Multilevel simulation of functionals of Bernoulli random variables with application to basket credit derivatives

K. Bujok[*], B. M. Hambly and C. Reisinger

Mathematical Institute, Oxford University
24–29 St Giles, Oxford, OX1 3LB, UK
Email: {bujok, hambly, reisinge@maths.ox.ac.uk}

September 12, 2013

Abstract

We consider N Bernoulli random variables, which are independent conditional on a common random factor determining their probability distribution. We show that certain expected functionals of the proportion L_N of variables in a given state converge at rate $1/N$ as $N \to \infty$. Based on these results, we propose a multi-level simulation algorithm using a family of sequences with increasing length, to obtain estimators for these expected functionals with a mean-square error of ϵ^2 and computational complexity of order ϵ^{-2}, independent of N. In particular, this optimal complexity order also holds for the infinite-dimensional limit. Numerical examples are presented for tranche spreads of basket credit derivatives.

Key words: Multilevel Monte Carlo simulation, large deviations principle, exchangeability, basket credit derivatives

1 Introduction

This article is concerned with the efficient numerical estimation of expectations of functionals of a large number, N, of exchangeable Bernoulli random variables. The objective of this work is thus two-fold: to analyse the order of convergence in $1/N$ of expected functionals as N tends to infinity, and to derive estimators for these expectations for which the computational complexity is asymptotically independent of N.

We begin by analysing the convergence in the case of general Lipschitz and smooth functions, p, of the average of N exchangeable Bernoulli random variables as N goes to infinity. We then consider the case when p has a certain piecewise linear structure and show that the convergence order is the same as in the smooth case. These results are relevant, for instance, if one wants to approximate the result for large but finite N by its limit. A number of applications come from the credit risk literature. In [14], Vasicek derives an expression for the limiting distribution of portfolio losses in a Normal factor model, where default of a firm is indicated by its value process being below a default

[*]K. Bujok acknowledges support from EPSRC and Nomura via a CASE award, and from the Oxford-Man Institute.

barrier at maturity of the debt. In the large portfolio limit, the randomness comes solely from a common market factor, while a law of large numbers holds for idiosyncratic components conditionally on this factor. Bush *et al.*, in [5], extend this to a dynamic set-up where it is seen that the density of the limit empirical measure of firm values satisfies a stochastic partial differential equation (SPDE) and can be used to approximate tranche spreads of basket credit derivatives; [4] gives an extension to jump diffusion models while [9] include extensions to heterogeneity and self-exciting defaults rendering the resulting equations non-linear. Further studies focus particularly on the tail of the limiting loss distribution, see [7], [12] and the references therein.

A driving practical motivation for investigating the limiting behaviour is that the original sequence of random variables is costly to simulate, because of the large number N of underlying processes, often required over large time horizons. Moreover, often many Monte Carlo samples are necessary for sufficiently accurate estimation of, for instance, expected tranche losses of credit basket. This paper takes a different tack and develops a simulation method where the computational complexity is asymptotically independent of N. A small tweak of the algorithm can also be used to approximate the limit obtained when N goes to infinity.

More concretely, it turns out that an interpretation of the multi-level Monte Carlo approach (see [11]) in the present context allows us to construct estimators based on sequences with increasing lengths and a number of samples which decreases faster than the length increases, such that the overall computational complexity is essentially no larger than for fixed small N.

A conceptually similar though distantly related approach is used in [3], where the multilevel idea is applied to a sequence of martingales to estimate a dual upper bound for the value of an early exercise option. In that setting they are able to show, as we do here, that the achievable complexity is not substantially larger than that of a non-nested simulation. The general problem of estimating conditional expectations through nested multilevel simulation is addressed in [6]. There, further extrapolation is used to reduce the bias of estimators, while here we will propose an improved estimator which reduces the variance of higher level estimators.

This article is organised as follows. In Section 2, we introduce the setting and outline the main convergence results, explaining how they can be used to construct efficient estimators. The first key result on the convergence order of expected functionals is proved in Section 3, with numerical illustrations from an example of a basket credit derivative presented in Section 4. In Section 5, we introduce in detail two multilevel simulation methods and derive bounds on their computational complexity to achieve a prescribed accuracy. Finally, in Section 6 we present numerical results illustrating the efficiency gains achieved through multilevel simulation in this context and Section 7 discusses possible extensions.

2 Set-up and main results

In this article, we are concerned with the behaviour of "loss" variables describing the fraction of N random variables in a certain state, and expected functionals of this loss variable, as N goes to infinity. The application we have in mind, and for which we will present numerical illustrations, is that of a basket of defaultable firms, and then the loss is the fraction of firms which default over a certain period.

More precisely, on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, consider a sequence of Bernoulli random variables Y_i, $i \in \mathbb{N}$, and a random variable L taking its values in $[0,1]$. If required we write $\Omega = \Omega_Y \times \Omega_L$ where canonically we could take $\Omega_Y = \{0,1\}^{\mathbb{N}}$ and $\Omega_L = [0,1]$. The probability measure $\mathbb{P}$ is constructed as follows. The random variable L is generated according to its marginal law $\mathbb{P}_L$ and then, conditional on $\mathcal{F}_L$, the σ-algebra generated by L, the Y_i are independent random variables with law given by

$$\mathbb{P}[Y_i = 1 | \mathcal{F}_L] = L. \tag{2.1}$$

To re-iterate, the Bernoulli random variables are conditionally independent given a common factor. Thus for each $n \in \mathbb{N}$

$$\mathbb{P}(Y_1 = y_1, \ldots, Y_n = y_n, L \in B) = \int_B l^{s_n} (1-l)^{n-s_n} \mathbb{P}_L(L \in dl), \quad \forall y_i \in \{0,1\}, B \subset [0,1]$$

where $s_n = \sum_{i=1}^n y_i$. We will often write $\mathbb{P}_{|L} = \mathbb{P}(. \mid \mathcal{F}_L)$ for the conditional law of the Y_i given $\mathcal{F}_L$ and $\mathbb{E}_{|L}$ for the associated conditional expectation. In the setting of defaultable firms, $Y_i = 1$ iff the i-th firm defaults, and L is a global factor modelling the common tendency of firms to default. We define the loss variable to be the proportion of Bernoulli variables in state 1

$$L_N = \frac{1}{N} \sum_{i=1}^N Y_i. \tag{2.2}$$

We consider a Lipschitz function p and random variables P and P_N defined as

$$\begin{aligned} P &\equiv p(L), && (2.3) \\ P_N &\equiv p(L_N). && (2.4) \end{aligned}$$

In particular, we will study p of the form

$$p(l) \equiv [l - K_1]^+ - [l - K_2]^+ = \begin{cases} 0 & l \le K_1, \\ l - K_1 & K_1 \le l \le K_2, \\ K_2 - K_1 & l \ge K_2, \end{cases} \tag{2.5}$$

where $[x]^+ = \max(x, 0)$ denotes the positive part and $0 \le K_1 < K_2 \le 1$ are constants. In credit derivative pricing, the particular shape of the function p in (2.5) measures the losses in a certain *tranche* with attachment point K_1 and detachment point K_2, and its expectation is the building block for formulae for CDO tranche spreads. A typical CDO pool consists of $N = 125$ firms, while typical loan or mortgage books can have substantially more obligors, and it is therefore practically relevant to understand the behaviour of expected functionals for large N and to devise computationally efficient estimators.

By a conditional version of the strong law of large numbers and the continuity of p

$$\begin{aligned} L_N &\to L \quad \text{for } N \to \infty, \ \mathbb{P}_{|L} - a.s., && (2.6) \\ P_N &\to P \quad \text{for } N \to \infty, \ \mathbb{P}_{|L} - a.s. && (2.7) \end{aligned}$$

This convergence will also hold in $L^2(\Omega_Y, \mathbb{P}_{|L})$ (see Lemma 3.1).

We study here the convergence rate of $P_N - P$ and will prove the following two results. The first statement for Lipschitz and smooth functions p is a relatively straightforward consequence of (2.1) and the easily computable L^2 convergence rate of L_N. The second result shows that for a specific p which is only piecewise smooth we can still obtain the same convergence order as in the smooth case and with explicitly computable bounds.

Theorem 2.1. *Let P and P_N be defined by* (2.3) *and* (2.4)*, respectively, and assume that p is Lipschitz with constant c_p. We have that*

$$|\mathbb{E}[P_N - P]| \;\le\; \frac{c_p}{2\sqrt{N}}, \tag{2.8}$$

$$Var[P_N - P] \;\le\; \frac{c_p^2}{4N}. \tag{2.9}$$

If, moreover, p is differentiable and the derivative has Lipschitz constant C_p, then

$$|\mathbb{E}[P_N - P]| \;\le\; \frac{C_p}{8N}. \tag{2.10}$$

Theorem 2.2. *For p defined in (2.5), if the cumulative density function (CDF) F_L of L is Lipschitz at $K_1 > 0$ and $K_2 < 1$ with Lipschitz constant c_L, i.e.,*

$$|F_L(K_j) - F_L(l)| \le c_L\,|K_j - l| \tag{2.11}$$

for $j = 1, 2$ and all $l \in [0, 1]$, then

$$|\mathbb{E}[P_N - P]| \le \frac{4c_L\sqrt{\pi}}{N}.$$

Note that if L has a density function which is bounded, then the CDF is certainly Lipschitz. The fact that we only need the Lipschitz property at K_1 and K_2 will be useful for the applications considered later.

Taking the two Theorems together, order 1 for the convergence of expectations also follows for piecewise smooth p which are Lipschitz overall, provided F_L is Lipschitz.

These Theorems show that expected functionals for large or infinite N can be successively approximated by those with smaller N. Combining this with a control variate idea leads to multilevel simulation with a substantial variance reduction for large N. Specifically, the above results imply that for Lipschitz p we have $|\mathbb{E}[P_N - P_{MN}]| \le c_1/\sqrt{N}$ and $Var[P_N - P_{MN}] \le c_2/N$ for any positive integer M with some constants c_1 and c_2. We can consider a sequence $N_l = M^l$, $l \in \mathbb{N}$, with corresponding $L^{(l)} = L_{N_l}$ and $P^{(l)} = P_{N_l}$. Translating the central idea in [11] to this setting, we use the decomposition

$$\mathbb{E}[P^{(l)}] = \mathbb{E}[P^{(0)}] + \sum_{k=1}^{l} \mathbb{E}[P^{(k)} - P^{(k-1)}] \tag{2.12}$$

and estimate every summand $\mathbb{E}[P^{(k)} - P^{(k-1)}]$ separately by defining estimators

$$Z_l \equiv n_l^{-1} \sum_{j=1}^{n_l} \left(P^{(l,j)} - P_c^{(l,j)}\right), \tag{2.13}$$

where 'c' denotes a 'coarse' estimator on level l, i.e., using only N_{l-1} instead of N_l Bernoulli random variables, precisely,

$$P^{(l,j)} \;=\; p(L^{(l,j)}), \quad \text{where } L^{(l,j)} = N_l^{-1} \sum_{i=1}^{N_l} Y_i^{(l,j)}, \tag{2.14}$$

$$P_c^{(l,j)} \;=\; p(L_c^{(l,j)}), \quad \text{where } L_c^{(l,j)} = N_{l-1}^{-1} \sum_{i=1}^{N_{l-1}} Y_i^{(l,j)}, \tag{2.15}$$

where $Y_i^{(l,j)}$, $j = 1, \ldots, n_l$, are independent samples of Y_i for fixed level l and independent across levels. They are constructed from a loss factor $L^{(l,j)}$ (with the same distribution as L, independent across l and j) in the same way that Y_i is constructed from L.

The number of samples on each level, n_l, can be chosen to obtain an optimal allocation of computational cost for a given overall mean-square error (MSE). The general construction in [11] immediately gives the following result.

Proposition 2.1 (cf. [11], Theorem 3.1)**.** *Let P, $P^{(l)}$ as above. If there exist independent estimators Z_l based on n_l Monte Carlo samples, and positive constants $\alpha, \beta, c_1, c_2, c_3$ such that $\alpha \geq \frac{1}{2}$ and*

i) $\left|\mathbb{E}[P^{(l)} - P]\right| \leq c_1 M^{-\alpha l}$

ii) $\mathbb{E}[Z_l] = \begin{cases} \mathbb{E}[P^{(0)}], & l = 0 \\ \mathbb{E}[P^{(l)} - P^{(l-1)}], & l > 0 \end{cases}$

iii) $\mathbb{V}[Z_l] \leq c_2\, n_l^{-1} M^{-\beta l}$

iv) $C_l \leq c_3\, n_l\, N_l$*, where C_l is the computational complexity of Z_l*

then there exists a positive constant c_4 such that for any $\varepsilon < e^{-1}$ there are values K and n_l for which the multilevel estimator

$$G_K = \sum_{l=0}^{K} Z_l, \tag{2.16}$$

has a mean-square-error with bound

$$MSE \equiv \mathbb{E}\left[(G_K - E[P])^2\right] < \varepsilon^2$$

with a computational complexity C with bound

$$C \leq \begin{cases} c_4\, \varepsilon^{-2}, & \beta > 1, \\ c_4\, \varepsilon^{-2} (\log \varepsilon)^2, & \beta = 1, \\ c_4\, \varepsilon^{-2-(1-\beta)/\alpha}, & 0 < \beta < 1. \end{cases}$$

The above result is meaningful only in situations where it is not possible or practical to sample from L directly, as otherwise $\mathbb{E}[P] = \mathbb{E}[p(L)]$ could be computed with complexity $O(\epsilon^{-2})$ in the standard Monte Carlo way.

Moreover, in some situations it is not $p(L)$ which is of interest, but $p(L_N)$ for large but finite N, and then it is essential to have a method to estimate $\mathbb{E}[L_N]$ in a complexity which does not increase sharply in N.

For instance, take N given and estimate P_N with the standard (i.e., single level) Monte Carlo estimator

$$\widehat{P}_N \quad \equiv \quad \frac{1}{n} \sum_{j=1}^{n} p\left(\frac{1}{N} \sum_{i=1}^{N} Y_i^{(j)}\right),$$

where n is the number of samples and the $(Y_i^{(j)})$, for different j, are independent samples of (Y_i). Then $\mathbb{E}[\widehat{P}_N] = \mathbb{E}[P_N]$ and $Var[\widehat{P}_N] = \frac{1}{n} Var[P_N]$, where it follows from

$$Var[P] \leq 2\left(Var[P_N] + Var[P_N - P]\right),$$

under the conditions of either Theorem 2.1 or Theorem 2.2, that

$$Var[P_N] \geq Var[P]/2 - Var[P_N - P] \geq Var[P]/2 - c_1/N \geq c_2 Var[P],$$

for N sufficiently large and some constants c_1, c_2 independent of N. That is to say, the variance of P_N and subsequently that of the estimator $\widehat{P}_N$ is bounded below with a positive number independent of N. Hence, if a MSE of ϵ^2 is required for $\mathbb{E}[P_N]$, the complexity is

$$C \geq nN \geq c_2 N/\epsilon^2,$$

i.e., increases (at least) linearly in N. (A similar argument shows that this is also an upper bound.)

If one wants to use $\widehat{P}_N$ not as an estimator to $\mathbb{E}[P_N]$ but $\mathbb{E}[P]$, a bias occurs and

$$MSE = \mathbb{E}[\widehat{P}_N - P]^2 + Var[\widehat{P}_N] = \mathbb{E}[P_N - P]^2 + Var[\widehat{P}_N] = O(N^{-2/\alpha}) + O(n^{-1}),$$

assuming the bias is of order α as in Proposition 2.1. To reduce the bias and hence the error, N has to be increased simultaneously with n. More precisely, for MSE ϵ^2 it is optimal to choose $N = O(\epsilon^{-1/\alpha})$ and $n = O(\epsilon^{-2})$, leading to a computational complexity

$$C = O(nN) = O\left(\epsilon^{-2-1/\alpha}\right).$$

The following Corollary addresses both cases of large finite and infinite N and improves on the convergence rates of the standard Monte Carlo estimator.

Corollary 2.1. *Let P_N and P be as in (2.3) and (2.4), and assume p is Lipschitz.*

1. *There is a multilevel estimator for $\mathbb{E}[P]$ with MSE ϵ^2 with computational complexity $C \leq c_4 (\log \epsilon)^2 \epsilon^{-2}$.*
2. *For all N, there is a multilevel estimator for $\mathbb{E}[P_N]$ with MSE ϵ^2 with computational complexity $C \leq c_4 (\log \epsilon)^2 \epsilon^{-2}$, where c_4 is independent of N.*

Note that only order 1/2 is required for the convergence of expectations in Proposition 2.1, i), and that the complexity is then dictated by β, the case $\beta = 1$ implied by Theorem 2.1 for all Lipschitz payoffs being a boundary case.

The estimators for both $\mathbb{E}[P]$ and $\mathbb{E}[P^{(L)}] = \mathbb{E}[P_{N_L}]$, for $N_L = M^L$ fixed, are given by (2.16). In the first case, the maximum level K and the number of samples n_l on each level have to be increased successively as part of the simulation algorithm until a desired MSE is reached, as explained in [11]. In the second case, a similar procedure can be used but K is not increased further once the desired level L is reached. By construction, at that point, the total MSE is small enough that no additional samples need to be generated. This algorithm is formalised at the start of Section 6.

For the specific p as in (2.5), we can exploit the piecewise linearity of p to construct multilevel estimators with even better complexity, by making the following observations:

The summands in (2.12) are unchanged if we replace $P^{(k-1)} = p(L^{(k-1)})$ with any of $p(L_m^{(k-1)})$ for $m = 1, \ldots, M$, where

$$L_m^{(k-1)} \equiv \frac{1}{N_{k-1}} \sum_{i=1}^{N_{k-1}} Y_{i+(m-1)N_{k-1}}. \tag{2.17}$$

This is a direct consequence of the exchangeability. Now,

$$L^{(k)} = \frac{1}{M} \sum_{m=1}^{M} L_m^{(k-1)} \tag{2.18}$$

and, if all $L_m^{(k-1)}$ lie in the same interval $[0, K_1]$, $(K_1, K_2]$ or $(K_2, 1]$, also $P^{(k)} = \overline{P}^{(k-1)}$, where

$$\overline{P}^{(k-1)} \equiv \frac{1}{M} \sum_{m=1}^{M} P_m^{(k-1)} = \frac{1}{M} \sum_{m=1}^{M} p(L_m^{(k-1)}), \tag{2.19}$$

since p is linear in these intervals. Because of $\mathbb{E}[P^{(k-1)}] = \mathbb{E}[\overline{P}^{(k-1)}]$, we can now write

$$\mathbb{E}[P^{(l)}] = \mathbb{E}[P^{(0)}] + \sum_{k=1}^{l} \mathbb{E}[P^{(k)} - \overline{P}^{(k-1)}], \tag{2.20}$$

and estimate the individual terms in the sum independently in the multilevel spirit, i.e., with estimators

$$\overline{Z}_l \equiv n_l^{-1} \sum_{j=1}^{n_l} \left(P^{(l,j)} - \overline{P}^{(l,j)} \right), \tag{2.21}$$

where $P^{(l,j)}$ is defined as in (2.14), but instead of $P_c^{(l,j)}$ we use

$$\overline{P}^{(l,j)} \quad = \quad M^{-1} \sum_{m=1}^{M} p(L_m^{(l,j)}), \quad \text{where } L_m^{(l,j)} = N_{l-1}^{-1} \sum_{i=1}^{N_{l-1}} Y_{i+(m-1)N_{l-1}}^{(l,j)}, \tag{2.22}$$

and where the rest of the set-up is as earlier.

There is only a variance contribution from a specific sample of the k-th term if at least two $P_m^{(k-1)}$ lie in different intervals. For large k, the probability of this is small, and we will be able to show the following result.

Theorem 2.3. *For p as in (2.5), let $P^{(l)}$ as in Proposition 2.1 and $\overline{P}^{(l-1)}$ as in (2.19). If the CDF F_L of L is Lipschitz with Lipschitz constant c_L, then*

$$Var[P^{(l)} - \overline{P}^{(l-1)}] \quad \leq \quad \frac{c_2}{N_l^{3/2}}, \tag{2.23}$$

where $c_2 = c_L \, 4\sqrt{M\pi}(\sqrt{2} + \sqrt{M})\sqrt{\frac{7}{8}(M^2 + 6M + 1)}$.

Here and throughout the paper we give explicit expressions for the constants. These should not be regarded as optimal in any sense.

Corollary 2.2. *For Lipschitz F_L and p as in (2.5), there is a constant c_5 and multilevel estimators for $\mathbb{E}[P]$ and $\mathbb{E}[P_N]$ with MSE ϵ^2 with computational complexity $C \leq c_5 \, \epsilon^{-2}$.*

Note that we have managed to remove the logarithmic factor present in Corollary 2.1 and that c_5 does not depend on N.

3 Proof of convergence rates

We first prove Theorem 2.1 which contains statements in the general and smooth case. The rest of this section is devoted to the proof of Theorem 2.2 dealing with a specific non-smooth payoff relevant to our application.

Lemma 3.1. *Let P_N and P be as in (2.3) and (2.4), and assume p is Lipschitz with constant c_p. Then*

$$\mathbb{E}_{|L}[(P_N - P)^2] \le \frac{c_p^2}{4N}.$$

Proof. Since the function p in (2.3) is assumed Lipschitz and $\mathbb{E}_{|L}[L_N] = L$, we have

$$\mathbb{E}_{|L}[(P_N - P)^2] \le c_p^2 \; \mathbb{E}_{|L}[(L_N - L)^2] = c_p^2 \; Var[L_N|\mathcal{F}_L] = \frac{c_p^2}{N} \; Var[Y_i|\mathcal{F}_L] = \frac{c_p^2}{N} \; L(1-L).$$

For $L \in [0,1]$, $L(1-L) \le \frac{1}{4}$, which gives the result. □

Proof of Theorem 2.1. Equation (2.9) follows directly from Lemma 3.1, and then, by Cauchy-Schwarz,

$$|\mathbb{E}[P - P_N]| \le \sqrt{\mathbb{E}[\mathbb{E}_{|L}[(P_N - P)^2]]} \le \frac{c_p}{2\sqrt{N}}.$$

For differentiable p, we can write

$$\mathbb{E}[p(L) - p(L_N)] = \mathbb{E}[p'(L)(L - L_N)] + \mathbb{E}[r(L, L_N)],$$

with some remainder r, where the first term on the left-hand side is

$$\mathbb{E}[\mathbb{E}_{|L}[p'(L)(L - L_N)]] = 0.$$

If p has a Lipschitz derivative,

$$|p(x) - p(y) - p'(x)(x-y)| \le \frac{1}{2} C_p (x-y)^2$$

for all $0 \le x, y \le 1$ and the remainder term satisfies

$$|\mathbb{E}[r(L, L_N)]| \le \frac{C_p}{8N},$$

from which (2.10) follows. □

Now, we turn to the proof of Theorem 2.2 and show a few Lemmas first. We divide the ranges of L and L_N into the three intervals $I_1 = [0, K_1]$, $I_2 = (K_2, 1]$ and $I_3 = (K_1, K_2]$, in each of which the function p from (2.3) is linear; the point being that the probability of L and L_N lying in different intervals is small for large N, and the expected difference of $P - P_N$ is small if they are in the same interval. The following Lemmas quantify this.

Lemma 3.2. *For $j = 1, 2$, we have*

$$\mathbb{P}_{|L}(L \in I_j, L_N \in I_j^c) \le 1_{L \in I_j} \, e^{-N(L-K_j)^2}, \tag{3.1}$$

$$\mathbb{P}_{|L}(L \in I_j^c, L_N \in I_j) \le 1_{L \in I_j^c} \, e^{-N(L-K_j)^2}. \tag{3.2}$$

Proof. This is a standard large deviations result. By Theorem 2.2.3 in [8], p. 27, and Remark (c) thereafter, for $\mathcal{F}_L$-independent and identically distributed random variables $(Y_i)_{1\le i\le N}$ with $\mathbb{E}[Y_i \mid \mathcal{F}_L] = L$, we obtain that if $0 < L \le K_j$,

$$\mathbb{P}_{|L}(L_N > K_j) \le e^{-Ng(L,K_j)},$$

and if $K_j < L < 1$,

$$\mathbb{P}_{|L}(L_N \le K_j) \le e^{-Ng(L,K_j)},$$

where the rate function $g(L,K_j)$ is given on p. 35 in [8] as

$$g(L,K_j) = K_j \log\left(\frac{K_j}{L}\right) + (1-K_j) \log\left(\frac{1-K_j}{1-L}\right),$$

since Y_i are Bernoulli distributed random variables with $P(Y_i = 1 \mid \mathcal{F}_L) = L$. It is straightforward to check that for all $L \in (0,1)$

$$g(L,K_j) \ge (K_j - L)^2. \tag{3.3}$$

Hence, by (3.3), for $0 < L \le K_j$

$$\mathbb{P}_{|L}(L_N > K_j) \le e^{-Ng(L,K_j)} \le e^{-N(K_j-L)^2}, \tag{3.4}$$

and similarly for $K_j < L < 1$. These estimates are clearly true for the degenerate cases $L = 0$ and $L = 1$. From this the result follows. $\square$

Lemma 3.3. *Let p be as in (2.5). If A_N is the event that L_N and L are in the same interval and A_N^c its complement, then*

$$\mathbb{E}[(P_N - P)1_{A_N}] = -\mathbb{E}[(L_N - L)1_{A_N^c} 1_{\{L\in I_3\}}]. \tag{3.5}$$

Proof. By splitting the range of L into the different intervals,

$$\begin{aligned}
\mathbb{E}[(P_N - P)1_{A_N}] &= \sum_{j=1}^{3} \mathbb{E}[(P_N - P)1_{A_N} 1_{\{L\in I_j\}}] \\
&= \mathbb{E}[(L_N - L)1_{A_N} 1_{\{L\in I_3\}}] \\
&= -\mathbb{E}[(L_N - L)1_{A_N^c} 1_{\{L\in I_3\}}],
\end{aligned}$$

where we have used in the second line that $P_N = P$ if both L_N and L lie in either I_1 or I_2 and that $P_N - P = L_N - L$ in I_3; in the last line that $\mathbb{E}_{|L}[L_N - L] = 0$ and $1_{A_N} + 1_{A_N^c} = 1$. $\square$

Lemma 3.4. *Let A_N^c be as in Lemma 3.3. If the CDF F_L of L is Lipschitz at K_j, $j = 1,2$, with constant c_L, then*

$$\mathbb{E}\left[\left(\mathbb{P}_{|L}[A_N^c]\right)^{\frac{1}{2}}\right] \le \frac{c_L\, 4\sqrt{\pi}}{\sqrt{N}}. \tag{3.6}$$

Proof. Let $I_1^c = (K_1, 1]$ and $I_2^c = [0, K_2]$ be the complements in $[0,1]$ of I_1 and I_2, then

$$A_N^c \subseteq \{L \in I_1, L_N \in I_1^c\} \cup \{L \in I_2, L_N \in I_2^c\} \cup \{L \in I_1^c, L_N \in I_1\} \cup \{L \in I_2^c, L_N \in I_1^c\}$$

and therefore

$$\begin{aligned} \mathbb{P}_{|L}[A_N^c] \;\; &\leq \;\; \mathbb{P}_{|L}[L \in I_1,\ L_N \in I_1^c] + \mathbb{P}_{|L}[L \in I_2,\ L_N \in I_2^c] + \mathbb{P}_{|L}[L \in I_1^c,\ L_N \in I_1] \\ &\quad + \mathbb{P}_{|L}[L \in I_2^c,\ L_N \in I_2]. \end{aligned}$$

By (3.1), (3.2) we have

$$\mathbb{P}_{|L}[A_N^c] \leq 2\ \left(e^{-N(L-K_1)^2} + e^{-N(L-K_2)^2}\right),$$

and we obtain

$$\mathbb{E}\left[\left(\mathbb{P}_{|L}[A_N^c]\right)^{\frac{1}{2}}\right] \;\; \leq \;\; 2^{\frac{1}{2}}\ \left(\mathbb{E}\left[e^{-N\frac{(L-K_1)^2}{2}}\right] + \mathbb{E}\left[e^{-N\frac{(L-K_2)^2}{2}}\right]\right). \tag{3.7}$$

If we extended F_L by 0 and 1 from $[0,1]$ to $\mathbb{R}$ then, for $j = 1, 2$, we have

$$\begin{aligned} \mathbb{E}\left[e^{-N\frac{(L-K_j)^2}{2}}\right] \;\; &= \;\; \int_{-\infty}^{\infty} e^{-N\frac{(l-K_j)^2}{2}}\ dF_L(l) && (3.8) \\ &= \;\; N\ \int_{-\infty}^{\infty} (l-K_j) e^{-N\frac{(l-K_j)^2}{2}} F_L(l)\ dl && (3.9) \\ &\leq \;\; NF_L(K_j) \int_{-\infty}^{\infty} (l-K_j) e^{-N\frac{(l-K_j)^2}{2}}\ dl + c_L\ N \int_{-\infty}^{\infty} (l-K_j)^2 e^{-N\frac{(l-K_j)^2}{2}}\ dl \\ &= \;\; \frac{c_L\sqrt{2\pi}}{\sqrt{N}}, \end{aligned}$$

where we used the Lipschitz property of the CDF after (3.9) and then integrated exactly. The result follows directly by insertion in (3.7). □

Proof of Theorem 2.2. By the tower property of conditional expectations and Jensen's inequality, we have

$$|\mathbb{E}[(P_N - P)\ 1_{A_N^c}]| \leq \mathbb{E}[|\mathbb{E}_{|L}[(P_N - P)\ 1_{A_N^c}]|]. \tag{3.10}$$

Then Cauchy-Schwarz gives

$$|\mathbb{E}[(P_N - P)\ 1_{A_N^c}]| \leq \mathbb{E}_{|L}\left[\left(\mathbb{E}[(P_N - P)^2]\right)^{\frac{1}{2}}\ \left(\mathbb{P}_{|L}[A_N^c]\right)^{\frac{1}{2}}\right]. \tag{3.11}$$

By Lemmas 3.1 and 3.4, we obtain

$$|\mathbb{E}[(P_N - P)\ 1_{A_N^c}]| \leq \frac{c_L 2\sqrt{\pi}}{N}. \tag{3.12}$$

Similarly, using Lemma 3.3 and the same argument as above,

$$|\mathbb{E}[(P_N - P)\ 1_{A_N}]| \leq \frac{c_L 2\sqrt{\pi}}{N},$$

from which the statement follows.

□

4 An application and numerical results

To illustrate the theoretical rate of convergence, we study numerical results for expected tranche losses of a synthetic CDO for an increasing size N of the underlying CDS pool.

We consider a structural factor model (see, e.g., [13, 4]), where the *distance-to-default* of the i-th firm, $i = 1, \dots, N$, evolves according to

$$X_t^i = X_0^i + \beta t + \sqrt{1-\rho}\, W_t^i + \sqrt{\rho}\, B_t + J_t, \quad t > 0, \tag{4.1}$$

where $\rho \in [0,1)$, β given. Here, B is assumed to be a standard Brownian motion and $J_t = \sum_{k=1}^{CP_t} \Pi_k$, where CP_t is a Poisson process with intensity λ and Π_k are independent Normals with mean μ_Π and variance σ_Π^2, while all W^i are independent standard Brownian motions and independent of B and J. Thus B and J model factors affecting the whole market, whereas W^i are idiosyncratic effects.

The i-th firm is considered to be in default if its distance-to-default is below 0 at any one of the observation times $T_j = jq$, $q = 0.25$ (quarterly), up to $T_{20} = T = 5$, the assumed maturity of the debt here. We introduce the default time τ_i and Bernoulli random variable Y_i indicating default of the i-th firm before T, by

$$\begin{aligned} \tau^i &= \inf\left(\{t \in \{T_1, \dots, T_M\} : X_t^i \le 0\} \cup \{\infty\}\right), \\ Y_i &= 1_{\{\tau^i \le T\}}. \end{aligned} \tag{4.2}$$

For the numerical experiments, the initial values X_0^i are drawn independently from a Normal distribution,

$$X_0^i \sim N(\mu_{X_0}, \sigma_{X_0}^2),$$

where the mean $\mu_{X_0} = 4.6$ and standard deviation $\sigma_{X_0} = 0.8$ are obtained from a calibration to iTraxx data as detailed in [4], as are $\rho = 0.13$, $\lambda = 0.04$, $\mu_\Pi = -0.5$ and $\sigma_\Pi^2 = 0.17$.

That the definition of Y_i in (4.2) fits into the initial set-up is a consequence of the exchangeability of X_t^i in (4.1). If we define

$$\overline{X}_t^i \equiv \begin{cases} X_t^i, & t < \tau^i, \\ 0, & t \ge \tau^i, \end{cases}$$

then the $\overline{X}_T^i$ are still exchangeable. Hence, by de Finetti's Theorem (see [10]), there exists a random measure α on $\mathbb{R}$ such that a.s.

$$\alpha(B) = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^N 1_{\overline{X}_T^i \in B}$$

for all Borel sets B. Conditional on α, the $\overline{X}_T^i$ and Y_i are i.i.d. The link to the random variable L is established by defining

$$L \equiv \alpha(\{0\}) = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^N 1_{\{\overline{X}_T^i = 0\}} = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^N Y_i = \lim_{N\to\infty} L_N.$$

Clearly, Y_i takes values in $\{0,1\}$ and $\mathbb{P}[Y_i = 1 | \mathcal{F}_L] = \mathbb{E}_{|L}[Y_i] = \mathbb{E}_{|L}[L_N] = L$.

It is shown in [4] that the above random measure α is the sum of L times a Dirac measure located at 0 and a continuous part which satisfies a stochastic partial differential equation. To generated (approximate) samples of L, we numerically solve the SPDE by a combined Monte Carlo finite difference method (see again [4]) to generate samples of the random measure, and use this to compute L. So, on this instance, there is an alternative – albeit very costly – way of simulating L directly, and we use this to investigate the relevant properties of L empirically.

Specifically, in view of the conditions of Theorem 2.2, we illustrate the numerically computed CDF F_L of L for different parameters in Figure 1. It appears that F_L is Lipschitz in $(0,1)$ but that the derivative at 0 and 1 can become very large in certain parameter ranges for $\mu_0 = \mu_{X_0}$ and overall instantaneous correlation

$$\rho_A = (\rho + \zeta)/(1 + \zeta), \quad \zeta = \lambda(\mu_\Pi^2 + \sigma_\Pi^2), \tag{4.3}$$

between X_t^i and X_t^j (see [4]).

For large values of μ_0, the probability of defaults becomes very small and the density of L is concentrated around 0. For ρ_A approaching 1, all Y_i become identical and therefore either all or none of the firms default, such that here the density of L is concentrated at 0 and 1. In the degenerate case $\rho_A = 0$ (i.e., $\rho = \lambda = 0$), L is deterministic, the measure is atomic and F_L a step function.

The empirical evidence thus suggests that F_L is Lipschitz in the range $(0,1)$. Given that Theorem 2.2 only requires the Lipschitz property at interior values K_j, the conditions appear to be satisfied and the Theorem to apply in this setting. Even in situations where F_L has a bounded derivative at 0 and 1, the fact that only the Lipschitz constants from K_1 and K_2 enter into the estimates gives us substantially smaller bounds.

We now move on to present numerical results for the payoff function p from (2.5) illustrating the convergence as the number of firms N goes to infinity. We consider portfolios consisting of $N_k = M^k = 5^k$ companies for $k = 1, \ldots, 7$.

To include a recovery value of defaulted firms in the model, we rescale L_N by $(1-R)$, where $R = 0.4$ is the recovery rate. Equivalently, we pick $(K_1, K_2) = (1-R)^{-1}(a,d)$ in (2.3) and $(a,d) \in \{(0, 0.03), (0.03, 0.06), (0.06, 0.09), (0.09, 0.12), (0.12, 0.22), (0.22, 1)\}$ as the attachment and detachment points for iTraxx tranches, and then study $(1-R)p(L_N)$.

A straightforward Monte Carlo estimator for expected tranche losses $\mathbb{E}[P^{(k)}]$ is then given by

$$\widehat{G}_k = \frac{1}{n}\sum_{j=1}^{n}(1-R)p(L^{(k,j)}), \tag{4.4}$$

$$L^{(k,j)} = \frac{1}{N_k}\sum_{i=1}^{N_k} Y_i^{(j)}, \tag{4.5}$$

where $(Y_i^{(j)})$ are independent samples of Y_i, i.e., corresponding to independent paths for B, W and J. There is no time discretisation error as (4.1) can be sampled directly. However, it turns out to be computationally prohibitively expensive to choose n, the number of samples, large enough to produce estimators with sufficiently small RMSE to allow us to distinguish between $\widehat{G}_k$ and $\widehat{G}_{k+1}$ for large k.

We therefore use the multilevel simulation approach outlined in Section 2 and detailed further in Section 5. The point is that the differences $G_{k+1} - G_k$ are simulated directly in

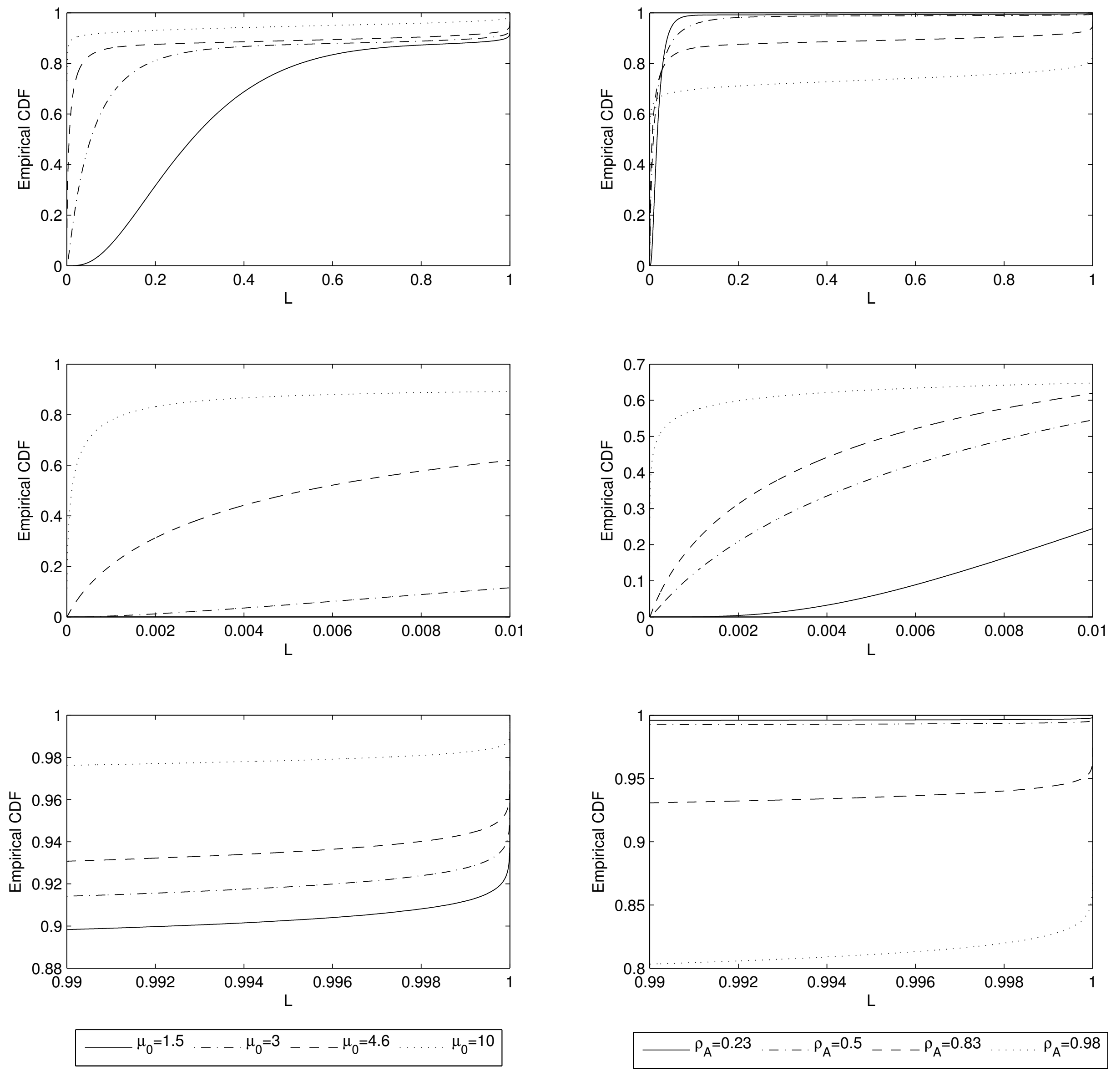

Figure 1: Top row: Empirical CDF F_L for different values of $\mu_0 = \mu_{X_0}$ (left) and different ρ_A (right). The values of ρ_A are arrived at by (4.3) from $(\rho, \lambda) \in \{(0.03, 0.001), (0.1, 0.002), (0.35, 0.0035), (0.35, 0.0351), (0.8, 0.1)\}$. All other parameters are fixed as given in the text. The plots in the second and third rows are zoomed into the ranges of L close to 0 and 1, respectively.

the multilevel approach. Therefore, we approximate $|G - G_k|$, where $G \equiv \lim_{k\to\infty} G_k$, by

$$S_k = |G_k - G_K| = \left| \sum_{l=k+1}^{K} Z_l \right| \tag{4.6}$$

for $k < K$, where Z_l is an estimator for $\mathbb{E}[P^{(l)} - P^{(l-1)}]$ as used in the construction of G_k in (2.16) (precisely, we used the estimator Z_l defined later in (2.13)). The difference between S_k and $|G - G_k|$ for $k = K - 1$ is given by $G_{K-1} - G_K \approx (G_{K-1} - G)(1 - 1/M)$ and for $k = K - 2$ by $G_{K-2} - G_K \approx (G_{K-2} - G)(1 - 1/M^2)$. Given $M = 5$ in our examples, the error due to this approximation will be seen to be smaller than the estimation error.

The results are shown in Figure 2. We plot the logarithm of S_k to base M, together with the sample standard deviation of the the multilevel estimators G_k (see (2.16)) and

$$y_k = -k + y_0, \tag{4.7}$$

where y_0 is a suitably chosen constant, to verify the predicted convergence order empirically. The data points appear to be in good agreement with first order convergence.

5 Analysis of the multilevel method

In this section, we describe and analyse a multilevel simulation approach for the estimation of expected functionals of the form (2.3) and (2.4), the latter with a particular emphasis on the case of large N.

The multilevel Monte Carlo method proposed by Giles in [11] estimates the expected value of a functional of the solution to a stochastic differential equation obtained by a timestepping scheme. It performs computations on different refinement levels l with time steps $h_l = h_0 M^{-l}$ for $M > 1$, such as to minimise the overall computational time of the Monte Carlo estimator for prescribed mean square error (MSE). Since the MSE consists of a Monte Carlo error (variance) and a discretisation error (bias), the method controls both the number of samples n_l on level l, to bound the Monte Carlo variance of order $O(n_l^{-1})$, and the finest L with time step h^{-L} on which to approximate the SDE, in order to reduce the bias. The multilevel method is based on two premises: Monte Carlo estimators for an increasing number of time steps converge at a certain order in h_l, and the computational cost needed to calculate an estimator increases with $n_l h_l^{-1}$. In this approach, estimators obtained with a smaller number of time steps are used as control variates for estimators with a larger number of time steps, which significantly decreases the computation time.

To obtain a complexity result for an estimator of $\mathbb{E}[P]$ with P from (2.3), we substitute h_l by N_l^{-1} in Theorem 3.1 of [11] and immediately obtain Proposition 2.1 from Section 2.

By direct inspection, for the construction of Z_l from (2.13), Assumption ii) holds in Proposition 2.1. From Theorem 2.1, we know that i) holds with $\alpha = 1/2$ for general Lipschitz p. Clearly, the computational effort to compute Z_l is proportional to $n_l N_l$ as required in iv). Finally, iii) holds by the following simple application of Lemma 3.1.

Proposition 5.1. *Let $P^{(l)} = P_{N_l}$ as per (2.4), where p is Lipschitz with constant c_p, then*

$$Var[P^{(l)} - P^{(l-1)}] \quad \le \quad c_p^2 \, \frac{M+1}{2N_l}. \tag{5.1}$$

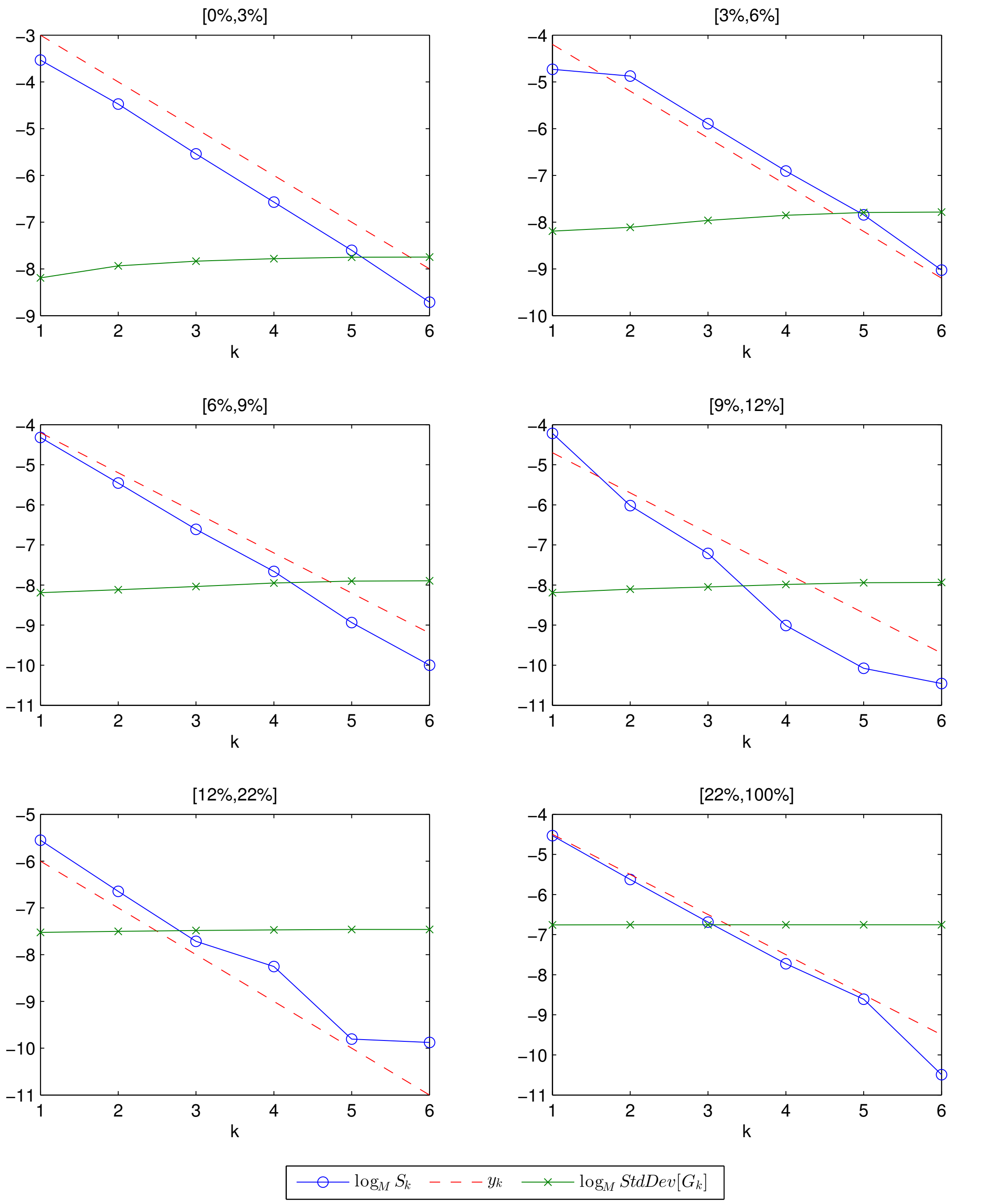

Figure 2: Shown here is $\log_M S_k$, where S_k given by (4.6) is an estimator for $|\mathbb{E}[P^{(k)} - P]|$. The various plots are for tranches ranging from [0%-3%] to [22%-100%], of a CDO basket consisting of $N_k = M^k = 5^k$ companies, where $k = 1, \ldots, 6$. The comparison with the predicted trend y_k from (4.7) confirms the first order convergence. Included is also the standard deviation of the estimated tranche loss G_k.

Proof. This follows directly from

$$\begin{aligned}\mathbb{E}_{|L}[(P^{(l)} - P^{(l-1)})^2] &= \mathbb{E}_{|L}[((P^{(l)} - P) - (P^{(l-1)} - P))^2] \\ &\leq 2\left(\mathbb{E}_{|L}[(P^{(l)} - P)^2] + \mathbb{E}_{|L}[(P^{(l-1)} - P)^2]\right) \qquad (5.2)\end{aligned}$$

by Lemma 3.1 and taking expectations over L. □

We have therefore proven the first statement of Corollary 2.1.

In practice, it is also relevant to be able to compute $\mathbb{E}[P_N]$ efficiently for finite N. It is clear that for fixed N, the complexity is bounded by $c\,\epsilon^{-2}$ for some $c > 0$, but for a naïve (single-level) estimator the constant c will increase with N, as detailed in Section 2. From the proof of Theorem 3.1 in [11] it is clear, however, that there is a multilevel estimator with *a priori* bounded upper level K which satisfies the second statement in Corollary 2.1.

We now discuss the multilevel estimator $\overline{Z}_l$, based on the faster decay rate 3/2 for piecewise linear payoffs in Theorem 2.3, which we prove subsequently.

It is clear that $\overline{Z}_l$ satisfies ii) in Proposition 2.1 and that the computational complexity is still bounded as required per iv). In fact, as the main computational cost is typically in sampling Y_i, the computational complexity is virtually identical to that of Z_l. In particular, if we evaluate (2.14) by using (2.18) and the already computed (2.22), the difference in evaluating Z_l and $\overline{Z}_l$ is an $O(M)$ cost, i.e., independent of N_l. Now, given Theorem 2.3, we have that

$$Var[\overline{Z}_l] \leq c\,n_l^{-1} M^{-3/2\,l}, \qquad (5.3)$$

for some c, such that we are in the first regime in the complexity result of Proposition 2.1, i.e., we have optimal complexity order.

We have not commented so far on the (optimal) selection of M. The choice of $M = 5$ in Section 4 was to some extent dictated by the application of a CDO basket where the target size is $N = 125 = 5^3$, and therefore for $M = 5$ this N is reached exactly for level $K = 3$. For different M, or indeed for N which is not an integer power of an integer M, one can adapt the method easily by choosing K as the largest integer such that $M^{K-1} < N$, and then estimate the correction between $\mathbb{E}[P^{(K-1)}]$ and $\mathbb{E}[P_N]$ by a last estimator Z_K. Such considerations are obviously irrelevant for the estimation of $\mathbb{E}[P]$, and there the choice of M is entirely dictated by complexity issues.

The total error is a combination of the bias, dictated by the number of Bernoulli random variables N_K on the finest level and therefore largely independent of L, and the variance of the individual estimators Z_l or $\overline{Z}_l$. The effect of increasing M is that the variance of Z_l may increase, but conversely the number of levels required to reach a given N will decrease and therefore the total number of random variables which need to be simulated may be lower. There is a discussion in [11] on the optimal selection, with a heuristic calculation for $\beta = 1$, suggesting an optimal value of 6 or 7, which is then lowered to 4 in computations to incorporate a sufficient number of levels for a reliable estimation of the variance on course levels. For a faster decay of the variance, $\beta = 3/2$, the optimal M can be expected to be smaller, and therefore $M = 5$ seems a sensible choice, although we did not test this systematically.

The remainder of this section is devoted to the proof of Theorem 2.3.

Lemma 5.1. *Assume the CDF F_L of L is Lipschitz with constant c_L. Let $B^{(l)}$ be the event that $L^{(l)}$ lies in the same interval as $L^{(l-1)}$, $B^{(l),c}$ its complement, then*

$$\mathbb{E}\left[\left(\mathbb{P}_{|L}[B^{(l),c}]\right)^{\frac{1}{2}}\right] \le \frac{C}{\sqrt{N_l}},$$

where $C = c_L\, 4\sqrt{\pi}(\sqrt{2}+\sqrt{M})$.

Proof. Let $A^{(l)}$ again be the event that $L^{(l)}$ and L are in the same interval, $A^{(l),c}$ its complement. Then from

$$B^{(l),c} \quad \subseteq \quad \left(A^{(l)} \cap A^{(l-1),c}\right) \bigcup \left(A^{(l),c} \cap A^{(l-1)}\right) \bigcup \left(A^{(l),c} \cap A^{(l-1),c}\right)$$

follows

$$\begin{aligned}
\mathbb{P}_{|L}[B^{(l),c}] &\le \mathbb{P}_{|L}[A^{(l)} \cap A^{(l-1),c}] + \mathbb{P}_{|L}[A^{(l),c} \cap A^{(l-1)}] + \mathbb{P}_{|L}[A^{(l),c} \cap A^{(l-1),c}] \\
&\le 2\,\mathbb{P}_{|L}[A^{(l),c}] + \mathbb{P}_{|L}[A^{(l-1),c}],
\end{aligned}$$

which leads to

$$\mathbb{E}\left[\left(\mathbb{P}_{|L}[B^{(l),c}]\right)^{\frac{1}{2}}\right] \le \sqrt{2}\,\mathbb{E}\left[\left(\mathbb{P}_{|L}[A^{(l),c}]\right)^{\frac{1}{2}}\right] + \mathbb{E}\left[\left(\mathbb{P}_{|L}[A^{(l-1),c}]\right)^{\frac{1}{2}}\right].$$

By Lemma 3.4, we obtain the result. □

Lemma 5.2. *For P, $P^{(l)}$, $P^{(l-1)}$ and $\overline{P}^{(l-1)}$ as above, p Lipschitz with constant 1,*

$$\mathbb{E}_{|L}[(P^{(l)} - P)^4] \le \frac{3}{16N_l^2}\left(1 + \frac{4}{3N_l}\right) \le \frac{7}{16N_l^2}, \tag{5.4}$$

$$\mathbb{E}_{|L}[(P^{(l)} - P^{(l-1)})^4] \le \frac{C}{N_l^2}, \tag{5.5}$$

$$\mathbb{E}_{|L}[(P^{(l)} - \overline{P}^{(l-1)})^4] \le \frac{C}{N_l^2}, \tag{5.6}$$

where $C = \frac{7}{8}(M^2 + 6M + 1)$.

Proof. See Appendix A. □

Proof of Theorem 2.3. Let $E^{(l)}$ be the event that all $L_m^{(l-1)}$ lie in the same interval, $1 \le m \le M$, and $E^{(l),c}$ its complement, then

$$\mathbb{E}[(P^{(l)} - \overline{P}^{(l-1)})^2] = \mathbb{E}[(P^{(l)} - \overline{P}^{(l-1)})^2 1_{E^{(l)}}] + \mathbb{E}[(P^{(l)} - \overline{P}^{(l-1)})^2 1_{E^{(l),c}}].$$

By (2.18) and linearity of p in each interval, we have

$$\mathbb{E}[(P^{(l)} - \overline{P}^{(l-1)})^2 1_{E^{(l)}}] = 0.$$

By Cauchy-Schwartz, we have

$$\mathbb{E}_{|L}[(P^{(l)} - \overline{P}^{(l-1)})^2 1_{E^{(l),c}}] \le \left(\mathbb{E}_{|L}[(P^{(l)} - \overline{P}^{(l-1)})^4]\right)^{\frac{1}{2}} \left(\mathbb{P}_{|L}[E^{(l),c}]\right)^{\frac{1}{2}},$$

hence,

$$\mathbb{E}[(P^{(l)} - \overline{P}^{(l-1)})^2 1_{E^{(l),c}}] \le \mathbb{E}\left[\left(\mathbb{E}_{|L}[(P^{(l)} - \overline{P}^{(l-1)})^4]\right)^{\frac{1}{2}} \left(\mathbb{P}_{|L}[E^{(l),c}]\right)^{\frac{1}{2}}\right].$$

By Lemma 5.2, we have that

$$\left(\mathbb{E}_{|L}[(P^{(l)} - \overline{P}^{(l-1)})^4]\right)^{\frac{1}{2}} \le \frac{\sqrt{c_1}}{N_l}, \tag{5.7}$$

where $c_1 = \frac{7}{8}(M^2 + 6M + 1)$.

If we denote by $B_m^{(l)}$ the event that $L_m^{(l-1)}$ and $L^{(l)}$ lie in the same interval, then

$$E^{(l),c} = \bigcup_{m=1}^{M} B_m^{(l),c}$$

and therefore

$$\mathbb{P}_{|L}(E^{(l),c}) \quad \le \quad \sum_{m=1}^{M} \mathbb{P}_{|L}(B_m^{(l),c}) \quad = \quad M\,\mathbb{P}_{|L}(B^{(l),c}).$$

By Lemma 5.1, this gives

$$\mathbb{E}\left[\left(\mathbb{P}_{|L}[E^{(l),c}]\right)^{\frac{1}{2}}\right] \le \frac{c_2}{N_l},$$

where $c_2 = c_L\, 4\sqrt{M\pi}(\sqrt{2} + \sqrt{M})$. Together with (5.7), we obtain the result. □

6 Multilevel tests

In this section, we present multilevel simulation results based on the estimators from the previous section and illustrating the theoretical findings from there. We return to the example from Section 4 and estimate expected tranche losses for credit baskets with an increasing number of firms $N_l = M^l$.

For the estimator Z_l from (2.13), an upper bound for the variance – although not a sharp one – is analytically known from (5.1) and we could use that to determine the number n_l of samples on level l which is required to bring the variance contribution under a desired threshold. For the improved estimator $\overline{Z}_l$ from (2.21), however, the bound in (5.3) contains the unknown Lipschitz constant of the CDF of F_L via Theorem 2.3. In order to determine the optimal allocation n_l^*, we use the following algorithm as per [11]. In contrast to there, the upper level K is fixed here which simplifies the stopping criterion somewhat.

1. Start with $k = 1$.
2. Estimate the variance V_k of a single sample using $n_k = 10^4$ realisations.
3. Calculate the optimal number of samples, n_l^*, for $l = 0, 1, \dots, k$, using

$$n_l^* = \left\lceil \gamma^{-2}\sqrt{V_l\, N_l^{-1}} \left(\sum_{j=1}^{k} \sqrt{V_j\, N_j}\right)\right\rceil, \tag{6.1}$$

   where γ^2 is a chosen upper bound of $Var[G_K]$.

4. Draw extra samples for each level according to n_l^*.

5. If $k < K$, set $k = k + 1$ and go to 2.

6. If $k = K$, finish.

Remark 6.1. *As per [11], choosing n_l^* by (6.1), guarantees that the variance $Var[G_K]$ is bounded by γ^2, since*

$$Var[G_K] \;=\; \sum_{l=1}^{K} (n_l^*)^{-1} V_l \leq \sum_{l=1}^{K} \left(\gamma^{-2} \sqrt{V_l N_l^{-1}} \sum_{j=1}^{K} \sqrt{V_j N_j} \right)^{-1} V_l \;=\; \gamma^2.$$

A side effect is that, for $k < K$, the variance is smaller than for $k = K$, since

$$Var[G_k] = \sum_{l=1}^{k} (n_l^*)^{-1} V_l < \gamma^2 \frac{\sum_{l=1}^{k} \sqrt{V_l N_l}}{\sum_{l=1}^{K} \sqrt{V_l N_l}}.$$

Hence, if we compute estimators G_k for all k as a by-product of G_K, the variance is the smallest for G_1 and then for G_k, $k = 2, \ldots, K$, gradually reaches the upper bound γ^2. This effect can be observed in Figure 3.D.

In Figure 3 we show results for the same parameter setting as in Section 4 and only for the equity tranche. Results from other tests were very similar and did not show any noteworthy additional effects. In order to easily see the rate of convergence in 3.A., we plot the logarithm of V_l to base M, together with

$$f_k = -\beta\, k + f_0 \tag{6.2}$$

for different values of β. The estimated slope is $\widehat{\beta} \approx 1$ for the original estimator and $\widehat{\beta} \approx 3/2$ for the improved estimator, which agrees with the theoretical findings. The order of convergence of $|E[P^{(l)} - P^{(l-1)}]|$ is $\widehat{\alpha} \approx 1$, which also agrees with the previous results. As can be observed in Figure 3.C, he number of samples ranges from 150 millions for $k = 1$ to 34000 for $k = 7$. The improved estimator gives further reductions in computational time: the total number of samples ranges now from 35 millions for $k = 1$ to only 350 for $k = 7$. The standard deviation of G_k is an increasing function of k, and is less than or equal to the chosen upper bound $\gamma = 4 \times 10^{-6}$.

7 Conclusions and extensions

A main focus of this paper was the construction of an efficient simulation algorithm for functionals of a large number of exchangeable random variables. For a specific set-up, we were able to demonstrate optimal complexity order by theoretical analysis and numerical illustrations.

Discussion

The results from the previous section show that the computational savings can be significant in situations of practical relevance. As seen from Figure 3.C, already for $N = 125$ (i.e., $k = 3$), the size of a CDO basket, the required number n_3 of samples on this level

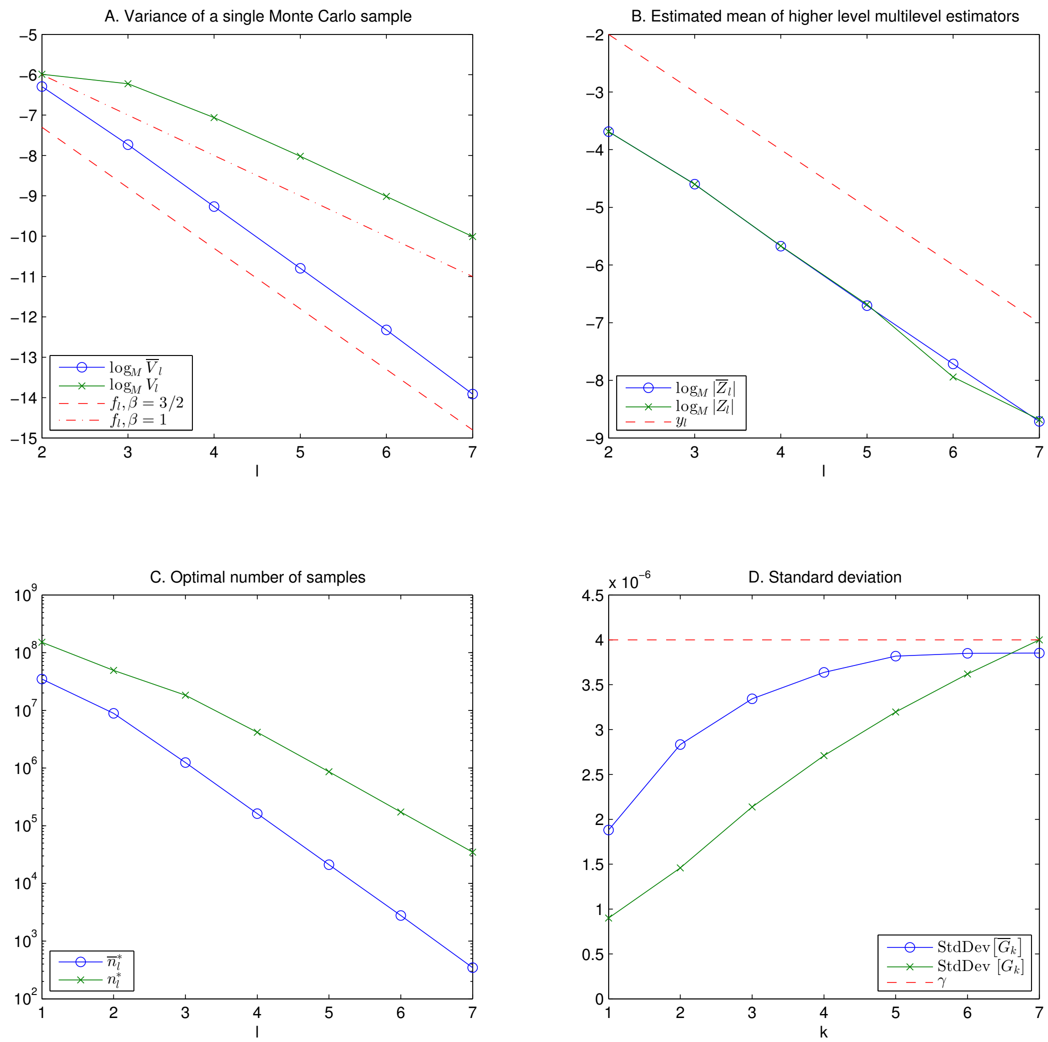

Figure 3: Multilevel results for the expected loss in the equity tranche of a CDO basket consisting of N_k companies, $N_k = M^k = 5^k$, $k = 1, \ldots, 7$. Overlined quantities refer to the estimator $\overline{Z}_l$ from (2.21), all others to the standard estimator Z_l from (2.13). A. Variance of a single Monte Carlo sample, V_l and $\overline{V}_l$, together with a predicted trend, f_l, given by (6.2), where $\beta = 1$ or $\beta = 3/2$. B. Mean at level l, Z_l and $\overline{Z}_l$, and a trend, y_l, defined by (4.7), with slope -1. C. Optimal number of simulations in both cases, n_l^* and $\overline{n}_l^*$, calculated according to (6.1) for $k = K = 7$. D. Standard deviation of multilevel estimators G_k defined in (2.16), and similar for $\overline{G}_k$, with their chosen upper bound, γ.

is reduced by about two orders of magnitude compared to the number of samples for $k = 1$, n_1. It is roughly this number which would be required for a standard (i.e., single level) estimator on level 3 for a variance comparable to the one achieved by the multilevel estimator at substantially lower cost.

Extensions – random recovery and random factor loadings

There is ample empirical evidence that a basic factor model such as the one described in Section 4 does not adequately reproduce observed market spreads of credit derivatives and other stylised facts of credit markets. Two effects that have so far been neglected are credit contagion (i.e., the default of one firm has an impact on the credit worthiness and dependence structure of others) and the dependence of recovery values on the wider credit environment. We focus here on the latter effect and follow [2] for a model that captures this dependence.

Consider thus the total loss as given by

$$\widetilde{L}_N = \frac{1}{N}\sum_{i=1}^{N} l_i\, Y_i,$$

where l_i represent a random *loss-given-default* for company i and Y_i are default indicators as previously. It is sometimes convenient to write

$$l_i = l_{\max} \cdot (1 - R_i)\,,$$

where $l_{\max}$ is a (constant) notional maximum loss and $R_i \in [0,1]$ is the (random) recovery rate of the i-th firm. In keeping with our general framework, we assume that the R_i are identically distributed and independent conditional on $\mathcal{F}_L$.

For continuous payoffs p, we still have

$$\begin{aligned} \widetilde{L}_N &\to \mathbb{E}[\widetilde{L}_N] =: \widetilde{L} \quad \text{for } N \to \infty,\ \mathbb{P}_{|L} - a.s., && (7.1)\\ \widetilde{P}_N := p(\widetilde{L}_N) &\to \mathbb{E}[\widetilde{P}_N] =: \widetilde{P} \quad \text{for } N \to \infty,\ \mathbb{P}_{|L} - a.s. && (7.2) \end{aligned}$$

The L^2-convergence is described in the following.

Corollary 7.1 (to Theorem 2.1)**.** *Let $\widetilde{P}$ and $\widetilde{P}_N$ be given by (7.2), and assume that p is Lipschitz with constant c_p. We have that*

$$\begin{aligned} |\mathbb{E}[\widetilde{P}_N - \widetilde{P}]| &\le \frac{c_p\ l_{\max}}{\sqrt{N}}, && (7.3)\\ Var[\widetilde{P}_N - \widetilde{P}] &\le \frac{c_p^2\ l_{max}^2}{N}. && (7.4) \end{aligned}$$

Proof. In the same way as the proof of Lemma 3.1,

$$\mathbb{E}_{|L}[(\widetilde{P}_N - \widetilde{P})^2] \le c_p^2\ \mathbb{E}_{|L}[(\widetilde{L}_N - \widetilde{L})^2] = c_p^2\ Var[\widetilde{L}_N|\mathcal{F}_L] = \frac{c_p^2}{N}\ Var[l_i Y_i|\mathcal{F}_L] \le \frac{c_p^2\ l_{\max}^2}{N},$$

which gives the result for the variance. The result for the expectation follows again immediately. □

The order 1/2 for the convergence of the expectations and order 1 for the variances is sufficient to be able to apply Corollary 2.1 to establish the $\epsilon^2 \log^2 \epsilon$ complexity for MSE ϵ^2 of the multilevel method. The following numerical tests indicate that the order 1/2 is not sharp and indeed we expect order 1 for sufficient regularity of the payoffs and/or distribution function of L. The proof of this becomes more technical than in the pure Bernoulli case because we lose the explicit form of the characteristic function. Thus, and because of the irrelevance of this for the convergence speed of the multilevel method, we do not pursue this further here.

We now consider a particular model similar to the one in [2] and give a numerical illustration. Specifically, let

$$\begin{aligned} X_t^i &= X_0^i + \beta t + \sqrt{1-\rho}\, W_t^i + \sqrt{\rho}\, B_t + J_t, \quad t > 0, &(7.5)\\ R_t^i &= \Phi(\mu^R + \beta^R t + \sigma^R \sqrt{1-\rho^R}\, \xi_t^i + \sigma^R \sqrt{\rho^R}\, B_t + J_t), &(7.6) \end{aligned}$$

where the processes in the first line are defined as in (4.1), and in the second line (ξ^i) is a standard Brownian motion independent of everything else, while $\mu^R, \beta^R,\ \sigma^R > 0$ and $0 \le \rho^R \le 1$ are constants. Φ is the cumulative density of the standard normal, but could be replaced by any increasing function $f : \mathbb{R} \to [0,1]$. This has the effect that the recovery rate is positively correlated with the market factors B and J, with some idiosyncratic noise, and thus there is a negative dependence between recovery rates and default frequencies. See, for instance, [1] for an early but influential study of this empirical fact.

The above model is not precisely contained in the set-up of Corollary 7.1, because the recovery rates processes (7.6) are not independent conditional on L (as a result of the different exposure of R^i to B and J compared to X^i). However, both the X^i and R^i are independent conditional on B and J, and therefore if $\mathcal{F}_L$ in the proof of Corollary 7.1 is replaced by $\mathcal{F}_B \times \mathcal{F}_J$, the filtration generated by the common factors (a larger filtration than $\mathcal{F}_L$), the result still follows.

In the numerical simulations, we choose the values of μ_R, β_R and σ_R such that, for all $i = 1, \ldots, N$, $R_0^i = 0.4$, $\mathbb{E}[l_i Y_i \,|\, \tau^i \le T] \approx 0.7$ and $Var[l_i Y_i \,|\, \tau^i \le T] \approx 0.16$, compared to Section 6, where the recovery rate is constant at 0.4. In particular, we have $\mu_R = -0.25$, $\beta_R = 1.5$ and $\sigma_R = 0.9$. Also, we assume that $\rho_R = \rho$. All other parameters are the same as in the tests in Sections 4 and 6 (see the paragraphs after (4.1) for the model set-up and parameter values).

The results in Fig. 4 are presented in the same format as Fig. 3 earlier for the constant recovery rate. There is clear evidence that the convergence of the variance and mean are still both of first order in $1/N$, where N is the basket size.

Another extension, also proposed in [2], are *random factor loadings* of the type

$$X_t^i = X_0^i + \beta t + \nu\, W_t^i + a(B_t)\, B_t, \quad t > 0, \tag{7.7}$$

where a is a given deterministic function and $\nu^2 = 1 - Var(a(B_t)B_t)$. This model can capture contagion effects where, for decreasing a, firm values are more closely correlated to the common market factor in bad times. Such an extension fits directly in the general framework developed earlier in this paper, assuming the technical conditions on the cumulative density F_L hold where needed (see in particular Theorem 2.2).

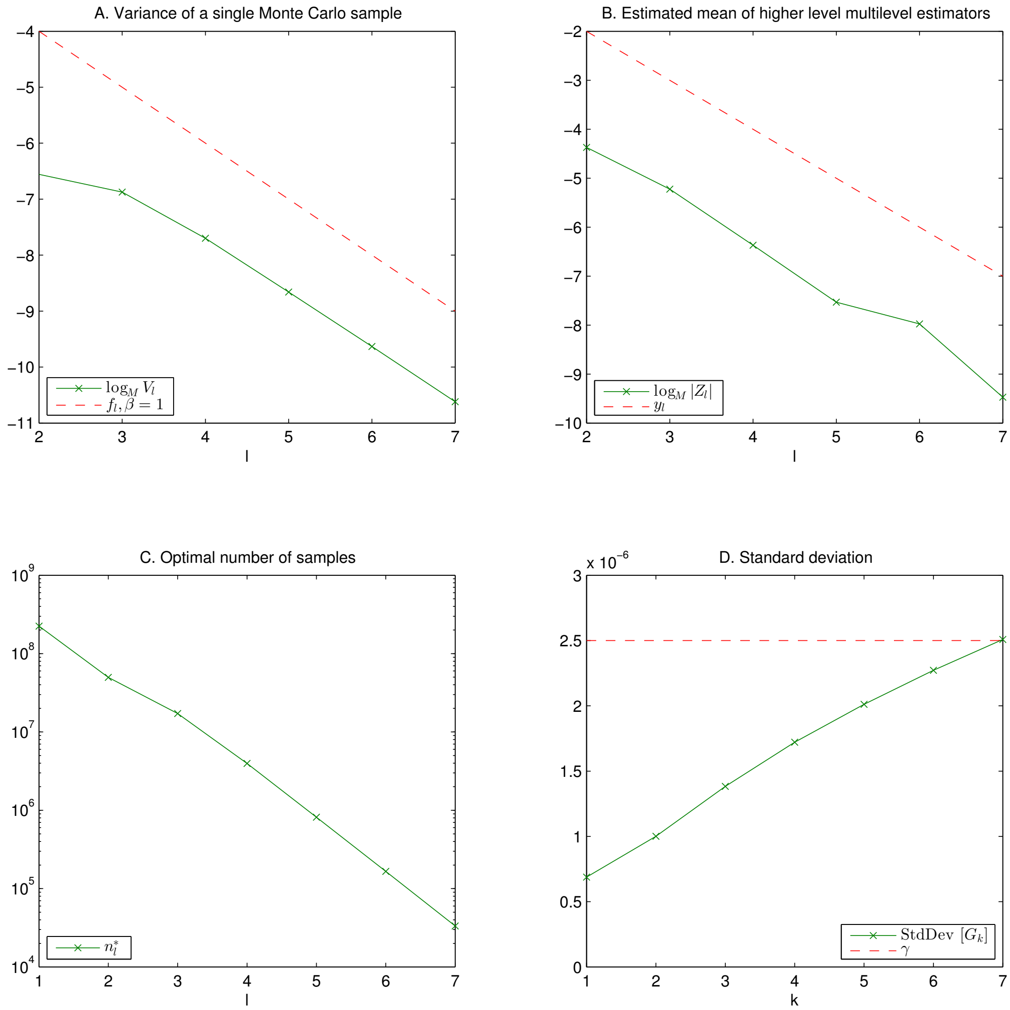

Figure 4: Multilevel results for the expected loss in the equity tranche of a CDO basket consisting of N_k companies, $N_k = M^k = 5^k$, $k = 1, \ldots, 7$, where the recovery rate is random and given by (7.6). The quantities refer to the standard estimator Z_l from (2.13). A. Variance of a single Monte Carlo sample, V_l together with a predicted trend, f_l, given by (6.2), where $\beta = 1$. B. Mean at level l, Z_l and a trend, y_l, defined by (4.7), with slope -1. C. Optimal number of simulations, n_l^*, calculated according to (6.1) for $k = K = 7$. D. Standard deviation of multilevel estimators G_k defined in (2.16) with chosen upper bound, γ.

Outlook

We would expect there to be scope to apply the presented nested simulation approach to a wider range of settings beyond the particular application studied here. An interesting extension would be to the model from [9], where the analysis requires further tools accounting especially for the heterogeneity of the basket, resulting in non-exchangeability. While our motivation comes from credit baskets and some of the later results are specific to piecewise linear functionals encountered in the valuation of basket credit derivatives, we hope there to be a wider relevance of the main approach to the simulation of certain functionals arising in large interacting particle systems and elsewhere.

Acknowledgement: We thank Mike Giles for suggesting the improved estimator, and two anonymous referees for their helpful suggestions which improved the presentation of the results.

References

[1] E.I. Altman, B. Brady, A. Resti, and A. Sironi, *The Link between Default and Recovery Rates: Theory, Empirical Evidence, and Implications*, J. Business, 78 (6), 2203–2208, November 2005.

[2] L. Andersen and J. Sidenius, *Extensions to the Gaussian copula: random recovery and random factor loadings* , J. Credit Risk, 1 (1), 29–70, Winter 2004/05.

[3] D. Belomestny, J. Schoenmakers, and F. Dickmann, *Multilevel dual approach for pricing American style derivatives*, Finance Stoch., DOI 10.1007/s00780-013-0208-5, 2013.

[4] K. Bujok and C. Reisinger, *Numerical valuation of basket credit derivatives in structural jump-diffusion models*, J. Comp. Fin., 15(4), 115–158, 2012.

[5] N. Bush, B. Hambly, L. Jin, H. Haworth, and C. Reisinger, *Stochastic evolution equations in portfolio credit modelling*, SIAM J. Finan. Math., 2, 627–664, 2011.

[6] N. Chen, P. Glynn, and Y. Liu, *Computing functions of conditional expectation via multilevel nested simulation*, Conference Presentation at MCQMC 2012.

[7] A. Dembo, J.-D. Deuschel, and D. Duffie, *Large portfolio losses*, Finance Stoch., 8, 3–16, 2004

[8] A. Dembo and O. Zeitouni, *Large deviations techniques and applications*, Jones and Bartlett Pulishers, Boston, first edition, 1993.

[9] K. Giesecke, K. Spiliopoulos, R.B. Sowers, and J.A. Sirignano, *Large portfolio asymptotics for loss from default*, Math. Finance, DOI 10.1111/mafi.12011, 2012.

[10] O. Kallenberg, *Probabilisitc symmetries and invariance principles*, Springer, New York, 2005.

[11] M. Giles, *Multi-level Monte Carlo path simulation*, Oper. Res., 56, 607–617, 2008.

[12] P. Glasserman, W. Kang, and P. Shahabuddin, *Large deviations in multifactor portfolio credit risk*, Math. Finance, 17(3), 345–379, 2007.

[13] A. Lipton, *Assets with jumps*, RISK, 15(9), 149–153, 2002.

[14] O. Vasicek, *Limiting loan loss probability distribution*, KMV Corporation, Document Number: 999–0000–046, 1991.

A Moment computations

Proof. [of Lemma 5.2] We begin by showing (5.4) and then deduce (5.5) and (5.6). We have

$$| P^{(l)} - P | \leq | L^{(l)} - L |,$$

where

$$L^{(l)} = \frac{1}{N_l} \sum_{i=1}^{N_l} Y_i.$$

Hence, we get

$$\mathbb{E}_{|L}[(P^{(l)} - P)^4] \leq \mathbb{E}_{|L}[(L^{(l)} - L)^4] = \mathbb{E}_{|L}\left[\left(\frac{1}{N_l}\sum_{i=1}^{N_l}(Y_i - L)\right)^4\right].$$

As L is $\mathcal{F}_L$-measurable and the Y_i are independent and identically distributed given $\mathcal{F}_L$ with $\mathbb{E}_{|L}[Y_i - L] = 0$, we have

$$\begin{aligned}
\mathbb{E}_{|L}[(L^{(l)} - L)^4] &= \frac{1}{N_l^4}\mathbb{E}_{|L}\left[\sum_{i=1}^{N_l}(Y_i - L)^4 + 6\sum_{i\neq j}(Y_i - L)^2(Y_j - L)^2\right] \\
&= \frac{1}{N_l^3}\left((1-L)^4 L + L^4(1-L)\right) + \frac{3(N_l - 1)}{N_l^3}\left((1-L)^2 L + L^2(1-L)\right)^2 \\
&= \frac{3L^2(1-L)^2}{N_l^2} + \frac{L(1-L)(1-6L(1-L))}{N_l^3}.
\end{aligned}$$

Using the fact that $0 \leq L(1-L) \leq 1/4$ we have the required bound in (5.4).

For (5.5), observe that there are many ways of estimating this fourth moment; we choose the following

$$\begin{aligned}
(P^{(l)} - P^{(l-1)})^4 &= \left((P^{(l)} - P) - (P^{(l-1)} - P)\right)^4 \\
&\leq 2\left((P^{(l)} - P)^4 + (P^{(l-1)} - P)^4\right) + 12(P^{(l)} - P)^2(P^{(l-1)} - P)^2.
\end{aligned}$$

Therefore, using Cauchy-Schwarz on the last term and applying (5.4) we have

$$\begin{aligned}
\mathbb{E}_{|L}[(P^{(l)} - P^{(l-1)})^4] &\leq 2\left(\mathbb{E}_{|L}[(P^{(l)} - P)^4] + \mathbb{E}_{|L}[(P^{(l-1)} - P)^4]\right) \\
&\qquad + 12\,\mathbb{E}_{|L}[(P^{(l)} - P)^4]^{1/2}\,\mathbb{E}_{|L}[(P^{(l-1)} - P)^4]^{1/2} \\
&\leq \frac{7}{8N_l^2} + \frac{7}{8N_{l-1}^2} + \frac{42}{8N_l N_{l-1}}
\end{aligned}$$

as required to obtain (5.5).

Finally, (5.6) follows from

$$
\begin{aligned}
\mathbb{E}_{|L}[(P^{(l)} - \overline{P}^{(l-1)})^4] &= \mathbb{E}_{|L}\left[\left(\frac{1}{M}\sum_{m=1}^{M}(P^{(l)} - P_m^{(l-1)})\right)^4\right] \\
&\leq \mathbb{E}_{|L}\Big[\frac{1}{M}\sum_{m=1}^{M}(P^{(l)} - P^{(l-1)})^4\Big] \\
&= \mathbb{E}_{|L}[(P^{(l)} - P^{(l-1)})^4],
\end{aligned}
$$

and an application of (5.5). □