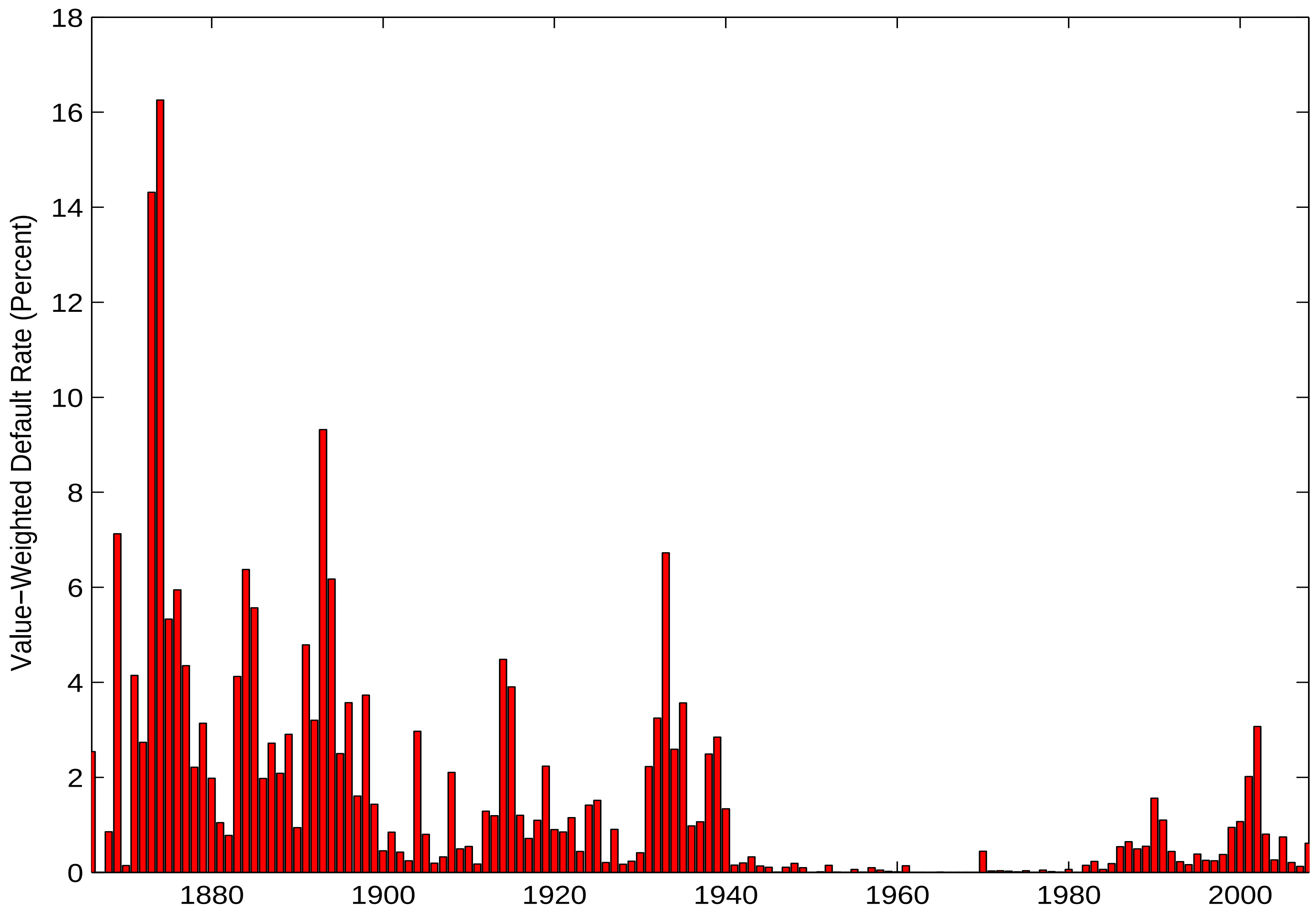
Optimal Importance Sampling For Dynamic Portfolio Credit Risk

Kay Giesecke
Management Science & Engineering
Stanford University
giesecke@stanford.edu
www.stanford.edu/∼giesecke

Joint work with Alex Shkolnik, Stanford

Corporate defaults cluster

Value-weighted default rate 1865–2008, US nonfinancial



Correlated default risk

Important applications

- Risk management of credit portfolios
  - Likelihood of large losses
  - Portfolio risk measures: VaR etc.
  - Capital allocation
  - Hedging
- Optimization of credit portfolios
- Risk analysis, valuation, and hedging of portfolio credit derivatives

Default timing

- Consider a portfolio of n defaultable assets
  - Default stopping times τ^i relative to $(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathbb{F}$
  - Default indicators $N^i_t = I(\tau^i \leq t)$
  - Vector of default indicators $N = (N^1, \ldots, N^n)$
- The **portfolio default process** $C = 1_n \cdot N$ counts defaults
  - At the center of many applications

Bottom-up model of default timing

- Name i defaults at intensity λ^i
  - $N^i - \int_0^{\cdot}(1 - N_s^i)\lambda_s^i ds$ is a martingale
  - λ^i represents the conditional default rate: for small $\Delta > 0$

$$\lambda_t^i \Delta \approx \mathbb{P}(i \text{ defaults during } (t, t+\Delta] \,|\, \mathcal{F}_t)$$

- The vector process $\lambda = (\lambda^1, \ldots, \lambda^n)$ is the modeling primitive
  - Component processes are correlated: diffusion, common or correlated or feedback jumps
  - Large literature
  - We don't assume a specific model

Model computation

- For fixed $T > 0$ and $C_T = 1_n \cdot N_T$ we require
  - $\mathbb{P}(C_T \geq m)$ for large m and its constituent sensitivities
  - Quantiles of $\mathbb{P}(C_T \geq m)$ and other risk measures
- Semi-analytical transform techniques
  - Limited to (one-) factor doubly-stochastic models of N
  - Numerical transform inversion can be challenging
- Monte Carlo simulation
  - Much wider scope
  - **Rare-event problem**

Conventional importance sampling

- Change measure from $\mathbb{P}$ to $\mathbb{Q}$ and simulate N under $\mathbb{Q}$; then

$$\mathbb{E}(f(N_T)) = \mathbb{E}_{\mathbb{Q}}(Z_T f(N_T))$$

- Radon-Nikodym density $Z_T = d\mathbb{P}/d\mathbb{Q}$ is given by

$$Z_T = \prod_{i=1}^{n} \exp\left(\int_0^T \log\left(\frac{\lambda^i_{s-}}{\nu^i_{s-}}\right) dN^i_s - \int_0^{T\wedge\tau^i} (\lambda^i_s - \nu^i_s) ds \right)$$

where $\nu = (\nu^1, \ldots, \nu^n)$ is the $\mathbb{Q}$-intensity of N

- Challenges
  - Find $\mathbb{Q}$ that minimizes $\mathbb{E}_{\mathbb{Q}}((Z_T f(N_T))^2)$: depends on λ
  - Evaluation of $Z_T f(N_T)$: discretization leads to bias
  - Computationally expensive for large n

Our IS strategy

- Our approach addresses these challenges; it has two parts
  1. Construct a time-inhomogeneous, continuous-time Markov chain $M \in \{0,1\}^n$ with the property that

     $$M_t \stackrel{law}{=} N_t$$

  2. Develop optimal $\mathbb{Q}$ for M, exploiting

     $$\mathbb{E}(f(N_T)) = \mathbb{E}(f(M_T)) = \mathbb{E}_{\mathbb{Q}}(Z_T f(M_T))$$

- Advantages
  - Optimal $\mathbb{Q}$ does not depend on the specific structure of λ
  - IS estimator $Z_T f(M_T)$ can be evaluated exactly
  - Computationally inexpensive even for large n

Mimicking Markov chain

- **Proposition** (G-Kakavand-Mousavi-Takada, 2009)
  Let M be a Markov chain that takes values in $\mathbb{S} = \{0,1\}^n$, starts at 0_n, has no joint transitions in any of its components and whose ith component has transition rate $p_n^i(\cdot, M)$ where

  $$p_n^i(t,B) = \mathbb{E}(\lambda_t^i I(N_t^i = 0) \,|\, N_t = B)$$

  for $B \in \mathbb{S}$. Then

  $$\mathbb{P}(M_t = B) = \mathbb{P}(N_t = B)$$

- Calculation of $p_n^i(t,B)$ feasible for many models λ in the literature

Rare-event regime

- For $\mu \in (0,1)$, consider $\xi_n = \{J_T \geq \mu n\}$, where $J = 1_n \cdot M$
- When is (ξ_n) rare, i.e., when $\mathbb{P}(\xi_n) = \mathbb{P}(C_T \geq \mu n) \to 0$ as $n \to \infty$
  - Seek IS scheme that is provably effective at estimating $\mathbb{P}(\xi_n)$
- Let $\mathbb{G}$ be the filtration generated by M; J has $\mathbb{G}$-intensity $p_n(\cdot, M)$, where $p_n(t,B) = \sum_{i=1}^n p_n^i(t,B)$
- Let K be a $\mathbb{G}$-Poisson process stopped at its nth jump with rate

$$\beta_n = \sup_{t \leq T, B \in \mathbb{S}} p_n(t,B)$$

- K dominates J in the sense that

$$\mathbb{P}(J_T \geq \mu n) \leq \mathbb{P}(K_T \geq \mu n) = \mathbb{P}(S_{\lceil \mu n \rceil} \leq T)$$

where $(S_m)_{1 \leq m \leq n}$ is the increasing sequence of event times of K

Rare-event regime

- The CLT indicates that $\mathbb{P}(S_{\lceil \mu n \rceil} \leq T) \to 0$ if $\beta_n < \lceil \mu n \rceil / T$
- Thus, (ξ_n) is rare if the rates $p_n^i(t,B)$ of the Markov chain M mimicking N satisfy

$$p_n(t,B) < \lceil \mu n \rceil / T \quad \text{for all } B \in \mathbb{S} \text{ and } t \in [0,T]$$

- A similar argument shows that (ξ_n) is not rare if

$$p_n(t,B) \geq \lceil \mu n \rceil / T \quad \text{for all } B \in \mathbb{S} \text{ and } t \in [0,T]$$

- We will show that the optimal measure $\mathbb{Q}$ for estimating $\mathbb{P}(C_T \geq \lceil \mu n \rceil)$ shifts the rates of M so that the $(\mathbb{G}, \mathbb{Q})$-intensity of J is at the threshold $\lceil \mu n \rceil / T$

Measure change for M

- We would like to choose $\mathbb{Q}$ to minimize the second moment under $\mathbb{Q}$ of the IS estimator $Z_T I(J_T \geq m)$ of $\mathbb{P}(C_T \geq m)$

- For $\theta > 0$, consider the family of $\mathbb{Q}$-rates of M given by

$$q_n^i(t,B) = \theta \, \frac{p_n^i(t,B)}{p_n(t,B)} \qquad \left(\text{so } \frac{q_n^i(t,B)}{\sum_{i=1}^n q_n^i(t,B)} = \frac{p_n^i(t,B)}{\sum_{i=1}^n p_n^i(t,B)} \right)$$

- The Radon-Nikodym derivative takes the form

$$Z_T(\theta) = \exp(\theta(T \wedge S_n) - J_T \log \theta + D_T)$$

where D_T, the $\mathbb{P}$-density of the $(S_m)_{1 \leq m \leq J_T}$ over $[0,T]$ wrt. Lebesgue measure, does not depend on θ:

$$D_T = \int_0^T \log p_n(s-, M_{s-}) dJ_s - \int_0^T p_n(s, M_s) ds$$

Measure change for M

- The second moment of the estimator $Z_T(\theta)I(J_T \geq m)$ is

$$\begin{aligned} \mathbb{E}_{\mathbb{Q}}(Z_T^2(\theta)I(J_T \geq m)) &= \mathbb{E}(Z_T(\theta)I(J_T \geq m)) \\ &\leq \exp(\theta T - m \log \theta)\mathbb{E}(\exp(D_T)I(J_T \geq m)) \end{aligned}$$

  where the inequality holds for all $\theta \geq 1$

- Minimizing the second moment is difficult, but minimizing the bound over $\theta \geq 1$ is easy: the minimizer θ^* is given by

$$\theta^* = \theta^*_m = m/T$$

- This suggests the IS estimator $Z_T(\lceil \mu n \rceil / T)I(J_T \geq \lceil \mu n \rceil)$ for $\mathbb{P}(C_T \geq \lceil \mu n \rceil)$

Measure change for M

- The counting process J has $(\mathbb{G}, \mathbb{Q}(\theta))$-intensity

$$q_n(t, M_t) = \sum_{i=1}^{n} q_n^i(t, M_t) = \theta I(S_n > t)$$

  and is hence a Poisson process with rate θ stopped at S_n

- For large n, the $\mathbb{Q}(\lceil \mu n \rceil / T)$-mean of J_T is

$$\mathbb{E}_{\mathbb{Q}}(J_T) = \frac{\lceil \mu n \rceil}{T} \mathbb{E}_{\mathbb{Q}}(T \wedge S_n) \approx \lceil \mu n \rceil$$

  and so the event $\{J_T \geq \lceil \mu n \rceil\}$ is not rare under $\mathbb{Q}(\lceil \mu n \rceil / T)$

- The $\mathbb{P}$-rates $p_n^i(t, B)$ of M are shifted to $q_n^i(t, B)$ so that the $\mathbb{Q}$-intensity $q_n(\cdot, M) = \sum_{i=1}^{n} q_n^i(\cdot, M)$ of J is just at the rare-event boundary $\lceil \mu n \rceil / T$

Asymptotic optimality

- **Theorem.** Suppose $p_n(t,B) < \lceil \mu n \rceil / T$, i.e. $\mathbb{P}(C_T \geq \lceil \mu n \rceil) \to 0$. If, for $\alpha_n = \inf_{t \leq T, B \in \mathbb{S}} p_n(t,B)$ and $\beta_n = \sup_{t \leq T, B \in \mathbb{S}} p_n(t,B)$,

$$\limsup_n \frac{\beta_n}{\alpha_n} = 1$$

then the second moment of the IS estimator

$$Y_n = Z_T(\lceil \mu n \rceil / T) I(J_T \geq \lceil \mu n \rceil)$$

of the probability

$$\mathbb{E}_{\mathbb{Q}}(Y_n) = \mathbb{P}(C_T \geq \lceil \mu n \rceil)$$

satisfies

$$\limsup_{n \to \infty} \frac{\log \mathbb{E}_{\mathbb{Q}}(Y_n^2)}{\log \mathbb{E}_{\mathbb{Q}}(Y_n)} = 2$$

Asymptotic optimality

Discussion

- The condition prevents the oscillations of the $(\mathbb{G}, \mathbb{P})$ intensity $p_n(t, M)$ of J from becomming arbitrarily large as n grows
- There are other optimal measures $\mathbb{Q}$
  - Take $\mathbb{Q}$ to be the measure associated with the rates

$$q_n^i(t, B) = \theta \frac{1 - B^i}{n - 1_n \cdot B}$$

  - The $\mathbb{Q}$-intensity of J is $q_n(t, M_t) = \theta I(S_n > t)$, as above
  - The estimator is asymptotically efficient, albeit under more stringent hypotheses on the rate bounds

Derivative estimators

- For some parameter $\theta > 0$, let $\mathbb{P}_\theta$ be a probability measure equivalent to $\mathbb{P}$ under which the default process $N = (N^1, \ldots, N^n)$ has intensity $\lambda^\theta = (\theta\lambda^1, \lambda^2, \ldots, \lambda^n)$
- Under mild conditions,

$$\partial_\theta \mathbb{E}_\theta(f(N_T)) = \mathbb{E}_\theta(f(M_T)\partial_\theta \log h_\theta)$$

  - The chain M has $\mathbb{P}_\theta$-transition rate $(\theta p_n^1, p_n^2, \ldots, p_n^n)$
  - h_θ is the density of $\mathbb{P}_\theta$ with respect to some probability measure μ on $\mathcal{G}_T$ independent of θ

$$\partial_\theta \log h_\theta = \frac{1}{\theta}\left(M_T^1 - \theta \int_0^T p_n^1(s, M_s)ds\right)$$

which we recognize as $1/\theta$ times the compensated $(\mathbb{G}, \mathbb{P}_\theta)$-jump martingale associated with M^1, evaluated at T

Derivative estimators

- **Theorem.** Under the conditions of the previous Theorem, the IS estimator

$$Y_n = Z_T(\lceil \mu n \rceil / T) I(J_T \geq \lceil \mu n \rceil) \left(M_T^1 - \int_0^T p_n^1(s, M_s) ds \right)$$

of the sensitivity

$$\partial_\theta \mathbb{P}_\theta (C_T \geq \lceil \mu n \rceil)|_{\theta=1}$$

is asymptotically optimal in the sense that

$$\limsup_{n \to \infty} \frac{\log \mathbb{E}_{\mathbb{Q}}(Y_n^2)}{\log |\mathbb{E}_{\mathbb{Q}}(Y_n)|} = 2$$

Extensions

- It is straightforward to extend the analysis to include a random loss at default
- The analysis can also be extended to risk measures such as value at risk or expected shortfall
  - Under technical conditions, variance reduction for the probability leads to variance reduction for the risk measure
  - Asymptotic optimality

Numerical results

Self-exciting intensity model for $n = 100$

- Suppose the $\mathbb{P}$-intensities $\lambda_t^i = X_t^i + c^i(t, N_t)$
  - Extends Jarrow & Yu (2001), Kusuoka (1999), Yu (2007)
  - Feedback specification $c^i(t, B) = \sum_{j \neq i}^n \beta^{ij} B^j$ can be varied
  - Analytical solutions not known
- Suppose the idiosyncratic factor follows the $\mathbb{P}$-Feller diffusion

$$dX_t^i = \kappa_i(\theta_i - X_t^i)dt + \sigma_i \sqrt{X_t^i} dW_t^i$$

  where $(W^1, \ldots, W^n)$ is a standard $\mathbb{P}$-Brownian motion
- Parameters selected randomly (high credit quality)
  - Practice: estimation from derivative prices or default history

Numerical results

- **Proposition.** The $\mathbb{P}$-rates of the mimicking chain M are

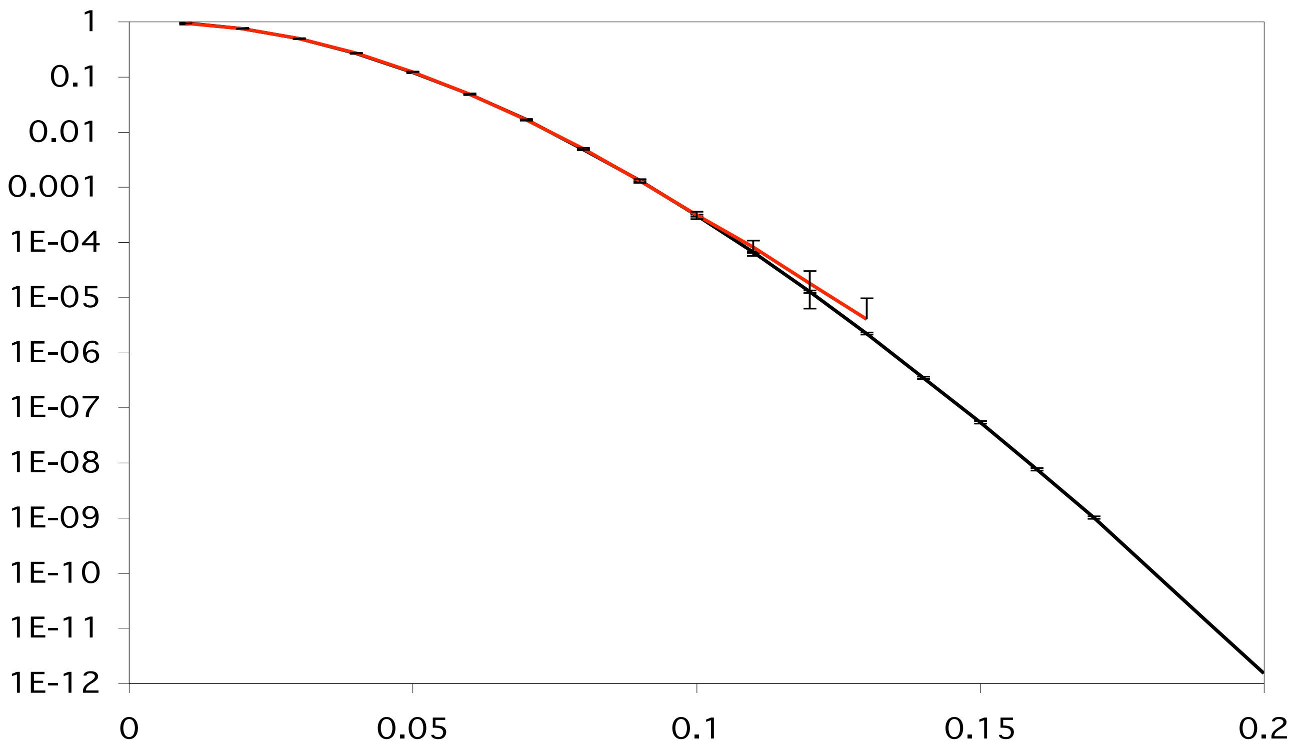
$$\begin{aligned} p_n^i(t,B) &= \mathbb{E}(\lambda_t^i I(N_t^i = 0) \,|\, N_t = B) \\ &= \frac{4X_0^i \gamma_i^2 \exp(\gamma_i t)}{(\gamma_i - \kappa_i + (\gamma_i + \kappa_i)\exp(\gamma_i t))^2} \\ &- \frac{\theta_i \kappa_i}{\sigma_i^2} \frac{(\kappa_i^2 - \gamma_i^2)(\exp(\gamma_i t) - 1)}{\gamma_i - \kappa_i + (\gamma_i + \kappa_i)\exp(\gamma_i t)} + \sum_{j \neq i} \beta^{ij} B^j \end{aligned}$$

  for $B^i = 0$, where $\gamma_i = \sqrt{\kappa_i^2 + 2\sigma_i^2}$

- Generalizations
  - Can add compound Poisson jumps without reducing tractability
  - General affine jump diffusion dynamics

Numerical results

$\mathbb{P}(C_1 \geq \mu)$ for $n = 100$, plain MC (500K, red) vs. IS (10K each)



Variance ratios for $\mathbb{P}(C_1 \geq \mu n)$, $n = 100$

μn	$\mathbb{P}(C_1 \geq \mu n)$	Var IS (10K)	Var Plain (500K)	Var Ratio
3	0.4906572	0.1862536	0.2499024	1.341732
4	0.2660202	0.0828141	0.1962571	2.369851
5	0.1196635	0.02467735	0.1080868	4.379999
6	0.04823202	0.005089037	0.0459401	9.027268
7	0.01679726	0.0007392553	0.01622778	21.95152
8	0.004851613	7.61035e-05	0.004923527	64.69515
9	0.00132148	6.290501e-06	0.001300307	206.7096
10	0.0003043514	3.951607e-07	0.0003099045	784.2493
11	6.616864e-05	2.031115e-08	8.199344e-05	4036.868
12	1.267209e-05	8.217485e-10	1.799971e-05	21904.16
13	2.199999e-06	2.737301e-11	3.999992e-06	146129.1
14	3.484863e-07	7.542041e-13	0	0
15	5.439357e-08	1.922568e-14	0	0

Variance ratios for $\partial_\theta \mathbb{P}_\theta(C_1 \geq \mu n)|_{\theta=1}$, $n = 100$

μn	$\partial_\theta \mathbb{P}_\theta(C_1 \geq \mu n)\|_{\theta=1}$	Var IS (10K)	Var Plain (500K)	Var Ratio
3	0.006661911	0.01548746	0.01839772	1.187911
4	0.005118402	0.006294804	0.01210119	1.92241
5	0.00311368	0.001831398	0.006499334	3.548838
6	0.00178735	0.000434359	0.002888021	6.648926
7	0.0006891087	6.235889e-05	0.00109927	17.62812
8	0.0002516899	7.208854e-06	0.0003520437	48.83490
9	8.134683e-05	6.290217e-07	0.0001003328	159.5061
10	2.350587e-05	4.752067e-08	2.949409e-05	620.6581
11	6.302068e-06	2.934012e-09	1.173715e-05	4000.376
12	1.118905e-06	1.08919e-10	3.933792e-06	36116.67
13	2.478318e-07	4.301901e-12	2.618956e-09	60879.04
14	4.080089e-08	1.185663e-13	0	0
15	6.813895e-09	3.233037e-15	0	0

Conclusions

- Exact and asymptotically optimal IS scheme for intensity-based models of portfolio credit risk
  - Probability of large losses
  - Constituent sensitivities
  - Risk measures
- Broadly applicable since largely model-independent
  - Multi-factor doubly-stochastic models
  - Multi-factor frailty models
  - Self-exciting models

Conclusions

- We address the rare-event simulation problem for intensity-based models of portfolio credit risk
  - Bassamboo & Jain (2006): biased IS estimators for standard doubly-stochastic model
  - Giesecke, Kakavand, Mousavi & Takada (2009): unbiased selection/mutation estimators for a wide range of models but no optimality certificate
- Our results complement the IS schemes developed for copula-based models of portfolio credit risk
  - Bassamboo, Juneja & Zeevi (2008)
  - Chen & Glasserman (2008)
  - Glasserman & Li (2005)

References

Bassamboo, Achal & Sachin Jain (2006), Efficient importance sampling for reduced form models in credit risk, *in* L. F.Perrone, F. P.Wieland, J.Liu, B. G.Lawson, D. M.Nicol & R. M.Fujimoto, eds, 'Proceedings of the 2006 Winter Simulation Conference', IEEE Press, pp. 741–748.

Bassamboo, Achal, Sandeep Juneja & Assaf Zeevi (2008), 'Portfolio credit risk with extremal dependence: Asymptotic analysis and efficient simulation', *Operations Research* **56**(3), 593–606.

Chen, Zhiyong & Paul Glasserman (2008), 'Fast pricing of basket default swaps', *Operations Research* **56**(2), 286–303.

Giesecke, Kay, Hossein Kakavand, Mohammad Mousavi & Hideyuki Takada (2009), Exact and efficient simulation of correlated defaults. Working Paper, Stanford University.

Glasserman, Paul & Jingyi Li (2005), 'Importance sampling for portfolio credit risk', *Management Science* **51**(11), 1643–1656.

Jarrow, Robert A. & Fan Yu (2001), 'Counterparty risk and the pricing of defaultable securities', *Journal of Finance* **56**(5), 555–576.

Kusuoka, Shigeo (1999), 'A remark on default risk models', *Advances in Mathematical Economics* **1**, 69–82.

Yu, Fan (2007), 'Correlated defaults in intensity based models', *Mathematical Finance* **17**, 155–173.