SIAM J. FINANCIAL MATH.
Vol. 1, pp. 868–896

Exact and Efficient Simulation of Correlated Defaults*

K. Giesecke†, H. Kakavand†, M. Mousavi†, and H. Takada‡

Abstract. Correlated default risk plays a significant role in financial markets. Dynamic intensity-based models, in which a firm default is governed by a stochastic intensity process, are widely used to model correlated default risk. The computations in these models can be performed by Monte Carlo simulation. The standard simulation method, which requires the discretization of the intensity process, leads to biased simulation estimators. The magnitude of the bias is often hard to quantify. This paper develops an exact simulation method for intensity-based models that leads to unbiased estimators of credit portfolio loss distributions, risk measures, and derivatives prices. In a first step, we construct a Markov chain that matches the marginal distribution of the point process describing the binary default state of each firm. This construction reduces the original estimation problem to one involving a Markov chain expectation. In a second step, we estimate the Markov chain expectation using a simple acceptance/rejection scheme that facilitates exact sampling. To address rare event situations, the acceptance/rejection scheme is embedded in an overarching selection/mutation scheme, in which a selection mechanism adaptively forces the chain into the regime of interest. Numerical experiments demonstrate the effectiveness of the method for a self-exciting model of correlated default risk.

Key words. portfolio credit risk, Markovian projection, rare-event simulation, acceptance/rejection, selection/mutation

AMS subject classifications. 60G55, 60J27, 60J75, 90-08, 65C05

DOI. 10.1137/090778055

1. Introduction.

Correlated default risk is one of the most pervasive threats in financial markets. Confronting this threat is a daily business for credit investors such as banks making loans to individuals and corporations or fixed income managers allocating assets in the credit markets. These investors must measure the aggregate default risk in their asset portfolios and devise strategies to mitigate that risk. These tasks typically involve estimating the risk capital, to cushion potential default losses at high confidence levels, and estimating the prices of portfolio credit derivatives, which are financial instruments that provide insurance against correlated default risk.

Risk management and derivatives pricing applications require a stochastic model of correlated default timing. Intensity-based models are widely used for this purpose. In these models, a portfolio constituent firm defaults at an inaccessible stopping time whose stochastic structure is governed by an intensity, or conditional default rate. The intensity follows

*Received by the editors November 23, 2009; accepted for publication (in revised form) September 27, 2010; published electronically November 30, 2010. This work was supported by a Mizuho-DL Financial Technology grant. This paper was presented at seminars at the Chinese University of Hong Kong, Columbia University, Hong Kong University of Science and Technology, Mizuho-DL Financial Technology, the 2009 ICBI Global Derivatives and Risk Management Conference, and the 15th INFORMS Applied Probability Society Conference.
http://www.siam.org/journals/sifin/1/77805.html

†Department of Management Science & Engineering, Stanford University, Stanford, CA 94305-4026 (giesecke@stanford.edu, mousavi@stanford.edu, hossein@stanford.edu).

‡Mizuho-DL Financial Technology, Tokyo (hideyuki-takada@fintec.co.jp).

a stochastic process that reflects the information revealed over time, including the value of exogenous risk factors and the state of other firms in the economy. The intensity processes are correlated across firms, to incorporate the dependence between firm defaults. While many intensity models have been developed in the literature, model computation remains challenging. The scope of semianalytical transform methods is limited, including mainly doubly stochastic models. In these models, firm intensities are driven by common risk factors. Conditional on a realization of these factors, default times are independent of one another.

Monte Carlo simulation is an alternative tool for performing computations in intensity-based models. It can be applied to models outside the scope of transform methods or to address applications for which transform methods are unsuitable. A standard simulation method, which applies to most intensity models and is thus widely used, exploits a time-change result for point processes due to [58]. Meyer showed that a nonexplosive counting process can be transformed into a standard Poisson process by a change of time given by the counting process compensator, or cumulative intensity. This implies that the first jump time of the process is equal in law to the first hitting time of the compensator process to a standard exponential variable. This insight provides a recipe for simulating a default time with given intensity process: generate a path of the cumulative intensity and record its first hitting time to a level drawn independently from a standard exponential distribution. The resulting default times have the correct joint distribution, as implied by the correlated evolution of firm intensity processes.

While widely applicable, the time-scaling scheme may lead to biased simulation estimators. This is because it may not be possible to construct the full path of the continuous-time stochastic process followed by the time integral of the intensity. Often, the path must be approximated on a discrete-time grid. Further, it may be difficult to draw exact samples of the values of the integrated intensity at the grid points, because the joint distribution of the integrated intensities across firms, from which one needs to sample, is rarely computationally tractable. In this case, the values of an integrated intensity must be approximated by first simulating the continuous-time intensity process on the discrete-time grid, and then integrating the discretized values. If the intensity values cannot be sampled from their joint transition law, then the SDEs that describe the joint dynamics must be discretized by the Euler or some higher order scheme. Due to the multiple layers of approximations, it is hard to quantify the magnitude of the discretization bias in the resulting simulation estimators. While the bias can be reduced by increasing the number of discretization time steps, this comes at the expense of increasing the time required to generate a replication. Since the additional computational effort per firm replication is scaled by the number of firms in the portfolio, this can quickly lead to a computational burden that is prohibitive for the large portfolios that occur in practice.

This paper develops an exact sampling method that leads to unbiased simulation estimators for intensity-based models. The scope of the method is more limited than that of the time-scaling method but wider than that of the transform methods. It comprises a broad range of models proposed in the literature, including doubly stochastic and self-exciting formulations. The method has two parts. We first construct an inhomogeneous, continuous-time Markov chain M whose value at t has the same distribution as the value at t of the point process N describing the binary default state of each portfolio constituent. The construction reduces the problem of estimating the expectation of $f(N_t)$ to that of estimating the expec-

tation of $f(M_t)$. Unlike N, the mimicking chain M has deterministic interarrival intensities, and this facilitates the exact sampling of M using a simple acceptance/rejection scheme. As a result, we obtain an unbiased estimator of the expectation of $f(M_t)$.

For portfolios of high-quality names, most replications produce few if any defaults, so the computational effort required to obtain accurate estimators may be very large. This is especially true for estimators of large loss probabilities or tail risk measures, which are at the center of risk management applications. To address this problem, we embed the acceptance/rejection scheme in an overarching selection/mutation scheme developed by [21]. On a discrete-time grid we evolve a collection of particles, i.e., copies of the mimicking Markov chain M, using the acceptance/rejection scheme. At each time step, particles are randomly selected by sampling with replacement, placing more weight on particles with a larger number of transitions in the previous period. The new generation of selected particles is then evolved over the next period, at whose end a selection takes place again. The selection procedure adaptively forces the chain into the regime of interest and therefore reduces variance. The resulting estimators inherit the unbiasedness of the plain acceptance/rejection estimators.

Numerical experiments demonstrate the effectiveness of the method for a portfolio of 100 names. We analyze a self-exciting model, in which firm intensities follow correlated Feller jump-diffusion processes that jump whenever a default event occurs. We find that the exact method requires significantly less computation time than the conventional time-scaling method, for all levels of accuracy. The root mean square errors of the simulation estimators converge much faster for the exact method. The selection/mutation scheme is found to offer substantial variance reduction.

1.1. Related literature. While many alternative intensity-based models of correlated default risk have been developed in the literature, there is surprisingly little work on simulation methods for these models. The authors of [27] review time-scaling and other approaches. [24] provides an inverse transform method for simulating the first to occur of a given set of events. This scheme is exact and can be used to sample the default times sequentially; it leads to unbiased estimators of an expectation of a function of the vector of default times of the constituent names. The method developed in this paper leads to unbiased estimators of an expectation of a function of the future value of the vector point process indicating the default state of the constituent names.

Giesecke, Kakavand, and Mousavi [38] develop an exact method for the related problem of simulating a one-dimensional, real-valued point process. They project the point process onto its own filtration and then sample it in this coarser filtration. The sampling is based on the intensity in the subfiltration, which is deterministic between arrival times and therefore facilitates the use of exact schemes. This projection method leads to unbiased estimators of an expectation of a function of the path of the one-dimensional point process and a skeleton of the driving state process.

Bassamboo and Jain [5] propose an asymptotically optimal importance sampling scheme to estimate the probability of large portfolio losses in a doubly stochastic intensity model with affine risk factor processes. Their approach exploits the conditional independence of firm defaults in the doubly stochastic setting. The implementation of the estimators relies on the time-scaling scheme.

Carmona, Fouque, and Vestal [11] use a selection/mutation scheme to estimate the distri-

bution of portfolio loss in a structural model of correlated default risk. Here, a firm defaults when its market value hits a given barrier. Firm values follow correlated stochastic volatility processes. Carmona and Crepey [10] numerically contrast the performance of the selection/mutation and importance sampling schemes when estimating the distribution of portfolio loss in a Markov chain model.

There are several papers on variance reduction schemes for copula-based models of correlated default risk. In a copula-based model the firm dependence structure is specified by a copula function that maps firm-level default probabilities to the joint default probability. [6], [14], [49], [40], [42] develop importance sampling schemes that exploit the conditional independence of firm defaults that is also a feature of the copula models. Glasserman and Li [41] examine a related scheme for a mixed Poisson model of portfolio credit risk.

1.2. Structure of this article. Section 2 formulates the problem, reviews conventional simulation approaches, and outlines the exact method. Section 3 discusses the construction of the mimicking Markov chain. Section 4 develops two algorithms for estimating expectations associated with the mimicking chain. Section 5 constructs the mimicking Markov chain for a broad range of models proposed in the literature. Section 6 provides a numerical case study that demonstrates the effectiveness of the method. An appendix contains the proofs.

2. Preliminaries.

2.1. Default point processes. Consider a portfolio of n firms that are subject to default risk. The random default times of these firms are modeled by almost surely distinct stopping times $\tau^i > 0$, which are defined on a complete probability space $(\Omega, \mathcal{F}, P)$ with right-continuous and complete information filtration $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$. In risk management applications, P is the statistical probability, while in derivatives pricing applications, P is a risk-neutral pricing measure. Associated with the τ^i are indicator processes N^i given by $N^i_t = I(\tau^i \leq t)$, where $I(A)$ is the indicator function of an event $A \in \mathcal{F}$. For each i, there is a strictly positive, integrable, and progressively measurable process λ^i such that the variables

$$N^i_t - \int_0^t I(\tau^i > s)\lambda^i_s ds \tag{1}$$

form a martingale relative to $\mathbb{F}$. This means that λ^i_t is the $\mathcal{F}_t$-conditional default rate of firm i at time $t < \tau^i$, measured in events per unit of time. We refer to the process λ^i as the default intensity of firm i, recognizing that this may involve an innocuous abuse of terminology, as λ^i need not drop to 0 at τ^i. The intensities follow correlated stochastic processes that need not be specified at this point. The correlation among the intensities reflects the default dependence structure of the portfolio constituent firms.

Our goal is to calculate $E[f(N_T)]$ for a suitable real-valued function f on $\{0,1\}^n$ and a horizon $T > 0$, where $N = (N^1, \ldots, N^n)$ is the vector of firm default indicator processes. Examples include the single-name probabilities $P(\tau^i > T)$, joint probabilities $P(\cap_{i\in S}\{\tau^i > T\})$ for subsets S of firms, the distribution $P(C_T = k)$ of the default counting process $C = 1_n \cdot N$, where 1_n is an n-vector of ones, and the option $E[(C_T - K)^+]$. If taken under the statistical probability measure, $P(C_T = k)$ is the fundamental quantity for the risk management of portfolios of corporate debt. If evaluated under a risk-neutral pricing measure, the option

$E[(C_T - K)^+]$ is a key quantity required to price portfolio credit derivatives, as illustrated in [31].

For clarity in the exposition, our formulation has not made explicit the role of the financial loss at a default. This is without loss of generality in case the loss $\ell = (\ell^1, \ldots, \ell^n)$ is deterministic as in [5], [6], [10], [11], [14], [15], [49], [40] and others. In that case, ℓ can be incorporated into the function f, to express the probability that the portfolio loss $\ell \cdot N_T$ at T exceeds a level, or an option on $\ell \cdot N_T$, as $E[f(N_T)]$. To treat a loss ℓ that is random but independent of N as in [42], the algorithms developed below require only minor modifications. Independence is a reasonable assumption for lack of data bearing on the correlation between ℓ and N.

2.2. Conventional simulation schemes. To estimate $E[f(N_T)]$ by simulation, we need to sample the variable N_T. This is straightforward if the intensity λ is deterministic and N is a Poisson process. It is also straightforward if λ follows a stochastic process that is adapted to the subfiltration of $\mathbb{F}$ generated by N, in which case λ is deterministic between default times. In these cases, the classical thinning, or acceptance/rejection (A/R) scheme of [54], can be used to sample the jump times of N exactly. This scheme leads to unbiased simulation estimators of $E[f(N_T)]$. It involves the sampling of candidate arrival times from a dominating Poisson process, and an acceptance test. However, if λ is not adapted to the filtration generated by N, then it evolves randomly also between events, and a dominating Poisson process may not exist. In this case, one may be able to use an inverse transform scheme to sample the default times exactly. This requires an explicit expression of the probabilities $P(\tau^i > T)$. It may also be possible to apply the inverse scheme sequentially, as in [24] and [27]. This requires that the conditional distributions of the interarrival times of C and the defaulter identities be tractable and also the ability to sample the state variables determining these distributions.

A more widely applicable method exploits a time-change result for point processes due to [58]. Consider a counting process Z with jumps of size one and compensator $\widehat{Z}$ that is continuous and increases to ∞ almost surely. Meyer proved that Z is a standard Poisson process under a change of time defined by $\widehat{Z}$, relative to the time-changed filtration. Thus, the first jump time of Z is equal in law to $\inf\{t : \widehat{Z}_t \geq \mathcal{E}\}$, where $\mathcal{E}$ is a standard exponential random variable. This provides a recipe for simulating the first jump time of Z: generate a path of $\widehat{Z}$ and record its hitting time to the level $\mathcal{E}$, drawn independently from a standard exponential distribution. To apply this recipe to generating a path of N^i for a given λ^i, we let Z be a counting process with compensator $\widehat{Z}_t = \int_0^t \lambda_s^i ds$ and set $N^i = \min(Z, 1)$. We generate a path of $\int_0^t \lambda_s^i ds$ and draw $\mathcal{E}$ to obtain a sample of τ^i as the hitting time $\inf\{t : \int_0^t \lambda_s^i ds \geq \mathcal{E}\}$. As the λ^i are correlated across i, the $\int_0^t \lambda_s^i ds$ must be drawn from the appropriate joint distribution.

While the time-scaling scheme has a wide scope, it suffers from an important shortcoming: it usually leads to biased estimators of $E[f(N_T)]$. This is because, aside from very special cases, it is impossible to generate the full path of the continuous-time stochastic process $\int_0^t \lambda_s^i ds$. The path must be approximated on a discrete-time grid. Even worse, it may not be possible to draw exact samples of the values of $\int_0^t \lambda_s^i ds$ at the grid points, because the distribution of $(\int_0^t \lambda_s^1 ds, \ldots, \int_0^t \lambda_s^n ds)$ from which one needs to sample is rarely known or computationally tractable. This forces one to approximate the values of $\int_0^t \lambda_s^i ds$ by first

approximating the continuous-time intensity process λ^i on the discrete-time grid and then integrating the discretized values. If the intensity values cannot be sampled exactly from their joint transition law, then the SDE that describes the dynamics of $(\lambda^1, \dots, \lambda^n)$ must be discretized by the Euler or some other scheme.

Due to the multiple approximations and the multiple dimensions n of N, it is hard to quantify the magnitude of the discretization bias in the estimators of $E[f(N_T)]$. The bias can be reduced by increasing the number of discretization time steps, but this also increases the computational cost of a replication. Reducing the bias to an acceptable level may require a prohibitively large computational effort, since the dimension n of N is often large in practice, and the effort scales with n. Further, it is hard to determine the optimal trade-off between the number of discretization time steps and the number of simulation trials because the convergence rate of the bias is unknown.

2.3. Exact simulation. Below we develop an alternative simulation method that eliminates the need to discretize the vector process $(\lambda^1, \dots, \lambda^n)$ and that leads to unbiased estimators of $E[f(N_T)]$. The method has two parts. We first construct a time-inhomogeneous continuous-time Markov chain $M = (M^1, \dots, M^n) \in \{0,1\}^n$ with the property that $M_T = N_T$ in distribution for each fixed T. This construction is explained in section 3. It reduces the problem of estimating the general point process expectation $E[f(N_T)]$ to the simpler problem of estimating the Markov chain expectation $E[f(M_T)]$. Estimators of $E[f(M_T)]$ are obtained by exact sampling of M_T using a thinning scheme, as explained in section 4.

3. Mimicking Markov chain.

3.1. Construction. Throughout, we let 0_n denote an n-vector of zeros.

Proposition 3.1. *Suppose that the default indicator process N has intensity λ. Let M be a Markov chain on $[0,\infty)$ that takes values in $\{0,1\}^n$, starts at 0_n, has no joint transitions in any of its components, and whose ith component has transition rate $h^i(\cdot, M)$, where*

$$h^i(t,B) = E(\lambda_t^i I(\tau^i > t) \,|\, N_t = B), \quad B \in \{0,1\}^n. \tag{2}$$

Then $M_T = N_T$ in distribution for each $T \geq 0$.

Proposition 3.1 shows that a component transition function $h^i(t,B)$ of the mimicking Markov chain M is given by the projection of the primitive firm intensity $\lambda_t^i I(\tau^i > t)$ onto the value of the default process $N_t = B$, which indicates the state at time t of each firm in the portfolio. The indicator $I(\tau^i > t)$ guarantees that $h^i(t,B)$ vanishes if $B^i = 1$, i.e., in a state where firm i is in default. In the special case where N is a priori a Markov point process, $\lambda_t^i = q^i(t, N_t)$ for some function q^i on $\mathbb{R}_+ \times \{0,1\}^n$, and $h^i(t,B) = q^i(t,B)(1-B^i)$. In general, N is not a priori a Markov point process. Then the conditional expectation (2) is nontrivial, and further steps are required to calculate it for the given process λ; see section 5. In any case, Proposition 3.1 leads to a Markov point process M whose value at t has the same distribution as N_t.

The fact that the construction leads to a Markov process is a consequence of the conditioning set in the conditional expectation (2). Rather than conditioning on the sigma-field $\sigma(N_s : s \leq t)$ generated by the path of N during $[0,t]$, which seems natural at first, the conditional expectation (2) is taken with respect to the final value N_t only. The conditioning on

the path does not in general produce a Markov point process. Nevertheless, the conditional expectation $E(\lambda_t^i I(\tau^i > t) \,|\, \sigma(N_s : s \le t))$ is meaningful: it defines the intensity of N^i in the subfiltration generated by N. As shown in [38], this observation can be used to develop an alternative projection scheme for the exact simulation of a point process. That scheme samples a point process in its own filtration, based on the intensity in the subfiltration. It is appropriate for estimating an expectation $E[g(U_t : t \le T)]$ that involves the path of a univariate point process $U \in \mathbb{R}$. For computational reasons, the scheme is less well suited to sampling a vector point process such as $N = (N^1, \dots, N^n) \in \{0,1\}^n$. The method developed in this paper targets the vector process N but is restricted to expectations of the form $E[f(N_T)]$. This is because the auxiliary chain M matches the distribution of N_T for fixed T only.

Proposition 3.1 extends a univariate construction in [8, Chapter II, exercise E8], which is refined and applied by [2], [18], and [56] to the calibration of (univariate) intensity-based, top-down models of portfolio credit risk. These papers construct a mimicking Markov chain for a nonterminating counting process taking values in $\mathbb{N}_0$. Their setting is different from ours even if $n = 1$, because the counting processes N^i that we consider take values in $\{0,1\}$. Lopatin [55] suggests a multivariate version. Bentata and Cont [7] analyze the construction of a mimicking Markov process for a semimartingale that may be discontinuous.

3.2. Markov point process. To prepare the design of simulation algorithms for M in section 4 below, we consider the mimicking chain M as a Markov point process relative to its own right-continuous and complete filtration $\mathbb{G} = (\mathcal{G}_t)_{t \ge 0}$ generated by M. The construction of M implies that, for a suitable real-valued function g on $\{0,1\}^n$, the process $g(M) - \int_0^\cdot \mathscr{A}_s g(M_s) ds$ is a martingale in the filtration $\mathbb{G}$, where

$$\mathscr{A}_t g(B) = \sum_{i=1}^n h^i(t,B)(g(B^{[i]}) - g(B)), \quad B \in \{0,1\}^n,$$

is the generator of M at t. Here, $B^{[i]}$ denotes the vector B whose ith element B^i is replaced by $1 - B^i$. It follows that the process

$$M^i - \int_0^\cdot h^i(s, M_s) ds \tag{3}$$

is a martingale with respect to $\mathbb{G}$. Thus, the component counting process M^i has intensity $h^i(\cdot, M)$ in the filtration $\mathbb{G}$. Recall that $h^i(\cdot, B)$ vanishes for any $B \in \{0,1\}^n$ whose ith element is equal to 1, and compare with the Doob–Meyer decomposition (1) of the firm default indicator process N^i in the reference filtration $\mathbb{F}$. By Proposition 3.1, the distributions of N_t^i and M_t^i agree, and so do the distributions of $C_t = 1_n \cdot N_t$ and $1_n \cdot M_t$. Let h be the n-vector of the functions h^i. From the martingale property of (3),

$$1_n \cdot M - \int_0^\cdot 1_n \cdot h(s, M_s) ds \tag{4}$$

is a $\mathbb{G}$-martingale as well. Therefore, the counting process $1_n \cdot M$ has intensity $1_n \cdot h(\cdot, M)$ relative to the filtration $\mathbb{G}$. As indicated in Table 1, that intensity is the counterpart to the

Table 1
Indicator and counting processes and their Markovian counterparts.

Filtration	Indicator process	Component process	Component intensity	Counting process	Counting intensity
$\mathbb{F}$	N	N^i	$\lambda^i(1-N^i)$	$1_n \cdot N$	$\sum_{i=1}^n \lambda^i(1-N^i)$
$\mathbb{G}$	M	M^i	$h^i(\cdot, M)$	$1_n \cdot M$	$1_n \cdot h(\cdot, M)$

intensity Λ of $C = 1_n \cdot N$ in the reference filtration $\mathbb{F}$, given by

$$\Lambda = \sum_{i=1}^{n} (1 - N^i)\lambda^i. \tag{5}$$

We have

$$1_n \cdot h(t, M_t) = E(\Lambda_t \,|\, N_t = M_t) = \sum_{k=0}^{n-1} H(t,k) I(T_k \le t < T_{k+1}) \tag{6}$$

almost surely. Here, $(T_k)_{k=0,1,\ldots,n}$ is the sequence of event times of $1_n \cdot M$ starting at 0, which is strictly increasing almost surely because there are no joint transitions in M almost surely (Proposition 3.1), and $H(t,k)$ is the $\mathcal{G}_{T_k}$-measurable interarrival intensity function given by

$$H(t,k) = 1_n \cdot h(t, M_{T_k}), \quad t \ge T_k, \ k = 0, 1, \ldots, n. \tag{7}$$

Formula (7) implies that the interarrival intensities of the mimicking Markov counting process $1_n \cdot M$ evolve deterministically through time. This is a key property that our simulation algorithms are going to exploit. Note that the original model, which is formulated in the filtration $\mathbb{F}$, has more complicated interarrival intensity dynamics. This can be seen from formula (5), which indicates that the interarrival $\mathbb{F}$-intensities of $C = 1_n \cdot N$ follow stochastic processes whenever the firm $\mathbb{F}$-intensities λ^i do.

Because $H(t,k)$ is $\mathcal{G}_{T_k}$-measurable, the random variable $T_{k+1} - T_k$ is equal in $\mathcal{G}_{T_k}$-conditional distribution to the first jump time of a time-inhomogeneous Poisson process starting at T_k with intensity function $H(t,k)$ for $t \ge T_k$. Thus, for all $s > 0$ the conditional survival function of the interarrival times of $1_n \cdot M$ satisfies

$$\begin{aligned} P(T_{k+1} - T_k > s \,|\, \mathcal{G}_{T_k}) &= P(1_n \cdot M_{T_k+s} = k \,|\, M_{T_k}) \\ &= \exp\left(- \int_{T_k}^{T_k+s} H(t,k)\, dt \right), \quad k = 0, 1, \ldots, n-1. \end{aligned} \tag{8}$$

Let $I_k \in \{1, 2, \ldots, n\}$ be the $\mathcal{G}_{T_k}$-measurable random variable identifying the component of M in which the kth transition takes place, $k = 1, \ldots, n$. Noting that the sigma-field $\mathcal{G}_{T_k-}$ is generated by the random variables $(T_m, I_m)_{m \le k-1}$ and T_k, and using an argument similar to the one applied by [8, Theorem II.15], we see that

$$P(I_k = i \,|\, \mathcal{G}_{T_k-}) = \frac{h^i(T_k, M_{T_{k-1}})}{H(T_k, k-1)}, \quad i, k = 1, 2, \ldots, n. \tag{9}$$

4. Exact simulation algorithms. We wish to evaluate the expectation $E[f(N_T)]$ for suitable functions f on $\{0,1\}^n$ and fixed $T > 0$. The key insight is that Proposition 3.1 reduces the problem of evaluating $E[f(N_T)]$ to the problem of evaluating $E[f(M_T)]$ for the mimicking Markov chain M. Since the number of portfolio constituents n can be 100 or even larger, we estimate $E[f(M_T)]$ by Monte Carlo simulation of M rather than through alternative numerical methods that would be plagued by the high dimensionality of the state space $\{0,1\}^n$. This section discusses two exact simulation algorithms for this purpose.

4.1. Sequential acceptance/rejection scheme. We simulate the mimicking chain M by sequentially generating the event times and identities (T_k, I_k) introduced above. The generation of event times is based on the interarrival intensities (7) of $1_n \cdot M$, which evolve deterministically through time. Given an event time, the corresponding identity is drawn from the discrete distribution (9).

The inverse or time-scaling methods can be used to generate T_{k+1} from formula (8) for the conditional survival function $S_k(t)$ of the interarrival time $T_{k+1} - T_k$. Draw $U \sim U(0,1)$ and calculate the inverse $\inf\{s > 0 : S_k(s) \leq U\}$. While allowing for exact sampling, this procedure requires us to evaluate the function $S_k(t)$ at many points t in order to determine the inverse at U. Depending on the structure of the function $H(s,k)$, this may be numerically intensive since $S_k(t)$ involves the time-integral of $H(s,k)$.

We prefer an alternative acceptance/rejection (A/R) scheme, which is based on the classical A/R, or thinning, scheme of [54]. This scheme requires the evaluation of $H(\cdot,k)$ only at a set of candidate times for T_{k+1}. The candidate times are generated from a Poisson process whose rate dominates $H(T_k+s,k)$ for s in some interval. A candidate time c is accepted with a probability given by the ratio of $H(T_k+c,k)$ to the dominating Poisson rate. The tighter the dominating bound on $H(T_k+s,k)$, the fewer candidate times need to be generated. Therefore, the dominating Poisson process is redetermined at least at each acceptance or rejection of a candidate time.

Algorithm 4.1. *To generate a sample path of M over $[0,T]$, perform the following:*

1. *Initialize $t = 0$, $k = 0$, $T_0 = 0$, and $M_0 = 0_n$.*
2. *Stop if $t \geq T$.*
3. *Find $J(k) = J(t,k)$ and $K(k) = K(t,k)$ such that $H(t+s,k) \leq J(k)$, $0 \leq s \leq K(k)$.*
4. *Draw a random variable $\mathcal{E}$ from the exponential distribution with parameter $J(k)$.*
   - *If $\mathcal{E} > K(k)$, then set $t = t + K(k)$ and go to step 2.*
   - *If $\mathcal{E} \leq K(k)$, then draw $U \sim U(0,1)$. If $UJ(k) \leq H(t+\mathcal{E},k)$, then set $k = k+1$ and $t = T_k = t + \mathcal{E}$. Else set $t = t + \mathcal{E}$ and go to step 2.*
5. *Draw a random variable I from the discrete distribution*

$$P(I = i) = \frac{h^i(T_k, M_{T_{k-1}})}{H(T_k, k-1)}, \quad i = 1, 2, \ldots, n,$$

and let Q_k be the n-vector with the Ith component equal to one and the rest equal to zero. Set $M_{T_k} = Q_1 + \cdots + Q_k$. Go to step 2.

Algorithm 4.1 is applied to generate a collection of R sample paths $M^1, \ldots, M^R$ of the mimicking Markov chain M over $[0,T]$. Thanks to Proposition 3.1, for suitable functions f on $\{0,1\}^n$ we can estimate the expectation $E[f(N_T)]$ by $\frac{1}{R}\sum_{r=1}^{R} f(M_T^r)$. In particular, the

distribution of the portfolio default count $P(C_T = k)$ is estimated by

$$P_R(C_T = k) = \frac{1}{R}\sum_{r=1}^{R} I(1_n \cdot M_T^r = k), \quad k = 0, 1, \ldots, n. \tag{10}$$

Algorithm 4.1 also leads to estimators of firm-level probabilities $P(\tau^i > T)$ and related quantities, such as $P(\cap_{i\in S}\{\tau^i > T\})$ for subsets S of firms. The setting may simply be $n = 1$, with only a single firm of interest. In this case, (10) is an estimator of $P(N_T^1 = k)$ for $k = 0, 1$, while step 5 of the algorithm is redundant. In the general case of $n > 1$, estimates of $P(\tau^i > T)$ are obtained as a byproduct: with e^i denoting an n-vector with its ith component equal to one and the rest equal to zero, we take $f(M_T^r) = I(e^i \cdot M_T^r = 0)$.

The estimators $E[f(N_T)]$ generated by Algorithm 4.1 are unbiased, because the A/R scheme generates exact samples of the mimicking chain M, and $E[f(N_T)] = E[f(M_T)]$. The A/R scheme applies to M because this process has deterministic interarrival intensities that are usually easy to bound. The A/R scheme does not generally apply to the original default indicator process N, because this process has stochastic interarrival intensities that are usually not bounded by a constant almost surely.

Algorithm 4.1 may be inefficient in some situations. This occurs, for example, when the portfolio constituents have small default probabilities, which is typical for investment grade portfolios of highly rated issuers, and we are interested in estimating tail probabilities of N_T. In this situation, a prohibitively large number of replications may be required to estimate these probabilities accurately with Algorithm 4.1.

4.2. Selection/mutation scheme. To reduce variance, we embed the sequential thinning mechanism into a selection/mutation (S/M) scheme. Let $T > 0$ be the simulation horizon. Partition the interval $[0, T]$ into m subintervals of length T/m. Let V be the discrete-time Markov chain given by

$$V_p = M_{pT/m}, \quad p = 0, 1, \ldots, m.$$

We consider a collection of "particles" $\{V_p^r\}_{r=1,2,\ldots,R}$ that are evolved on the discrete-time grid $p = 0, 1, \ldots, m$, all starting from the same state 0_n at $p = 0$. At a time step p, we use the sequential A/R Algorithm 4.1 to independently *mutate* (evolve) each particle V_p^r during $(p, p+1]$ according to the transition rates determined by Proposition 3.1. Then, before entering the next mutation step, we *select* particles according to the number of transitions during $(p, p+1]$. The selection is done probabilistically, by sampling with replacement. The selection probability increases with the number of transitions during $(p, p+1]$, so the selection favors particles with transitions. The total number of particles R is kept constant, and the selected particles are then evolved over the next period. The final estimator "corrects" for the selections performed at each time step.

Algorithm 4.2. *To generate an estimate of $P(C_T = k)$, perform the following:*

1. *Initialize $V_0^r = W_0^r = 0_n$ for $1 \le r \le R$.*
2. *For each $0 \le p \le m - 1$, repeat the following steps:*
   Selection.
   Fix a constant $\delta > 0$. From the set of particles $(W_p^r, V_p^r)_{1\le r\le R}$ at p
   - *Compute the normalizing constant $\eta_p = \frac{1}{R}\sum_{r=1}^{R} \exp[\delta 1_n \cdot (V_p^r - W_p^r)]$.*

- *Select, independently and with replacement, R new particles as follows. Select particle r with probability*

$$\frac{1}{R\eta_p}\exp[\delta 1_n\cdot(V_p^r-W_p^r)]. \tag{11}$$

Denote the R selected particles by $(\hat{W}_p^r,\hat{V}_p^r)_{1\le r\le R}$.

Mutation.

Evolve the particle $(\hat{W}_p^r,\hat{V}_p^r)$ to (W_{p+1}^r,V_{p+1}^r), independently for each $1\le r\le R$.

- *Set $W_{p+1}^r=\hat{V}_p^r$.*
- *Obtain V_{p+1}^r by generating transitions from $\hat{V}_p^r$ using Algorithm 4.1.*

3. *For $k=0,1,\dots,n$, calculate the estimator of $P(C_T=k)$ as*

$$P_{\delta,m,R}(C_T=k)=\frac{\eta_0\cdots\eta_{m-1}}{R}\sum_{r=1}^{R}I(1_n\cdot V_m^r=k)\exp\left(-\delta 1_n\cdot W_m^r\right). \tag{12}$$

Algorithm 4.2 is a variant of an S/M scheme for estimating rare-event probabilities for time-inhomogeneous Markov chains developed and analyzed by [21]. From their Theorem 2.3 and the fact that the A/R Algorithm 4.1 facilitates exact sampling from the marginal distribution of M in the mutation step of the scheme, we conclude that the estimator (12) is unbiased in the sense that $E[P_{\delta,m,R}(C_T=k)]=P(C_T=k)$ for fixed δ,m,R.

The algorithm requires the selection of the number of particles R, the number of time steps m, and the value of δ. The parameter δ specifies the exponential weight function $\exp[\delta 1_n\cdot(V_p^r-W_p^r)]$, which determines the probability distribution used for the sampling with replacement in the particle selection step of the algorithm.[1] Here, $1_n\cdot(V_p^r-W_p^r)$ is the number of transitions (defaults) of the Markov chain particle r during period p. For $\delta=0$, each particle has the same probability of being selected. For $\delta>0$, the selection probability increases with the number of transitions. The larger δ is, the relatively greater is the focus on particles with a larger number of transitions. In the extreme case, the particle with the largest number of transitions is selected R times.

Thus, for a positive δ the selection step favors particles with a greater number of transitions. It tends to replace particles with few transitions with those that experienced more transitions. As a result, with each selection step the particles are forced further into the regime of interest, i.e., a scenario with a large number of defaults during the simulation interval $[0,T]$. The number of time steps m determines the number of selections performed during $[0,T]$. All else being equal, the larger the m, the faster the particles transition to the rare event regime.[2] The estimator (12) accounts for the selections performed at each time step: compare with the estimator (10) generated by the A/R scheme. The required adjustment to the estimator (10) follows from formula (3.13) in [21] and is governed by the form of the weight function.

The parameters δ and m need to be chosen appropriately to guarantee variance reduction for the event of interest. However, the optimal configuration problem for a general setting

[1] Other specifications of the weight function can be envisioned. The formulation we adopted from [21] has computational (memory) advantages: it does not require us to keep track of the full path history of each particle, since only the most recent transitions are relevant for the selection.

[2] Note that m, unlike δ, has a direct impact on the memory requirement for the S/M scheme, which is $2Rn+m$ for our weight function specification.

has not yet been addressed to our knowledge. In practice, simple experiments can lead to a reasonable configuration for a given setting. We explain this in the context of our numerical case study in section 6.

There are alternative approaches to variance reduction. Relative to an importance sampling (IS) scheme, the S/M algorithm eliminates the need to determine and simulate from an importance measure under which the event of interest is not rare anymore. In the S/M scheme, we simulate the chain M under the reference measure using the A/R algorithm and the transition rate functions h^i. The chain is adaptively forced into the regime of interest by the selection mechanism. As shown by [21], the selection mechanism can be interpreted as a twisting of Feynman–Kac particle path measures, a measure change analogous to the one underpinning IS.[3] An advantage of an IS formulation is that it often enables one to establish certain asymptotic optimality properties of the IS estimator.[4] These properties formally prove the effectiveness of the estimator in a rare-event regime and lead to an optimal configuration of the algorithm. However, for the time-inhomogeneous Markov chain considered here, the optimal IS scheme has not yet been worked out, to our knowledge.[5]

5. Calculating the projection. The practical implementation of the exact simulation method requires the construction of the mimicking Markov chain M for the intensity model λ at hand. This construction amounts to the calculation of the conditional expectation $E(\lambda_t^i I(\tau^i > t) \,|\, N_t = B)$ defining the transition rates $h^i(t, B)$ of M; see Proposition 3.1. This section calculates this expectation for a range of doubly stochastic, frailty, and self-exciting models and therefore extends the scope of the exact simulation method to many models proposed in the literature. The calculation relies, intuitively speaking, on Bayes' rule and leads to an explicit expression for the transition rate $h^i(t, B)$ in terms of the transform

$$\phi(t, u, z, Z) = E\left[\exp\left(-u\int_0^t Z_s ds - zZ_t\right)\right],$$

where Z is a nonnegative stochastic process and u and z are reals. This transform can be computed in closed form for a wide range of processes Z, including affine jump-diffusion processes. For any choice of Z that we consider below, we assume that $\partial_z \phi(t, u, z, Z)|_{z=0}$ exists and is finite. Below, B denotes $(B^1, \ldots, B^n) \in \{0, 1\}^n$.

5.1. Doubly stochastic models. We begin with a simple model in which firms default independently of one another. The calculation of the corresponding Markov transition rate $h^i(t, B)$ serves as a stepping stone for the calculation in models with a nontrivial default dependence structure.

Proposition 5.1. *Suppose that N is doubly stochastic[6] with intensities $\lambda^i = X^i$ for mutually independent nonnegative adapted processes X^i. Then, for $B^i = 0$, we have*

$$h^i(t, B) = -\frac{\partial_z \phi(t, 1, z, X^i)|_{z=0}}{\phi(t, 1, 0, X^i)}.$$

[3] Carmona and Crepey [10] discuss this analogy further.

[4] See [3] for an excellent discussion.

[5] However, see [43], [44] and the references in these papers for important results in this direction. See [5] for an optimal IS scheme for doubly stochastic intensity models λ.

[6] Saying that N is doubly stochastic means that the intensity λ is a function of an adapted process Z and that, given a path of Z, the components N^i are independent inhomogeneous Poisson processes, each stopped at its first jump time and having (conditionally deterministic) intensity λ^i.

We generalize to a model in which a firm is exposed to an idiosyncratic risk factor X^i and a systematic risk factor Y that is common to all firms. The random variation of Y generates correlated movements in firms' conditional default probabilities. Conditional on a realization of Y, firms default independently of one another. This and related formulations have been used extensively in theoretical and empirical analyses of correlated default risk; see, for example, [12], [13], [19], [23], [25], [28], [30], [32], [46], [47], [51], [59], and [60].

Proposition 5.2. *Suppose that N is doubly stochastic with intensities $\lambda^i = X^i + \alpha^i Y$ for mutually independent nonnegative adapted processes $X^1, \dots, X^n, Y$ and nonnegative factor loadings α^i. Then, for $B^i = 0$, we have*

$$h^i(t,B) = -\frac{\partial_z \phi(t,1,z,X^i)|_{z=0}}{\phi(t,1,0,X^i)} - \alpha^i \frac{\sum_{k=0}^{2^n-1} c_k(t) \partial_z \phi(t,b_k,z,Y)|_{z=0}}{\sum_{k=0}^{2^n-1} c_k(t)\phi(t,b_k,0,Y)}, \tag{13}$$

where the deterministic functions $c_k(t)$ and the constants b_k are determined by the relation

$$\sum_{k=0}^{2^n-1} c_k(t) e^{-b_k v} = \prod_{j=1}^{n} \left[B^j - (2B^j - 1)\phi(t,1,0,X^j) e^{-\alpha^j v}\right], \quad v > 0. \tag{14}$$

The multiplication of the n terms on the right-hand side of (14) results in a sum of at most 2^n terms, and typically fewer as $B^j = 0$ for some j. The $c_k(t)$ are the coefficients of the summands. Each constant b_k is a sum of values α^j for certain j; note that $b_0 = 0$ and $b_{2^n-1} = \sum_{j=1}^n \alpha^j$. An algorithm for the efficient computation of the $c_k(t)$ and the b_k is based on the recursive scheme of [1].

We can extend to a doubly stochastic model with multiple common factors, allowing the description of a more sophisticated firm dependence structure.

Proposition 5.3. *Suppose that N is doubly stochastic with intensities $\lambda^i = X^i + \alpha^i \cdot Y$ for mutually independent nonnegative adapted processes $X^1, \dots, X^n$ and $Y = (Y_1, \dots, Y_q)$ and nonnegative factor loadings $\alpha^i = (\alpha^i_1, \dots, \alpha^i_q)$. Then, for $B^i = 0$, we have*

$$\begin{aligned} h^i(t,B) = &-\frac{\partial_z \phi(t,1,z,X^i)|_{z=0}}{\phi(t,1,0,X^i)} \\ &- \frac{\sum_{l=1}^q \alpha^i_l \sum_{k=0}^{2^n-1} c_k(t) \partial_{z_l} \phi(t,b_{k1},z_1,Y_1)\cdots\phi(t,b_{kq},z_q,Y_q)|_{z_1=z_2=\cdots=0}}{\sum_{k=0}^{2^n-1} c_k(t)\phi(t,b_{k1},0,Y_1)\cdots\phi(t,b_{kq},0,Y_q)}, \end{aligned} \tag{15}$$

where the deterministic functions $c_k(t)$ and the q-vector b_k of constants are determined by the relation (14), where v is a q-vector of positive constants and the products $b_k v$ and $\alpha^j v$ are interpreted as dot products.

5.2. Frailty models. Doubly stochastic models ignore the impact of a default on the intensities of the surviving firms. This impact is channeled through the network of informational and contractual relationships in the economy. For instance, for U.S. corporate defaults, Duffie et al. [29] find strong evidence for the presence of frailty, or unobservable common or correlated risk factors. The uncertainty about the value of a frailty generates an additional channel of default correlation, above and beyond the "doubly stochastic channel." It also leads to additional dynamical effects in constituent intensities, in that a default causes a jump in the

intensities of any other firms that depend on the same frailty. These jump effects are due to Bayesian learning in the reference filtration $\mathbb{F}$.

We generalize the complete information model of Proposition 5.3 to include a firm's exposure to unobservable frailty risk factors. This extends the reach of the exact method to the model specifications in a substantial frailty literature, which includes [16], [22], [29], [34], [35], [50], [57], [62], and others.

Proposition 5.4. *Suppose that, relative to a complete information filtration $\mathbb{H} \supset \mathbb{F}$, N is doubly stochastic with intensities $X^i + \alpha^i \cdot Y$ for mutually independent nonnegative processes $X^1, \dots, X^n$ and $Y = (Y_1, \dots, Y_q)$ that are adapted to $\mathbb{H}$, and nonnegative factor loadings $\alpha^i = (\alpha^i_1, \dots, \alpha^i_q)$. Assume that all but one risk factor Y_m are adapted to the observation filtration $\mathbb{F}$.*[7] *Then $h^i(t, B)$ satisfies (15).*

Proposition 5.4 states that the transition rates of the mimicking Markov chain in a model with an unobservable common risk factor Y_m agree with those in the corresponding complete information doubly stochastic model. Then, by Proposition 3.1, the distributions of the default indicator N_t must agree in these two model specifications. This may seem surprising at first: the specifications generate different intensity processes, so one would expect that they imply different distributions for N_t. However, while the intensities in the two models are different, they admit the same projections onto N. This is because the $\mathbb{F}$-intensity λ^i in the frailty model is the optional projection onto $\mathbb{F}$ of the complete information $\mathbb{H}$-intensity $X^i + \alpha^i \cdot Y$. Now the conclusion follows from iterated expectations. The proof of Proposition 5.4 formalizes this intuition.

While the presence of frailty makes no difference for the unconditional distribution $P(N_t = B)$ of the default indicator N_t, it is important to note that it does influence the conditional distributions $P(N_t = B \,|\, \mathcal{F}_s)$ for $t > s > 0$. The reason is that the sigma-fields $\mathcal{F}_s$ representing the observable information available at time s are different for frailty and complete information models. In the frailty model of Proposition 5.4, $\mathcal{F}_s$ does not contain the path of the frailty risk factor Y_s over $[0, s]$, while in the complete information model of Proposition 5.3 it does.

5.3. Self-exciting models. The impact of a default on the intensities of the surviving firms can also be attributed to contagion, by which the distress of a firm is propagated to other firms. The authors of [4] find strong evidence for the presence of contagion in U.S. corporate defaults, after controlling for other channels of default correlation, including exposure to observable and unobservable risk factors. This empirical evidence can be addressed with a self-exciting model, in which the intensity of a firm responds to the default of another firm. Formulations of this type have been considered by [17], [20], [33], [36], [37], [45], [48], [52], [53], [55], [63], and others.

Proposition 5.5. *Suppose that N has intensities $\lambda^i = X^i + c^i(\cdot, N)$, where $X = (X^1, \dots, X^2)$ solves $dX_t = \mu(t, X_t)dt + \sigma(t, X_t)dW_t + dJ_t$ for a standard Brownian motion W, a point process J with arrival intensity $\gamma(t, X_t)$, and fixed jump sizes and suitable functions (μ, σ, γ) such that the X^i are nonnegative and independent, and where $c : \mathbb{R}_+ \times \{0, 1\}^n \to \mathbb{R}_+$ is a*

[7] Saying that N is doubly stochastic relative to a complete information filtration $\mathbb{H}$ means that the doubly stochastic property holds if one artificially enlarges the reference observation filtration $\mathbb{F}$ to make all risk factor processes adapted. Note that N is not $\mathbb{F}$-doubly stochastic.

bounded function. Assume that $\phi(t,-1,0,X^i)$ *is finite. Then, for* $B^i=0$*, we have*

$$h^i(t,B)=-\frac{\partial_z\phi(t,1,z,X^i)|_{z=0}}{\phi(t,1,0,X^i)}+c^i(t,B).$$

The impact function c^i specifies the dependence of firm i's intensity on the state of the other firms in the portfolio. By convention, $c^i(\cdot,B)=g^i(\cdot,B^1,\dots,B^{i-1},B^{i+1},\dots,B^n)$ for some bounded $g^i:\mathbb{R}_+\times\{0,1\}^{n-1}\to\mathbb{R}_+$. Examples of the impact function include $c^i(t,B)=\sum_{j\neq i}\beta^{ij}(t)B^j$ for nonnegative, deterministic, and bounded functions $\beta^{ij}(t)$ that model the impact on firm i of firm j's default. A parsimonious "mean-field" model is obtained by setting $\beta^{ij}(t)=\frac{1}{n-1}$. To allow for nonlinear dependence on events, we can specify a bounded nonnegative function φ^i and take $c^i(t,B)=\varphi^i(\sum_{j\neq i}\beta^{ij}(t)B^j)$ for deterministic functions $\beta^{ij}(t)$ that are not required to be nonnegative.

The proof of Proposition 5.5 indicates that we can also treat an alternative multiplicative formulation $\lambda^i=X^ic^i(\cdot,N)$. In this case the impact function acts as a scaling to the "baseline hazard" X^i. We could take, for instance, $c^i(t,B)=\exp(\sum_{j\neq i}\beta^{ij}(t)B^j)$ for deterministic real-valued functions $\beta^{ij}(t)$.

6. Numerical results.

This section demonstrates the utility of the exact method through numerical experiments. We consider a variant of the self-exciting model treated by Proposition 5.5.

6.1. Model.

For nonnegative constants β^{ij}, consider the specification

$$\lambda^i_t=X^i_t+\sum_{j\neq i}\beta^{ij}N^j_t, \tag{16}$$

where the risk factors X^i follow mutually independent Feller diffusions:

$$dX^i_t=\kappa_i(\theta_i-X^i_t)dt+\sigma_i\sqrt{X^i_t}dW^i_t,\quad X^i_0>0. \tag{17}$$

Here, κ_i is a parameter controlling the speed of mean-reversion of X^i, θ_i is the level of mean reversion, and σ_i controls the diffusive volatility of X^i. The process $(W^1,\dots,W^n)$ is a standard Brownian motion. The parameter β^{ij} determines the impact on firm i of firm j's default. The corresponding jump terms generate correlation between the firm intensities. The matrix (β^{ij}) governs the default dependence structure.

For each constituent firm $i=1,\dots,n$, we initialize the risk factor X^i_0 at its long-run mean θ_i. The parameters are selected randomly. We draw κ_i from $U[0.5,1.5]$ and θ_i from $U[0.001,0.051]$. We take $\sigma_i=\min(\sqrt{2\kappa_i\theta_i},\bar\sigma_i)$, where $\bar\sigma_i$ is drawn from $U[0,0.2]$. We draw β^{ij} from $U[0,0.01]$ for each $j=1,\dots,n$. In practice, the parameters are calibrated from market rates of derivatives referenced on the constituent issuers and on the portfolio, as in [30] or [59].

The formulation (16)–(17) generalizes the specifications in [48], [53], and [63] to include a diffusion term that modulates the intensity of a firm between arrivals. In the absence of the diffusion term, the intensity is piecewise deterministic, so that N can be simulated exactly using the classical A/R scheme of [54]. The exact method developed here extends the reach of this scheme to the richer model (16)–(17).

6.2. Mimicking chain. Proposition 5.5 determines the rate $h^i(t,B)$ of the mimicking Markov chain in terms of the risk factor transform ϕ, its partial derivative $\partial_z\phi$, and the parameters β^{ij}. The transform ϕ takes the well-known exponentially affine form

$$\phi(t,u,z,X^i) = \exp(a^i(t,u,z) + b^i(t,u,z)X_0^i), \tag{18}$$

where, for $\gamma_i = \sqrt{\kappa_i^2 + 2\sigma_i^2 u}$, the coefficient functions

$$a^i(t,u,z) = \frac{2u(1-\exp(\gamma_i t)) - z(\gamma_i + \kappa_i + (\gamma_i - \kappa_i)\exp(\gamma_i t))}{\sigma_i^2 z(\exp(\gamma_i t)-1) + \gamma_i - \kappa_i + (\gamma_i + \kappa_i)\exp(\gamma_i t)}, \tag{19}$$

$$b^i(t,u,z) = \frac{2\kappa_i\theta_i}{\sigma_i^2}\log\frac{2\gamma_i\exp((\gamma_i+\kappa_i)t/2)}{\sigma_i^2 z(\exp(\gamma_i t)-1) + \gamma_i - \kappa_i + (\gamma_i+\kappa_i)\exp(\gamma_i t)}. \tag{20}$$

The derivative $\partial_z\phi$ of ϕ is also available in closed form. Proposition 5.5 then implies, for $B^i = 0$ and evaluating γ_i at $u = 1$, the formula

$$h^i(t,B) = \frac{4X_0^i\gamma_i^2\exp(\gamma_i t)}{(\gamma_i - \kappa_i + (\gamma_i+\kappa_i)\exp(\gamma_i t))^2} - \frac{\theta_i\kappa_i}{\sigma_i^2}\frac{(\kappa_i^2-\gamma_i^2)(\exp(\gamma_i t)-1)}{\gamma_i - \kappa_i + (\gamma_i+\kappa_i)\exp(\gamma_i t)} + \sum_{j\neq i}\beta^{ij}B^j.$$

With parameter values selected as explained above, the function $h^i(t,\cdot)$ is decreasing. This suggests an adaptive rule for setting the bound $J(k)$ and the interval length $K(k)$ in step 3 of Algorithm 4.1. The first candidate time S for T_{k+1} is generated using $H(T_k,k) = \sum_{i=1}^n h^i(T_k, M_{T_k})$ as a bound, where the interval length is taken to be the time to the simulation horizon. If that time is rejected, we generate the next candidate time using the value $H(S,k)$ as a bound, taking the interval length to be the remaining time to the simulation horizon. The value $H(S,k)$ is computed in any case for the acceptance test. We proceed according to this rule until a candidate time is accepted.

6.3. Estimators. We contrast the estimators generated by the exact scheme with those generated by the time-scaling method described in section 2.2. The time-scaling method requires paths of the continuous-time stochastic processes $\int_0^t \lambda_s^i ds$ for $i = 1,\dots,n$, where the λ^i follow jump-diffusion processes that are correlated through common jumps. We must discretize the time interval and simulate the joint integral process dynamics on this discrete-time grid. Since the joint law of the integrals is not known, we first simulate the joint intensity dynamics on the discrete-time grid and then integrate. To generate the values of λ^i, we generate the values of X^i by sampling from the noncentral chi-squared distribution that describes the transition law of X^i, and add the value β^{ij} when another firm j defaults. While the sampling from the chi-squared distribution leads to exact values of the X^i, it tends to be more time-consuming than an alternative Euler scheme. On the other hand, the Euler scheme introduces additional discretization bias because it does not facilitate exact sampling. To analyze the trade-off between computation time and bias, we implement both the exact sampling from the chi-squared transition law and the modified Euler scheme in [39, (3.66)].

To compare the estimators generated by the different simulation methods, we consider the root mean square error (RMSE), given by $\sqrt{\mathrm{SE}^2 + \mathrm{Bias}^2}$. The standard error SE is estimated as the sample standard deviation of the simulation output divided by the square root of the

Table 2

Simulation results under the self-exciting model (16)–(17) for $E[(C_1 - 3)^+]$. "Time-scaling (χ^2)" refers to the time-scaling method using the exact sampling of the values of X^i from the noncentral transition law. "Time-scaling (Euler)" refers to the time-scaling method using the modified Euler scheme to sample the X^i. The true value was estimated to be 1.0145, based on 5,000,000 trials with the exact scheme.

Method	Trials	Steps	Estimate	Bias	SE	RMSE	Time (min)
Exact A/R	5K	N/A	1.0394	0	0.0239	0.0239	0.10
	7.5K	N/A	1.0247	0	0.0193	0.0193	0.15
	10K	N/A	0.9854	0	0.0165	0.0165	0.20
	50K	N/A	1.0161	0	0.0073	0.0073	1.01
	100K	N/A	1.0083	0	0.0052	0.0052	2.03
	1M	N/A	1.0138	0	0.0016	0.0016	20.29
Time-scaling (χ^2)	1K	32	0.8744	0.0893	0.0650	0.1104	21.20
	2.5K	50	0.9400	0.0796	0.0420	0.0899	85.30
	5K	71	1.0538	0.0735	0.0246	0.0775	257.30
	7.5K	87	1.0873	0.0697	0.0199	0.0725	483.50
	10K	100	1.0805	0.0660	0.0171	0.0682	767.60
Time-scaling (Euler)	1K	32	0.8570	0.2094	0.0474	0.2147	4.83
	2.5K	50	0.8448	0.1763	0.0277	0.1785	12.27
	5K	71	0.9254	0.1324	0.0232	0.1344	36.57
	7.5K	87	0.9394	0.1070	0.0193	0.1087	86.54
	10K	100	0.9354	0.0860	0.0186	0.0880	194.27

number of trials. The bias is given by the difference between the expectation of the estimator and the true value. The bias of the estimator generated by the exact method is zero. The bias of the estimator generated by the time-scaling method with a specific number of time steps can be estimated using a large number of trials to estimate the expectation of the estimator, and then taking the difference with the true value, estimated using the exact method with a large number of trials.

We estimate the (undiscounted) value of a call $E[(C_T - K)^+]$ on the default count at T. This option is a basic building block for the valuation of portfolio credit derivatives; see [31]. We take the number of reference names $n = 100$, which is the standard portfolio size for many traded portfolio derivatives, $T = 1$ year, and $K = 3$. Table 2 reports the simulation results. The bias in the table is estimated using 50,000 trials. The number of discretization time steps in the time-scaling method is set equal to the square-root of the number of simulation trials.[8] The experiments were performed on a desktop PC with an Intel 3.4 GHz processor and 1 GB of RAM, running Windows XP Professional. The methods were implemented in MATLAB.

Figure 1, which shows the convergence of the RMSE graphically, indicates the substantial performance advantages of the exact method. The exact method requires the shortest computation time to achieve a given accuracy. It also has the fastest convergence rate. This rate is of order $O(1/\sqrt{t})$, where t is the computation time, the optimal rate of unbiased schemes.

[8]It is unclear how to allocate the computational budget of the time-scaling method between the number of time steps and the number of replications. The square-root rule is adopted from [9] and others. It is motivated by the results in [26], which show that for first order methods it is asymptotically optimal to increase the number of time steps in a manner proportional to the square root of the number of replications. However, the optimal constant of proportionality is not known.

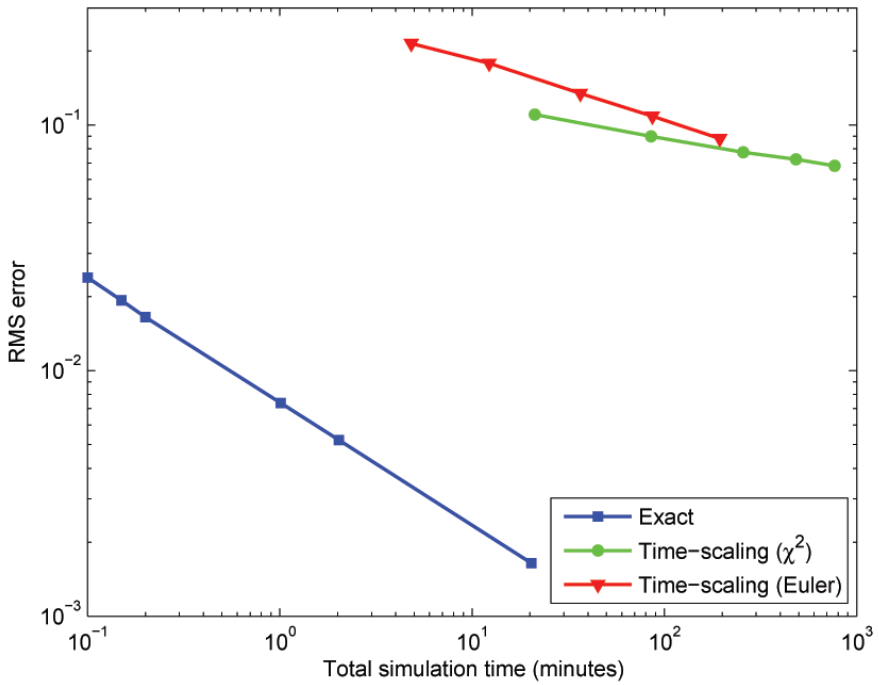

Figure 1. *Convergence of the RMSEs of the exact A/R and time-scaling methods under the self-exciting model (16)–(17) for $E[(C_1 - 3)^+]$.*

There is an alternative exact scheme for the model (16)–(17). This scheme is based on the repeated application of the first-to-default time and identity simulation algorithm developed by [24]. This scheme requires greater computational effort than the A/R scheme, which is easily explained in case $n = 2$. In the first-to-default scheme, one starts by generating $T_1 = \tau^1 \wedge \tau^2$ by the inverse transform method from $P(T_1 > t) = \phi(t, 1, 0, X^1)\phi(t, 1, 0, X^2)$. Given T_1, one then samples the identity of the first defaulter from the discrete distribution defined by $\gamma(T_1, i)/(\gamma(T_1, 1) + \gamma(T_1, 2))$ for $i = 1, 2$. Here $\gamma(t, 1) = -\phi(t, 1, 0, X^2)\partial_z\phi(t, 1, z, X^1)|_{z=0}$; a similar expression holds for $\gamma(t, 2)$. This second step is similar to step 5 in Algorithm 4.1. In a third step, one draws the second default time T_2. To do this, note that, given $\mathcal{F}_{T_1}$, the time $T_2 - T_1$ is equal in law to the first jump time of a doubly stochastic Poisson process started at T_1, with $\mathbb{F}$-intensity $(X_t^2 + \beta^{21})_{t \geq T_1}$ (this assumes $T_1 = \tau^1$). Thus one needs to sample from $P(T_2 - T_1 > t \,|\, \mathcal{F}_{T_1}) = E(\exp(-\int_{T_1}^{T_1+t}(X_s^2 + \beta^{21})ds) \,|\, X_{T_1}^2, T_1)$. This can be done by the inverse method, once $X_{T_1}^2$ is drawn from the conditional distribution of X_t^2 given $T_1 = \tau^1 = t$. However, as shown by [24], this conditional distribution is known only in terms of its transform, making it relatively costly to sample from it. The A/R scheme avoids this third step and also tends to require less computational effort to generate the T_i.

6.4. Variance reduction. We demonstrate the effectiveness of the S/M Algorithm 4.2 for variance reduction under the self-exciting model (16)–(17). We estimate $P(C_T = k)$ for the test portfolio described in section 6.1. To measure the variance reduction offered by Algorithm 4.2, we compute variance ratios for each of several values of k. Each variance ratio is calculated

by estimating the variance of the estimator generated by the plain A/R Algorithm 4.1 and dividing it by the estimated variance of the estimator generated by Algorithm 4.2. The variance is estimated by the sample variance of the simulation output.[9]

To run Algorithm 4.2, we need to select the number of particles R, the number of selections m performed during $[0,T]$, and the parameter δ of the selection probability (11). The number of particles R is analogous to the number of trials in a standard simulation. In practice, this quantity is determined by the desired accuracy of the estimator, or the available computational budget, for fixed m. However, the theoretically optimal allocation of the computational budget between m and R has not yet been worked out to our knowledge. Therefore, we will illustrate the influence of R and m on the estimator through experiments, treating each variable separately.

Also the choice of δ is difficult. Intuitively, we want to pick δ so as to minimize the relative error of the estimator, given by the variance of the estimator divided by the product of the estimator and the square root of the number of simulation trials. Again, the theoretically optimal choice of δ has not yet been worked out, to our knowledge. Therefore, we approximate the optimal δ through experiments. Specifically, we discretize a range of values of δ, run the simulation for each grid value using a small number of particles, and select the value of δ that produces the smallest relative error. The δ so chosen increases with the target event count k, because the selections must place greater weight on particles with transitions, the smaller the probability of interest.

Table 3 shows the results for $T = 1$, $m = 4$ selections, and $R = 10{,}000$ particles. Table 4 reports the results when only $R = 1{,}000$ particles are used. To provide a meaningful comparison between the two methods, the number of trials in the estimation using the plain A/R Algorithm 4.1 is chosen such that the total time required to estimate $P(C_T = k)$ is approximately the same as that required by the S/M Algorithm 4.2 for the given R and m (excluding the fixed time it takes to select δ). We see that the smaller the probability, the larger the variance ratio. Further, for sufficiently small probabilities, the smaller the number of particles R, the higher the variance ratio. Figure 2 graphs the estimated probabilities $P(C_1 = k)$ reported in Tables 3 and 4. It indicates the relative benefits of the S/M scheme for each value of R.

Next we analyze the role of the number of selections m. Figure 3 graphs the estimated variance ratios for each of several values of m, fixing the number of particles $R = 1{,}000$. The variance ratio increases with the number of selections if the probability of interest is only moderately small. For smaller event probabilities, this may not be the case anymore. The intuition is as follows. The computation time required by the S/M scheme increases with m. The increase in computation time is relatively larger the larger k becomes, i.e., the smaller the probability of interest. Now, by the design of our experiments, the number of trials that can be completed by the plain A/R scheme increases with m and k, so that the absolute variance of the corresponding estimator may decrease faster than the variance of the estimator generated by the S/M scheme. However, note that, for a fixed number of particles, increasing m always reduces the variance of the simulation estimator in absolute terms. This

[9] In the case of Algorithm 4.2, the sample variance is a biased estimator of the variance. This is because the samples are not independent due to the selections performed in the scheme.

Table 3

Variance reduction ratios for estimating $P(C_1 = k)$ under the self-exciting model (16)–(17). In the S/M Algorithm 4.2, $m = 4$, and δ is chosen such that the relative error of the simulation estimator is minimized. The number of trials in the estimation using the plain A/R Algorithm 4.1 is chosen such that the total time required to estimate $P(C_T = k)$ is approximately the same as that required by the S/M scheme, excluding the fixed time it takes to select δ. (*) indicates that the event of interest was not observed in any of the trials.

	S/M scheme				Plain A/R scheme		
k	δ	Particles	Estimate	SE	Trials	Estimate	VarRatio
8	0.55	10,000	0.02364596	0.00096165	15,806	0.02240	2.54
9	0.7	10,000	0.01253552	0.00065596	16,452	0.01349	3.08
10	0.75	10,000	0.00587429	0.00044357	16,452	0.00626	3.16
11	0.8	10,000	0.00337833	0.00020285	17,742	0.00293	7.10
12	0.8	10,000	0.00162340	0.00014706	17,742	0.00220	10.14
13	0.85	10,000	0.00068818	0.00007336	18,065	0.00066	12.33
14	0.85	10,000	0.00029433	0.00002062	18,387	0.00027	63.88
15	1.05	10,000	0.00016310	0.00001609	19,032	0.00011	121.80
16	1.05	10,000	0.00006790	0.00000471	19,032	0.00005	236.92
17	1.15	10,000	0.00002597	0.00000352	19,355	(*)	
18	1.15	10,000	0.00000970	0.00000171	19,355	(*)	
19	1.15	10,000	0.00000500	0.00000054	19,355	(*)	
20	1.15	10,000	0.00000203	0.00000020	19,355	(*)	
21	1.15	10,000	0.00000106	0.00000008	19,355	(*)	
22	1.3	10,000	0.00000039	0.00000005	19,677	(*)	

Table 4

Variance reduction ratios for estimating $P(C_1 = k)$ under the self-exciting model (16)–(17). In the S/M Algorithm 4.2, $m = 4$, and δ is chosen such that the relative error of the simulation estimator is minimized. The number of trials in the estimation using the plain A/R Algorithm 4.1 is chosen such that the total time required to estimate $P(C_1 = k)$ is approximately the same as that required by the S/M scheme, excluding the fixed time it takes to select δ. (*) indicates that the event of interest was not observed in any of the trials.

	S/M scheme				Plain A/R scheme		
k	δ	Particles	Estimate	SE	Trials	Estimate	VarRatio
8	0.55	1,000	0.02433732	0.00310121	1,343	0.02010	2.05
9	0.7	1,000	0.01110821	0.00156666	1,407	0.00498	2.02
10	0.75	1,000	0.00738774	0.00108022	1,439	0.00625	5.33
11	0.8	1,000	0.00340511	0.00083321	1,600	0.00500	7.16
12	0.8	1,000	0.00156213	0.00029693	1,600	0.00125	14.16
13	0.85	1,000	0.00073482	0.00019565	1,600	0.00125	32.62
14	0.85	1,000	0.00024352	0.00005751	1,600	0.00188	565.20
15	1.05	1,000	0.00009061	0.00001921	1,726	0.00058	1562.75
16	1.05	1,000	0.00009384	0.00001385	1,759	0.00057	2951.99
17	1.15	1,000	0.00006344	0.00000372	1,790	(*)	
18	1.15	1,000	0.00002131	0.00000152	1,823	(*)	
19	1.15	1,000	0.00000971	0.00000091	1,887	(*)	
20	1.15	1,000	0.00000041	0.00000014	1,887	(*)	
21	1.15	1,000	0.00000082	0.00000008	1,918	(*)	
22	1.3	1,000	0.00000031	0.00000004	1,983	(*)	

is indicated by Figure 4, which graphs the estimated probabilities $P(C_1 = k)$ for each of several values of m, for fixed $R = 1{,}000$.

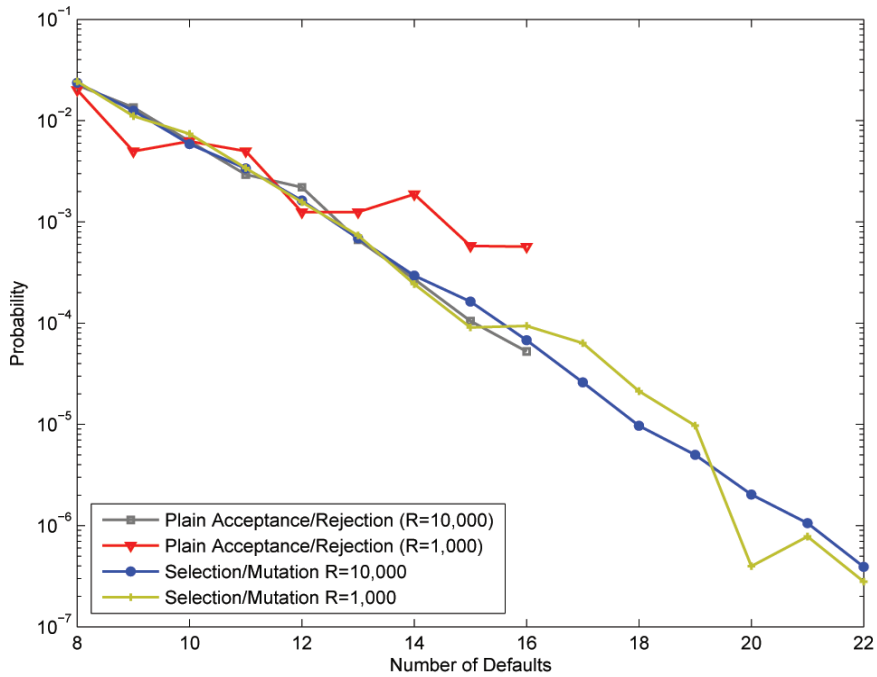

Figure 2. *Estimated probabilities $P(C_1 = k)$ under the self-exciting model (16)–(17) for each of several values of R. The number of selections $m = 4$.*

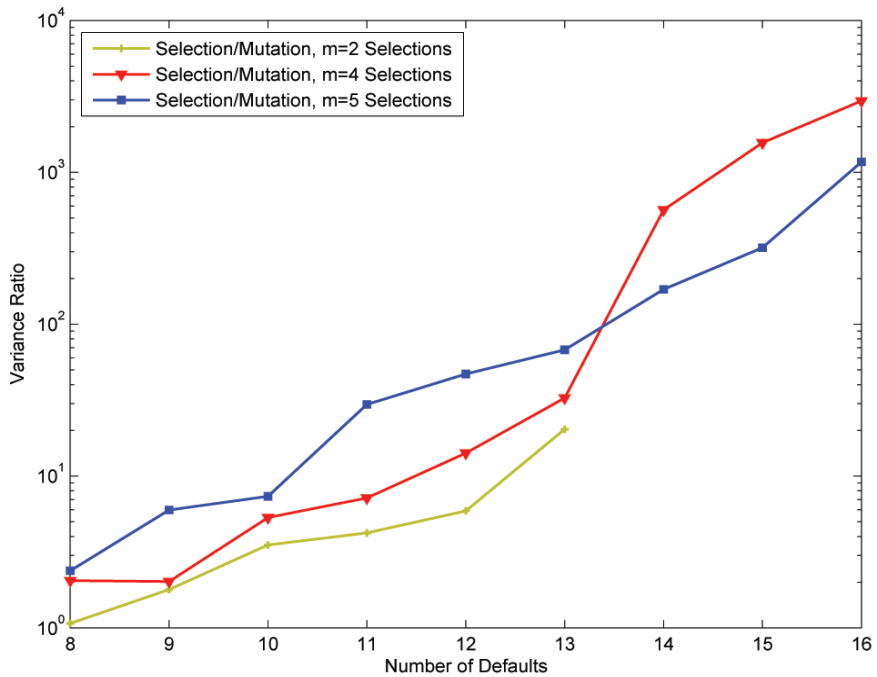

Figure 3. *Estimated variance reduction ratios under the self-exciting model (16)–(17) for each of several values of m. The number of particles $R = 1{,}000$.*

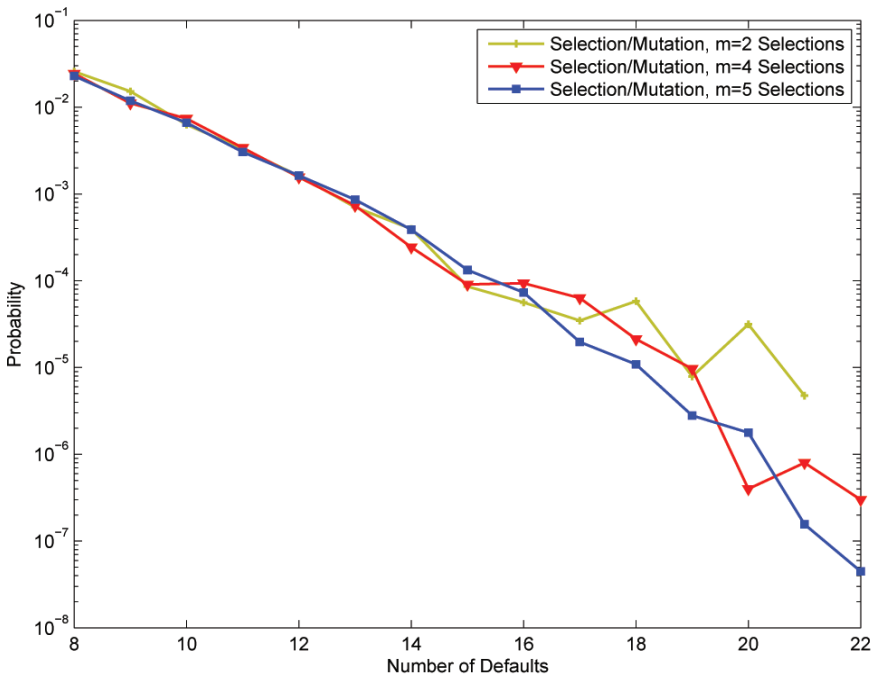

Figure 4. *Estimated probabilities $P(C_1 = k)$ under the self-exciting model (16)–(17) for each of several values of m. The number of particles $R = 1{,}000$.*

6.5. Potential extensions. There are several potential variations and extensions of the formulation (16)–(17). As explained in section 5.3, the additive specification of the feedback term in the intensity dynamics (16) could be replaced by a multiplicative specification. This would lead to different self-exciting dynamics and a greater degree of flexibility in designing the feedback behavior, without reducing the analyticity of h^i. The Feller diffusion risk factor dynamics (17) could be extended to include a compound Poisson jump term. This extension would allow for discontinuous movements of the intensity between defaults, while requiring only a minor modification of the coefficient functions (20) and (19) based on the results of [25]. The rate h^i would still take a closed form. More generally, the Feller diffusion dynamics (17) could be replaced by more general affine jump-diffusion dynamics. The transform (18) would remain exponentially affine in the state, while the coefficient functions would be given as solutions to a system of ODEs.

7. Conclusion. This paper develops a simulation method for dynamic intensity-based models of correlated default risk. The method generates unbiased estimators of credit portfolio loss distributions, risk measures, and prices of derivative securities that are referenced on a portfolio of defaultable assets. It reduces the simulation problem to one of a simple Markov chain expectation. This problem can be treated with exact methods. An overarching selection/mutation scheme reduces variance in rare-event situations. The method is widely applicable to many intensity models in the literature. Numerical experiments demonstrate its effectiveness and highlight its advantages over alternative methods.

The simulation method has potential applications in other areas that deal with the arrival of correlated events. These include, in particular, applications in reliability, where intensity-based models have long been used to analyze the reliability of systems of interdependent components whose failure times are correlated.

Appendix. Proofs.

Proof of Proposition 3.1. For all $B \in \{0,1\}^n$ and $t > 0$, we have

$$\begin{aligned} I(N_t = B) &= I(N_0 = B) + \sum_{0<s\le t} \big[I(N_s = B) - I(N_{s-} = B)\big] \\ &= I(N_0 = B) + \sum_{i=1}^{n} \int_0^t \big[I(N_s = B) - I(N_{s-} = B)\big] dN_s^i \\ &= I(N_0 = B) + \sum_{i=1}^{n} \int_0^t \big[I(N_{s-} + e^i = B) I(B^i = 1) - I(N_{s-} = B)\big] dN_s^i, \end{aligned} \tag{21}$$

where e^i is an n-vector with ith component equal to 1 and the rest equal to 0. Next we recall the Doob–Meyer decomposition (1) of the submartingale N^i into the sum of a martingale M^i and a process $\int_0^{\cdot} \lambda_s^i (1 - N_s^i) ds$. Since the integrand in the integral

$$\int_0^t \big[I(N_{s-} + e^i = B) I(B^i = 1) - I(N_{s-} = B)\big] dM_s^i$$

is bounded and predictable, the integral defines a martingale with initial value equal to zero. Thus, after we take expectation on both sides of (21) and apply Fubini's theorem, we obtain

$$P(N_t = B) = P(N_0 = B) + \sum_{i=1}^{n} \int_0^t E\left(\big[I(N_s = B - e^i) I(B^i = 1) - I(N_s = B)\big] \lambda_s^i (1 - N_s^i)\right) ds.$$

Now we differentiate both sides of this equation with respect to t. By the definition of the deterministic functions $h^i(t, B)$, we obtain

$$\partial_t P(N_t = B) = \sum_{i=1}^{n} I(B^i = 1) P(N_t = B - e^i) h^i(t, B - e^i) - \sum_{i=1}^{n} P(N_t = B) h^i(t, B).$$

But this equation coincides with the backward Kolmogorov equation,

$$\partial_t P(M_t = B) = \sum_{i=1}^{n} I(B^i = 1) P(M_t = B - e^i) h^i(t, B - e^i) - \sum_{i=1}^{n} P(M_t = B) h^i(t, B),$$

which describes the time evolution of the distribution of the Markov chain M with transition rates $h^i(t, B)$. Thus, the probabilities $P(M_t = B)$ and $P(N_t = B)$ satisfy the same ODE. Since the solution to this ODE is unique and $M_0 = N_0 = 0_n$ by construction, we conclude that M_t and N_t must have the same distribution for all $t \ge 0$. ∎

Proof of Proposition 5.1. By the independence of the random variables τ^i, we get

$$h^i(t, B) = E(X_t^i (1 - N_t^i) \,|\, N_t^i = B^i) = (1 - B^i) E(X_t^i \,|\, \tau^i > t).$$

By Bayes' formula and iterated expectations,

$$E(X_t^i \,|\, \tau^i > t) = \frac{E(X_t^i I(\tau^i > t))}{P(\tau^i > t)} = \frac{E(X_t^i P(\tau^i > t \,|\, (X_s^i)_{s\le t}))}{E(P(\tau^i > t \,|\, (X_s^i)_{s\le t}))}. \tag{22}$$

The doubly stochastic property of τ^i implies that

$$P(\tau^i > t \,|\, (X_s^i)_{s\le t}) = \exp\left(-\int_0^t X_s^i ds\right),$$

whose expectation equals $\phi(t,1,0,X^i)$. This gives the denominator of the right-hand side of (22). For the numerator,

$$E(X_t^i P(\tau^i > t \,|\, (X_s^i)_{s\le t})) = E\left(X_t^i \exp\left(-\int_0^t X_s^i ds\right)\right) = -\partial_z \phi(t,1,z,X^i)|_{z=0},$$

and this completes the proof. ∎

Proof of Proposition 5.2. By the independence of the idiosyncratic factors X^i,

$$h^i(t,B) = (1-B^i)E(X_t^i \,|\, \tau^i > t) + (1-B^i)\alpha^i E(Y_t \,|\, N_t = B),$$

where the first summand is treated as in the proof of Proposition 5.1. It remains to calculate the second summand. By Bayes' formula and iterated expectations,

$$E(Y_t \,|\, N_t = B) = \frac{E(Y_t I(N_t = B))}{P(N_t = B)} = \frac{E(Y_t P(N_t = B \,|\, (Y_s)_{s\le t}))}{E(P(N_t = B \,|\, (Y_s)_{s\le t}))}.$$

Given a path of the common factor Y, the intensities λ^i are independent of one another, and so are the components N^i of the process N. Thus,

$$\begin{aligned} P(N_t = B \,|\, (Y_s)_{s\le t}) &= \prod_{j=1}^n P(N_t^j = B^j \,|\, (Y_s)_{s\le t}) \\ &= \prod_{j=1}^n \left[B^j - (2B^j - 1)P(\tau^j > t \,|\, (Y_s)_{s\le t})\right]. \end{aligned} \tag{23}$$

By iterated expectations, the doubly stochastic property, and the independence of the processes Y and X^j, we get

$$\begin{aligned} P(\tau^j > t \,|\, (Y_s)_{s\le t}) &= E\left[\exp\left(-\int_0^t (\alpha^j Y_s + X_s^j) ds\right) \,\middle|\, (Y_s)_{s\le t}\right] \\ &= \exp\left(-\alpha^j \int_0^t Y_s ds\right) E\left[\exp\left(-\int_0^t X_s^j ds\right)\right] \\ &= \exp\left(-\alpha^j \int_0^t Y_s ds\right) \phi(t,1,0,X^j). \end{aligned} \tag{24}$$

Now, expanding the n-fold product on the right-hand side of (23) and using (24), we see that there are deterministic functions $c_k(t)$ and constants b_k satisfying (14) such that

$$P(N_t = B \,|\, (Y_s)_{s\le t}) = \sum_{k=0}^{2^n-1} c_k(t) \exp\left(-b_k \int_0^t Y_s ds\right).$$

Taking expectation on both sides of this equation leads to

$$P(N_t = B) = \sum_{k=0}^{2^n-1} c_k(t)\phi(t, b_k, 0, Y).$$

A similar argument is applied to $E(Y_t P(N_t = B \,|\, (Y_s)_{s\le t}))$. ∎

Proof of Proposition 5.3. Apply the argument used in the proof of Proposition 5.2. ∎

Proof of Proposition 5.4. For clarity in the exposition, we consider the case $q = 2$ and suppose that the observation filtration $\mathbb{F}$ is the right-continuous and complete filtration generated by the processes $X^1, \dots, X^n$, N, and Y_1. Thus, the common factor Y_1 is observable (adapted to the filtration $\mathbb{F}$), while the common factor Y_2 is a frailty (not adapted to $\mathbb{F}$). An intensity λ^i of N^i relative to the observation filtration $\mathbb{F}$ is given by the optional projection (see [61, Chapter VI, p. 375]) of the complete information intensity $X^i + (\alpha_1^i, \alpha_2^i) \cdot (Y_1, Y_2)$ onto $\mathbb{F}$. Since X^i and Y_1 are adapted to $\mathbb{F}$, we have

$$\lambda_t^i = X_t^i + \alpha_1^i Y_{1t} + \alpha_2^i E(Y_{2t} \,|\, \mathcal{F}_t)$$

almost surely, for each $t \ge 0$. Then

$$h^i(t, B) = (1 - B^i)\big\{E(X_t^i \,|\, N_t = B) + \alpha_1^i E(Y_{1t} \,|\, N_t = B) + \alpha_2^i E(U_t \,|\, N_t = B)\big\},$$

where $U_t = E(Y_{2t} \,|\, \mathcal{F}_t)$. Since the X^i are independent of one another, the first expectation on the right-hand side of this equation can be analyzed as in the proof of Proposition 5.1. The second expectation is treated as in the proof of Proposition 5.2, by conditioning on the path of $Y = (Y_1, Y_2)$ over $[0, t]$ and using the doubly stochastic property of N in the complete information filtration. The third expectation can be calculated by an analogous conditioning argument by noting that

$$E(U_t \,|\, N_t = B) = E(E(Y_{2t} \,|\, \mathcal{F}_t) \,|\, N_t = B) = E(Y_{2t} \,|\, N_t = B)$$

since $\sigma(N_t) \subset \mathcal{F}_t$. The sum of these three expectations gives (15) for $q = 2$. ∎

Proof of Proposition 5.5. We have

$$h^i(t, B) = (1 - B^i)\{E(X_t^i \,|\, N_t = B) + c^i(t, B)\}.$$

To calculate the conditional expectation, we apply a measure change argument developed by [17]. For clarity in the exposition, we consider the case $n = 2$ and take $c^i(\cdot, 0_2) = 0$. The general case can be treated by the same argument. We have

$$E(X_t^i \,|\, N_t = 0_2) = \frac{E(X_t^i I(N_t = 0_2))}{P(N_t = 0_2)} = \frac{E^*(X_t^i \exp(-\int_0^t (\lambda_s^1 + \lambda_s^2) ds))}{E^*(\exp(-\int_0^t (\lambda_s^1 + \lambda_s^2) ds))},$$

where E^* denotes the expectation operator relative to the absolutely continuous probability measure P^* on $\mathcal{F}_t$ defined by the density

$$\frac{dP^*}{dP} = I(N_t = 0_2) \exp\left(\int_0^t (\lambda_s^1 + \lambda_s^2) ds\right).$$

The condition $\phi(t, -1, 0, X^i) < \infty$ guarantees that P^* is well-defined. Under P^*, the event $\{N_t = 0_2\}$ has measure 1. Girsanov's theorem for absolutely continuous measure changes as in [61, Chapter III.8, Theorem 41] along with Lévy's theorem imply that the standard P-Brownian motion W driving the X^i's remains a standard Brownian motion under P^* on $[0, t]$, relative to the filtration $\mathbb{F}$ augmented by the P^*-null sets. The intensity of the point process J remains $\gamma(t, X_t)$ relative to P^* and the augmented filtration, because the components of J do not have jumps in common with the components of N almost surely. Therefore, the dynamics of the X^i are invariant under the measure change. Taking $i = 1$, it follows that

$$\begin{aligned} E^*\left(X_t^1 \exp\left(-\int_0^t (\lambda_s^1 + \lambda_s^2) ds\right)\right) &= E\left(X_t^1 \exp\left(-\int_0^t X_s^1 ds\right)\right) E\left(\exp\left(-\int_0^t X_s^2 ds\right)\right) \\ &= -\partial_z \phi(t, 1, z, X^1)|_{z=0} \phi(t, 1, 0, X^2). \end{aligned}$$

An analogous expression holds for $i = 2$. Similarly,

$$E^*\left(\exp\left(-\int_0^t (\lambda_s^1 + \lambda_s^2) ds\right)\right) = \phi(t, 1, z, X^1)\phi(t, 1, 0, X^2),$$

implying that

$$E(X_t^i \mid N_t = 0_2) = -\frac{\partial_z \phi(t, 1, z, X^i)|_{z=0}}{\phi(t, 1, 0, X^i)}. \tag{25}$$

Next, we use a similar argument to calculate

$$E(X_t^i \mid N_t^i = 0) = \frac{E(X_t^i I(N_t^i = 0))}{P(N_t^i = 0)} = \frac{E^i(X_t^i \exp(-\int_0^t \lambda_s^i ds))}{E^i(\exp(-\int_0^t \lambda_s^i ds))},$$

where E^i denotes the expectation operator relative to the absolutely continuous probability measure P^i on $\mathcal{F}_t$ defined by the density

$$\frac{dP^i}{dP} = I(N_t^i = 0) \exp\left(\int_0^t \lambda_s^i ds\right).$$

Under P^i, the event $\{N_t^i = 0\}$ has measure 1. As reasoned in the case of P^*, Girsanov's theorem implies that the dynamics of (X^1, X^2) are invariant under the measure change. Taking $i = 1$ and recalling that $c^1(s, N_s)$ takes the form $g^1(s, N_s^2)$,

$$\begin{aligned} E^1\left(X_t^1 \exp\left(-\int_0^t \lambda_s^1 ds\right)\right) &= E^1\left(X_t^1 \exp\left(-\int_0^t X_s^1 ds - \int_0^t c^1(s, N_s) ds\right)\right) \\ &= -\partial_z \phi(t, 1, z, X^1)|_{z=0} E^1\left(\exp\left(-\int_0^t g^1(s, N_s^2) ds\right)\right). \end{aligned}$$

An analogous expression holds for $i = 2$. Applying a similar argument to the expectation $E^i(\exp(-\int_0^t \lambda_s^i ds))$, we find that $E(X_t^i \,|\, N_t^i = 0) = E(X_t^i \,|\, N_t = 0_2)$. Bayes' rule, along with this relation, then shows that also $E(X_t^1 \,|\, N_t = (0,1)) = E(X_t^1 \,|\, N_t^1 = 0)$ and $E(X_t^2 \,|\, N_t = (1,0)) = E(X_t^2 \,|\, N_t^2 = 0)$. Thus, we have shown that

$$E(X_t^i \,|\, N_t = B) = -\frac{\partial_z \phi(t,1,z,X^i)|_{z=0}}{\phi(t,1,0,X^i)}$$

for all $B \in \{0,1\}^2$ with $B^i = 0$. ∎

Acknowledgments. We thank seminar participants for their comments, and Apaar Sadhwani for research assistance. We also thank Jose Blanchet, Rama Cont, Jean-Pierre Fouque, Josselin Garnier, Paul Glasserman, Martin Haugh, Baeho Kim, and Steve Kou for comments. We are grateful to two anonymous referees for their insightful reviews.

REFERENCES

[1] L. Andersen, J. Sidenius, and S. Basu, *All your hedges in one basket*, Risk, 16 (2003), pp. 67–72.
[2] M. Arnsdorf and I. Halperin, *BSLP: Markovian bivariate spread-loss model for portfolio credit derivatives*, J. Comput. Finance, 12 (2008), pp. 77–100.
[3] S. Asmussen and P. Glynn, *Stochastic Simulation: Algorithms and Analysis*, Springer-Verlag, New York, Berlin, 2007.
[4] S. Azizpour, K. Giesecke, and G. Schwenkler, *Exploring the Sources of Default Clustering*, working paper, Department of Management Science and Engineering, Stanford University, Palo Alto, CA, 2008.
[5] A. Bassamboo and S. Jain, *Efficient importance sampling for reduced form models in credit risk*, in Proceedings of the 2006 Winter Simulation Conference, L. F. Perrone, F. P. Wieland, J. Liu, B. G. Lawson, D. M. Nicol, and R. M. Fujimoto, eds., IEEE Press, Piscataway, NJ, 2006, pp. 741–748.
[6] A. Bassamboo, S. Juneja, and A. Zeevi, *Portfolio credit risk with extremal dependence: Asymptotic analysis and efficient simulation*, Oper. Res., 56 (2008), pp. 593–606.
[7] A. Bentata and R. Cont, *Mimicking the Marginal Distributions of a Semimartingale*, working paper, Columbia University, New York, 2009.
[8] P. Brémaud, *Point Processes and Queues—Martingale Dynamics*, Springer-Verlag, New York, 1980.
[9] M. Broadie and O. Kaya, *Exact simulation of stochastic volatility and other affine jump diffusion processes*, Oper. Res., 54 (2006), pp. 217–231.
[10] R. Carmona and S. Crepey, *Particle methods for the estimation of Markovian credit portfolio loss distributions*, Int. J. Theor. Appl. Finance, (2010), to appear.
[11] R. Carmona, J.-P. Fouque, and D. Vestal, *Interacting particle systems for the computation of rare credit portfolio losses*, Finance Stoch., 13 (2009), pp. 613–633.
[12] A. Chapovsky, A. Rennie, and P. Tavares, *Stochastic intensity modeling for structured credit exotics*, Int. J. Theor. Appl. Finance, 10 (2007), pp. 633–652.
[13] S. Chava and R. Jarrow, *Bankruptcy prediction with industry effects*, Rev. Finance, 8 (2004), pp. 537–569.
[14] Z. Chen and P. Glasserman, *Fast pricing of basket default swaps*, Oper. Res., 56 (2008), pp. 286–303.
[15] Z. Chen and P. Glasserman, *Sensitivity estimates for portfolio credit derivatives using Monte Carlo*, Finance Stoch., 12 (2008), pp. 507–540.
[16] P. Collin-Dufresne, R. Goldstein, and J. Helwege, *How Large Can Jump-to-Default Risk Premia Be? Modeling Contagion via the Updating of Beliefs*, working paper, Carnegie Mellon University, Pittsburgh, PA, 2009.
[17] P. Collin-Dufresne, R. Goldstein, and J. Hugonnier, *A general formula for the valuation of defaultable securities*, Econometrica, 72 (2004), pp. 1377–1407.

[18] R. Cont and A. Minca, *Extracting Portfolio Default Rates from CDO Spreads*, working paper, Columbia University, New York, 2008.
[19] S. Das, D. Duffie, N. Kapadia, and L. Saita, *Common failings: How corporate defaults are correlated*, J. Finance, 62 (2007), pp. 93–117.
[20] M. Davis and V. Lo, *Modeling default correlation in bond portfolios*, in Mastering Risk, Volume 2: Applications, C. Alexander, ed., Prentice–Hall, Englewood Cliffs, NJ, 2001, pp. 141–151.
[21] P. Del Moral and J. Garnier, *Genealogical particle analysis of rare events*, Ann. Appl. Probab., 15 (2005), pp. 2496–2534.
[22] M. Delloye, J.-D. Fermanian, and M. Sbai, *Dynamic frailties and credit portfolio modeling*, Risk, 19 (2006), pp. 101–109.
[23] G. di Graziano and C. Rogers, *A dynamic approach to the modelling of correlation credit derivatives using Markov chains*, Int. J. Theor. Appl. Finance, 12 (2009), pp. 45–62.
[24] D. Duffie, *First-to-Default Valuation*, working paper, Stanford University, Palo Alto, CA, 1998.
[25] D. Duffie and N. Garleanu, *Risk and valuation of collateralized debt obligations*, Fin. Anal. J., 57 (2001), pp. 41–59.
[26] D. Duffie and P. Glynn, *Efficient Monte Carlo estimation of security prices*, Ann. Appl. Probab., 4 (1995), pp. 897–905.
[27] D. Duffie and K. J. Singleton, *Simulating Correlated Defaults*, working paper, Graduate School of Business, Stanford University, Palo Alto, CA, 1999.
[28] D. Duffie, L. Saita, and K. Wang, *Multi-period corporate default prediction with stochastic covariates*, J. Fin. Econ., 83 (2006), pp. 635–665.
[29] D. Duffie, A. Eckner, G. Horel, and L. Saita, *Frailty correlated default*, J. Finance, 64 (2009), pp. 2089–2123.
[30] A. Eckner, *Computational techniques for basic affine models of portfolio credit risk*, J. Comput. Finance, 15 (2009), pp. 63–97.
[31] E. Errais, K. Giesecke, and L. R. Goldberg, *Affine point processes and portfolio credit risk*, SIAM J. Financial Math., 1 (2010), pp. 642–665.
[32] P. Feldhütter, *An Empirical Investigation of an Intensity Based Model for Pricing CDO Tranches*, working paper, Copenhagen Business School, Copenhagen, Denmark, 2007.
[33] R. Frey and J. Backhaus, *Pricing and hedging of portfolio credit derivatives with interacting default intensities*, Int. J. Theor. Appl. Finance, (2008), to appear.
[34] G. Giampieri, M. Davis, and M. Crowder, *A hidden Markov model of default interaction*, Quant. Finance, 5 (2005), pp. 27–34.
[35] K. Giesecke and L. Goldberg, *Sequential defaults and incomplete information*, J. Risk, 7 (2004), pp. 1–26.
[36] K. Giesecke and S. Weber, *Credit contagion and aggregate loss*, J. Econom. Dynam. Control, 30 (2006), pp. 741–761.
[37] K. Giesecke, L. Goldberg, and X. Ding, *A top-down approach to multi-name credit*, Oper. Res., (2009), to appear.
[38] K. Giesecke, H. Kakavand, and M. Mousavi, *Exact simulation of point processes with stochastic intensities*, Oper. Res., (2010), to appear.
[39] P. Glasserman, *Monte Carlo Methods in Financial Engineering*, Springer-Verlag, New York, Berlin, 2003.
[40] P. Glasserman and J. Li, *Importance sampling for portfolio credit risk*, Management Sci., 51 (2005), pp. 1643–1656.
[41] P. Glasserman and J. Li, *Importance sampling for a mixed Poisson model of portfolio credit risk*, in Proceedings of the 2003 Winter Simulation Conference, S. Chick, P. J. Sanchez, D. Ferrin, and D. J. Morrice, eds., IEEE Press, Piscataway, NJ, 2006, pp. 267–275.
[42] P. Glasserman, W. Kang, and P. Shahabuddin, *Fast simulation of multifactor portfolio credit risk*, Oper. Res., 56 (2008), pp. 1200–1217.
[43] P. Glynn and D. Iglehart, *Importance sampling for stochastic simulations*, Management Sci., 35 (1989), pp. 1367–1392.
[44] P. Heidelberger, V. Nicola, and P. Shahabuddin, *Bounded relative error in estimating transient measures of highly dependable non-Markovian systems*, ACM Trans. Model. Computer Simul., 4 (1994), pp. 137–164.

[45] A. Herbertsson and H. Rootzen, *Pricing k-th-to-default-swaps under default contagion*, J. Comput. Finance, 12 (2008), pp. 49–78.
[46] S. Inglis and A. Lipton, *Factor models for credit correlation*, Risk, 20 (2007), pp. 110–115.
[47] S. Inglis, A. Lipton, I. Savescu, and A. Sepp, *Dynamic credit models*, Stat. Interface, 1 (2008), pp. 211–227.
[48] R. A. Jarrow and F. Yu, *Counterparty risk and the pricing of defaultable securities*, J. Finance, 56 (2001), pp. 555–576.
[49] M. Joshi and D. Kainth, *Rapid and accurate development of prices and greeks for nth to default credit swaps in the Li model*, Quant. Finance, 4 (2004), pp. 266–275.
[50] S. J. Koopman, A. Lucas, and A. Monteiro, *The multi-stage latent factor intensity model for credit rating transitions*, J. Econometrics, 142 (2008), pp. 399–424.
[51] S. Kou and X. Peng, *Default Clustering and Valuation of Collateralized Debt Obligations*, working paper, Columbia University, New York, 2009.
[52] H. Kraft and M. Steffensen, *Bankruptcy, counterparty risk, and contagion*, Rev. Finance, 11 (2007), pp. 209–252.
[53] S. Kusuoka, *A remark on default risk models*, Adv. Math. Econ., 1 (1999), pp. 69–82.
[54] P. Lewis and G. Shedler, *Simulation of nonhomogeneous Poisson processes by thinning*, Naval Logistics Quart., 26 (1979), pp. 403–413.
[55] A. Lopatin, *A Simple Dynamic Model for Pricing and Hedging Heterogenous CDOs*, working paper, Numerix, New York, 2009.
[56] A. Lopatin and T. Misirpashaev, *Two-dimensional Markovian model for dynamics of aggregate credit loss*, Adv. Econometrics, 22 (2008), pp. 243–274.
[57] A. McNeil and J. Wendin, *Bayesian inference for generalized linear mixed models of portfolio credit risk*, J. Empirical Finance, 14 (2007), pp. 131–149.
[58] P.-A. Meyer, *Démonstration simplifée d'un théorème de Knight*, in Séminaire de Probabilités V, Lecture Notes in Math. 191, Springer-Verlag, Berlin, 1971, pp. 191–195.
[59] A. Mortensen, *Semi-analytical valuation of basket credit derivatives in intensity-based models*, J. Derivatives, 13 (2006), pp. 8–26.
[60] E. Papageorgiou and R. Sircar, *Multiscale intensity models and name grouping for valuation of multi-name credit derivatives*, Appl. Math. Finance, 15 (2007), pp. 73–105.
[61] P. Protter, *Stochastic Integration and Differential Equations*, Springer-Verlag, New York, Berlin, 2004.
[62] P. Schönbucher, *Information-Driven Default Contagion*, working paper, ETH Zürich, Zürich, Switzerland, 2004.
[63] F. Yu, *Correlated defaults in intensity based models*, Math. Finance, 17 (2007), pp. 155–173.