COLEMAN FUNG
RISK MANAGEMENT
RESEARCH CENTER

Working Paper # 2013 -02

Efficient Monte Carlo Counterparty Credit Risk Pricing and Measurement

Samin Ghamami, University of California at Berkeley
Bo Zhang, IBM Thomas J. Watson Research Center, NY

January 15, 2013

**University of California
Berkeley**

Efficient Monte Carlo Counterparty Credit Risk Pricing and Measurement

Samim Ghamami* and Bo Zhang†

January 15, 2013

Abstract

Counterparty credit risk (CCR), a key driver of the 2007-08 credit crisis, has become one of the main focuses of the major global and U.S. regulatory standards. Financial institutions invest large amounts of resources employing Monte Carlo simulation to measure and price their counterparty credit risk. We develop efficient Monte Carlo CCR frameworks by focusing on the most widely used and regulatory-driven CCR measures: expected positive exposure (EPE), credit value adjustment (CVA), and effective expected positive exposure (eEPE). Our numerical examples illustrate that our proposed efficient Monte Carlo estimators outperform the existing crude estimators of these CCR measures substantially in terms of mean square error (MSE). We also demonstrate that the two widely used sampling methods, the so-called Path Dependent Simulation (PDS) and Direct Jump to Simulation date (DJS), are not equivalent in that they lead to Monte Carlo CCR estimators which are drastically different in terms of their MSE.

1 Introduction

Counterparty credit risk (CCR) is the risk that a party to an OTC derivative contract may default prior to the expiration of the contract and fail to make the required contractual payments, (see [3] for the basic CCR definitions). Counterparty credit risk has been widely considered as one of the key drivers of the 2007-09 credit crisis, and it has become one of main focuses of the major global and U.S. regulatory frameworks; Basel III[1]and the Dodd-Frank Act of 2009-10, (see, for instance, [1]). It is well known that pricing and measuring counterparty credit risk is computationally extremely intensive; financial institutions (derivative dealers) invest large

*Department of Economics, University of California, Berkeley, CA 94720-3880, USA, email: samim_ghamami@berkeley.edu

†Business Analytics and Mathematical Sciences Department, IBM Thomas J. Watson Research Center, Yorktown Heights, NY, 10598 USA, email:bozhang@gatech.edu.

[1]Basel III is a global regulatory standard on bank capital adequacy, stress testing and market liquidity risk agreed upon by the members of the Basel Committee on Banking Supervision in 2010-11, and scheduled to be introduced from 2013 until 2018.

amounts of resources to develop and maintain Monte Carlo simulation "engines" to manage their counterparty risk, (see [16], [11], and [3]).

In this paper we develop efficient Monte Carlo frameworks for pricing and measuring counterparty risk. More specifically, we focus on efficient Monte Carlo estimation of the most widely used and regulatory-driven CCR measures, expected positive exposure (EPE), credit value adjustment (CVA), and effective EPE (eEPE), as defined below. Efficiency criteria under consideration are variance, bias, and computing time of the Monte Carlo estimators. Our proposed Monte Carlo estimators of EPE, CVA, and eEPE outperform the existing "crude" estimators of these CCR measures substantially in terms of mean square error (MSE). To the best of our knowledge, this paper is the first to consider efficiency improvement for Monte Carlo CCR estimation.

Counterparty credit exposure [3], denoted by V, of a financial institution against its counterparty, is the larger of zero and the market value of the portfolio of OTC derivative contracts the financial institution holds with its counterparty. To effectively introduce our efficient Monte Carlo procedures we consider credit exposures in the absence of the commonly used risk mitigants, i.e., collateral and netting agreements. This simple setting facilitates the effective communication of our main results.

EPE is a widely used counterparty credit risk measure for regulatory and economic capital calculations, (see Chapters 2 and 11 of [11]). It is defined as follows,

$$\mathrm{EPE} \equiv \int_0^T E[V_t]dt, \tag{1}$$

where $E[V_t]$ is the expected value of the (credit) exposure at time $t \geq 0$, and $T > 0$ denotes the time to maturity of the longest transaction in the OTC derivative portfolio.

Effective EPE (eEPE), another widely used regulatory and economic capital-related counterparty risk measure [11] is defined as follows in the CCR literature:

$$\mathrm{eEPE}_{dst} \equiv \sum_{i=1}^{n} \max_{1 \leq j \leq i} E[V_j]\Delta_i. \tag{2}$$

This definition is based on a discrete time grid, $0 \equiv t_0 < t_1 < ... < t_n \equiv T$ with $\Delta_i = t_i - t_{i-1}$, $i = 1, ..., n$. We prefer and propose the following continuous version of eEPE:

$$\mathrm{eEPE} \equiv \int_0^T \max_{0 \leq u \leq t} E[V_u]dt, \tag{3}$$

which is consistent with the definition of EPE and has the advantage of not requiring an a priori specification of a discrete time grid. Our results in Section 5 apply to eEPE as well as eEPE_{dst}.

eEPE is the "conservative" version of EPE that accounts for *roll-over* risk. Roll-over risk refers to the following scenario. Expiration of some of the short-term trades in the OTC derivatives portfolio before T would decrease some of the $E[V_i]$ and so EPE. However, it is likely that these short-term trades are replaced by new ones. When these replacements are not captured by the Monte Carlo CCR "engine", EPE is underestimated, (see [16]).

CVA, which is the difference between the risk free portfolio value and the true counterparty default risky portfolio value, (see [15]), has become one of the main focuses of the Basel III; derivative dealers are required to calculate CVA *charges* for each of their counterparties on a frequent basis.

Let τ, a positive random variable, denote the default time of the counterparty. It can be shown that CVA, the price of the counterparty credit risk, is equal to the risk neutral expected discounted loss, i.e.,

$$\text{CVA} \equiv E[(1-R)D_\tau V_\tau \mathbf{1}\{\tau \leq T\}], \tag{4}$$

where $1\{A\}$ is the indicator of the event A, $D_t = B_0/B_t$ is the stochastic discount factor at time t, B_t is the value of the money market account at time t, and R is the financial institution's recovery rate, (see, for instance, Chapter 7 of [11] for a derivation of this formula). Hereafter we suppress the dependence of the CVA on the recovery rate, R. When V and τ are independent, we refer to CVA as independent CVA. Let F denote the cumulative distribution function of τ. Independent CVA can be written as follows,

$$\text{CVA}_I \equiv E[D_\tau V_\tau \mathbf{1}\{\tau \leq T\}] = \int_0^T E[D_t V_t] dF_t, \tag{5}$$

where the last equality follows from conditioning on τ, the independence of V and τ, and the independence of D and τ. We focus on efficient Monte Carlo estimation of independent CVA in this paper. [2]

EPE, effective EPE, and independent CVA are estimated based on the Monte Carlo estimation of expected exposures, $E[V_t]$, and expected discounted exposures, $E[D_t V_t]$. Section 2 summarizes the common features of the Monte Carlo CCR framework widely used by financial institutions and introduces the notion of *Marginal Matching.* Marginal matching enables one to differentiate the two widely used CCR sampling methods, *Path Dependent Simulation* (PDS) and *Direct Jump to Simulation* date (DJS). These two terms were first introduced by Pykhtin and Zhu in 2006 [16]. Practitioners choose either of the sampling methods arbitrarily.[3] We illustrate that PDS and DJS-based CCR estimators have drastically different MSE; their computing time also may be not be equal. Section 3 introduces an efficient Monte Carlo framework for estimating EPE. Using our results in Section 3, we introduce efficient Monte Carlo estimators of independent CVA in Section 4. Section 5 considers efficient Monte Carlo estimation of eEPE. Our numerical examples indicate that employing our Monte Carlo CCR schemes leads to substantial MSE reduction.

2 Monte Carlo Counterparty Credit Risk Estimation

Contract level credit exposure at time $t > 0$ is the maximum of the contract's market value and zero, $\max\{C_t, 0\}$, where C_t denotes the time-t value of the derivative contract. Consider a

[2]Wrong (right) way risk are referred to as cases where credit exposures are negatively (positively) correlated with the credit quality of the counterparty, (see [6], [3], and [12]).

[3]One of the authors' former employer is a large investment bank.

financial institution that holds a portfolio of k OTC derivative contracts with its counterparty. Counterparty level credit exposure is

$$V_t = \sum_{i=1}^{k} \max\{C_t^i, 0\}, \tag{6}$$

where C_t^i denotes the time-t value of the i'th derivative contract in the OTC derivatives portfolio. When risk mitigants are employed, V_t is defined differently. For instance, in the presence of netting agreements, credit exposure becomes, (see [15]),

$$V_t = \max\{\sum_{i=1}^{k} C_t^i, 0\}. \tag{7}$$

A typical Monte Carlo counterparty risk engine of a derivatives dealer estimates various types of CCR measures based on sampling from the credit exposure process on a time grid, $0 < t_1 < ... < t_n = T$, where T denotes the maturity of the longest transaction in a portfolio of OTC derivatives and $t_1, ..., t_n$ are sometimes referred to as *valuation points*. Set $V_i \equiv V_{t_i}$.

Some of the CCR measures are *static* in the sense that they are defined based on a given fixed time point. Expected exposure (EE) at time t_i, is simply $E[V_i]$. Also, VaR type of measures for a given valuation point t_i is referred to as *potential future exposure*. Derivatives dealers use Monte Carlo simulation to estimate EE and PFE for all the given valuation points $t_1, ..., t_n$ on a frequent basis, (see [11] and [15] for more details on CCR measures). Note that CCR measures considered in this paper, EPE, CVA, and eEPE, are *dynamic* in the sense that they depend on the time evolution of the credit exposure process.

In what follows we first summarize the simulation of the credit exposure process. Then, we introduce the notion of Marginal Matching in sampling from the time evolution of the credit exposure process.

2.1 Simulating the Credit Exposure Process

We assume that credit exposure is a stochastic process $\{V_t\ ; t \geq 0\}$ defined on a given filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0 \leq t \leq \infty}, P)$. Given (6) and (7), V_t can be viewed as a function of the stochastic processes that drive the values of the derivative contracts, $C_t^1, ..., C_t^k$. In risk management, these underlying stochastic processes are usually referred to as risk factors, e.g., interest rates, commodity prices, and equity prices. To generate a Monte Carlo realization of V_t, for a fixed $t > 0$, first, the underlying risk factors should be sampled from up to time $t > 0$. Next, given the Monte Carlo realization of the risk factors up to time $t > 0$, the derivative contracts C_t^i should be valued. This two-step procedure generates a single Monte Carlo realization of V_t. It is a risk management common practice to use the physical probability measure in the first step and risk-neutral measure in the second. This applies to Monte Carlo estimation of EPE and eEPE. However, since CVA is usually viewed as the market price of counterparty credit risk, risk-neutral measure is usually used in both steps. Depending on the complexity of the payoff function of the derivative contracts, the valuation step could take straightforward Black-Scholes-type analytical calculations, or it could demand approximations that depending on the desired level of accuracy might be computationally intensive. These approximations could also involve Monte Carlo simulation: Nested Monte Carlo refers to the use of a second layer of Monte Carlo simulation in the valuation step of the above procedure, (see [10]), and regression-based Monte Carlo (see [2]) uses ideas from regression-based Monte Carlo American option pricing, (see Chapter 8 of [7]).

2.2 Marginal Matching

Let $X = (X_1, ..., X_n)$ denote a random vector with distribution function F_X. Let $\omega_X \equiv (E[h_1(X_1)], ..., E[h_n(X_n)])$ for some functions $h_1, ..., h_n$. And let $\theta_X \equiv g(\omega_X)$ for a function g that maps ω_X from R^n to R. Two simple examples of θ_X are as follows,

$$\sum_{i=1}^{n} E[h(X_i)] \quad \text{and} \quad \max\{E[h(X_1)], ..., E[h(X_n)]\},$$

that is θ_X is defined based on the marginal distribution of (functions of) $X_1, ..., X_n$. Let $Y = (Y_1, ..., Y_n)$ denote another random vector with distribution function F_Y such that,

$$X \neq^d Y \ , \quad X_i =^d Y_i \text{ for all } i = 1, ..., n, \tag{8}$$

where $=^d$ denotes "being equal in distribution". Simply note that since the marginal distributions of X and Y *match*, $\theta_X = \theta_Y$. Now, suppose that θ_X is to be estimated with Monte Carlo simulation. Given (6), samples can be drawn from F_X or F_Y. Let $\hat{\theta}_{X,m}$ and $\hat{\theta}_{Y,m}$ denote Monte Carlo estimators of θ_X based on m simulation runs when samples are drawn from F_X and F_Y, respectively. Obviously,

$$\hat{\theta}_{X,m} \neq^d \hat{\theta}_{Y,m},$$

and so between $\hat{\theta}_{X,m}$ and $\hat{\theta}_{Y,m}$, i.e., when deciding on whether to sample from F_X or F_Y, the estimator with lower mean square error (MSE) should be chosen.

Example: Finite-Dimensional Distributions of Brownian Motion Let $\{X_t \ ; t \geq 0\}$ denote a Brownian motion with drift μ and volatility parameter σ. Consider the random vector $X = (X_1, ..., X_n) \equiv (X_{t_1}, ..., X_{t_n})$ on the time grid, $0 < t_1 < t_2 < ... < t_n$. That is, following the basic definition of a Brownian motion, X is a multivariate normal random vector with $E[X_{t_i}] = \mu t_i$ and $\text{Var}(X_{t_i}) = \sigma^2 t_i$, and $\text{cov}(X_{t_i}, X_{t_j}) = t_i > 0$ for $t_i < t_j$. Now, let $Y = (Y_1, ..., Y_n)$ denote a multivariate normal random vector whose marginal distributions match that of X but with $\text{cov}(Y_i, Y_j) = 0$, i.e., components of Y are independent.

Stochastic Models of the Risk Factors Let $\{R_t\ ; t \geq 0\}$, representing the dynamics of a risk factor, denote a stochastic process defined on a given filtered probability space, $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0\leq t\leq\infty}, P)$. In this paper we assume that $\{R_t\ ; t \geq 0\}$ is a Gauss-Markov process (see Chapter 5 of [13]) or a Geometric Brownian motion (GBM). Many of the widely used continuous time stochastic processes in finance and economics are in this class. Note that Gauss-Markov processes and GBM have this property that $\mathrm{cov}(R_s, R_t) > 0$ for any $0 < s < t$. Consider the finite dimensional distribution of R on a time grid, $t_1, ..., t_n$ and set $R_i \equiv R_{t_i}$. Suppose that $R = (R_1, ..., R_n)$ can be sampled from *exactly* in the sense that the distribution of the simulated R is precisely that of the R process at times $t_1, ..., t_n$; examples are Brownian motion, Ornstein-Uhlenbeck processes, and GBM, whose simulations involve generating positively correlated normal random variables. Let $\tilde{R} = (\tilde{R}_1, ..., \tilde{R}_n)$ denote a random vector for which $R \neq^d \tilde{R}$ but $R_i =^d \tilde{R}_i$ for all $i = 1, ..., n$ and $\mathrm{cov}(\tilde{R}_i, \tilde{R}_j) = 0$ for all $i \neq j$. That is, simulation of $\tilde{R}_1, ..., \tilde{R}_n$ can be done by generating n uncorrelated normal random variables.

PDS Sampling versus DJS Sampling In the CCR literature when counterparty risk measures are estimated based on sampling from the finite-dimensional distributions of the underlying risk factors, the sampling is referred to as *Path Dependent Simulation* (PDS sampling). Otherwise, when the notion of marginal matching is used, the sampling is referred to as *Direct Jump to Simulation date* (DJS). For instance, in the Brownian motion example above, sampling from X and Y when estimating θ_X-type estimands are referred to as PDS and DJS sampling, respectively. In Monte Carlo estimation of CCR measures, PDS and DJS sampling have been widely considered equivalent. We have also observed that practitioners often choose either of the sampling methods arbitrarily. One of the contributions of this paper is to differentiate DJS and PDS in terms of the mean square error of the estimators of EPE, eEPE, and CVA.

3 Efficient Monte Carlo Estimation of EPE

In this section we consider efficient Monte Carlo estimation of EPE,

$$\mathrm{EPE} = \int_0^T E[V_t]dt,$$

where V denotes the credit exposure process, and $T > 0$ represents the expiration time of the longest maturity derivative contract in an OTC derivatives portfolio. Consider a time grid, $0 \equiv t_0 < t_1 < ... < t_n \equiv T$, with a fixed n. Set $\Delta_i \equiv t_i - t_{i-1}$ and $V_i \equiv V_{t_i}$, $i = 1, ..., n$. Let $\hat{\theta}_{b,m,n,k}$ denote a class of Monte Carlo estimators of EPE defined as follows,

$$\hat{\theta}_{b,m,n,k} = \sum_{i=1}^{n} \bar{V}_i \Delta_i,$$

where $\bar{V}_i = \sum_{j=1}^{m} V_{ij}/m$ and $V_{i1}, ..., V_{im}$ represent the m simulation samples at valuation point t_i. The subscript b refers to the biased nature of the estimators, and the subscript k could take

p and d, referring to PDS and DJS based simulation of the credit exposure process, respectively. As mentioned in Section 2.1, simulating the credit exposure process involves sampling from the underlying risk factors. Hereafter, PDS and DJS-based simulations of the credit exposure process refer to the cases where the underlying risk factors are sampled from based on their finite dimensional distributions (PDS sampling) and based on the notion of marginal matching (DJS sampling), respectively. Note that,

$$\text{MSE}(\hat{\theta}_{b,m,n,k}) = \text{Var}(\sum_{i=1}^{n} \bar{V}_i \Delta_i) + \left(\sum_{i=1}^{n} E[\bar{V}_i]\Delta_i - \int_0^T E[V_t]dt \right)^2 .$$

We assume that Monte Carlo realizations of V_i are unbiased estimates of $E[V_i]$, $i = 1, ..., n$. This implies that the bias part of the MSE of $\hat{\theta}_{b,m,n,k}$ is not affected by the choice of the sampling method (PDS or DJS). In Section 3.1, we assume that n, the number of "valuation points" is fixed, and we compare the efficiency of $\hat{\theta}_{b,m,n,p}$ and $\hat{\theta}_{b,m,n,d}$ in terms of variance and computing time both for *path independent* and *path dependent* derivatives. Next, we introduce our efficient biased, yet consistent Monte Carlo estimators of EPE. In Section 3.3. we introduce efficient unbiased estimators of EPE. Numerical examples in Section 3.4 indicate that our proposed estimators substantivally outperform the "crude" estimators of EPE in terms of the mean square error.

3.1 Comparing PDS and DJS-based Monte Carlo Estimation of EPE

Suppose that the credit exposure process, V, defined on a given filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0 \le t \le \infty}, P)$, is driven by a single risk factor, denoted by S, which is a Gauss-Markov process or a GBM. Let $(\mathcal{F}_t)_{0 \le t \le \infty}$ denote the filtration generated by $\{S_t \ ; t \ge 0\}$. Consider the simple setting where V denotes the contract level exposure and a financial institution takes a long position in a maturity-T derivative contract with its counterparty. Let Π_T denote the payoff function of the derivative contract. It is well known from martingale pricing that

$$V_t = C_t = n_t E\left[\frac{\Pi_T}{n_T} | \mathcal{F}_t\right], \tag{9}$$

where n is a numeraire (stochastic discount factor). We would like to compare the efficiency of $\hat{\theta}_{b,m,n,p}$ and $\hat{\theta}_{b,m,n,d}$ in terms of variance and computing time for *path independent* and *path dependent* derivatives. Without any loss of generality, consider the well-known setting where $\{S_t \ ; t \ge 0\}$ is a GBM, $S_t = S_0 e^{X_t}$, and $\{X_t \ ; t \ge 0\}$ is a Brownian motion with drift μ and volatility σ. Consider a given payoff function with fixed maturity $T > 0$. In the path independent case, $V_t = C_t = n_t E\left[\frac{\Pi_T}{n_T} | S_t\right] \equiv f(S_t)$. That is, credit exposure is considered as a function of the risk factor.[4] Consider the time grid, $0 \equiv t_0 < t_1 < ... < t_n \equiv T$ and set $V_i \equiv V_{t_i}$. In the path

[4]Consider, for instance, the payoff function $\Pi_T = (S_T - K)^+$ of a maturity-T GBM-driven vanilla call option with strike K. Assuming zero short rate, $C_t = E[(S_T - K)^+ | S_t] = E[(S_t S_{T-t} - K)^+ | S_t]$. Note that the function f in $f(S_t) \equiv E[(S_t S_{T-t} - K)^+ | S_t]$, which is well-defined for all values of $t \ge 0$ given the payoff function Π_T with a fix maturity T, is in fact a function of t and S_t. In Section 3, for notational simplicity, we suppress the dependence of f on t in the definition $C_t = n_t E[\frac{\Pi_T}{n_T} | S_t] \equiv f(S_t)$.

dependent case we assume that $V_i = g(S_1, ..., S_i)$, where g is a function from R^i to R. Hereafter, for notational simplicity, we suppress the dependence of V on the stochastic discount factor.

Path Independent Case Set $\theta \equiv \sum_{i=1}^{n} E[V_i]\Delta_i$. Recall that,

$$\hat{\theta}_{b,m,n,k} = \sum_{i=1}^{n} \bar{V}_i \Delta_i,$$

where $\bar{V}_i$ is the m-simulation-run average of $V_{i1}, ..., V_{im}$. With $V_i = f(S_i)$ and $S_i = S_0 e^{X_i}$, Monte Carlo estimation of θ requires sampling from the multivariate normal random vector, $X = (X_1, ..., X_n)$. This is the so-called PDS sampling method. An alternative sampling method, using the notion of marginal matching, is to sample from the multivariate normal random vector, $Y = (Y_1, ..., Y_n)$, whose components are uncorrelated but marginal distributions match those of X. This is the so-called DJS method. To be more specific, in DJS sampling, S_i is generated from time zero. That is, generate Y_i, a normal random variable with mean μt_i and variance $\sigma^2 t_i$, and set $S_i = S_0 e^{Y_i}$. In PDS sampling, V_i's are sampled based on generating the sample path of the GBM sequentially at $i = 1, ..., n$. That is, to generate a realization of V_i, S_i is generated given the previously generated value of S_{i-1}. [5] Note that since for any given $t > 0$, V_t is a function of $S_t = S_0 e^{X_t}$, DJS-based simulation of the exposure process implies that $\text{cov}(V_i, V_j) = 0$ for any $i \neq j$, $i, j = 1, ..., n$.

We now show that PDS-based simulation of the exposure process implies that $\text{cov}(V_i, V_j) > 0$ for any $i \neq j$. First consider the case where V_t is the time-t value of a path independent maturity-T derivative contract with payoff function Π_T, which is driven by a single risk factor denoted by S. That is, $V_t = C_t = E[\Pi_T | S_t] \equiv f(S_t)$, for a function f. For any $0 < u < t$ we have

$$\text{cov}(V_u, V_t) = \text{cov}(f(S_u), f(S_t)) = E\left[\text{cov}(f(S_u), f(S_t)|S_u)\right] + \text{cov}\left(f(S_u), E[f(S_t)|S_u]\right),$$

where the last equality follows from the conditional covariance formula (see Chapter 3 of [17]). It is easy to check that the first term on the right hand side above is zero. Consider the second term and note that

$$E[f(S_t)|S_u] = E[E[\Pi_T|S_t]|S_u] = E[\Pi_T|S_u] = f(S_u),$$

and so we conclude that for any $0 < u < t$,

$$\text{cov}(V_u, V_t) = \text{Var}(V_u) > 0.$$

[5] More specifically, to sample from S_i generate $\tilde{X}_i$ and set $S_i = S_{i-1} e^{\tilde{X}_i}$, where $\tilde{X}_i$ is a normal random variable with mean $\mu \Delta_i$ and variance $\sigma^2 \Delta_i$.

Now, consider the general case where the credit exposure at time $t > 0$ is the maximum of zero and the time-t value of a derivative contract. That is, $V_t = \max\{C_t, 0\}$, where $C_t = E[\Pi_T|S_t] \equiv f(S_t)$ as defined before.[6] We now argue that for any $0 < u < t$,

$$\mathrm{cov}(V_u, V_t) > 0, \tag{10}$$

when the payoff function is a monotone function of the risk factor S. Again, conditioning on S_u and using conditional covariance formula gives

$$\mathrm{cov}(V_u, V_t) = \mathrm{cov}\left(V_u, E[V_t|S_u]\right) = \mathrm{cov}\left(\max\{f(S_u), 0\}, E[\max\{f(S_t), 0\}|S_u]\right).$$

First consider the first term $\max\{f(S_u), 0\}$ inside the covariance function on the right hand side above. Note that since f is a monotone function, $\max\{f(S_u), 0\} \equiv \tilde{f}(S_u)$ is also a monotone function of S_u. Next, consider the second term $E[\max\{f(S_t), 0\}|S_u]$. Note that when S is a Gauss-Markov process, for any $0 < u < t$ we have $S_t = S_u + S_{t-u}$, where S_u and S_{t-u} are independent random variables. Also, when S is a GBM, for any $0 < u < t$ we have $\log(S_t) = \log(S_u) + \log(S_{t-u})$, where S_u and S_{t-u} are independent random variables. This follows from the independent and stationary increments properties of Gauss-Markov processes[7] and that their finite dimensional distributions are multivariate normal. This implies that $E[\max\{f(S_t), 0\}|S_u]$ is a monotone function of S_u. To see this, consider the case where f is an increasing function. Increasing S_u will increase $S_t = S_u + S_{t-u}$ ($S_t = S_u S_{t-u}$ when S is a GBM); this increases $\max\{f(S_t), 0\}$. So, $E[\max\{f(S_t), 0\}|S_u] \equiv \tilde{h}(S_u)$ also becomes an increasing function of S_u. A similar argument can be used when f is a decreasing function. Consequently, we can write $\mathrm{cov}(V_u, V_t) = \mathrm{cov}(\tilde{f}(S_u), \tilde{h}(S_u))$, where $\tilde{f}$ and $\tilde{h}$ are both either increasing or decreasing functions of S_u. Using Chebyshev's algebraic inequality (see, for instance, Proposition 2.1 in [5]) gives $\mathrm{cov}(V_u, V_t) = \mathrm{cov}(\tilde{f}(S_u), \tilde{h}(S_u)) > 0$.

The monotonicity assumption of the payoff function is satisfied for most of the actively traded OTC derivative contracts; well-known exceptions are Barrier[8] and Lookback options, (see, for instance, [14]). Under this monotonicity assumption which leads to (10), it is not difficult to see that

$$\mathrm{Var}(\hat{\theta}_{b,m,n,d}) \leq \mathrm{Var}(\hat{\theta}_{b,m,n,p}). \tag{11}$$

The above inequality holds since

$$\mathrm{Var}(\hat{\theta}_{b,m,n,d}) = \sum_{i=1}^{n} \frac{\mathrm{Var}(V_i)\Delta_i^2}{m} \leq \sum_{i=1}^{n} \frac{\mathrm{Var}(V_i)\Delta_i^2}{m} + \frac{2}{m^2}\sum_{i<j} \mathrm{cov}(V_i, V_j)\Delta_i\Delta_j = \mathrm{Var}(\hat{\theta}_{b,m,n,p}). \tag{12}$$

[6]For instance, consider the case where C_t represents the time-t value of an interest rate swap. Then, C_t can be negative for some $t > 0$.

[7]Note that when S is a GBM, logarithm of S is a Brownian motion, which is a Gauss-Markov process.

[8]More specifically, the payoff function of up-and-in and down-and-out European barrier call options are monotone functions of the underlying security prices. This monotonicity assumption does not hold for up-and-out and down-and-in European barrier call options, (see Chapter 6 of [14] and the references there).

Path Dependent Case We now consider the *path dependent* case. For instance, suppose that V_t is time t value of a maturity-T arithmetic Asian option, where the payoff at the time T is a function of $S_1, ...S_n$. That is, $V_i = g(S_1, ..., S_i)$, where g is a function from R^i to R. The DJS sampling method is to make $V_i = g(S_1, ..., S_i)$ and $V_j = g(S_1, ..., S_j)$, $i < j$, uncorrelated random variables. That is, sample from $S_1, ..., S_i$ to generate a single realization of V_i. To generate V_j, start again from time zero, and sample from $S_1, ..., S_i, ...S_j$. Under this DJS-type sampling method, V_i and V_j become uncorrelated, $\text{cov}(V_i, V_j) = 0$. In the PDS-type sampling, given the Monte Carlo realization of V_i, to generate V_j, one uses the previously generated $S_1, ..., S_i$ and only samples from $S_{i+1}, ..., S_j$. In this case V_i and V_j are dependent. Using conditional covariance formula and arguments similar to the ones used in the path independent case, it can be shown that $\text{cov}(V_i, V_j) > 0$. More specifically, it can be shown that $\text{cov}(V_i, V_j) > 0$ holds without any restriction on the payoff function and the underlying risk factors when V_t coincides with the time-t value of the derivative contract C_t. In the more general case where credit exposure is the maximum of zero and C_t, under the monotonicity assumption of the payoff function and risk factors being Gauss-Markov or GBM, it can be shown that $\text{cov}(V_i, V_j) > 0$.

To compare the efficiency of the DJS and PDS-based estimators of θ in the path dependent case, computing time is also to be considered in parallel with variance of the estimators. Suppose that the computational time to calculate $\hat{\theta}_{b,m,n,k}$ is proportional to the number of random variables that are to be generated. Let $\text{ct}(\hat{\theta}_{b,m,n,k})$ denote the computational effort associated with $\hat{\theta}_{b,m,n,k}$. Note that,

$$\frac{\text{ct}(\hat{\theta}_{b,1,n,d})}{\text{ct}(\hat{\theta}_{b,1,n,p})} \approx n \quad \text{and} \quad \frac{\text{Var}(\hat{\theta}_{b,1,n,p})}{\text{Var}(\hat{\theta}_{b,1,n,d})} \approx n. \tag{13}$$

To see why (13) holds note that to calculate $\hat{\theta}_{b,1,n,d}$, $\frac{n(n+1)}{2}$ random variables are to be generated while $\hat{\theta}_{b,1,n,p}$ requires generating n random variables, (assuming that the calculation of $E[\Pi^A|F_i]$ does not require generating additional random variables). Also, note that as can be seen from (12), variance of the PDS-based estimator is of order n^2 because of the covariance terms while the DJS-based estimator has a variance of order n. Now, one should select the estimator with the lower

variance per replication $\times$ *expected computing time*,

(see [9] for the formal formulation of this useful criterion in comparing alternative Monte Carlo estimators) . So, we conclude that for the path dependent case Monte Carlo estimators of EPE, i.e., $\hat{\theta}_{b,m,n,d}$ and $\hat{\theta}_{b,m,n,p}$, have a similar performance for fixed and sufficiently large n.

3.2 Efficient Monte Carlo EPE Estimation: Biased Estimators

In this subsection, we suppress the subscript b in $\hat{\theta}_{b,m,n,k}$ and instead write $\hat{\theta}_{m,n,k}$ for notational simplicity. We would like to find the number of valuation points, n, and the number of simulation runs at each valuation point, m, to minimize $\text{MSE}(\hat{\theta}_{m,n,k})$,

$$\text{MSE}(\hat{\theta}_{m,n,k}) = \text{Var}(\hat{\theta}_{m,n,k}) + (E[\hat{\theta}_{m,n,k}] - \text{EPE})^2.$$

given a fixed computational budget, denoted by s, that is proportional to, mn. Also, $k = p$, and d refer to PDS and DJS-based simulation of the credit exposure process on a time grid $0 \equiv t_0 < t_1 < ... < t_n \equiv T$. That is, as shown in the previous section, under PDS sampling and DJS sampling, $\mathrm{cov}(V_i, V_j) > 0$ and $\mathrm{cov}(V_i, V_j) = 0$, respectively, for any $i \neq j$, $i, j = 1, ..., n$.

To formulate and solve this optimization problem, we specify the order of the variance and bias of the Monte Carlo estimator of eEPE, $\hat{\theta}_{m,n,k}$. Note that from the basic results on the endpoint Reimann sum approximation of the integral time-discretization bias is of order $1/n$. We are not concerned with deriving sharp estimates of the orders of variance. In fact, our numerical examples indicate that choosing approximately optimal m and n using even very rough approximates for the orders of variance and bias leads to substantial MSE reduction compared to industry practice.

Suppose that the time grid is equidistant, i.e., $\Delta_i \equiv \Delta = \frac{T}{n}$. We assume that $E[V_t^2] < \infty$ for all $t \in [0, T]$. It is not difficult to show that,

$$\mathrm{Var}(\hat{\theta}_{m,n,d}) = O(\frac{1}{mn}). \tag{14}$$

To see this,[9] consider $M > 0$ such that $E[V_t^2] \leq M$ for $t \in (0, T]$. Note that,

$$\mathrm{Var}(\hat{\theta}_{m,n,d}) = \Delta^2 \sum_{i=1}^{n} \frac{\mathrm{Var}(V_i)}{m} \leq (\frac{T}{n})^2 \sum_{i=1}^{n} \frac{E(V_i^2)}{m} \leq \frac{MT^2}{mn}.$$

Now, consider the variance of the PDS-based estimator, $\hat{\theta}_{m,n,p}$,

$$\mathrm{Var}(\hat{\theta}_{m,n,p}) = \Delta^2 \sum_{i=1}^{n} \frac{\mathrm{Var}(V_i)}{m} + \Delta^2 \frac{2}{m^2} \sum_{i=1}^{n} \sum_{j<i} \mathrm{cov}(V_i, V_j).$$

As shown before, the first term above is $O(\frac{1}{mn})$. Also, under PDS sampling, the credit exposure process is simulated according to its finite dimensional distributions for which the covariance terms are positive. So, the second term is $O(\frac{1}{m^2})$. This gives,

$$\mathrm{Var}(\hat{\theta}_{m,n,p}) = O(\frac{1}{mn} + \frac{1}{m^2}). \tag{15}$$

Note that if a function belongs to $O(\frac{1}{mn} + \frac{1}{m^2})$, it also belongs to $O(\frac{1}{m})$. Therefore, we have

$$\mathrm{Var}(\hat{\theta}_{m,n,p}) = O(\frac{1}{m}). \tag{16}$$

This second and more familiar approximate order of variance can be viewed as follows.[10] For $m = 1$, $\hat{\theta}_{1,n,p}$ converges to a positive constant as $n \to \infty$. Then, we have $\mathrm{Var}(\hat{\theta}_{m,n,p}) = O(\frac{1}{m})$ converging to zero as $m \to \infty$.

[9] The Landau symbol, O, in $f(x, y) = O(g(x, y))$ means that $f(x, y)/g(x, y)$ stays bounded in some limit, say $x, y \to 0$ or $x, y \to \infty$.

[10] See page 365 of [7].

PDS-Based Biased Efficient Estimator of EPE We choose the number of valuation points, n, and number of simulation runs at each valuation point, m, to minimize the mean square error of the PDS-based estimator, $\hat{\theta}_{m,n,p}$, under a fixed computational budget proportional to mn. Approximating the variance of $\hat{\theta}_{m,n,p}$ using (15) leads to the following optimization problems,

$$\min_{m,n} \left(\frac{c_{p,1}}{mn} + \frac{c_{p,2}}{m^2} + \frac{c_2}{n^2} \right) \quad \text{subject to} \quad s = c_3 mn, \tag{17}$$

for some constants, $c_{p,1}, c_{p,2}, c_2$, and c_3. MSE of $\hat{\theta}_{m,n,p}$ is minimized at,

$$m = cs^{\frac{1}{2}} \ \text{ and } \ n = \tilde{c}s^{\frac{1}{2}}, \tag{18}$$

for constants c and $\tilde{c}$. If we approximate the variance of $\hat{\theta}_{m,n,p}$ using (16) in the MSE minimization problem, the solution becomes,

$$m = cs^{\frac{2}{3}} \ \text{ and } \ n = \tilde{c}s^{\frac{1}{3}}. \tag{19}$$

Our numerical examples indicate that (18) and (19) lead to very similar simulation performance in practical settings.

DJS-Based Biased Efficient Estimator of EPE Let c_d denote a constant. Given (14), we approximate $\mathrm{Var}(\hat{\theta}_{m,n,d})$ with $\frac{c_d}{mn}$ in the MSE minimization problem for the DJS-based estimator,

$$\min_{m,n} \left(\frac{c_d}{mn} + \frac{c_2}{n^2} \right) \quad \text{subject to} \quad s = c_3 mn,$$

to which the trivial optimal solution is $m = 1$ and $n = \hat{c}s$ fo some constant $\hat{c}$. We note that estimating the various constant parameters appearing in all the above mentioned MSE minimization problems is not possible in practice. In our numerical examples we simply set all these constant parameters equal to 1.

3.3 Efficient Monte Carlo EPE Estimation: Unbiased Estimators

In this section we derive unbiased estimators of EPE. Specifically, we eliminate the time discritization bias at the expense of introducing additional randomness. To control the variance that would be increased as the result of this new source of randomness, we use stratified sampling. Let τ denote a $[0, T]$ Uniform random variable that is independent of the credit exposure, V. We have,

$$\mathrm{EPE} = TE[V_\tau], \tag{20}$$

which simply follows from conditioning on τ, i.e., using $E[V_\tau] = E[E[V_\tau|\tau]]$, independence of V and τ, and noting that $f(t) = \frac{1}{T}$, $t \in [0, T]$, is the probability density function of τ. Now, consider the following identity,

$$\mathrm{EPE} = TE[V_\tau] = T\sum_{i=1}^{n} E[V_\tau|\tau \in A_i]p_i = \sum_{i=1}^{n} E[V_\tau|\tau \in A_i]\Delta_i, \tag{21}$$

where $A_i = [0, t_i)$, $p_i \equiv P(\tau \in A_i) = \frac{\Delta_i}{T}$, on the time grid, $0 \equiv t_0 < t_1 < ... < t_n \equiv T$, and $\Delta_i = t_i - t_{i-1}$. Our proposed unbiased estimators of EPE use the identity (21) by estimating the conditional expectations, $E[V_\tau|\tau \in A_i]$,

$$\hat{\theta}_{u,m,n,k} = \sum_{i=1}^{n} \bar{V}_{\tau_i}\Delta_i, \tag{22}$$

where $\tau_i \equiv \tau|\tau \in A_i$, $\bar{V}_{\tau_i} = \sum_{j=1}^{m} V_{\tau_{ij}}/m$, and $\tau_{i1}, ..., \tau_{im}$ are i.i.d. copies of τ_i. That is, to draw a single realization of V_{τ_i}, we first sample from τ conditional on $\tau \in A_i$. Note that τ_i is a $[t_{i-1}, t_i]$ Uniform random variable. Next, given this realization of τ_i, we generate V_{τ_i}. The subscript $k = p$ and d refer to PDS and DJS sampling, respectively.[11]That is, PDS-based simulation in calculating $\hat{\theta}_{u,m,n,p}$ implies that $\text{cov}(V_{\tau_i}, V_{\tau_j}) > 0$ for $i \neq j$, $i, j = 1, ..., n$, and DJS-based simulation in calculating $\hat{\theta}_{u,m,n,d}$ implies that $\text{cov}(V_{\tau_i}, V_{\tau_j}) = 0$ for $i \neq j$. This immediately implies $\text{Var}(\hat{\theta}_{u,m,n,d}) \leq \text{Var}(\hat{\theta}_{u,m,n,p})$. Consider a more general setting that allows different numbers of simulation runs for each stratum. That is, let m_i denote the number of runs used to estimate $E[V_\tau|\tau \in A_i]$ and $N = m_1 + ... + m_n$ denote the total number simulation runs. Note that our setting with equidistant strata and $m_i \equiv m$, for $i = 1, ..., n$ coincides with proportional stratified sampling which uses $m_i = Np_i$, (see [18] for results on proportional stratification). This is because τ is a $[0, T]$ Uniform random variable. In this paper we do not address further possible improvements of our unbiased stratified sampling-based estimators of EPE by attempting to find optimal $m_1, ..., m_n$ and n under fixed computational budgets. Our numerical examples indicate that using our unbiased stratified sampling-based estimators by setting $m_i \equiv m$ and choosing m and n as specified in subsection 3.2 leads to substantial MSE reduction when compared to crude biased Monte Carlo estimators of EPE.

Proposition 1 below shows that $\hat{\theta}_{u,m,n,d}$ and the biased DJS-based estimator of EPE, $\hat{\theta}_{b,m,n,d}$, are asymptotically equivalent in terms of MSE. This equivalence is further confirmed by our numerical experiments (see the next subsection) in practical settings with fixed and finite computational budgets proportional to mn. In addition, our numerical examples presented in the next subsection show that the unbiased PDS-based estimator of EPE, $\hat{\theta}_{u,m,n,p}$, outperforms the efficient biased PDS-estimator, $\hat{\theta}_{b,m,n,p}$, introduced in the previous section.

Proposition 1. *Consider the credit exposure process, $\{V_t \; ; t \geq 0\}$, defined on a given filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0 \leq t \leq \infty}, P)$. Suppose that biased and unbiased Monte Carlo estimators of EPE calculated under DJS-sampling,*

$$\hat{\theta}_{b,m,n,d} = \sum_{i=1}^{n} \bar{V}_i \Delta_i, \quad \textit{and} \quad \hat{\theta}_{u,m,n,d} = \sum_{i=1}^{n} \bar{V}_{\tau_i}\Delta_i. \tag{23}$$

are defined on an equi-distant time grid, $0 \equiv t_0 < t_1 < ... < t_n \equiv T$, where $\Delta_i \equiv t_i - t_{i-1} = T/n \equiv \Delta$, $\tau_i \equiv \tau|\tau \in A_i$ and $A_i = [t_{i-1}, t_i)$. Let $\bar{V}_i$ and $\bar{V}_{\tau_i}$ denote the averages of m Monte

[11] Recall that the biased estimators of EPE, $\hat{\theta}_{m,n,k}$, $k = p, d$, are based on Right Reiman sum approximation of the integral of the expected exposures in the EPE formula. Our proposed unbiased estimators $\hat{\theta}_{u,m,n,k}$, $k = p, d$, can simply be viewed as a Reiman sum approximation of the EPE where each expected exposure is evaluated at a randomly selected point within each subinterval.

Carlo realizations of V_i*, and* V_{τ_i}*, respectively. That is, the total number of simulation runs is* $N = mn$*. We assume that* $E[V_i^2] < \infty$*, for all* $i = 1, ..., n$*. Asymptotic performance of* $\hat{\theta}_{b,m,n,d}$ *and* $\hat{\theta}_{u,m,n,d}$ *is equivalent in the following sense,*

$$\lim_{n\to\infty} nMSE(\hat{\theta}_{b,m,n,d}) = n\,Var(\hat{\theta}_{u,m,n,d}) = c\int_0^T Var(V_t)dt, \tag{24}$$

where c *is a constant.*

3.4 Numerical Examples

In this section we use simple numerical examples to illustrate the efficiency of our proposed Monte Carlo estimators of EPE. We consider contract level exposure in a simple setting where V_t denotes the value of a geometric Brownian motion driven forward contract at time $t > 0$. That is, we assume that the underlying security price process following a geometric Brownian motion $\{S_t\ ; t \geq 0\}$, where $S_t = S_0 e^{X_t}$ with $\{X_t;\ t \geq 0\}$ being a Brownian motion with drift μ, and volatility σ. In this case, at any time t the value of a forward contract coincides with the security price at that time S_t and thus EPE can be computed analytically. This enables us to calculate the MSE exactly. We consider six different Monte Carlo estimators of EPE in our numerical examples.

Let $\hat{\theta}_{c,p}$ and $\hat{\theta}_{c,d}$ denote the "crude" and biased Monte Carlo estimators of EPE under PDS and DJS sampling, respectively. That is,

$$\hat{\theta}_{c,k} = \sum_{i=1}^{n} \bar{V}_i \Delta_i, \tag{25}$$

where $\Delta_i = t_i - t_{i-1}$, $0 \equiv t_0 < t_1 < ... < t_n \equiv T$, $k = p, d$, and $\bar{V}_i$ is the m-simulation-run average of V_i. We shall shortly specify the choice of the valuation points.

Let $\hat{\theta}_{e,b,p}$ and $\hat{\theta}_{e,b,d}$ denote the efficient and biased Monte Carlo estimators of EPE under PDS and DJS sampling, respectively. In particular, their statistical efficiency is a result of solving the MSE minimization problems in Section 3.2 to derive the (approximately) optimal number of points on the time grid, n, and simulation runs at each of these time points, m, given a fixed computational budget proportional to mn.

Let $\hat{\theta}_{u,p}$ and $\hat{\theta}_{u,d}$ denote the unbiased stratified sampling-based Monte Carlo estimators of EPE under PDS and DJS sampling, respectively. That is,

$$\hat{\theta}_{u,k} = \sum_{i=1}^{n} \bar{V}_{\tau_i} \Delta_i, \tag{26}$$

where $\bar{V}_{\tau_i} = \sum_{j=1}^{m_i} V_{\tau_{ij}}/m_i$ with $\tau_i \equiv \tau | \tau \in A_i$, $A_i = [t_{i-1}, t_i]$, and $k = p, d$.

We set $T = 1$. The crude estimators of EPE are calculated based on 12 valuation points, $n = 12$, at 1, 2, 3, 4, 8, 12, 18, 21, 24, 36, 49 weeks and 1 year. We note that one year, $T = 1$, with the number of valuation points fixed at 12, is a setting widely used by financial institutions.[12] The time grid used to calculate our efficient estimators of EPE is equidistant, i.e., $\Delta_i \equiv \Delta = T/n$. Computational budget, s, is fixed at $12,000$ and $120,000$, respectively. To calculate $\hat{\theta}_{e,b,p}$ under these fixed computational budgets, the solution, (19) with both c and $\tilde{c}$ set to 1, to the MSE minimization problem of Section 3.2 is used.[13] This gives, $n = 23$ and $m = 524$ for $s = 12,000$, and $n = 50$, and $m = 2433$ for $s = 120,000$. Similarly, to calculate $\hat{\theta}_{e,b,d}$, we use the solution to the MSE minimization problem, (3.2). That is, we set $n = 12,000$ and $m = 1$ for $s = 12,000$, and $n = 120,000$ and $m = 1$ for $s = 120,000$. In calculating the stratified sampling estimators of EPE, $\hat{\theta}_{u,p}$ and $\hat{\theta}_{u,d}$, we do not address the problem of deriving the optimal values of n, and $m_1, ..., m_n$. Instead, we simply use the setting of $\hat{\theta}_{e,b,p}$ and $\hat{\theta}_{e,b,d}$, respectively. That is, to calculate $\hat{\theta}_{u,p}$, we set $n = 23,\ m = 524$, and $n = 50,\ m = 2433$, under $s_1 = 12,000$ and $s_2 = 120,000$, respectively. And to calculate $\hat{\theta}_{u,d}$, we set $n = 12,000,\ m = 1$, and $n = 120,000,\ m = 1$, under $s_1 = 12,000$ and $s_2 = 120,000$, respectively.

Tables 1 to 4 illustrate that our proposed estimators of EPE lead to substantial MSE reduction when compared to the "crude" Monte Carlo estimators. Comparing the MSE of the PDS-based estimators, $\hat{\theta}_{c,p}$, $\hat{\theta}_{e,b,p}$, and $\hat{\theta}_{u,p}$, we find that our proposed stratified sampling-based estimator of EPE leads to an MSE reduction by a factor of up to 100; this unbiased estimator also dominates the efficient biased estimator of EPE, in some cases quite substantially (see Tables 3 and 4). Comparing MSE of the DJS-based Monte Carlo estimators of EPE, $\hat{\theta}_{c,d}$, $\hat{\theta}_{e,b,d}$, and $\hat{\theta}_{u,d}$, we observe that the stratified sampling-based estimator of EPE and our efficient biased EPE estimator perform similarly, which suggests that the asymptotic equivalence result in Proposition 1 can hold for even a moderate number of valuation points. Both efficient DJS estimators lead to substantial MSE reduction when compared to the corresponding crude estimator of EPE. Finally, we note that the variance and MSE for the crude estimators do not change much as the computational budget increases from 12,000 to 120,000, whereas those of efficient estimators reduce by up to an order of ten. This contrast yields the simple, yet useful insight that the number of valuation points should vary as the computational budget varies.

	EPE	Variance	MSE	CPU Time
$\hat{\theta}_{c,p}$	34.6559	.047219	.48478	.00380
$\hat{\theta}_{c,d}$	34.6522	.005028	.43768	.00162
$\hat{\theta}_{e,b,p}$	34.1802	.077212	.1117	.00253
$\hat{\theta}_{e,b,d}$	33.9955	.004785	.004786	.00174
$\hat{\theta}_{u,p}$	33.9964	.072068	.072064	.00518
$\hat{\theta}_{u,d}$	33.9956	.004865	.004866	.00335

Table 1: $S_0 = 30, \mu = .2, \sigma = .3,\ s = 12,000$

[12] There is no mathematical basis for this arrangement of valuation points. It is believed that since some trades have "short" expiration times, having more valuation points earlier would increase the accuracy of the estimators of CCR measures.

[13] Our numerical examples indicate that the solutions (19) and (18) derived based on $\mathrm{Var}(\hat{\theta}_{b,m,n,p}) = O(\frac{1}{m})$ and $\mathrm{Var}(\hat{\theta}_{b,m,n,p}) = O(\frac{1}{mn} + \frac{1}{m^2})$, respectively, lead to estimators which perform similarly in terms of their MSE.

	EPE	Variance	MSE	CPU Time
$\hat{\theta}_{c,p}$	34.652	.004791	.4372	.03887
$\hat{\theta}_{c,d}$	34.6521	.000501	.43303	.01564
$\hat{\theta}_{e,b,p}$	34.0798	.016741	.024026	.02299
$\hat{\theta}_{e,b,d}$	33.9948	.000483	.000483	.02409
$\hat{\theta}_{u,p}$	33.9957	.015533	.015533	.04420
$\hat{\theta}_{u,d}$	33.9945	.000486	.000486	.03426

Table 2: $S_0 = 30, \mu = .2, \sigma = .3,\ s = 120,000$

	EPE	Variance	MSE	CPU Time
$\hat{\theta}_{c,p}$	57.7556	.16106	23.5389	.00389
$\hat{\theta}_{c,d}$	57.7598	.01628	23.4351	.00189
$\hat{\theta}_{e,b,p}$	54.1296	.23369	1.6954	.00270
$\hat{\theta}_{e,b,d}$	52.9238	.015853	.015862	.00189
$\hat{\theta}_{u,p}$	52.9226	.217	.21698	.00516
$\hat{\theta}_{u,d}$	52.9198	.015796	.015796	.00390

Table 3: $S_0 = 30, \mu = 1, \sigma = .3,\ s = 12,000$

	EPE	Variance	MSE	CPU Time
$\hat{\theta}_{c,p}$	57.7579	.016112	23.4159	.03891
$\hat{\theta}_{c,d}$	57.7591	.001616	23.4136	.01661
$\hat{\theta}_{e,b,p}$	53.4783	.047841	.35899	.02412
$\hat{\theta}_{e,b,d}$	52.9212	.001563	.001564	.02627
$\hat{\theta}_{u,p}$	52.9189	.045783	.045781	.04657
$\hat{\theta}_{u,d}$	52.9203	.001565	.001565	.03598

Table 4: $S_0 = 30, \mu = 1, \sigma = .3,\ s = 120,000$

4 Efficient Monte Carlo Estimation of Independent CVA

To present our results on efficient Monte Carlo CVA_I estimation, we suppress the dependence of CVA on the stochastic discount factor by assuming zero short rate,

$$\text{CVA}_I = E\left[E[V_\tau \mathbf{1}\{\tau \leq T\}|\tau]\right] = \int_0^T E[V_t]dF_t, \tag{27}$$

where F denotes the cumulative distribution function of τ, which is assumed to be known (market observable) from, for instance, credit default swap spreads of the counterparty, (see, for instance, [12]).

Note that independent CVA can be viewed as the weighted average of the expected exposure with the weights being default probabilities. Therefore, our results from Section 3 on efficient estimation of EPE apply here.

Efficient Biased Estimators of CVA$_I$ We can employ our MSE minimization formulation to first specify the approximately optimal n and m under a fixed computational budget, and then estimate CVA$_I$ with

$$\xi_{b,k} = \sum_{i=1}^{n} \bar{V}_i \Delta F_i, \tag{28}$$

where $k = p, d$ denotes PDS and DJS sampling, respectively, $\bar{V}_i = \sum_{j=1}^{m} V_{ij}/m$ as defined in Section 3, and $\Delta F_i \equiv F(t_i) - F(t_{i-1})$. (We have suppressed the dependence of $\xi_{b,k}$ on m and n, i.e., $\xi_{b,k} \equiv \xi_{b,m,n,k}$.)

Efficient Unbiased Estimators of CVA$_I$ Note that

$$E[V_\tau \mathbf{1}\{\tau \leq T\}] = \sum_{i=1}^{n} E[V_\tau | \tau \in A_i] P(\tau \in A_i), \tag{29}$$

where stratum i is $A_i = [t_{i-1}, t_i)$. Let m_i, $i = 1, 2, ..., n$ denote the number of simulation runs used to estimate $E[V_i]$, where $V_i \equiv V_{t_i}$, $t_0 \equiv 0$, and $t_n = T$. Also, $N = \sum_{i=1}^{n} m_i$ denotes the total number of simulation runs used in estimating CVA$_I$. Using τ as the stratification variable and the identity (29), the stratified sampling estimator of CVA$_I$ is

$$\xi_{u,k} = \sum_{i=1}^{n} \bar{V}_{\tau_i} p_i, \tag{30}$$

where $k = p, d$ denotes PDS and DJS sampling, respectively. Also, $p_i \equiv P(\tau \in A_i) = \Delta F_i$, $\tau_i \equiv \tau | \tau \in A_i$, and $\bar{V}_{\tau_i} = \sum_{j=1}^{m_i} V_{\tau_{ij}}/m_i$. That is, to draw a single realization of V_{τ_i}, we first sample from τ conditional on $\tau \in A_i$; next, given this realization of τ_i, we generate V_{τ_i}. In terms of computing time, $\xi_{b,k}$ requires generating N realizations of V_i and $\xi_{u,k}$ requires N additional samples from the truncated τ based on the strata defined above. Note that since generating V_i is computationally much more intensive than the truncated τ, $\xi_{b,k}$ outperforms $\xi_{u,k}$ merely marginally in terms of the computational time.

Similar to our numerical examples in Section 3.4, we have observed that $\xi_{u,p}$ outperforms $\xi_{b,p}$ in terms of mean square error.[14] In what follows we compare the MSE of the DJS-based biased and unbiased estimators of CVA$_I$, $\xi_{b,d}$ and $\xi_{u,d}$. Lemma 1 below compares the asymptotic performance of $\xi_{b,d}$ and $\xi_{u,d}$. The proof of Lemma 1 is similar to Proposition 1, and so it is omitted.

Lemma 1. *Consider the proposed estimators of CVA$_I$, $\xi_{b,d}$ and $\xi_{u,d}$ as defined in (28) and (30), respectively. Suppose that proportional sampling is used, i.e., $m_i = Np_i$, and $\sum_{i=1}^{n} m_i = N$, $i =$*

[14] Assuming that τ is an exponential random variable, the results of our numerical examples for $\xi_{u,p}$ and $\xi_{b,p}$ are very similar to those in Section 3.4, and so are omitted.

$1, ..., n$. *We assume that* $E[V_i^2] < \infty$, $i = 1, ..., n$. *Note that DJS sampling gives* $cov(V_i, V_j) = 0$ *for all* $i \neq j$ *and* $i, j = 1, ..., n$. *Then the following holds,*

$$\lim_{n\to\infty} n\,Var(\xi_{u,d}) = nMSE(\xi_{b,d}) = c\int_0^T Var(V_t)dF(t) \tag{31}$$

where c is a constant and F is the cumulative distribution function of τ. *That is,* $\xi_{b,d}$ *and* $\xi_{u,d}$ *perform similarly in terms of asymptotic MSE.*

It is useful to also compare the MSE of $\xi_{b,d}$ and $\xi_{u,d}$ in the practical finite-n settings. Note that $\text{MSE}(\xi_{u,d}) = \text{Var}(\xi_{u,d})$ because $\xi_{u,d}$ is unbiased; we have

$$\text{MSE}(\xi_{u,d}) - \text{MSE}(\xi_{b,d}) = \frac{1}{N}\sum_{i=1}^{n}\left(\text{Var}(V_\tau|\tau \in A_i) - \text{Var}(V_i)\right)p_i - \left(\sum_{i=1}^{n} E[V_i]\Delta F_i - \text{CVA}_I\right)^2, \tag{32}$$

where to derive the variance of $\xi_{b,d}$, we note that marginal matching is used such that $\text{cov}(V_i, V_j) = 0$, i.e., using DJS sampling. Also, note that $\text{Var}(\bar{V}_i) = \text{Var}(V_i)/Np_i$. For $\text{Var}(\xi_{u,d})$ we have used the standard results on proportional stratified sampling, (see [18]). For the "finite" n case, (32) implies that depending on the functional form of $\text{Var}(V_i)$, the density of τ, and the numerical scheme to approximate the integral in the independent CVA formula when calculating $\xi_{b,d}$, either $\xi_{b,d}$ or $\xi_{u,d}$ could outperform the other in terms of MSE. However, our various numerical results indicate that in finite-n settings the stratified sampling estimator, $\xi_{u,d}$, usually outperforms $\xi_{b,d}$; in fact, it can lead to substantial MSE reduction. The following example is one instance of such numerical results.

A Numerical Example Suppose that $\{V_t\ ; t \geq 0\}$ is a geometric Brownian motion, $V_t = V_0 e^{X_t}$, where $\{X_t\ ; t \geq 0\}$ is a Brownian motion with drift μ and volatility σ. Also, let τ, counterparty's default time, be an exponential random variable with mean $1/\lambda$. Note that $\text{Var}(\xi_{u,d})$ and $\text{MSE}(\xi_{b,d})$ are easily analytically computable; set $\alpha_1 \equiv \mu + \sigma^/2 - \lambda$, $\alpha_2 \equiv \frac{\lambda}{\alpha_1}$, $\beta_1 \equiv 2\mu + 2\sigma^2 - \lambda$, $\beta_2 \equiv \frac{\lambda}{\beta_1}$. Again, consider the time grid, $0 \equiv t_0 < t_1 < ... < t_n \equiv T$. Then, $\text{CVA}_I = \alpha_2 V_0(\exp(\alpha_1 T) - 1)$, $E[V_\tau|\tau \in A_i] = \frac{\alpha_2 V_0}{p_i}(\exp(\alpha_1 t_i) - \exp(\alpha_1 t_{i-1}))$, and $E[V_\tau^2|\tau \in A_i] = \frac{\beta_2 V_0^2}{p_i}(\exp(\beta_1 t_i) - \exp(\beta_1 t_{i-1}))$.

	$\text{Var}(\xi_{u,d})$	$\text{Var}(\xi_{b,d})$	$\text{Bias}^2(\xi_{b,d})$	$\text{MSE}(\xi_{b,d})/\text{Var}(\xi_{u,d})$
$n = 10,\ N = 1 \times 10^4$	1.9476×10^{-7}	2.1502×10^{-7}	2.4544×10^{-5}	127.1294
$n = 20,\ N = 2 \times 10^4$	9.7377×10^{-8}	1.0243×10^{-7}	2.2763×10^{-5}	234.8168
$n = 50,\ N = 5 \times 10^4$	3.8950×10^{-8}	3.8143×10^{-8}	2.1727×10^{-5}	558.8347
$n = 100,\ N = 1 \times 10^5$	1.9475×10^{-8}	1.9677×10^{-8}	2.1387×10^{-5}	1.0992×10^3

Table 5: $T = 1,\ V_0 = 1,\ \mu = 0,\ \sigma = 0.2,\ \lambda = .1$

Therefore, we suggest not to eliminate all the randomness resulting from τ, as $\xi_{b,d}$ does at the expense of introducing bias. We recommend $\xi_{u,d}$ which eliminates bias but leaves out some controlled randomness from τ. Note that in approximating the integral on the right side of the CVA_I formula (27), $\xi_{b,k}$ uses the Right Riemann sum. In our numerical examples we have also used Left Riemann sum and the Middle sum; in practice where V_t is the maximum of zero and the time t value of a portfolio of possibly thousands of OTC derivative contracts, the functional form of $\text{Var}(V_t)$ is not known and so a time-discretization biased-optimal numerical approximation procedure can not be chosen before estimating CVA.

We introduce a second unbiased estimator of CVA_I in the Appendix. This estimator, denoted by ξ_2, is computationally faster than the stratified sampling-based estimator at the expense of less controlled randomness. We compare the mean square error and computing time of ξ_2 and $\xi_{b,d}$ in the Appendix; there we identify conditions under which ξ_2 outperforms the biased estimator, $\xi_{b,d}$, in terms of MSE and computing time.

5 Efficient Monte Carlo Estimation of eEPE

In this section we discuss efficient Monte Carlo estimation of effective expected positive exposure, eEPE,

$$\text{eEPE} = \int_0^T \max_{0 \le u \le t} E[V_u]dt,$$

where $\{V_t \; ; t \ge 0\}$ denotes the credit exposure process, and T denotes the expiration time of the transaction with the longest maturity in a portfolio of OTC derivatives held by a financial institution with its counterparty. Consider the time grid, $0 \equiv t_0 < t_1 < ... < t_n \equiv T$. Set $\Delta_i \equiv t_i - t_{i-1}$, $i = 1, ..., n$. Monte Carlo estimators of eEPE are,

$$\hat{\theta}_{m,n,k} = \sum_{i=1}^{n} \max_{1 \le j \le i} \{\bar{V}_j\} \Delta_i, \tag{33}$$

where $\bar{V}_j$ denotes the m-simulation run average of the i.i.d. random variables, $V_{j1}, ..., V_{jm}$. The subscript $k = p$ and d denote PDS and DJS sampling, respectively. That is, under $k = p$ ($k = d$), V_j's are positively correlated (uncorrelated). Consider the mean square error of $\hat{\theta}_{m,n,k}$,

$$\text{MSE}(\hat{\theta}_{m,n,k}) \;\; = \;\; \text{Var}\left(\sum_{i=1}^{n} \max_{1 \le j \le i} \{\bar{V}_j\} \Delta_i\right) + \left(\sum_{i=1}^{n} E[\max_{1 \le j \le i} \{\bar{V}_j\}] \Delta_i - \text{eEPE}\right)^2. \tag{34}$$

It is useful to differentiate the following two sources of bias,

$$\left(\sum_{i=1}^{n} E[\max_{1 \le j \le i} \{\bar{V}_j\}] \Delta_i - \sum_{i=1}^{n} \max_{1 \le j \le i} E[\bar{V}_j] \Delta_i\right) - \left(\text{eEPE} - \sum_{i=1}^{n} \max_{1 \le j \le i} E[\bar{V}_j] \Delta_i\right). \tag{35}$$

That is, the first part of the bias is due to the presence of the maximum operator and the second part is time-discretization bias. Note that for a fixed n, variance of $\hat{\theta}_{m,n,k}$ converges to zero as $m \to \infty$. Now, consider Proposition 2 below whose proof is in the Appendix.

Proposition 2. *Let* $\{V_t;\ t \geq 0\}$ *denote the credit exposure process. Let,*

$$M_{n,m,k} \equiv \max\{\bar{V}_1, ..., \bar{V}_n\},$$

where $V_i \equiv V_{t_i}$ *on the time grid* $0 \equiv t_0 < t_1 < ... < t_n \equiv T$, *and* $\bar{V}_i = \sum_{j=1}^{m} V_{ij}/m$, $V_{i1}, ..., V_{im}$ *are i.i.d random variables. Also,* $k = d$ *and* $k = p$ *refer to the cases where* V_i *are uncorrelated and positively correlated, respectively, resulting from DJS and PDS-based simulation of* V. *Assume that* $E[V_i^2] < \infty$ *for all* $i = 1, ..., n$. *Let* $M_n \equiv \max\{E[V_1], ..., E[V_n]\}$. *Then, as* $m \to \infty$,

$$M_{n,m,k} \to M_n \quad a.s., \tag{36}$$

where a.s. stands for almost surely.

Note that dominated convergence theorem and Proposition 2 give $E[M_{n,m,k}] \to M_n$ as $m \to \infty$.[15] So, the first part of the bias

$$\sum_{i=1}^{n} E[\max_{1\leq j\leq i}\{\bar{V}_j\}]\Delta_i - \sum_{i=1}^{n} \max_{1\leq j\leq i} E[\bar{V}_j]\Delta_i$$

converges to zero as $m \to \infty$. That is, $\hat{\theta}_{m,n,d}$ and $\hat{\theta}_{m,n,p}$ are consistent estimators of eEPE$_{dst}$ for a fixed n. Since as $n \to \infty$ the time-discritization bias also converges to zero, $\hat{\theta}_{m,n,d}$ and $\hat{\theta}_{m,n,p}$ are also consistent estimators of eEPE.

In what follows we first show that for a fixed n and sufficiently large m, $\hat{\theta}_{m,n,d}$ outperforms $\hat{\theta}_{m,n,p}$ in terms of variance. Next, after specifying approximates for the order of variance and bias of $\hat{\theta}_{m,n,k}$, we formulate an MSE minimization problem over m and n given a fixed computational budget. Our numerical results indicate that our proposed estimators of eEPE, which use approximately optimal m and n, lead to substantial MSE reduction when compared to the crude estimators.

5.1 Comparing PDS and DJS-based Monte Carlo Estimators of eEPE

We are to compare the variance of $\hat{\theta}_{m,n,p}$ and $\hat{\theta}_{m,n,d}$ for a fixed n and sufficiently large m. Set

$$\theta \equiv \sum_{i=1}^{n} \max_{1\leq j\leq i} E[V_j]\Delta_i\ ,\ \hat{\theta}_{m,n,k} \equiv \sum_{i=1}^{n} \max_{1\leq j\leq i}\{\bar{V}_j\}\Delta_i,$$

where $k = p$ ($k = d$) refer to the cases where V_i and V_j for any $i \neq j$, are positively correlated (uncorrelated). In what follows we find it useful to append a second subscript m to $\bar{V}_i$ to emphasize that the average is based on m i.i.d random variables and a third subscript $k = d$ or p to indicate DJS or PDS.

Denote by $\delta_{i,j} \equiv E[V_i] - E[V_j]$ and $\delta \equiv \min\{|\delta_{i,j}| : i \neq j, i, j = 1, ..., n\}$. Without loss of generality, assume $\delta > 0$. Let $\sigma_{i,j,k}$ denote the standard deviation of $V_i - V_j$ under estimation method type k and $\sigma_{\max} \equiv \max\{\sigma_{i,j,k} : i, j = 1, ..., n, k = d, p\}$. For $i = 1, ..., n$, let τ_i

[15] Note that $M_{n,m,k} \leq \sum_{i=1}^{n} \bar{V}_i$ and Proposition 2 assumes $E[\bar{V}_i] = E[V_i] < \infty$.

denote the index for which $\max\{E[V_1], ..., E[V_i]\}$ is attained and $\tau_{i,m,k}$ be the index for which $\max\{\bar{V}_{1,m,k}, ..., \bar{V}_{i,m,k}\}$ is achieved. It then follows from these definitions that

$$\theta = \sum_{i=1}^{n} E[V_{\tau_i}]\Delta_i \quad \text{and} \quad \hat{\theta}_{m,n,k} = \sum_{i=1}^{n} \bar{V}_{\tau_{i,m,k},m,k}\Delta_i. \tag{37}$$

For $k = d$ or p and $i = 2, ..., n$, the probability that simulations do not yield the right τ_i can be bounded from above as follows

$$\begin{aligned}
P(\tau_{i,m,k} \neq \tau_i) &\leq \sum_{j \neq \tau_i, j=1,...,i} P(\bar{V}_{\tau_i,m,k} - \bar{V}_{j,m,k} < 0) \\
&= \sum_{j \neq \tau_i, j=1,...,i} P(\bar{V}_{\tau_i,m,k} - \bar{V}_{j,m,k} - \delta_{\tau_i,j} < -\delta_{\tau_i,j}) \\
&< \sum_{j \neq \tau_i, j=1,...,i} P(\bar{V}_{\tau_i,m,k} - \bar{V}_{j,m,k} - \delta_{\tau_i,j} < -\delta) \\
&< \sum_{j \neq \tau_i, j=1,...,i} P(|\bar{V}_{\tau_i,m,k} - \bar{V}_{j,m,k} - \delta_{\tau_i,j}| > \delta) \\
&\leq \sum_{j \neq \tau_i, j=1,...,i} \frac{\sigma^2_{\tau_i,j,k}}{m\delta^2} \qquad (38) \\
&\leq \frac{(i-1) \cdot \sigma^2_{\max}}{m\delta^2},
\end{aligned}$$

where (38) follows from the Chebyshev's inequality.

Consider the event $B_m = \{\tau_i = \tau_{i,m,d} = \tau_{i,m,p}, \text{for all } i = 1, ..., n\}$. It makes sense to call B_m the *desirable* event and B_m^c the *undesirable* event. Let $\hat{\theta}_{m,n,k,B_m}$ denote $\hat{\theta}_{m,n,k}$ conditional on the event B_m. We have that

$$\text{Var}(\hat{\theta}_{m,n,d,B_m}) < \text{Var}(\hat{\theta}_{m,n,p,B_m}). \tag{39}$$

This order can be established by first noting that

$$\hat{\theta}_{m,n,k,B_m} = \sum_{i=1}^{n} \bar{V}_{\tau_i,m,k}\Delta_i. \tag{40}$$

Then since V_i and V_j, for any $i \neq j$, are positively correlated (uncorrelated) under PDS (DJS) sampling, the variance of expression (40) is lower under DJS than under PDS.

Note that:

$$\begin{aligned}
P(B_m^c) &\leq \sum_{i=2}^{n} P(\tau_{i,m,d} \neq \tau_i) + \sum_{i=2}^{n} P(\tau_{i,m,p} \neq \tau_i) &(41) \\
&< 2\sum_{i=2}^{n} \frac{(i-1) \cdot \sigma^2_{\max}}{m\delta^2}, &(42)
\end{aligned}$$

The above argument leads to the following result.

Proposition 3. *Consider the desirable event* B_m *as defined above. First, conditional on this event,* (39) *holds. Secondly, the desirable event occurs asymptotically almost surely as* $m \to \infty$*. That is,* $\lim_{m\to\infty} P(B_m) = 1$*. More specifically,* $P(B_m^c)$ *goes to zero at rate* $1/m$ *as* $m \to \infty$*.*

Proposition 3 suggests that for sufficiently large m, $\text{Var}(\hat{\theta}_{m,n,d}) \leq \text{Var}(\hat{\theta}_{m,n,p})$. Our various numerical examples of Section 5.3 use $m > 400$; they all indicate that $\text{Var}(\hat{\theta}_{m,n,d}) \leq \text{Var}(\hat{\theta}_{m,n,p})$.

5.2 Efficient Monte Carlo Estimation of eEPE

Similar to our approach in subsection 3.2, we would like to find the number of valuation points, n, and the number of simulation runs at each valuation point, m, to minimize $\text{MSE}(\hat{\theta}_{m,n,k})$ given a fixed computational budget, s, that is proportional to, mn. To do so, we need to specify the order of the variance and bias of the Monte Carlo estimator of eEPE, $\hat{\theta}_{m,n,k}$. We are not concerned with deriving sharp estimates of the orders of variance and bias. In fact, our numerical examples indicate that choosing approximately optimal m and n using even very rough approximates for the orders of variance and bias lead to substantial MSE reduction. The following is used to formulate our MSE minimization problem: for $k = p$ or d,

$$\text{Var}(\hat{\theta}_{m,n,k}) \approx \frac{c_{1,k}}{m} \quad \text{and} \quad \text{Bias}(\hat{\theta}_{m,n,k}) \approx \frac{c_{2,k}}{m} + \frac{c_3}{n}, \tag{43}$$

for some constants $c_{1,k}, c_{2,k}, c_3$. The above approximation of the order of bias uses (35) and Proposition 2. Note that our rough approximate of the order of variance, applicable to both $\hat{\theta}_{m,n,d}$ and $\hat{\theta}_{m,n,p}$, does not depend on n. This is because of the presence of the maximum operators that leads to positive covariance terms. To see this, let $\bar{V}_i$ denote the m-simulation-run average of the i.i.d random variables, $V_{i1}, ..., V_{im}$, $i = 1, ..., n$, and consider an equidistant time grid with n time points, $\Delta = T/n$. Note that,

$$\text{Var}(\hat{\theta}_{m,n,k}) = \Delta^2 \text{Var}\left(\bar{V}_1 + \max\{\bar{V}_1, \bar{V}_2\} + ... + \max\{\bar{V}_1, ..., \bar{V}_n\}\right),$$

is equal to $\Delta^2 = \frac{T^2}{n^2}$ times the sum of n non-zero variance terms and $n(n-1)/2$ positive covariance terms both for $k = d$ and $k = p$. This leads to a result similar to (15) which can be further approximated by (16) as shown in subsection 3.2.

Given (43), we recommend solving the following MSE minimization problem to specify the approximately optimal m and n,

$$\min_{m,n} \left(\frac{c_1}{m} + (\frac{c_2}{m} + \frac{c_3}{n})^2\right) \quad \text{subject to} \quad s = cmn, \tag{44}$$

for some constants c_1, c_2, c_3, and c.

5.3 Numerical Examples

Our numerical examples presented below illustrate the efficiency of our proposed estimators of eEPE.[16] As in Section 3.4, we consider the simple forward contract where the underlying price

[16] We refer the reader to Section (D) of the Appendix for a discussion on eEPE_{dst} and numerical illustrations of Propositions 2 and 3.

process follows a geometric Brownian motion with drift μ and volatility σ. Let $\hat{\theta}_{c,p}$ and $\hat{\theta}_{c,d}$ denote the "crude" Monte Carlo estimators of eEPE under PDS and DJS sampling, respectively. That is,

$$\hat{\theta}_{c,k} = \sum_{i=1}^{n} \max_{1 \le j \le i} \bar{V}_j \Delta_i, \tag{45}$$

where $k = p, d$ and $\Delta_i = t_i - t_{i-1}$ and the t_i's are 1, 2, 3, 4, 8, 12, 18, 21, 24, 36, 49 weeks and 1 year, with $t_{12} = T = 1$ year. Let $\hat{\theta}_{e,p}$ and $\hat{\theta}_{e,d}$ denote the efficient Monte Carlo estimators of eEPE under PDS and DJS sampling, respectively, based on an equidistant time grid, i.e., expression (45) with $\Delta_i \equiv \Delta = T/n$) and resulting from solving the MSE minimization problem (44) (with constants c_i, $i = 1, 2, 3$, and c therein set to 1) in Section 5.2. In particular, under $s = 12,000$, the optimal $n = 29$ and $m = 414$, and under $s = 120,000$, the optimal $n = 62$ and $m = 1935$.

Our various numerical examples result in findings similar to those for the EPE estimation. For example, Tables 6 to 9, all based on 10^4 replications, show that the variance of the DJS-based estimators are much lower than that of the corresponding PDS-based estimators. Also, our proposed estimators of eEPE substantially outperform the crude Monte Carlo estimators in terms of MSE; for instance, MSE is reduced by a factor of 100 in Table 9.

	eEPE	Variance	MSE	CPU Time
$\hat{\theta}_{c,p}$	57.2278	.10931	22.4936	.00259
$\hat{\theta}_{c,d}$	57.2233	.034659	22.3768	.00168
$\hat{\theta}_{e,p}$	53.4344	.19358	1.0731	.00191
$\hat{\theta}_{e,d}$	53.4379	.011188	.89722	.00211

Table 6: $S_0 = 30, \mu = 1, \sigma = .25,\ s = 12,000$

	eEPE	Variance	MSE	CPU Time
$\hat{\theta}_{c,p}$	57.2277	.010824	22.3945	.02866
$\hat{\theta}_{c,d}$	57.2262	.003591	22.3734	.01427
$\hat{\theta}_{e,p}$	52.9363	.03962	.23301	.01817
$\hat{\theta}_{e,d}$	52.9354	.001083	.19367	.01545

Table 7: $S_0 = 30, \mu = 1, \sigma = .25,\ s = 120,000$

	eEPE	Variance	MSE	CPU Time
$\hat{\theta}_{c,p}$	81.0388	.24286	101.0233	.00279
$\hat{\theta}_{c,d}$	81.0309	.0843	100.7055	.00173
$\hat{\theta}_{e,p}$	72.8899	.3986	3.9705	.00221
$\hat{\theta}_{e,d}$	72.8885	.024652	3.5914	.00226

Table 8: $S_0 = 30, \mu = 1.5, \sigma = .25,\ s = 12,000$

	eEPE	Variance	MSE	CPU Time
$\hat{\theta}_{c,p}$	81.0332	.024156	100.692	.02929
$\hat{\theta}_{c,d}$	81.0302	.008395	100.6154	.0144
$\hat{\theta}_{e,p}$	71.8779	.083579	0.85454	.01935
$\hat{\theta}_{e,d}$	71.8807	.002425	0.77819	.01630

Table 9: $S_0 = 30, \mu = 1.5, \sigma = .25,\ s = 120,000$

6 Conclusion

It has become increasingly crucial for financial institutions to actively manage their counterparty credit risk. Proper counterparty credit risk management is challenging and computationally intensive. Monte Carlo simulation is often used for CCR pricing and measurement. We improve the existing widely used Monte Carlo CCR frameworks by substantially increasing the efficiency of Monte Carlo estimators of the key CCR measures: EPE, CVA, and eEPE. Introducing and using the notion of marginal matching, we show that the so-called path dependent simulation (PDS) method, which simulates the credit exposure process based on the finite dimensional distributions of the underlying risk factors, leads to CCR estimators whose variance is substantially larger than the variance of the CCR estimators calculated based on the so-called direct jump to simulation date (DJS) method. Taking into account the computational time in parallel with the mean square error, we demonstrate that DJS sampling is preferable to PDS sampling for path independent derivatives. For path dependent derivatives the two sampling methods are approximately equivalent. We show that the mean square error (MSE) of the crude Monte Carlo estimators of EPE, CVA, and eEPE can be substantially reduced by solving approximate MSE minimization problems that specify how to achieve an approximately optimal balance between bias squared and variance. Our proposed efficient estimators of EPE and CVA are in fact unbiased and derived using stratified sampling with the number of strata and simulation runs (allocated to each stratum) being chosen based on the solution to the aforementioned MSE minimization problems. Our various numerical examples illustrate that employing our proposed Monte Carlo frameworks will substantially increase the efficiency of the existing Monte Carlo CCR "engines".

Appendix

A Proof of Proposition 1

In this proof, for notational simplicity, we suppress the dependence of $\hat{\theta}_{b,m,n,d}$ and $\hat{\theta}_{u,m,n,d}$ on m and n. Note that,

$$n\text{MSE}(\hat{\theta}_{b,d}) = \frac{T}{m}\sum_{i=1}^{n}\text{Var}(V_i)\Delta_i + n(\sum_{i=1}^{n}E[V_i]\Delta_i - \int_0^T E[V_t]dt)^2, \tag{46}$$

where the first term on the right hand side of the above equality uses $\text{Var}(\bar{V}_i) = \text{Var}(V_i)/m$. So, $n\text{MSE}(\hat{\theta}_{b,d})$ converges to $c\int_0^T \text{Var}(V_t)dt$ as $n \to \infty$.

Now, consider $\text{Var}(\hat{\theta}_{u,d})$, and let $I_n \equiv I_n(\tau) \in \{1, ..., n\}$ denote the index of the stratum containing τ. Set $p_i = P(\tau \in A_i) = \frac{\Delta_i}{T}$. From standard results on stratified sampling we have,

$$\text{Var}(\hat{\theta}_{u,d}) = \frac{T^2}{mn}\sum_{i=1}^{n}\text{Var}(V_\tau|\tau \in A_i)p_i = \frac{T^2}{mn}E[\text{Var}(V_\tau|I_n)]. \tag{47}$$

Since $\int_0^T \text{Var}(V_t)dt = TE[\text{Var}(V_\tau|\tau)]$, to complete the proof, it suffice to show that, as $n \to \infty$,

$$E[\text{Var}(V_\tau|I_n)] \longrightarrow E[\text{Var}(V_\tau|\tau)]. \tag{48}$$

From the formula for the conditional variance, to show the convergence in (48), it suffice to show that, as $n \to \infty$,

$$E\left[(E[V_\tau|I_n])^2\right] \longrightarrow E\left[(E[V_\tau|\tau])^2\right]. \tag{49}$$

Set $X = E[V_\tau|\tau]$ and $X_n = E[V_\tau|I_n]$. Note that X_n is a martingale because as n increases I_n generate increasing family of sigma-algebras. We can use martingale convergence theorem (see Chapter 4 of [4]) to conclude that X_n converges to X almost surely as $n \to \infty$. Using continuous mapping theorem and dominated convergence theorem (see Chapter 1 of [4]) we conclude that, $E[X_n^2]$ converges to $E[X^2]$ almost surely, and so (49) holds. This completes the proof of Lemma 1. [17]

B A Second Unbiased Estimator of CVA_I

The following efficient estimator of independent CVA, similar to $\xi_{u,k}$, is unbiased, and it is derived using stratification and control variate method. Note that

$$E[V_\tau 1\{\tau \leq T\}] = E[V_\tau|\tau \leq T]p, \tag{50}$$

[17] The probabilistic arguments used in second part of the proof are similar to the ones used in the proof of Lemma 4.1 in [8].

where $p = P(\tau \leq T)$ is analytically computable. Let $\omega \equiv \tau | \tau \leq T$. That is, ω is a random variable distributed according to the distribution of τ conditional on the event $\tau \leq T$. To estimate the above conditional expectation we use ω as a control variable. Our second unbiased estimator of independent CVA based on m-simulation is,

$$\xi_2 = (\bar{V}_\omega + c^*(\bar{\omega} - E[\omega]))p, \tag{51}$$

where, from standard results on control variate method, $c^* = \text{cov}(V_\omega, \omega)/\text{Var}(\omega)$. And, $\bar{\omega}$ is the average of m Monte Carlo realizations of ω and $\bar{V}_\omega = m^{-1} \sum_{i=1}^m V_{\omega_i}$. Note that to sample from V_ω, first generate $\omega \equiv \tau | \tau \leq T$. Then, given this realization of ω, sample from V_ω, (note that ω and V are independent). The variance of ξ_2, which is based on m simulation runs, is,

$$\text{Var}(\xi_2) = \frac{p^2}{m}\left(\int_0^T \text{Var}(V_t) dF_\omega(t) + \text{Var}(E[V_\omega|\omega]) - \frac{\text{cov}^2(V_\omega, \omega)}{\text{Var}(\omega)}\right), \tag{52}$$

where the first two terms inside the parenthesis follow from the conditional variance formula,

$\text{Var}(V_\omega) = E[\text{Var}(V_\omega|\omega)] + \text{Var}(E[V_\omega|\omega])$ and the last term inside the parenthesis is due to the use of control variate method. Now, let us compare the performance of ξ_2 with the biased estimator of CVA_I under DJS sampling, $\xi_{b,d}$. Suppose that proportional sampling is used as in Lemma 1. Note that,

$$\begin{aligned} \text{Var}(\xi_2) - \text{MSE}(\xi_{b,d}) &= \frac{p}{m}\int_0^T \text{Var}(V_t) dF(t) - \frac{1}{N}\sum_{i=1}^n \text{Var}(V_i)\Delta F_i \\ &\quad + \frac{p^2}{m}\left(\text{Var}(E[V_\omega|\omega]) - \frac{\text{cov}^2(V_\omega, \omega)}{\text{Var}(\omega)}\right) - (\text{Bias})^2, \end{aligned} \tag{53}$$

where we have used $f_\omega(t) = f_\tau(t)/p$ for $t \in [0, T]$. Setting $m = Np$, simply note that for sufficiently large n and for cases where,

$$\text{cov}^2(V_\omega, \omega) \approx \text{Var}(\omega)\text{Var}(E[V_\omega|\omega]), \tag{54}$$

ξ_2 and $\xi_{b,d}$ perform similarly in terms of mean square error. However, for cases where $p = P(\tau \leq T)$ is small, which would also lead to reduction in computing time when calculating ξ_2, this estimator is preferable to $\xi_{b,d}$. Recall that ξ_2 and $\xi_{b,d}$ are calculated based on $m = Np$ and N simulation runs, respectively. (Note that from the conditional variance formula $\text{Var}(V_\omega) > \text{Var}(E[V_\omega|\omega])$, and so, there exists practical cases where (54) holds.)

To get a feel for this consider the simple example where V is a Brownian motion with drift μ and volatility σ and τ is an exponential random variable with mean $1/\lambda$. This example results in $\text{Var}(E[V_\omega|\omega]) = \mu^2\text{Var}(\omega)$ and $\text{cov}^2(V_\omega, \omega) = \mu^2\text{Var}(\omega)$. So, setting $m = Np = NP(\tau \leq T)$, we get $\text{Var}(\xi_3) = \text{MSE}(\xi_1)$ for large enough n, (as $n \to \infty$, the bias-squared term converges to zero with rate $1/n^2$ and the Riemann Stieltjes sum in (53) converges to the integral term in (53) with rate $1/n$). That is, for this example $\xi_{b,d}$ and ξ_2 perform similarly in terms of asymptotic MSE. However, taking in to account the computing time, ξ_2 should be preferred to $\xi_{b,d}$ when

p is small. Note that for finite n cases in the practical CCR settings, the bias term may not be negligible compared to variance, as seen in Table 5. And, so, for small p cases if a small simulation study reveals that $\text{cov}^2(V_\omega, \omega) \approx \text{Var}(\omega)\text{Var}(E[V_\omega|\omega])$, the estimator ξ_2 is preferable to $\xi_{b,d}$.

C Proof of Proposition 2

We first consider $M_{2,m,k}$. Let us assume that $M_2 = E[V_2]$ without loss of generality. Note that,

$$\max\{\bar{V}_1, \bar{V}_2\} - E[V_2] = \bar{V}_1 1\{\bar{V}_1 > \bar{V}_2\} + (\bar{V}_2 - E[V_2])1\{\bar{V}_2 > \bar{V}_1\} - E[V_2]1\{\bar{V}_1 > \bar{V}_2\}. \quad (55)$$

First, consider the indicator random variable, $1\{\bar{V}_1 > \bar{V}_2\}$; the dependence of $\bar{V}_i$ on m is suppressed for notational simplicity. Set $W^k \equiv V_1 - V_2$, where $k = d, p$ refer to the cases where V_1 and V_2 are uncorrelated and positively correlated, respectively. Note that $1\{\bar{V}_1 > \bar{V}_2\} \leq 1\{\bar{W^k} > E[W^k]\}$; $W_1^k, ..., W_m^k$ are i.i.d random variables and $\bar{W^k}$ is their average. It is well known that $1\{\bar{V}_1 > \bar{V}_2\} \to 0$ a.s. if and only if for all $\epsilon > 0$,

$$P(1\{\bar{V}_1 > \bar{V}_2\} > \epsilon \text{ i.o.}) = 0, \quad (56)$$

where i.o. stands for infinitely often. To see that (56) holds, note that,

$$P(1\{\bar{V}_1 > \bar{V}_2\} > \epsilon) \leq P(1\{\bar{W^k} > E[W^k]\} > \epsilon) \leq \frac{P(|\bar{W^k} - E[W^k]| > \tilde{\epsilon})}{\epsilon^2}, \quad (57)$$

for all $\tilde{\epsilon} > 0$. To derive the last inequality above the Chebyshev's inequality is used. Then, (57), almost sure convergence of $\bar{W^k} \to E[W^k]$ following from the strong law of large numbers (SLLN), and Kolmogorov's 0-1 law, (see Theorem 8.1 of [4]), give (56).

Now, consider the first term on the right side of (55). Given that $\bar{V}_1$ and $1\{\bar{V}_1 > \bar{V}_2\}$, almost surely converge to $E[V_1]$ and zero, respectively, it is not difficult to show that

$$\bar{V}_1 1\{\bar{V}_1 > \bar{V}_2\} \to 0 \ \text{ a.s.}.$$

To see this, it suffices to write

$$\bar{V}_1 1\{\bar{V}_1 > \bar{V}_2\} = (\bar{V}_1 - E[V_1])1\{\bar{V}_1 > \bar{V}_2\} + E[V_1]1\{\bar{V}_1 > \bar{V}_2\},$$

and use SLLN for the sequence of indicator random variables and $\bar{V}_1$. The last term on the right side of (55) converges to zero a.s. based on (56). Analogous arguments led to (56) show that the second term on the right side of (55) converges to zero a.s. This completes the proof for $n = 2$. Induction and analogous arguments are employed for the general case.

Suppose that $M_{n-1,m,k} \to M_{n-1}$, a.s. as $m \to \infty$. Assume that $M_n = E[V_n]$. Then, we need to show that a similar almost sure convergence holds for $M_{n,m,k}$. To see this, it suffices to note that, for all $\epsilon > 0$ and $\tilde{\epsilon} > 0$,

$$\begin{aligned} P(1\{\bar{V}_1 > \max\{\bar{V}_2, ..., \bar{V}_n\}\} > \epsilon) &\leq \frac{P\left((\bar{V}_1 - E[V_1]) - (\max\{\bar{V}_2, ..., \bar{V}_n\} - E[V_n]) > \tilde{\epsilon}\right)}{\epsilon^2} \\ &\leq \frac{P\left(|\bar{V}_1 - E[V_1]| > \tilde{\epsilon}\right)}{\epsilon^2} + \frac{P\left(|\max\{\bar{V}_2, ..., \bar{V}_n\} - E[V_n]| > \tilde{\epsilon}\right)}{\epsilon^2}, \end{aligned}$$

which is then used to show that

$$1\{\bar{V}_1 > \max\{\bar{V}_2, ..., \bar{V}_n\}\} \longrightarrow 0 \text{ a.s.}$$

This completes the proof.

D Numerical Examples for eEPE_{dst}

The numerical results presented in this section demonstrate the consistency of PDS and DJS estimators for eEPE_{dst} and the asymptotic efficiency of DJS over PDS. In particular, they support our Propositions 2 and 3.

We consider the simple forward contract and the underlying security price process following a geometric Brownian motion with initial value $S_0 = 30$, drift $\mu = 0.01$, and volatility $\sigma = 1$ here. We compare the crude PDS and DJS estimators $\hat{\theta}_{c,p}$ and $\hat{\theta}_{c,d}$ as defined in Section 5.3. Each estimation procedure is replicated $10,000$ times to produce the estimates.

In Tables 10 to 13, in addition to presenting the estimator value, variance, MSE (which, unlike that in Section 5.3, is defined with respect to the estimand of eEPE_{dst}), and CPU time, we also include a column named "WrongOrderProb", which gives the estimate for the probability that the indices at which the running maximums are achieved ever go wrong, i.e., using the notation in Section 5.1 of the main paper, $P(\tau_{i,m,k} \neq \tau_i, \text{for some } i)$, $k = p$ or d, corresponding to PDS and DJS respectively. The sum of these two probabilities provides an upper bound for $P(B_m^c)$ in the statement of Proposition 3. As these four tables show, this upper bound converges to zero as m increases, which implies $\lim_{m\to\infty} P(B_m^c) = 0$. Also, the bias of both estimators vanishes as m increases; this is consistent with Proposition 2.

	eEPE_{dst}	Variance	MSE	CPU Time	WrongOrderProb
$\hat{\theta}_{c,p}$	41.0514	20.6228	20.7011	0.000329	0.9613
$\hat{\theta}_{c,d}$	42.4019	7.901	10.5697	0.000431	1

Table 10: $m = 50$

	eEPE_{dst}	Variance	MSE	CPU Time	WrongOrderProb
$\hat{\theta}_{c,p}$	40.7481	2.1847	2.1849	0.00156	0.2789
$\hat{\theta}_{c,d}$	40.8833	0.77441	0.78761	0.00121	0.919

Table 11: $m = 500$

	eEPE_{dst}	Variance	MSE	CPU Time	WrongOrderProb
$\hat{\theta}_{c,p}$	40.7756	0.21446	0.2146	0.0173	0
$\hat{\theta}_{c,d}$	40.7708	0.07379	0.07379	0.0090	0.1987

Table 12: $m = 5000$

	eEPE_{dst}	Variance	MSE	CPU Time	WrongOrderProb
$\hat{\theta}_{c,p}$	40.7708	0.02156	0.02156	0.1708	0
$\hat{\theta}_{c,d}$	40.7682	0.00711	0.00711	0.0932	0.0001

Table 13: $m = 50,000$

References

[1] Bohme, M., Chiarella, D., Harle, P., Neukirchen, M., Poppensieker, T., and Raufuss, A., "Day of reckoning: new regulation and its impact on capital market businesses", *McKinsey Working Papaers on Risk*, (September 2011)

[2] Broadie, M., Du, Y., Moallemi, C. C., "Risk estimation via regression", Working paper, December (2011)

[3] Canabarro, E., and Duffie, D., "Measuring and Marking Counterparty Risk", *Asset/Liability Management for Financial Institutions*, Institutional Investor Books, (2003)

[4] Durrett, R., "Probability: Theory and Examples", 3rd Edition, Duxbury Advanced Series, (2005)

[5] Egozcue, M., Garcia, F., L., Wong, W. K., "On some covariance inequalities for monotonic and non-monotonic functions" *Journal of Inequalities in Pure and Applied Mathematics*, **vol.10** , (2009)

[6] Ghamami, S., Goldberg, L. R., "Stochastic Intensity Models of Wrong Way Risk: Wrong Way CVA Need Not Exceed Independent CVA", Working paper, December (2012)

[7] Glasserman, P., "Monte Carlo Methods in Financial Engineering", Springer, (2004)

[8] Glasserman, P., Heidelberger, P., and Shahabuddin, P., "Asymptotically optimal importance sampling and stratification for path dependent options", *Mathematical Finance*, **9** pp. 117-152, (1999)

[9] Glynn, P. W., and Whitt, W., "The efficiency of simulation estimators", *Operations Researcg*, **40**, pp. 505-520(1992)

[10] Gordy, M. B., and Sandeep J., "Nested Simulation in Portfolio Risk Measurement", *Management Science*, **vol. 56, no. 10**, pp. 1833-1848. (2010)

[11] Gregory, J., "Counterparty credit risk", Wiley Finance, (2010)

[12] Hull, J., and White, A., "CVA and wrong way risk", Forthcoming: *Financial Analyst Journal*, March 5, (2012)

[13] Karatzas, I., and Shreve, S. E. , "Brownian Motion and Stochastic Calculus", Springer, (1991)

[14] Musiela, M., and Rutkowski, M., "Martingale Methods in Financial Modelling", 2nd Edtion, Springer, (2010)

[15] Pykhtin, M., and Zhu, S., "A guide to modeling counterparty credit risk", *GARP Risk Review*, July/August, 16-22, (2007)

[16] Pykhtin, M., and Zhu, S., "Measuring counterparty credit risk for trading products under Basel II", *In the Basel Handbook*, 2nd edition, Risk books, London, (2006)

[17] Ross, S. M., "Introduction to Probability Models", Academic Press, 10th edition, (2009)

[18] Ross, S. M., "Simulation", 4th edition, Academic Press, (2006)