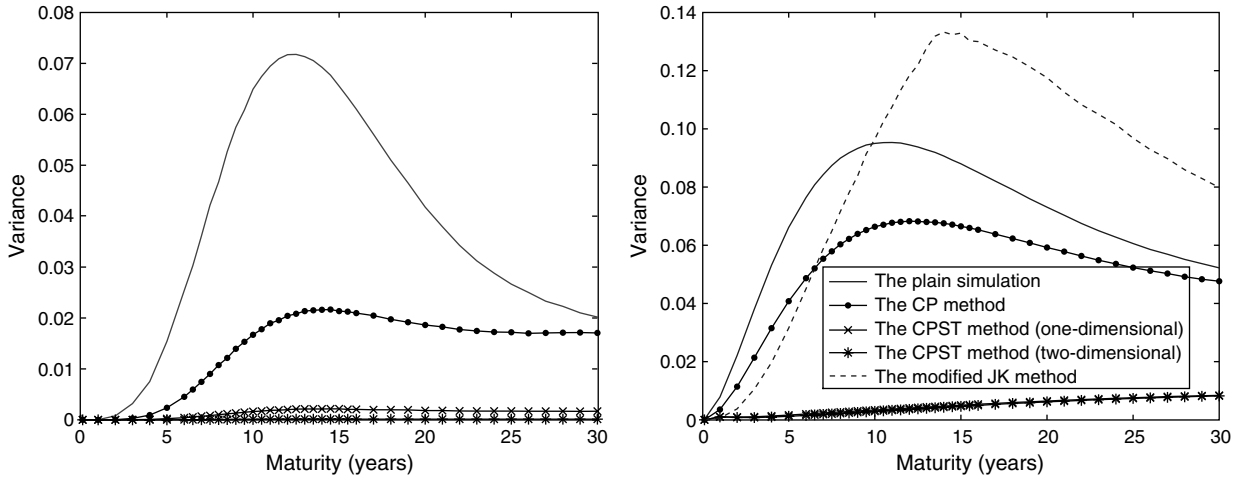


Figure 7. The variance of the estimated price of Swap A2 (left panel) and Swap A3 (right panel).



method, and the CPST method with one-dimensional stratified sampling applied to the single-factor models Swaps A4 and A5. Figure 7 shows corresponding results applying one- and two-dimensional stratification to the four-factor models Swaps A2 and A3. In all cases, the figures show substantial variance reduction from the combination of the CP method with stratified sampling.

The two panels in Figures 6 and 7 should be contrasted with those in Figure 5. In Figure 5, we saw that the conditional CP method was most effective when the factor loadings had varying signs; we suggested that this resulted from the offsetting effect on the default times, which tends to reduce the variability contributed by the factors. In both Figures 6 and 7, we see that stratification has a greater effect when the factor loadings all have the same sign (the right panel in each case). Thus, by removing variability from the factors, stratification contributes the greatest variance reduction in those cases in which the conditional CP method by itself is less effective.

Observe also in Figure 7 that in the case of mixed factor loadings (Swap A2), one-dimensional stratification has little benefit, and there is an evident difference between two-dimensional and one-dimensional stratification. In the case of positive loadings (Swap A3), one-dimensional stratification is as effective as two-dimensional stratification, and both improve markedly over the CP method without stratification. In all cases, the CPST methods outperform the modified JK method.

In gauging the effectiveness of the CP method (without or with stratified sampling), we need to consider the computing time required as well as the variance reduction achieved. In the independent assets case (Swap A1), the conditional probabilities q need only be calculated once and the time required for this calculation is negligible. Given the new probabilities, each replication of the CP method takes about 1–1.5 times longer than the plain Monte Carlo method. In the dependent case, the conditional probabilities need to be recalculated on each replication. With this additional work, each replication of the CP method takes about 4–5 times longer than the plain Monte Carlo method. The CP method with one-dimensional stratification takes about 5–6 times longer than the plain Monte Carlo method, and with two-dimensional stratification the time required is about 6–7 times longer. These comparisons are for the same total number of paths.

The additional computing time is offset by the variance reduction achieved except when defaults are very frequent. In the independent asset case (Swap A1), the CP method reduces variance by at least a factor of two for $T \leqslant 25$. In the dependent assets cases, the conditional CP method by itself improves efficiency in the examples with mixed factor loadings (Swap A2, $T \leqslant 6.5$, and Swap A4, $T \leqslant 14$), but not in the examples with all loadings positive (Swaps A3 and A5). With one-dimensional stratification, CPST improves efficiency for $T \leqslant 5$ in Swap A2, $T \leqslant 27$ for Swap A3, $T \leqslant 15$ for Swap A4, and $T \leqslant 22$ for Swap A5. Two-dimensional stratification extends the range for Swap A2 to $T \leqslant 8.5$ and has little effect on Swap A3, as we saw in Figure 7.

7. Concluding Remarks

In this paper, we have developed a variance reduction technique for Monte Carlo pricing of basket default swaps by analyzing and building on the method of Joshi and Kainth (2004). As in their original method, we sequentially increase the probability of default within the life of the swap as we generate the default times of the assets. However, in contrast to the original method, our choice of probabilities guarantees variance reduction and is, in a certain sense, optimal. We achieve further variance reduction through the strategic use of stratified sampling.

Although we have stressed the application of this method in the widely used Gaussian copula, it extends easily to other models' dependence, including multifactor versions of the various factor copula models in Laurent and

Gregory (2003). Also, eventhough we have focused on basket default swaps, similar ideas can be used for other portfolio credit derivatives; this is a topic of current investigation. Finally, Joshi and Kainth (2004) consider the calculation of price sensitivities as well as the prices themselves; a further analysis of this problem is another topic of current investigation.

Appendix. Proofs

A.1. Proof of Lemma 2

PROOF. The sequences $n - \sum_{j=1}^{i-1} Y_j$ and $N - i + 1$, $i = 1, \ldots, N+1$, are both decreasing. The first of these sequences starts at n; the second starts at $N > n$ and decreases to zero. If the two sequences are never equal, then the first one must reach zero first, at which point n defaults have occurred. If the sequences become equal at some i, then $\tilde{p}_i = 1$, so $Y_i = 1$ and they are equal again at $i+1$. This continues until $i = N+1$, at which point the sum of the default indicators is again n. Either way, the nth default occurs before T. □

A.2. Proof of Lemma 3

PROOF. Suppose that $Y_i = 1$. Then, the conditional distribution of τ_i under $\mathbf{P}$ is its conditional distribution given $\tau_1, \ldots, \tau_{i-1}$ and given $\tau_i \leqslant T$. This is the conditional distribution of $F_i^{-1}(U_i \mid \tau_1, \ldots, \tau_{i-1})$ given $U_i \leqslant p_i$, $p_i = \mathbf{P}(\tau_i \leqslant T \mid \tau_1, \ldots, \tau_{i-1})$, where U_i is uniform on $(0,1)$ and independent of $\tau_1, \ldots, \tau_{i-1}$. In other words, the conditional distribution of τ_i is the conditional distribution of $F_i^{-1}(U_i \mid \tau_1, \ldots, \tau_{i-1})$ given that U_i is uniformly distributed between 0 and p_i.

Under $\tilde{\mathbf{P}}$, we see from (4) that the conditional distribution of τ_i given $\tau_1, \ldots, \tau_{i-1}$ and given $Y_i = 1$ is the conditional distribution of $F_i^{-1}(p_i V_i/\tilde{p}_i \mid \tau_1, \ldots, \tau_{i-1})$, given that V_i is uniform between 0 and $\tilde{p}_i$. This is the same as the conditional distribution of $F_i^{-1}(U_i \mid \tau_1, \ldots, \tau_{i-1})$ given that U_i is uniformly distributed between 0 and p_i.

The argument for the case $Y_i = 0$ works similarly, except that now the conditional distribution of τ_i under both probability measures becomes the conditional distribution of $F_i^{-1}(U_i \mid \tau_1, \ldots, \tau_{i-1})$ given that U_i is uniformly distributed between p_i and 1. □

A.3. Proof of Theorem 1

PROOF. The second assertion is a consequence of the first. To establish (i), we calculate the likelihood ratio directly. In light of Lemma 2, we restrict attention to the event $\{\tau \leqslant T\}$. It is evident from (4) and the surrounding steps that, within this event, $\mathbf{P}$ and $\tilde{\mathbf{P}}$ have the same support, in the sense that any sequence of default times that could be generated by one sampling mechanism could also be generated by the other. Therefore, consider any nonnegative $t_1, \ldots, t_N$ with at least n of the t_i falling within $[0, T]$, and let $y_i = I(t_i \leqslant T)$, $i = 1, \ldots, N$. We can write the joint density or mass function of the default times under $\mathbf{P}$ as

$$\begin{aligned}&\mathbf{P}(Y_1 = y_1,\ \tau_1 \in dt_1, \ldots, Y_N = y_N,\ \tau_N \in dt_N)\\ &\quad = \mathbf{P}(Y_1 = y_1)\mathbf{P}(\tau_1 \in dt_1 \mid Y_1 = y_1)\mathbf{P}(Y_2 = y_2 \mid \tau_1 = t_1)\\ &\qquad \cdots \mathbf{P}(\tau_N \in dt_N \mid \tau_1 = t_1, \ldots, \tau_{N-1} = t_{N-1}, Y_N = y_N).\end{aligned}$$

Lemma 3 yields

$$\begin{aligned}&\tilde{\mathbf{P}}(Y_1 = y_1,\ \tau_1 \in dt_1, \ldots, Y_N = y_N,\ \tau_N \in dt_N)\\ &\quad = \tilde{\mathbf{P}}(Y_1 = y_1)\mathbf{P}(\tau_1 \in dt_1 \mid Y_1 = y_1)\tilde{\mathbf{P}}(Y_2 = y_2 \mid \tau_1 = t_1)\\ &\qquad \cdots \mathbf{P}(\tau_N \in dt_N \mid \tau_1 = t_1, \ldots, \tau_{N-1} = t_{N-1}, Y_N = y_N).\end{aligned}$$

The likelihood ratio is therefore given by

$$\prod_{i=1}^{N} \frac{\mathbf{P}(Y_i = y_i \mid \tau_1 = t_1, \ldots, \tau_{i-1} = t_{i-1})}{\tilde{\mathbf{P}}(Y_i = y_i \mid \tau_1 = t_1, \ldots, \tau_{i-1} = t_{i-1})}.$$

The ith factor in this product equals $p_i/\tilde{p}_i$ if $y_i = 1$ and it equals $(1-p_i)/(1-\tilde{p}_i)$ if $y_i = 0$. Thus, the ith factor is L_i and the likelihood ratio is indeed L. □

One could arrive at the same expression for the likelihood ratio by viewing the JK algorithm as applying a change of distribution to $U_1, \ldots, U_N$. Without importance sampling, these are independently and uniformly distributed over $(0,1)$. The JK algorithm makes U_i uniform over $(0, p_i)$ with probability $\tilde{p}_i$ and uniform over $(p_i, 1-p_i)$ with probability $1-\tilde{p}_i$. The advantage of the argument in Theorem 1 is that it follows from the key properties in (7) and Lemma 3, and does not rely on the mechanism used to implement (7); in particular, it does not rely on the use of the inverse transform method (in (5)) to generate the default times.

A.4. Proof of Proposition 1

PROOF. By Theorem 1, the JK method is unbiased, regardless of the order in which default times are generated. It therefore suffices to compare second moments in comparing the variances under different orderings.

Suppose that the assets are ordered in such a way that there are two consecutive assets, say the kth and $(k+1)$st, with default probabilities $p_k > p_{k+1}$. We will show that the second moment of the JK estimator will decrease after switching the order of the kth and $(k+1)$st assets.

Define events $\mathcal{A}_1 = \{\sum_{i=1}^{k-1} I(\tau_i \leqslant T) = n-1\}$, $\mathcal{A}_2 = \{\sum_{i=1}^{k-1} I(\tau_i \leqslant T) < n-1\}$, and $\mathcal{A}_3 = \{\sum_{i=1}^{k-1} I(\tau_i \leqslant T) \geqslant n\}$. Consider any realization of $\tau_1, \ldots, \tau_{k-1}$ and set $L_1^{k-1} = \prod_{i=1}^{k-1} L_i$, where L_i is the importance-sampling weight associated with asset i. The second moment is

$$\mathbf{E}(V^2(\tau_1, \ldots, \tau_N)L) = \sum_{i=1}^{3} \mathbf{E}(V^2(\tau_1, \ldots, \tau_N)L \mid \mathcal{A}_i)\mathbf{P}(\mathcal{A}_i).$$

Given $\mathscr{A}_3$, the importance sampling must have been suspended before the generation of the kth asset's default time. Therefore, switching the positions of the kth and $(k+1)$st assets does not affect the value of L, which is equal to L_1^{k-1}. Therefore, the second moment conditional on $\mathscr{A}_3$ remains the same.

Given $\mathscr{A}_1$, the second moment of the estimator is

$$\begin{aligned}
&\mathbf{E}(V^2(\tau_1,\dots,\tau_N)L \mid \mathscr{A}_1)\\
&\quad = L_1^{k-1}\frac{p_k}{\tilde{p}_k}\mathbf{E}(V^2(\tau_1,\dots,\tau_N)I(\tau_k \leqslant T) \mid \mathscr{A}_1)\\
&\qquad + L_1^{k-1}\frac{1-p_k}{1-\tilde{p}_k}\frac{p_{k+1}}{\tilde{p}_{k+1}}\mathbf{E}(V^2(\tau_1,\dots,\tau_N)\\
&\qquad\qquad \cdot I(\tau_k > T,\ \tau_{k+1} \leqslant T) \mid \mathscr{A}_1)\\
&\qquad + L_1^{k-1}\prod_{i=k}^{k+1}\frac{1-p_i}{1-\tilde{p}_i}\mathbf{E}\big(V^2(\tau_1,\dots,\tau_N)\\
&\qquad\qquad \cdot \prod_{j=k+2}^{N} L_j I(\tau_k > T,\ \tau_{k+1} > T) \mid \mathscr{A}_1\big), \qquad (14)
\end{aligned}$$

where $\tilde{p}_i = 1/(N-i+1)$ for $i = k, k+1$.

After switching the positions of the kth and $(k+1)$st assets, the second moment becomes

$$\begin{aligned}
&\mathbf{E}(V^2(\tau_1,\dots,\tau_N)L \mid \mathscr{A}_1)\\
&\quad = L_1^{k-1}\frac{p_{k+1}}{\tilde{p}_k}\mathbf{E}(V^2(\tau_1,\dots,\tau_N)I(\tau_{k+1} \leqslant T) \mid \mathscr{A}_1)\\
&\qquad + L_1^{k-1}\frac{1-p_{k+1}}{1-\tilde{p}_k}\frac{p_k}{\tilde{p}_{k+1}}\mathbf{E}(V^2(\tau_1,\dots,\tau_N)\\
&\qquad\qquad \cdot I(\tau_{k+1} > T,\ \tau_k \leqslant T) \mid \mathscr{A}_1)\\
&\qquad + L_1^{k-1}\prod_{i=k}^{k+1}\frac{1-p_i}{1-\tilde{p}_i}\mathbf{E}\big(V^2(\tau_1,\dots,\tau_N)\\
&\qquad\qquad \cdot \prod_{j=k+2}^{N} L_j I(\tau_k > T,\ \tau_{k+1} > T) \mid \mathscr{A}_1\big). \qquad (15)
\end{aligned}$$

We claim that

$$\begin{aligned}
&\mathbf{E}(V^2(\tau_1,\dots,\tau_N)I(\tau_k \leqslant T,\ \tau_{k+1} > T) \mid \mathscr{A}_1)\\
&\qquad \geqslant \mathbf{E}(V^2(\tau_1,\dots,\tau_N)I(\tau_k > T,\ \tau_{k+1} \leqslant T) \mid \mathscr{A}_1). \qquad (16)
\end{aligned}$$

We defer the proof of this claim until the end. By the choice of $\tilde{p}$ in the JK method, $\tilde{p}_i = 1/(N-i+1)$ for $i = k, k+1$, it is clear that

$$\tilde{p}_k = (1-\tilde{p}_k)\tilde{p}_{k+1},$$

whence the equality

$$\frac{p_k}{\tilde{p}_k} - \frac{1-p_{k+1}}{1-\tilde{p}_k}\frac{p_k}{\tilde{p}_{k+1}} = \frac{p_{k+1}}{\tilde{p}_k} - \frac{1-p_k}{1-\tilde{p}_k}\frac{p_{k+1}}{\tilde{p}_{k+1}}.$$

Subtracting Equation (15) from Equation (14) gives

$$\begin{aligned}
&L_1^{k-1}\left(\frac{p_k}{\tilde{p}_k} - \frac{1-p_{k+1}}{1-\tilde{p}_k}\frac{p_k}{\tilde{p}_{k+1}}\right)\\
&\quad \cdot \mathbf{E}(V^2(\tau_1,\dots,\tau_N)I(\tau_k \leqslant T,\ \tau_{k+1} > T) \mid \mathscr{A}_1)\\
&\quad + L_1^{k-1}\frac{p_k - p_{k+1}}{\tilde{p}_k}\mathbf{E}(V^2(\tau_1,\dots,\tau_N)I(\tau_k \leqslant T, \tau_{k+1} \leqslant T) \mid \mathscr{A}_1)\\
&\quad - L_1^{k-1}\left(\frac{p_{k+1}}{\tilde{p}_k} - \frac{1-p_k}{1-\tilde{p}_k}\frac{p_{k+1}}{\tilde{p}_{k+1}}\right)\\
&\quad \cdot \mathbf{E}(V^2(\tau_1,\dots,\tau_N)I(\tau_k > T,\ \tau_{k+1} \leqslant T) \mid \mathscr{A}_1) > 0.
\end{aligned}$$

Thus, the second moment is decreased by switching the order.

Given $\mathscr{A}_2$, $\tilde{p}_k$ has only one value, but $\tilde{p}_{k+1}$ has two choices:

$$\tilde{p}_k = \frac{n - \sum_{j<k} I(\tau_j \leqslant T)}{N-k+1},$$

$$\tilde{p}_{k+1} = \begin{cases} \pi_1 = \dfrac{n-1-\sum_{j<k} I(\tau_j \leqslant T)}{N-k} & \text{if } \tau_k \leqslant T,\\[2ex] \pi_2 = \dfrac{n-\sum_{j<k} I(\tau_j \leqslant T)}{N-k} & \text{if } \tau_k > T. \end{cases}$$

The second moment of the estimator is

$$\begin{aligned}
&\mathbf{E}(V^2(\tau_1,\dots,\tau_N)L \mid \mathscr{A}_2)\\
&\quad = L_1^{k-1}\frac{p_k}{\tilde{p}_k}\frac{p_{k+1}}{\pi_1}\mathbf{E}(V^2(\tau_1,\dots,\tau_N)\\
&\qquad\qquad \cdot \prod_{j=k+2}^{N} L_j I(\tau_k \leqslant T,\ \tau_{k+1} \leqslant T) \mid \mathscr{A}_2)\\
&\qquad + L_1^{k-1}\frac{p_k}{\tilde{p}_k}\frac{1-p_{k+1}}{1-\pi_1}\mathbf{E}\big(V^2(\tau_1,\dots,\tau_N)\\
&\qquad\qquad \cdot \prod_{j=k+2}^{N} L_j I(\tau_k \leqslant T,\ \tau_{k+1} > T) \mid \mathscr{A}_2\big)\\
&\qquad + L_1^{k-1}\frac{1-p_k}{1-\tilde{p}_k}\frac{p_{k+1}}{\pi_2}\mathbf{E}\big(V^2(\tau_1,\dots,\tau_N)\\
&\qquad\qquad \cdot \prod_{j=k+2}^{N} L_j I(\tau_k > T,\ \tau_{k+1} \leqslant T) \mid \mathscr{A}_2\big)\\
&\qquad + L_1^{k-1}\frac{1-p_k}{1-\tilde{p}_k}\frac{1-p_{k+1}}{1-\pi_2}\mathbf{E}\big(V^2(\tau_1,\dots,\tau_N)\\
&\qquad\qquad \cdot \prod_{j=k+2}^{N} L_j I(\tau_k > T,\ \tau_{k+1} > T) \mid \mathscr{A}_2\big). \qquad (17)
\end{aligned}$$

After switching the kth and $(k+1)$st assets, the second moment becomes

$$\begin{aligned}
&\mathbf{E}(V^2(\tau_1,\dots,\tau_N)L \mid \mathscr{A}_2)\\
&\quad = L_1^{k-1}\frac{p_k}{\tilde{p}_k}\frac{p_{k+1}}{\pi_1}\mathbf{E}\big(V^2(\tau_1,\dots,\tau_N)\\
&\qquad\qquad \cdot \prod_{j=k+2}^{N} L_j I(\tau_k \leqslant T,\ \tau_{k+1} \leqslant T) \mid \mathscr{A}_2\big)
\end{aligned}$$

$$
\begin{aligned}
&+ L_1^{k-1}\frac{1-p_{k+1}}{1-\tilde{p}_k}\frac{p_k}{\pi_2}\mathbf{E}\Big(V^2(\tau_1,\ldots,\tau_N)\\
&\qquad\qquad\cdot\prod_{j=k+2}^{N} L_j I(\tau_k\leqslant T,\ \tau_{k+1}>T)\mid\mathscr{A}_2\Big)\\
&+ L_1^{k-1}\frac{p_{k+1}}{\tilde{p}_k}\frac{1-p_k}{1-\pi_1}\mathbf{E}\Big(V^2(\tau_1,\ldots,\tau_N)\\
&\qquad\qquad\cdot\prod_{j=k+2}^{N} L_j I(\tau_k> T,\ \tau_{k+1}\leqslant T)\mid\mathscr{A}_2\Big)\\
&+ L_1^{k-1}\frac{1-p_k}{1-\tilde{p}_k}\frac{1-p_{k+1}}{1-\pi_2}\mathbf{E}\Big(V^2(\tau_1,\ldots,\tau_N)\\
&\qquad\qquad\cdot\prod_{j=k+2}^{N} L_j I(\tau_k> T,\ \tau_{k+1}> T)\mid\mathscr{A}_2\Big). \qquad (18)
\end{aligned}
$$

By noting that

$$
\begin{aligned}
(1-\tilde{p}_k)\pi_2 &= \left(1-\frac{n-\sum_{j<k}I(\tau_j\leqslant T)}{N-k+1}\right)\frac{n-\sum_{j<k}I(\tau_j\leqslant T)}{N-k}\\
&= \frac{n-\sum_{j<k}I(\tau_j\leqslant T)}{N-k+1}\\
&\quad\left(1-\frac{n-1-\sum_{j<k}I(\tau_j\leqslant T)}{N-k}\right)=\tilde{p}_k(1-\pi_1),
\end{aligned}
$$

switching the positions of the kth and the $(k+1)$st assets does not change the second moment conditional on $\mathscr{A}_2$.

We have shown that switching the positions of the kth and the $(k+1)$st assets when $p_k > p_{k+1}$ decreases the second moment of the JK estimator. Therefore, variance is minimized in the JK method by taking the assets in ascending order $p_1\leqslant p_2\leqslant\cdots\leqslant p_N$.

To conclude the proof, we need to establish (16). Using the relation between τ_i and U_i, we have

$$
\begin{aligned}
&\mathbf{E}(V^2(\tau_1,\ldots,\tau_N)I(\tau_k\leqslant T,\ \tau_{k+1}>T)\mid\mathscr{A}_1)\\
&\quad=\mathbf{E}(V^2(\tau_1,\ldots,\tau_N)I(U_k\leqslant p_k,\ U_{k+1}>p_{k+1})\mid\mathscr{A}_1)\\
&\quad\geqslant\mathbf{E}(V^2(\tau_1,\ldots,\tau_N)I(U_k\leqslant p_{k+1},\ U_{k+1}>p_k)\mid\mathscr{A}_1)\\
&\quad\geqslant\mathbf{E}(V^2(\tau_1,\ldots,\tau_N)I(U_k> p_k,\ U_{k+1}\leqslant p_{k+1})\mid\mathscr{A}_1)\\
&\quad=\mathbf{E}(V^2(\tau_1,\ldots,\tau_N)I(\tau_k> T,\ \tau_{k+1}\leqslant T)\mid\mathscr{A}_1).
\end{aligned}
$$

The second inequality is based on the following two facts: (1) Because $p_k > p_{k+1}$, then $h_k > h_{k+1}$, so $\tau_k(u) < \tau_{k+1}(u)\leqslant T$ for any $u\leqslant p_{k+1}$. Therefore τ, the nth default time, increases after switching the order. (2) $V^2(\tau_1,\ldots,\tau_N)$ decreases when τ increases. □

A.5. Proof of Proposition 2

Proof. To compare the variances of the estimators given by the JK method and the plain Monte Carlo method, it is enough to compare their second moments.

For a last-to-default swap, in the JK method, $\tilde{p}_i = 1$ for $i = 1,\ldots,N$, and

$$
L=\prod_{i=1}^{N}\frac{p_i}{\tilde{p}_i}<1.
$$

Therefore,

$$
\begin{aligned}
\tilde{\mathbf{E}}(V^2(\tau_1,\ldots,\tau_N)L^2) &= \mathbf{E}(V^2(\tau_1,\ldots,\tau_N)L)\\
&< \mathbf{E}(V^2(\tau_1,\ldots,\tau_N)),
\end{aligned}
$$

i.e., the JK method has smaller variance than plain Monte Carlo.

For an nth-to-default swap, with $n < N$, if the default probabilities of the assets are large enough or the life of the swap is long enough, the estimator given by the JK method has a larger variance than plain Monte Carlo sampling, as we now show by induction on N.

In the case $N = 2$, it suffices to consider $n = 1$, a first-to-default swap, for which the JK method yields

$$
\begin{aligned}
&\tilde{\mathbf{E}}(V^2(\tau_1,\ldots,\tau_N)L^2)\\
&\quad=\mathbf{E}(V^2(\tau_1,\ldots,\tau_N)L)\\
&\quad=\frac{p_1^2}{\tilde{p}_1}\mathbf{E}(V^2(\tau_1,\ldots,\tau_N)\mid\tau_1\leqslant T)\\
&\qquad+\frac{(1-p_1)^2}{1-\tilde{p}_1}\mathbf{E}(V^2(\tau_1,\ldots,\tau_N)p_2\mid\tau_1>T).
\end{aligned}
$$

In the plain Monte Carlo method,

$$
\begin{aligned}
\mathbf{E}(V^2(\tau_1,\ldots,\tau_N)) &= p_1\mathbf{E}(V^2(\tau_1,\ldots,\tau_N)\mid\tau_1\leqslant T)\\
&\quad+(1-p_1)\mathbf{E}(V^2(\tau_1,\ldots,\tau_N)\mid\tau_1>T).
\end{aligned}
$$

If $p_1\to 1$, then $p_1>\tilde{p}_1$, $(1-p_1)/(1-\tilde{p}_1)\to 0$, and $1-p_1\to 0$; therefore,

$$
\tilde{\mathbf{E}}(V^2(\tau_1,\ldots,\tau_N)L^2) > \mathbf{E}(V^2(\tau_1,\ldots,\tau_N))
$$

for all sufficiently large p_1. Therefore, the same holds if either the swap maturity or the default hazard rate is sufficiently large.

Now consider an arbitrary $N > 2$ and $n < N$. In the JK method,

$$
\begin{aligned}
&\tilde{\mathbf{E}}(V^2(\tau_1,\ldots,\tau_N)L^2)\\
&\quad=\frac{p_1^2}{\tilde{p}_1}\mathbf{E}\left(V^2(\tau_1,\ldots,\tau_N)\prod_{i=2}^{N}L_i\,\middle|\,\tau_1\leqslant T\right)\\
&\qquad+\frac{(1-p_1)^2}{1-\tilde{p}_1}\mathbf{E}\left(V^2(\tau_1,\ldots,\tau_N)\prod_{i=2}^{N}L_i\,\middle|\,\tau_1> T\right).
\end{aligned}
$$

In the plain Monte Carlo method,

$$
\begin{aligned}
\mathbf{E}(V^2(\tau_1,\ldots,\tau_N)) &= p_1\mathbf{E}(V^2(\tau_1,\ldots,\tau_N)\mid\tau_1\leqslant T)\\
&\quad+(1-p_1)\mathbf{E}(V^2(\tau_1,\ldots,\tau_N)\mid\tau_1>T).
\end{aligned}
$$

As $p_1 \to 1$, we need to compare the first term in the two expressions for the second moments of the two estimators. The first term in each case is a conditional expectation given $\tau_1 \leqslant T$. Conditional on $\tau_1 \leqslant T$, the two procedures (the JK method and plain Monte Carlo) reduce to problems that have $N-1$ assets (and require $n-1$ defaults) and are thus covered by the induction argument. Moreover, it follows from (2) that the conditional default probability $p_2 = \mathbf{P}(\tau_2 \leqslant T \mid \tau_1)$ approaches one as the marginal default probability $F_2(T) = \mathbf{P}(\tau_2 \leqslant T)$ approaches one. Thus, by induction,

$$\mathbf{E}\left(V^2(\tau_1,\ldots,\tau_N)\prod_{i=2}^{N} L_i \mid \tau_1 \leqslant T\right) > \mathbf{E}(V^2(\tau_1,\ldots,\tau_N) \mid \tau_1 \leqslant T)$$

once the marginal default probabilities are sufficiently large. We conclude that

$$\tilde{\mathbf{E}}(V^2(\tau_1,\ldots,\tau_N)L^2) > \mathbf{E}(V^2(\tau_1,\ldots,\tau_N))$$

once the marginal default probabilities are sufficiently large. □

A.6. Correlation Matrices Used in the Numerical Examples

The four-factor model in the example of Swap A2 uses the following randomly generated factor-loading matrix:

$$\mathbf{A} = \begin{pmatrix} -0.3955 & -0.6130 & 0.3032 & 0.3093 \\ 0.3984 & 0.4795 & 0.5043 & -0.0636 \\ 0.3316 & -0.2816 & -0.2323 & -0.3391 \\ 0.0706 & -0.2863 & 0.5232 & -0.2069 \\ -0.4649 & 0.0665 & -0.2604 & 0.5511 \\ 0.0769 & 0.3760 & 0.7403 & 0.5437 \\ 0.2009 & 0.6573 & 0.1423 & 0.2606 \\ -0.0588 & -0.3827 & -0.4279 & 0.4484 \\ 0.3364 & -0.4151 & 0.0155 & -0.1919 \\ -0.1921 & -0.0465 & -0.4700 & 0.3500 \end{pmatrix}.$$

The one-factor model in the example of Swap A4 uses

$$\mathbf{A} = \begin{pmatrix} 0.4564 \\ 0.4929 \\ -0.3811 \\ -0.3363 \\ -0.4127 \\ 0.3544 \\ 0.3412 \\ -0.4832 \\ -0.3756 \\ 0.3409 \end{pmatrix}.$$

The corresponding correlation matrix can be obtained by

$$\mathbf{\Sigma} = \mathbf{A}\mathbf{A}^\top + \mathbf{B}\mathbf{B}^\top,$$

where $\mathbf{B}$ is the supplementary matrix of $\mathbf{A}$.

A.7. Variance Reduction Ratios in the Numerical Examples

Here we supply estimated variance reduction ratios at various maturities in our examples:

Swap A1

T	0.5	1	2	5	10	15	20	30
CP	1,613.5	757.0	324.5	78.6	17.9	6.5	3.2	1.5

Swap A2

T	3	4	5	10	15	20	30
CP	10.5	7.9	6.5	3.9	3.1	2.2	1.2
CPST (one-dimensional)	11.0	8.1	6.7	3.9	3.1	2.3	1.2
CPST (two-dimensional)	14.3	12.1	10.5	6.5	4.7	3.0	1.4

(In Swap A2, the CP and CPST methods produce variance very close to zero for $T \leqslant 2$. This makes it difficult to measure their variance reduction ratios accurately, so we begin this table at $T = 3$.)

Swap A3

T	1	2	5	10	15	20	30
CP	2.2	1.9	1.6	1.4	1.3	1.2	1.1
CPST (one-dimensional)	7.5	22.5	39.9	27.9	17.4	11.3	6.2
CPST (two-dimensional)	7.8	26.7	53.4	32.6	17.9	11.8	6.3

Swap A4

T	0.5	1	2	5	10	15	20	30
CP	32.3	25.0	20.7	16.8	10.1	5.1	2.8	1.4
CPST	371.5	320.8	231.2	75.7	17.9	6.5	3.2	1.5

Swap A5

T	0.5	1	2	5	10	15	20	30
CP	5.3	4.4	3.6	3.0	2.5	2.1	1.8	1.4
CPST	120.3	194.8	183.8	80.5	27.8	12.8	7.2	3.6

Acknowledgments

This work was supported by a grant from Moody's Corporation and by NSF grants DMI0300044 and DMS0410234.

References

Andersen, L., J. Sidenius, S. Basu. 2003. All your hedges in one basket. *Risk* **16** 67–72.

Avranitis, A., J. Gregory. 2001. *Credit: The Complete Guide to Pricing, Hedging and Risk Management*. Risk Books, London, UK.

Bruyere, R., R. Cont, R. Copinot, L. Fery, J. Jaeck, T. Spitz, G. Smart. 2006. *Credit Derivatives and Structured Credit: A Guide for Investors*. Wiley, Chichester, UK.

Duffie, D., K. Singleton. 2003. *Credit Risk: Pricing, Measurement, and Management*. Princeton University Press, Princeton, NJ.

Feller, W. 1968. *An Introduction to Probability Theory and Its Applications*, Vol. I. Wiley, New York.

Glasserman, P. 2004. *Monte Carlo Methods in Financial Engineering*. Springer, New York.

Glasserman, P., J. Li. 2005. Importance sampling for portfolio credit risk. *Management Sci.* **51** 1643–1656.

Golub, G., C. F. Van Loan. 1996. *Matrix Computations*, 3rd ed. Johns Hopkins University Press, Baltimore, MD.

Gupton, G., C. Finger, M. Bhatia. 1997. *CreditMetrics Technical Document*. J. P. Morgan & Co., New York.

Joshi, M. 2005. Applying importance sampling to pricing single tranches of CDOs in a one-factor Li model. Working paper, Centre for Actuarial Sciences, Melbourne University, Melbourne, Australia.

Joshi, M., D. Kainth. 2004. Rapid and accurate development of prices and Greeks for *n*th to default credit swaps in the Li model. *Quantitative Finance*, Vol. 4. Institute of Physics Publishing, London, UK, 266–275.

Kalkbrener, M., H. Lotter, L. Overbeck. 2004. Sensible and efficient capital allocation for credit portfolios. *Risk* **17**(January) S19-–S24.

Laurent, J.-P., J. Gregory. 2003. Basket default swaps, CDOs, and factor copulas. Working paper, University of Lyon, Lyon, France.

Li, D. 2000. On default correlation: A copula function approach. *J. Fixed Income* **9** 43-–54.

Schönbucher, P. 2003. *Credit Derivatives Pricing Models: Model, Pricing and Implementation*. Wiley, Chichester, UK.