Efficient Monte Carlo Counterparty Credit Risk Pricing and Measurement

Samim Ghamami and Bo Zhang

Efficient Monte Carlo Counterparty Credit Risk Pricing and Measurement*

Samim Ghamami† and Bo Zhang‡

December 17, 2014

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 estimation 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.

Keywords: Risk Management, Counterparty Credit Risk, OTC Derivatives Market, Credit Value Adjustment, Efficient Monte Carlo Simulation, Basel II-III

1 Introduction and a Summary of Important CCR Measures

Counterparty credit risk (CCR) is the risk that a party to a derivative contract may default prior to the expiration of the contract and fail to make the required contractual payments, (see [5] for the basic CCR definitions). Counterparty credit risk has been widely considered as one of the key drivers of the 2007-08 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, [3]). It is well known that pricing and measuring counterparty credit risk is computationally extremely intensive; financial institutions (derivative dealers) invest large amounts of resources developing and maintaining Monte Carlo simulation "engines" to manage their counterparty risk, (see [24], [18], and [5]). While various aspects of counterparty credit risk have been subject of extensive research post 2007-08 financial crisis, statistical efficiency of the CCR estimators has received no attention in the literature.

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 develop efficient Monte Carlo counterparty credit risk estimators. Currently, CVA is a CCR measure that is only applied to bilateral derivatives transactions. However, EPE and EEPE are CCR measures applicable to both bilateral derivatives transactions and centrally-cleared derivatives transactions. Specifically, Basel Committee on Banking Supervision (BCBS) has devised regulatory capital charges on clearing member banks against their central counterparty credit risk; EPE and EEPE are components of these capital charges to be estimated via Monte Carlo simulation by large dealer banks.

Counterparty credit exposure [5], denoted by V, of a financial institution against one of its counterparties, is the larger of zero and the market value of the portfolio of derivatives contracts the financial institution holds with this counterparty. To effectively introduce our efficient Monte Carlo procedures, we consider credit exposures in the absence of the commonly used risk mitigants, collateral and netting agreements. This simple setting facilitates the 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 [18]). It is defined as follows,

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

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 derivatives portfolio.

Effective EPE (EEPE), another widely used regulatory and economic capital-related counterparty risk measure [18] is defined as follows in the CCR literature:

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

---

[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.

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:

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

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 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 [24]).

CVA, which is the difference between the risk free portfolio value and the true counterparty default risky portfolio value, (see [23]), has become one of the most important CCR measures for derivatives dealers and other large financial institutions. CVA is also one of the main components of the Basel III's counterparty credit risk capital framework.

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 [18] for a derivation of this formula). Hereafter we suppress the dependence of the CVA on the recovery rate, R. When V and τ are assumed to be 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, EEPE, and independent CVA are estimated based on the Reimann sum approximation of the integrals in (1), (3), and (5) and 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*, which enables us to define and differentiate the two widely used CCR sampling methods, *Path Dependent Simulation*

---

[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 [12], [5], and [19]).

(PDS) and *Direct Jump to Simulation* date (DJS). These two terms were first introduced by Pykhtin and Zhu in 2006 [24]. Practitioners do not usually evaluate and compare the statistical efficiency and computing time associated with the PDS and DJS-based Monte Carlo estimators of CCR measures. A recurring theme of Sections 3 through 5 of this paper is to illustrate that PDS and DJS-based CCR estimators have drastically different MSE.[3] Section 3 introduces an efficient Monte Carlo framework for estimating EPE, which also directly applies to efficient CVA_I estimation. Using our results in Section 3, we summarize our proposed Monte Carlo framework for efficient estimation of CVA_I in Section 4. Using our results in Section 3, 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. We would like to emphasize that Sections 4 and 5 should not be read independently. The main components of the proposed efficient CCR framework are developed in Section 3. As will be seen in the sequel, this is because EPE and CVA_I are both weighted sums of expected exposures, and the ideas developed for efficient EPE and CVA_I estimation have implications for efficient EEPE estimation. Appendix H discusses the scope of our study by considering cases that have not been explicitly formulated in the main body of the paper.

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 financial institution that holds a portfolio of k 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 derivatives portfolio. When risk mitigants are employed, V_t is defined differently. For instance, in the presence of netting agreements, credit exposure becomes, (see [23]),

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

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 derivatives and $t_1, ..., t_n$ are sometimes referred to as *valuation points*. Set $V_i \equiv V_{t_i}$.

---

[3]The first version of the paper had used the first edition of the excellent book by Gregory [17] as one of its references as at the time of the completion of our main results the second edition of the book by Gregory [18] had not been published. While [17] does not discuss statistical efficiency of CCR measure, [18] does. The statistical efficiency discussions of [18] are summarized at the Section 3.1.3's footnote.

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, Value at Risk (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 [18] and [23] for more details). Note that EPE, CVA, and EEPE, as defined in Section 1, are *dynamic* in the sense that they depend on the time evolution of the credit exposure process. This paper focuses on developing a Monte Carlo CCR framework that provably improves the efficiency of estimating these three risk measures. PFE and its corresponding dynamic measure Maximum PFE are other important CCR measures (see [18]), whose efficient Monte Carlo estimation is not addressed in this paper. Efficient Monte Carlo quantile estimation is a well-studied topic, (see Chapter 9 of [13] and the references therein). As will be illustrated in the sequel, the mathematical analysis of this paper has benefited from some similarities in the functional forms of EPE, CVA, and EEPE. The functional form of the maximum PFE, i.e., the maximum of the quantiles of the exposure process on a discrete time grid, and its Monte Carlo estimator has a distinct nature. We leave efficient Monte Carlo Maximum PFE estimation for future study.

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

Suppose 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 should be valued. This two-step procedure leads to 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 the 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 [16]), and regression-based Monte Carlo (see [4]) uses ideas from regression-based Monte Carlo American option pricing, (see Chapter 8 of [13]).

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 (8), 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 a 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 normals.

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 in the following class:[4] a Gauss-Markov process (see Chapter 5 of [20] or Chapter 3 of [13]) specified by

$$dR_t = (g_t + h_t R_t)dt + \sigma_t dB_t,$$

with g, h, and σ all deterministic functions of time and B a standard one-dimensional Brownian motion, a geometric Brownian motion (GBM),

---

[4]This assumption is only used in the proof of Proposition 1 and Proposition 2.

$$dR_t = \mu R_t dt + \sigma dB_t,$$

with given constants μ and σ, or a square-root diffusion specified as

$$dR_t = \alpha(b - R_t)dt + \sigma\sqrt{R_t}dB_t,$$

in which α and b are positive. Many of the widely used continuous time stochastic processes in finance and economics are in this class. 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, GBM, and the square-root diffusion specified above whose simulations involve generating positively correlated normal random variables. Let $\tilde{R} = (\tilde{R}_1, ..., \tilde{R}_n)$ denote a random vector for which $\tilde{R} \neq^d R$ but $\tilde{R}_i =^d R_i$ for all $i = 1, ..., n$ and $\text{cov}(\tilde{R}_i, \tilde{R}_j) = 0$ for all $i \neq j$. That is, simulation of $\tilde{R}_1, ..., \tilde{R}_n$ can be done by generating n uncorrelated or simply independent 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). 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 (see [24]). One of the main contributions of this paper is to differentiate DJS and PDS in terms of the mean square error of the estimators of EPE, CVA, and EEPE.

3   Efficient Monte Carlo Estimation of EPE

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

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

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,

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

where $\bar{V}_i \equiv \sum_{j=1}^{m} V_{ij}/m$ 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 6.1 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 Estimation of EPE

Suppose that the credit exposure process, V, is defined on a given filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0 \leq t \leq \infty}, P)$, where $(\mathcal{F}_t)_{0 \leq t \leq \infty}$ denote the filtration generated by the underlying risk factors. Consider the setting where V denotes the contract level exposure and a financial institution takes a 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

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

where n is a numeraire. Transactions between the financial institution and its counterparty for which $V_t = \max\{C_t, 0\} = C_t$ for all $0 < t \leq T$ are referred to as *unilateral transactions*, e.g. the financial institution takes a long position in a call option with its counterparty. Transactions for which $V_t = \max\{C_t, 0\} \neq C_t$ for some $0 < t \leq T$ are referred to as *bilateral transactions*, e.g. an interest rate swap between the financial institution and its counterparty.

The following simple example reviews simulation of the exposure process under PDS and DJS. Suppose that $\{S_t\,;t \geq 0\}$ is a GBM, $S_t = S_0 e^{X_t}$, and $\{X_t\,;t \geq 0\}$ is a Brownian motion with drift μ and volatility σ. Consider a unilateral transaction. Note that $V_t = C_t = n_t E\left[\frac{\Pi_T}{n_T}|S_t\right] \equiv f(S_t)$.

That is, credit exposure is considered as a function of the risk factor.[5] Consider the time grid, $0 \equiv t_0 < t_1 < ... < t_n \equiv T$ and let $V_i \equiv V_{t_i}$. 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}. 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 $\sigma^2 \Delta_i$. 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 $i \neq j$, $i, j = 1, ..., n$.

In what follows 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 for *path independent* and *path dependent* derivatives. We consider unilateral and bilateral transactions in both single risk-factor and multi-risk factor settings. That is, we consider two cases: a stylized setting where $(\mathcal{F}_t)_{0 \leq t \leq \infty}$ is the filtration generated by a single risk factor; we also consider the more general multi-risk factor settings.

3.1.1 Path Independent Case

The above mentioned example shows that under DJS, $\text{cov}(V_u, V_t) = 0$ for any $0 < u < t < T$. Proposition 1 and Proposition 2, whose proofs are given in the Appendix, consider this covariance function of the contract level credit exposure process under the PDS method for unilateral and bilateral transactions, respectively, and identify conditions under which $\text{cov}(V_u, V_t) > 0$ for any $0 < u < t < T$. Condition 2 of Proposition 1 below uses the well known changes of numeraire

---

[5]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 $t \geq 0$ 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 $C_t = n_t E[\frac{\Pi_T}{n_T}|S_t] \equiv f(S_t)$.

techniques of Geman-El Karoui-Rochet [10] for option type contracts with at most three distinct sources of randomness: stochastic short rate and a maximum of two risky assets. Well known examples of these contracts are options written on stocks or bonds, e.g. European options and exchange options.

Proposition 1. *Consider the credit exposure process, $\{V_t \; ; t \geq 0\}$, defined on a given filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0 \leq t \leq \infty}, P)$, and a T-maturity transaction between the financial institution and its counterparty that is unilateral, i.e. the credit exposure process is the price process, $V_t = C_t > 0$ for all $0 \leq t \leq T$, where C_t denotes the time-t value of the derivative contract with payoff Π_T. Then,*

$$cov(V_u, V_t) > 0,$$

for any $0 < u < t < T$ under any of the following conditions:

Condition 1: Numeraire is the money market account, B, with deterministic short rate, r, and Π_T is a function of $N \geq 1$ exogenously given risky assets.

Condition 2: Short rate is stochastic and the T-payoff function is a function of at most two risky assets as follows $\Pi_T = (\alpha_1 S_1(T) + \alpha_2 S_2(T))^+$, where α_1 and α_2 are any real numbers, and S_1 and/or S_2 are risky assets.

In the case of bilateral transactions, e.g. interest rate swaps, for which the exposure process satisfies $V_t = \max\{C_t, 0\} \neq C_t$ for some $0 < t \leq T$, where C_t denotes the time-t value of the derivative contract with payoff function Π_T, stronger assumptions are required to analytically show that $cov(V_u, V_t) > 0$ for any $0 < u < t < T$. This is shown in Proposition 2 below.

Proposition 2. *Consider the credit exposure process, $\{V_t \; ; t \geq 0\}$, defined on a given filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0 \leq t \leq \infty}, P)$, and a T-maturity transaction between the financial institution and its counterparty that is bilateral, i.e. the credit exposure process is the price process, $V_t = \max\{C_t, 0\} \neq C_t$ for some $0 < t \leq T$, where C_t denotes the time-t value of the derivative contract with payoff function Π_T. Then,*

$$cov(V_u, V_t) > 0$$

for any $0 < u < t < T$ under the following condition:

Numeraire is the money market account, B, with deterministic short rate, r, and Π_T is a monotone function of a single risky asset whose dynamics is modeled by a GBM, a Gauss-Markov process or a square-root diffusion.

The monotonicity assumption of the payoff function is satisfied for most of the actively traded OTC derivative contracts; well-known exceptions are Barrier[6] and Lookback options, (see, for instance, [22]).

---

[6]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 [22] and the references there).

Propositions 1 and 2 identify conditions for unilateral and bilateral transactions under which the credit exposure process satisfies $\text{cov}(V_u, V_t) > 0$ for any $0 < u < t < T$. This, then, implies that

$$\text{Var}(\hat{\theta}_{b,m,n,d}) \leq \text{Var}(\hat{\theta}_{b,m,n,p}). \tag{10}$$

Note that the above inequality holds since

$$\text{Var}(\hat{\theta}_{b,m,n,d}) = \sum_{i=1}^{n} \frac{\text{Var}(V_i)\Delta_i^2}{m} \leq \sum_{i=1}^{n} \frac{\text{Var}(V_i)\Delta_i^2}{m} + \frac{2}{m}\sum\sum_{i<j}\text{cov}(V_i, V_j)\Delta_i\Delta_j = \text{Var}(\hat{\theta}_{b,m,n,p}). \tag{11}$$

3.1.2 Path Dependent Case

Suppose that V_t is time t value of a maturity-T contract, where the payoff at the time T is a function of $S_1, ...S_n$, (for instance, an arithmetic Asian option). That is, $V_i = g(S_1, ..., S_i)$, where g is a function from $\mathcal{R}^i$ to $\mathcal{R}$. 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 $\text{cov}(V_i, V_j) > 0$, $i \neq j$. More specifically, it can be shown that $\text{cov}(V_i, V_j) > 0$ for unilateral and bilateral transactions under the first condition of Proposition 1 and Proposition 2's condition, respectively. That is, for the above mentioned covariance function to be positive, we need the numeraire money market account with deterministic short rate in the unilateral case. The bilateral case, additionally, requires monotonicity of the payoff function and its dependence on a single risk factor.

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.[7] More specifically, the estimator with the lower

variance per replication × expected computing time,

should be selected (see [15] for the formal formulation of this useful criterion in comparing alternative Monte Carlo estimators). Consider, for instance, arithmetic Asian options. Suppose

---

[7]In the path independent case computing time of DJS and PDS-based estimators of θ are roughly equal.

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,

$$\frac{\text{ct}(\hat{\theta}_{b,1,n,d})}{\text{ct}(\hat{\theta}_{b,1,n,p})} \approx n \quad \text{and} \quad \frac{\text{Var}(\hat{\theta}_{b,1,n,p})}{\text{Var}(\hat{\theta}_{b,1,n,d})} \approx n. \tag{12}$$

To see why (12) 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 (11), 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 . So, $\hat{\theta}_{b,m,n,d}$ and $\hat{\theta}_{b,m,n,p}$ have a similar performance for fixed and sufficiently large n. PDS and DJS-based estimators of other derivatives whose payoff depends on the path in a different form can be compared similarly.

We emphasize that what we have referred to as the "DJS method" for the path dependent case can be considered as a generalized version of what the conventional CCR literature often refers to as "DJS". While clarifying this point, Appendix G provides a further explanatory discussion of the DJS method. Appendix H, which discusses settings that have not been formulated in our study, illustrates the potential applicability of the DJS method for these more general cases.

3.1.3   Summary of Section 3.1

We summarize the result of Section 3.1 as it is used in the sequel and is directly applied to the efficient CVA_I estimation.[8] To compare the DJS and PDS-based estimators of EPE and CVA_I (both being viewed as weighted sums of expected exposures) variance and computing time of the Monte Carlo estimators are considered. The DJS method induces zero covariance between any two distinct time points of the simulated credit exposure process. So, it remains to look at this covariance function for the credit exposure process under the PDS method. When the dynamics of the risk factors are modeled by the class of continuous time stochastic processes considered in this paper, the covariance function of the credit exposure process under the PDS method becomes positive under conditions of Propositions 1 and 2 for unilateral and bilateral path independent derivatives transactions, respectively. Similar results hold for path dependent derivatives. That is, under conditions of Propositions 1 and 2, DJS-based estimators of EPE and CVA outperform the PDS-based estimators in terms of variance. For path independent derivatives PDS and DJS-based computing times are roughly equal. So, we recommend that

---

[8]Section 9.3.2 of Gregory [18] gives an overview of the PDS, referred to as pathwise method, and DJS, referred to as direct simulation, and broadly concludes that the pathwise method is more suitable for path-dependent derivatives. Section 12.4.4 of [18] discusses the PDS and DJS methods in Monte Carlo estimation of CVA where the counterparty's default is modeled by a structural-copula approach. In this context, [18] gives numerical results, Figure 12.10 and 12.11 of [18], illustrating superior statistical efficiency of DJS over PDS for Monte Carlo estimation of CVA when the counterparty's default is modeled by a structural-copula method.

the counterparty credit risk modeler uses DJS for path independent derivatives. DJS-based estimators usually have larger computing times for path dependent derivatives, and so the computing time should be considered in parallel with the variance. The useful criterion that does this is '*variance per replication × expected computing time*' as introduced in the previous section. There are widely traded path dependent derivatives for which PDS and DJS-based estimators of EPE and CVA perform approximately equally. For instance, for arithmetic Asian options the DJS and PDS-based estimators of EPE and CVA perform similarly.

There are contracts whose payoff function does not exactly match the mathematical conditions of Proposition 1 and 2. For those contracts, a small simulation study could compare the variance of the DJS and PDS-based estimators of EPE and CVA. The Appendix contains numerical examples for EPE estimation of a single interest rate swap, where we conclude that DJS outperforms PDS by at least an order of 10 in terms of variance while the computing times of both Monte Carlo estimators are roughly equal.

Hereafter, we assume that the credit exposure process V satisfies $cov(V_u, V_s) = 0$ and $cov(V_u, V_s) > 0$ when simulated under the DJS and PDS methods, respectively, for any $0 < u < t$.

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 $\mathrm{MSE}(\hat{\theta}_{m,n,k})$,

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

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, $cov(V_i, V_j) > 0$ and $cov(V_i, V_j) = 0$, 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 EPE, $\hat{\theta}_{m,n,k}$. Note that from basic results on endpoint Reimann sum approximation of integrals, 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 $t \in [0, T]$. First, we note that

$$\mathrm{Var}(\hat{\theta}_{m,n,d}) = O(\frac{1}{mn}). \tag{13}$$

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

---

[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 $x, y \to \infty$.

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

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

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

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 $O(\frac{1}{m})$. This gives,

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

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 (14) leads to the following optimization problems,

$$\min_{m,n} \left(\frac{c_{p,1}}{mn} + \frac{c_{p,2}}{m} + \frac{c_2}{n^2} \right) \quad \text{subject to} \quad s = c_3 mn, \tag{15}$$

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

$$m = c s^{\frac{2}{3}} \quad \text{and} \quad n = \tilde{c} s^{\frac{1}{3}}, \tag{16}$$

for constants c and $\tilde{c}$.

DJS-Based Biased Efficient Estimator of EPE Let c_d denote a constant. Given (13), 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 $n = \hat{c} s$ for 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.

14

Remark We do not claim originality in setting up an MSE minimization problem to derive an optimum balance between variance and bias squared; this can be seen in Chapter 6 of Glasserman [13] and the references there, particularly the paper by Duffie and Glynn [7]. Our contribution is that in our proposed efficient Monte Carlo CCR framework, choosing approximately optimal m and n via solving MSE minimization problems achieve substantial MSE reduction in Monte Carlo estimation of EPE, CVA, (and as will be seen in Section 5), and EEPE. This has neither appeared in the CCR literature nor been applied by practitioners. Moreover, our result that the efficient DJS-based estimator requires all its computational budget allocated to the number of valuation points is surprising. Consider a well defined continuous time stochastic process S whose finite dimensional distributions satisfy $cov(S_u, S_t) > 0$ for any $0 < u < t$. Suppose that Monte Carlo is to be used to estimate $\theta = E[\int_0^T S_t dt]$ for a given $T > 0$. The porposed efficient Monte Carlo estimator of θ employs the notion of marginal matching (as opposed to simulating the process based on the finite dimensional distribution of S) and uses 1 simulation run at each time point in a discrete time grid given a fixed computational budget all allocated to making the grid as fine as possible.

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,

$$\text{EPE} = TE[V_\tau], \tag{17}$$

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,

$$\text{EPE} = TE[V_\tau] = T \sum_{i=1}^n E[V_\tau|\tau \in A_i]p_i = \sum_{i=1}^n E[V_\tau|\tau \in A_i]\Delta_i, \tag{18}$$

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 $\Delta_i = t_i - t_{i-1}$. Assuming $t_i = iT/n$ for all $i = 1, ..., n$, our proposed unbiased estimators of EPE use the identity (18) by estimating the conditional expectations, $E[V_\tau|\tau \in A_i]$, i.e.,

$$\hat{\theta}_{u,m,n,k} = \sum_{i=1}^n \bar{V}_{\tau_i}\Delta_i, \tag{19}$$

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.[10]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 $m_i = N p_i$, (see [26] 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.

Comparing DJS-based Biased and Unbiased estimators Proposition 3 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.

Proposition 3. *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,*

$$\hat{\theta}_{b,m,n,d} = \sum_{i=1}^{n} \bar{V}_i \Delta_i, \quad and \quad \hat{\theta}_{u,m,n,d} = \sum_{i=1}^{n} \bar{V}_{\tau_i} \Delta_i. \tag{20}$$

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 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,

$$\lim_{n \to \infty} n MSE(\hat{\theta}_{b,m,n,d}) = n Var(\hat{\theta}_{u,m,n,d}) = c \int_0^T Var(V_t) dt, \tag{21}$$

where c is a constant.

---

[10]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.

Comparing PDS-based Biased and Unbiased estimators Analytically comparing the $\mathrm{MSE}(\hat{\theta}_{b,m,n,p})$ and $\mathrm{Var}(\hat{\theta}_{u,m,n,p})$ is quite difficult due to the presence of the covariance terms. 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.

4   Efficient Monte Carlo Estimation of Independent CVA

Independent CVA can be viewed as the weighted sum of expected exposures with the weights being default probabilities. Therefore, our results from Section 3 on efficient estimation of EPE immediately apply here (note that for EPE, the weights are subinterval lengths). To summarize 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,

$$\mathrm{CVA}_I = E\left[E[V_\tau \mathbf{1}\{\tau \leq T\}|\tau]\right] = \int_0^T E[V_t]dF_t, \tag{22}$$

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, e.g., [19]). The Remark at the end of this section discusses Monte Carlo CVA estimation with stochastic discounting. Appendix I discusses PDS and DJS-based estimation of CVA sensitivities with respect to the initial values of underlying risk factors using finite-difference approximations.

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

$$\xi_{b,k} = \sum_{i=1}^n \bar{V}_i \Delta F_i, \tag{23}$$

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

$$E[V_\tau \mathbf{1}\{\tau \leq T\}] = \sum_{i=1}^n E[V_\tau|\tau \in A_i]P(\tau \in A_i), \tag{24}$$

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 (24), the stratified sampling estimator of CVA_I is

17

$$\xi_{u,k} = \sum_{i=1}^{n} \bar{V}_{\tau_i} p_i, \tag{25}$$

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.

As mentioned before, proportional stratified sampling sets $m_i = NP(\tau \in A_i)$. For CVA$_I$ estimation, even if we assume A_i's to be equidistant strata, $P(\tau \in A_i)$'s are not equal in general. Therefore, proportional stratified sampling does not lead to an equal number of simulation runs at all the valuation points, as is the case for EPE estimation (see Section 3.3).

We have empirically observed that $\xi_{u,p}$ outperforms $\xi_{b,p}$ in terms of mean square error.[11] Proposition 4 below shows an asymptotic equivalence between $\xi_{b,d}$ and $\xi_{u,d}$. The proof of Proposition 4 is similar to Proposition 3, and so it is omitted. Also, similar to our numerical examples to be shown in Section 6.1, we have observed that DJS-based biased and unbiased estimators of CVA$_I$ are equivalent in terms of MSE for large n, which confirms the result of Proposition 4 below.

Proposition 4. *Consider the proposed estimators of CVA$_I$, $\xi_{b,d}$ and $\xi_{u,d}$ as defined in (23) and (25), respectively. Suppose that proportional sampling is used for both estimators, i.e., $m_i = Np_i$, and $\sum_{i=1}^{n} m_i = N$, $i = 1, ..., n$. 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,*

$$\lim_{n \to \infty} n \, Var(\xi_{u,d}) = n MSE(\xi_{b,d}) = c \int_0^T Var(V_t) dF(t) \tag{26}$$

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.

Similar to our proposed efficient EPE estimation framework, when the DJS method is chosen for sampling from the credit exposure process, we recommend using efficient biased CVA estimation where the approximately optimal m and n are derived via solving the MSE minimization problem discussed in Section 3.2. Note that the surprising approximately optimal solution sets $m = 1$ and allocates the total computational budget to the number of valuation points. When the PDS method simulates the credit exposure process, we suggest our unbiased stratified sampling based estimator of CVA$_I$, where the number of strata and simulation runs at each stratum are specified based on the solution to the MSE minimization problem as in Section 3.2.

---

[11]Assuming that τ is an exponential random variable, the results of our numerical examples for $\xi_{u,p}$ and $\xi_{b,p}$ are similar to those to be shown in Section 6.1, and so are omitted.

Remark This remark clarifies and discusses the potential application of the DJS method in Monte Carlo estimation of

$$\text{CVA}_I = \int_0^T E[D_t V_t] dF_t,$$

in the presence of stochastic discounting.[12] Let $D_t = e^{-\int_0^t r_u du}$, $0 < t \leq T$, and assume the risk neutral dynamics of the short rate is specified by a widely used short rate model, (see, e.g., Chapter 21 of [2]). Also, suppose that the valuation points of the Monte Carlo CCR "engine" are specified by the equidistant time grid $0 \equiv t_0 < t_1 < ... < t_n \equiv T$; $t_i - t_{i-1} = \Delta$ and that the time integral of the short rate in D is time-discretized on a finer grid $0 \equiv \tilde{t}_0 < \tilde{t}_1 < ... < \tilde{t}_{\tilde{n}} \equiv T$; $\tilde{t}_j - \tilde{t}_{j-1} = \tilde{\Delta} < \Delta$. Assume that $\Delta = c\tilde{\Delta}$; c being a positive integer. The DJS-based estimator of CVA_I seeks to make $D_i V_i$'s independent at any two distinct time points on the gird. Suppose that the $V_i \equiv V_{t_i}$ is a function of $S_1, ..., S_i$, which are dependent random variables, $i = 1, ..., n$. Note that $\{r_i \equiv r_{\tilde{t}_j}; j = 1, .., \tilde{n}\}$ represents the finite dimensional distribution of r on the finer grid. In the DJS-based single-simulation-run estimator of CVA_I, we are to have $cov(D_i V_i, D_j V_j) = 0$ for any $i \neq j$, $i, j = 1, .., n$. Achieving this zero correlation requires additional computational time due to the generation of additional random variables. To see this, consider two distinct valuation points on the grid, $t_k < t_l$. A single Monte Carlo realization of $D_k V_k$ requires generating $(r_1, r_2, ..., r_{ck})$ and $(S_1, ..., S_k)$. To sample from $D_l V_l$, the DJS method, instead of using the Monte Carlo realizations of the risk factor in the simulation of $D_k V_k$, starts anew and generates $(r_1, r_2, ..., r_{ck}, ..., r_{cl})$ and $(S_1, ..., S_k, ..., S_l)$, and so, $cov(D_k V_k, D_l V_l) = 0$. One should contrast this with the PDS method. In sampling from $D_l V_l$, the PDS method generates only $(r_{ck+1}, ..., r_{cl})$ and $(S_{k+1}, ..., S_l)$ along with the random variables generated previously in the simulation of $D_k V_k$. So, the DJS method is applicable in the context of Monte Carlo CVA estimation with stochastic discounting at an additional computational cost. Suppose the CCR modeler analytically, using Proposition 1 and 2 or their equivalents, or numerically, via a small simulation study, concludes that the PDS-based estimator has a higher variance, due to some positive covariance terms, $cov(D_i V_i, D_j V_j) > 0$ for $i \neq j$, $i, j = 1, .., n$, compared to that of the DJS-based estimator. Then, the DJS-based method is chosen only when the computing time has also been taken in to account. That is, the modeler chooses the DJS-based estimator of CVA_I only if it outperforms the PDS-based estimator in term of the following useful criterion,

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

as we discussed in Section 3.1.2.

5 Efficient Monte Carlo Estimation of EEPE

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

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

---

[12]We continue to assume that the recovery rate is constant, and to simplify the notation, we suppress the dependence of CVA to it.

where $\{V_t \; ; t \geq 0\}$ 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,

$$\hat{\theta}_{m,n,k} = \sum_{i=1}^{n} \max_{1 \leq j \leq i} \{\bar{V}_j\} \Delta_i, \tag{27}$$

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 $\hat{\theta}_{m,n,k}$,

$$\text{MSE}(\hat{\theta}_{m,n,k}) \;\; = \;\; \text{Var}\left(\sum_{i=1}^{n} \max_{1 \leq j \leq i} \{\bar{V}_j\} \Delta_i\right) + \left(\sum_{i=1}^{n} E[\max_{1 \leq j \leq i} \{\bar{V}_j\}]\Delta_i - \text{EEPE}\right)^2. \tag{28}$$

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

$$\left(\sum_{i=1}^{n} E[\max_{1 \leq j \leq i} \{\bar{V}_j\}]\Delta_i - \sum_{i=1}^{n} \max_{1 \leq j \leq i} E[\bar{V}_j]\Delta_i\right) - \left(\text{EEPE} - \sum_{i=1}^{n} \max_{1 \leq j \leq i} E[\bar{V}_j]\Delta_i\right). \tag{29}$$

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 5 below whose proof is in the Appendix.

Proposition 5. *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 $m \to \infty$,

$$M_{n,m,k} \to M_n \quad a.s., \tag{30}$$

where a.s. stands for almost surely.

Note that dominated convergence theorem and Proposition 5 give $E[M_{n,m,k}] \to M_n$ as $m \to \infty$.[13] So, the first part of the bias

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

---

[13]Note that $M_{n,m,k} \leq \sum_{i=1}^{n} \bar{V}_i$ and Proposition 5 assumes $E[\bar{V}_i] = E[V_i] < \infty$.

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.

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

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

where $k = p$ ($k = d$) refer to the cases where V_i and V_j for any $i \ne j$, 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 $\delta \equiv \min\{|\delta_{i,j}| : i \ne j, i,j = 1,...,n\}$. 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 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

$$\theta = \sum_{i=1}^{n} E[V_{\tau_i}]\Delta_i \quad \text{and} \quad \hat{\theta}_{m,n,k} = \sum_{i=1}^{n} \bar{V}_{\tau_{i,m,k},m,k}\Delta_i. \tag{31}$$

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

21

bounded from above as follows

$$P(\tau_{i,m,k} \neq \tau_i) \leq \sum_{j \neq \tau_i, j=1,\ldots,i} P(\bar{V}_{\tau_i,m,k} - \bar{V}_{j,m,k} < 0)$$

$$= \sum_{j \neq \tau_i, j=1,\ldots,i} P(\bar{V}_{\tau_i,m,k} - \bar{V}_{j,m,k} - \delta_{\tau_i,j} < -\delta_{\tau_i,j})$$

$$< \sum_{j \neq \tau_i, j=1,\ldots,i} P(\bar{V}_{\tau_i,m,k} - \bar{V}_{j,m,k} - \delta_{\tau_i,j} < -\delta)$$

$$< \sum_{j \neq \tau_i, j=1,\ldots,i} P(|\bar{V}_{\tau_i,m,k} - \bar{V}_{j,m,k} - \delta_{\tau_i,j}| > \delta)$$

$$\leq \sum_{j \neq \tau_i, j=1,\ldots,i} \frac{\sigma^2_{\tau_i,j,k}}{m\delta^2} \tag{32}$$

$$\leq \frac{(i-1) \cdot \sigma^2_{\max}}{m\delta^2},$$

where (32) follows from the Chebyshev's inequality.

Consider the event $B_m = \{\tau_i = \tau_{i,m,d} = \tau_{i,m,p}, \text{for all } i = 1,\ldots,n\}$. 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

$$\mathrm{Var}(\hat{\theta}_{m,n,d,B_m}) < \mathrm{Var}(\hat{\theta}_{m,n,p,B_m}). \tag{33}$$

This order can be established by first noting that

$$\hat{\theta}_{m,n,k,B_m} = \sum_{i=1}^{n} \bar{V}_{\tau_i,m,k} \Delta_i. \tag{34}$$

Then since V_i and V_j, for any $i \neq j$, are positively correlated (uncorrelated) under PDS (DJS) sampling, the variance of expression (34) is lower under DJS than under PDS.

Note that:

$$P(B_m^c) \leq \sum_{i=2}^{n} P(\tau_{i,m,d} \neq \tau_i) + \sum_{i=2}^{n} P(\tau_{i,m,p} \neq \tau_i) \tag{35}$$

$$< 2 \sum_{i=2}^{n} \frac{(i-1) \cdot \sigma^2_{\max}}{m\delta^2}, \tag{36}$$

The above argument leads to the following result.

Proposition 6. *Consider the desirable event B_m as defined above. First, conditional on this event, (33) 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 6 suggests that for sufficiently large m, $\mathrm{Var}(\hat{\theta}_{m,n,d}) \leq \mathrm{Var}(\hat{\theta}_{m,n,p})$. Our numerical examples in Section 6.2 use $m > 400$; they all indicate that $\mathrm{Var}(\hat{\theta}_{m,n,d}) \leq \mathrm{Var}(\hat{\theta}_{m,n,p})$.

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,

$$\text{Var}(\hat{\theta}_{m,n,k}) \approx \frac{c_{1,k}}{mn} + \frac{c_{2,k}}{m} \quad \text{and} \quad \text{Bias}(\hat{\theta}_{m,n,k}) \approx \frac{c_{3,k}}{m} + \frac{c_3}{n}, \tag{37}$$

for some constants $c_{1,k}, c_{2,k}, c_3$. The above approximation of the order of bias uses (29) and Proposition 5. Note that our rough approximate of the order of variance, applicable to both $\hat{\theta}_{m,n,d}$ and $\hat{\theta}_{m,n,p}$, is similar to (14). 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 (14).

Given (37), we recommend solving the following MSE minimization problem to specify the approximately optimal m and n,

$$\min_{m,n} \left(\frac{c_1}{mn} + \frac{c_2}{m} + (\frac{c_3}{m} + \frac{c_4}{n})^2\right) \quad \text{subject to} \quad s = cmn, \tag{38}$$

for some constants c_1, c_2, c_3, c_4, and c.

6   Numerical Examples

6.1   EPE estimation

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 $V_t \equiv S_t$ denotes the value of a geometric Brownian motion at time $t > 0$. That is, $S_t = S_0 e^{X_t}$ with $\{X_t; \ t \geq 0\}$ being a Brownian motion with drift μ, and volatility σ. This stylized example 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,

$$\hat{\theta}_{c,k} = \sum_{i=1}^{n} \bar{V}_i \Delta_i, \tag{39}$$

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,

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

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. 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. 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. Therefore, the number of simulation runs at each valuation point m is 1,000 and 10,000, respectively, for the crude estimator subject to these two budget values.

To calculate $\hat{\theta}_{e,b,p}$ under these fixed computational budgets, the solution, (16) with both c and $\tilde{c}$ set to 1, to the MSE minimization problem of Section 3.2 is used. 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.

Table 1 on page 26 illustrates that our proposed estimators of EPE lead to substantial MSE reduction when compared to the "crude" Monte Carlo estimators for a variety of parameter settings. 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 the third and fourth sections of Table 1). 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 3 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.

6.2 EEPE estimation

Our numerical examples presented below illustrate the efficiency of our proposed estimators of EEPE.[14] As in Section 6.1, we consider the geometric-Brownian-motion stylized example, where $V_t \equiv S_t = S_0 e^{X_t}$ with X being a 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,

$$\hat{\theta}_{c,k} = \sum_{i=1}^{n} \max_{1 \le j \le i} \bar{V}_j \Delta_i, \tag{41}$$

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. Also, the number of simulation runs $n = 1,000$ and $10,000$ respectively, for $s = 12,000$ and $120,000$.

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 (41) with $\Delta_i \equiv \Delta = T/n$) and resulting from solving the MSE minimization problem (38) (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, Table 2 on page 27, 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 the fourth section of Table 2.

7 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. Poor Monte

---

[14]We refer the reader to Section (F) of the Appendix for a discussion on $EEPE_{dst}$ and numerical illustrations of Propositions 5 and 6.

	EPE	Variance	MSE	CPU Time
Parameters: $S_0 = 30, \mu = .2, \sigma = .3, s = 12,000$				
$\hat{\theta}_{c,p}$	34.6559	0.047219	0.48478	0.00380
$\hat{\theta}_{c,d}$	34.6522	0.005028	0.43768	0.00162
$\hat{\theta}_{e,b,p}$	34.1802	0.077212	0.1117	0.00253
$\hat{\theta}_{e,b,d}$	33.9955	0.004785	0.004786	0.00174
$\hat{\theta}_{u,p}$	33.9964	0.072068	0.072064	0.00518
$\hat{\theta}_{u,d}$	33.9956	0.004865	0.004866	0.00335
Parameters: $S_0 = 30, \mu = .2, \sigma = .3, s = 120,000$				
$\hat{\theta}_{c,p}$	34.652	0.004791	0.4372	0.03887
$\hat{\theta}_{c,d}$	34.6521	0.000501	0.43303	0.01564
$\hat{\theta}_{e,b,p}$	34.0798	0.016741	0.024026	0.02299
$\hat{\theta}_{e,b,d}$	33.9948	0.000483	0.000483	0.02409
$\hat{\theta}_{u,p}$	33.9957	0.015533	0.015533	0.04420
$\hat{\theta}_{u,d}$	33.9945	0.000486	0.000486	0.03426
Parameters: $S_0 = 30, \mu = 1, \sigma = .3, s = 12,000$				
$\hat{\theta}_{c,p}$	57.7556	0.16106	23.5389	0.00389
$\hat{\theta}_{c,d}$	57.7598	0.01628	23.4351	0.00189
$\hat{\theta}_{e,b,p}$	54.1296	0.23369	1.6954	0.00270
$\hat{\theta}_{e,b,d}$	52.9238	0.015853	0.015862	0.00189
$\hat{\theta}_{u,p}$	52.9226	0.217	0.21698	0.00516
$\hat{\theta}_{u,d}$	52.9198	0.015796	0.015796	0.00390
Parameters: $S_0 = 30, \mu = 1, \sigma = .3, s = 120,000$				
$\hat{\theta}_{c,p}$	57.7579	0.016112	23.4159	0.03891
$\hat{\theta}_{c,d}$	57.7591	0.001616	23.4136	0.01661
$\hat{\theta}_{e,b,p}$	53.4783	0.047841	0.35899	0.02412
$\hat{\theta}_{e,b,d}$	52.9212	0.001563	0.001564	0.02627
$\hat{\theta}_{u,p}$	52.9189	0.045783	0.045781	0.04657
$\hat{\theta}_{u,d}$	52.9203	0.001565	0.001565	0.03598

Table 1: Monte Carlo EPE estimates for different parameter settings

	EEPE	Variance	MSE	CPU Time
Parameters: $S_0 = 30, \mu = 1, \sigma = .25,\ s = 12,000$				
$\hat{\theta}_{c,p}$	57.2278	0.10931	22.4936	0.00259
$\hat{\theta}_{c,d}$	57.2233	0.034659	22.3768	0.00168
$\hat{\theta}_{e,p}$	53.4344	0.19358	1.0731	0.00191
$\hat{\theta}_{e,d}$	53.4379	0.011188	0.89722	0.00211
Parameters: $S_0 = 30, \mu = 1, \sigma = .25,\ s = 120,000$				
$\hat{\theta}_{c,p}$	57.2277	0.010824	22.3945	0.02866
$\hat{\theta}_{c,d}$	57.2262	0.003591	22.3734	0.01427
$\hat{\theta}_{e,p}$	52.9363	0.03962	0.23301	0.01817
$\hat{\theta}_{e,d}$	52.9354	0.001083	0.19367	0.01545
Parameters: $S_0 = 30, \mu = 1.5, \sigma = .25,\ s = 12,000$				
$\hat{\theta}_{c,p}$	81.0388	.24286	101.0233	0.00279
$\hat{\theta}_{c,d}$	81.0309	0.0843	100.7055	0.00173
$\hat{\theta}_{e,p}$	72.8899	0.3986	3.9705	0.00221
$\hat{\theta}_{e,d}$	72.8885	0.024652	3.5914	0.00226
Parameters: $S_0 = 30, \mu = 1.5, \sigma = .25,\ s = 120,000$				
$\hat{\theta}_{c,p}$	81.0332	0.024156	100.692	0.02929
$\hat{\theta}_{c,d}$	81.0302	0.008395	100.6154	0.0144
$\hat{\theta}_{e,p}$	71.8779	0.083579	0.85454	0.01935
$\hat{\theta}_{e,d}$	71.8807	0.002425	0.77819	0.01630

Table 2: Monte Carlo EEPE estimates for different parameter settings

Carlo CCR estimation can lead to overestimation or underestimation of CCR risk. 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.

Our proposed efficient framework can be summarized in two steps. The counterparty credit risk modeler first needs to choose between the two credit exposure sampling methods: PDS or DJS. Introducing and using the notion of marginal matching, we identify conditions under which 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 since the computational time of the DJS-based estimator usually exceeds that of the PDS-based estimator, the two sampling methods could become approximately equivalent in some cases.

Next, in the second step, the CCR modeler needs to choose the number of valuation points

and simulation runs at each valuation point for efficient EPE, CVA_I, and EEPE estimations. 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. These MSE minimization problems can be easily solved after approximate orders of variance and bias for each of the above mentioned CCR measures under the PDS and DJS methods are derived.

For efficient EPE and CVA_I estimation, if the PDS method has been chosen in the first step above, we recommend employing our unbiased stratified sampling-based estimators of EPE and CVA_I. These unbiased estimators of EPE and CVA_I use stratified sampling with the number of strata and simulation runs (allocated to each stratum) being chosen based on the solution to the above mentioned MSE minimization problems.

Our numerical examples indicate that unbiased PDS-based estimators of EPE and CVA are preferable to the efficient biased PDS-based estimators of EPE and CVA. Our analytical results suggest that unbiased and biased DJS-based estimators of EPE and CVA are asymptotically equivalent. Our numerical examples confirm this approximate asymptotic equivalence.

Finally, an interesting case arises when the CCR modeler chooses the DJS method for estimating EPE and CVA_I. In this case, our proposed efficient EPE and CVA_I estimators use 1 simulation run at each valuation point and the total computational budget is allocated to making the discrete time grid (the set of valuation points) as fine as possible. Our various numerical examples illustrate that employing this two-step Monte Carlo CCR framework will substantially increase the efficiency of the existing Monte Carlo CCR "engines".

Appendix

A   Proof of Proposition 1

We first show that $\text{cov}(V_u, V_t) > 0$ holds under condition 1. That is, $V_t = C_t = B_t E[B_T^{-1}\Pi_T|\mathcal{F}_t]$. For any $0 < u < t$ we have

$$\text{cov}(V_u, V_t) = E\left[\text{cov}(V_u, V_t|\mathcal{F}_u)\right] + \text{cov}\left(V_u, E[V_t|\mathcal{F}_u]\right),$$

where the last equality follows from the conditional covariance formula (see Chapter 3 of [25]). It is easy to check that the first term on the right hand side above is zero. Consider the second term and note that

$$E[V_t|\mathcal{F}_u] = E\left[B_t E[\frac{\Pi_T}{B_T}|\mathcal{F}_t]|\mathcal{F}_u\right] = E\left[B_t \frac{\Pi_T}{B_T}|\mathcal{F}_u\right],$$

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

$$\text{cov}(V_u, V_t) = B_u B_t B_T^{-2}\text{Var}\left(E[\Pi_T|\mathcal{F}_u]\right) > 0.$$

The second part of the proof, which is based on condition 2, uses standard results on changes of numeraire techniques and Chebyshev's algebraic inequality (see, for instance, Proposition 2.1 in [9]).

Let C_t denote the time-t value of a derivative contract specified in Assumption 2. Recall Theorem 1 and Theorem 2 of [10], and note that

$$V_t = C_t = S_1(t)E_{Q^1}[(\alpha_1 + \alpha_2 Z(T))^+|\mathcal{F}_t]$$

where $Z = S_2/S_1$ and the subscript Q^1 refers to expectation under Q^{S_1}, i.e. S_1 is the numeraire. Note that

$$
\begin{aligned}
\mathrm{cov}(V_u, V_t) &= \mathrm{cov}\left(V_u, E[V_t|\mathcal{F}_u]\right) \\
&= \mathrm{cov}\left(S_1(u)E_{Q^1}\left[(\alpha_1 + \alpha_2 Z(T))^+|\mathcal{F}_u\right], E_{Q^1}\left[S_1(t)(\alpha_1 + \alpha_2 Z(T))^+|\mathcal{F}_u\right]\right),
\end{aligned}
$$

Consider the second term on the right hand side above, and suppose that the transition law of S_1 accepts the following specification

$$S_1(t) =^d \beta S_1(u)S_1(\tau), \tag{42}$$

where $\beta > 0$ is a constant and $t - u \equiv \tau$ (see the last part of the proof on transition law of the numeraire). We, then, have

$$E_{Q^1}\left[S_1(t)(\alpha_1 + \alpha_2 Z(T))^+|\mathcal{F}_u\right] = \beta S_1(u)E_{Q^1}\left[S_1(\tau)(\alpha_1 + \alpha_2 Z(T))^+|\mathcal{F}_u\right].$$

This gives

$$\mathrm{cov}(V_u, V_t) = \beta S_1^2(u)\mathrm{cov}\left(E_{Q^T}\left[(\alpha_1 + \alpha_2 Z(T))^+|\mathcal{F}_u\right], E_{Q^T}\left[S_1(\tau)(\alpha_1 + \alpha_2 Z(T))^+|\mathcal{F}_u\right]\right).$$

Note that both expectations above are monotone functions of $Z(u)$ and so using Chebyshev's algebraic inequality (see, for instance, Proposition 2.1 in [9]) gives $\mathrm{cov}(V_u, V_t) > 0$.

We now show that (42) holds for the class of stochastic processes considered in this paper for modeling the dynamics of risk factors. The dynamics of risky asset $S_1 > 0$ selected as the numeraire in the proof above is assumed to be modeled by a GBM or a square-root diffusion. In cases where the numeraire is a maturity-T zero-coupon bond, the second part of the proof above is modified as follows. We assume that the zero-coupon bond is modeled such that it possesses an affine term structure, i.e. it has the form $S_1(t) \equiv p(t, T) = e^{A-Br_t}$, where $A \equiv A(t, T)$ and $B \equiv B(t, T)$ are deterministic functions and the short rate r is modeled by a Gauss-Markov process or a square-root diffusion as specified before. It is, then, not difficult to see that $p(t, T) = e^{A-Br_t} =^d e^{\beta_1 r_u}e^{\beta_2 r_\tau}$, where β_1 and β_2 are constants and $0 < u < t < T$, $t - u \equiv \tau$. Using a similar approach shown in the second part of the proof above, we arrive at $\mathrm{cov}(V_u, V_t) > 0$.

B Proof of Proposition 2

Conditioning on $\mathcal{F}_u^S \equiv \mathcal{F}_u$ and using conditional covariance formula gives

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

for a well-defined function f. 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 $E[\max\{f(S_t), 0\}|\mathcal{F}_u]$.

Note that when S is a Gauss-Markov process, the transition law of S implies that for any $0 < u < t$, $S_t =^d \beta_1 S_u + \beta_2 S_{t-u}$, where S_u and S_{t-u} are independent random variables. Also, when S is a GBM, for any $0 < u < t$ we have $\log(S_t) =^d \log(S_u) + \log(S_{t-u})$, where S_u and S_{t-u} are independent random variables. This follows from the independent and stationary increments properties of Gauss-Markov processes[15] and that their finite dimensional distributions are multivariate normal. When S is a square-root diffusion, $dS_t = \alpha(b - S_t)dt + \sigma\sqrt{S_t}dB_t$ with B a standard one-dimensional Brownian motion and positive constants α and b, it can be shown that S_t given S_u is distributed as a positive constant times times a noncentral chi-square random variable with degrees of freedom that depends on α, σ, and b, and noncentrality parameter which is an increasing function of S_u, (see Chapter 3 of [13]).

This implies that under the class of risk factor models considered in the paper, $E[\max\{f(S_t), 0\}|\mathcal{F}_u]$ 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; this increases $\max\{f(S_t), 0\}$. So, $E[\max\{f(S_t), 0\}|\mathcal{F}_u] \equiv \tilde{h}(S_u)$ 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 gives $\mathrm{cov}(V_u, V_t) = \mathrm{cov}(\tilde{f}(S_u), \tilde{h}(S_u)) > 0$.

C Proof of Proposition 3

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,

$$n\mathrm{MSE}(\hat{\theta}_{b,d}) = \frac{T}{m}\sum_{i=1}^{n}\mathrm{Var}(V_i)\Delta_i + n(\sum_{i=1}^{n}E[V_i]\Delta_i - \int_0^T E[V_t]dt)^2, \tag{43}$$

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

Now, consider $\mathrm{Var}(\hat{\theta}_{u,d})$, 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,

---

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

$$\mathrm{Var}(\hat{\theta}_{u,d}) = \frac{T^2}{mn} \sum_{i=1}^{n} \mathrm{Var}(V_\tau | \tau \in A_i) p_i = \frac{T^2}{mn} E[\mathrm{Var}(V_\tau | I_n)]. \tag{44}$$

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

$$E[\mathrm{Var}(V_\tau | I_n)] \longrightarrow E[\mathrm{Var}(V_\tau | \tau)]. \tag{45}$$

From the formula for the conditional variance, to show the convergence in (45), it suffice to show that, as $n \to \infty$,

$$E\left[(E[V_\tau | I_n])^2\right] \longrightarrow E\left[(E[V_\tau | \tau])^2\right]. \tag{46}$$

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 [8]) 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 [8]) we conclude that, $E[X_n^2]$ converges to $E[X^2]$ almost surely, and so (46) holds. This completes the proof of Proposition 3. [16]

D   Proof of Proposition 5

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

$$\max\{\bar{V}_1, \bar{V}_2\} - E[V_2] = \bar{V}_1 1\{\bar{V}_1 > \bar{V}_2\} + (\bar{V}_2 - E[V_2])1\{\bar{V}_2 > \bar{V}_1\} - E[V_2]1\{\bar{V}_1 > \bar{V}_2\}. \tag{47}$$

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 $\epsilon > 0$,

$$P(1\{\bar{V}_1 > \bar{V}_2\} > \epsilon \text{ i.o.}) = 0, \tag{48}$$

where i.o. stands for infinitely often. To see that (48) holds, note that,

$$P(1\{\bar{V}_1 > \bar{V}_2\} > \epsilon) \leq P(1\{\bar{W}^k > E[W^k]\} > \epsilon) \leq \frac{P(|\bar{W}^k - E[W^k]| > \tilde{\epsilon})}{\epsilon^2}, \tag{49}$$

for all $\tilde{\epsilon} > 0$. To derive the last inequality above the Chebyshev's inequality is used. Then, (49), 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 [8]), give (48).

---

[16]The probabilistic arguments used in second part of the proof are similar to the ones used in the proof of Lemma 4.1 in [14].

	EPE	Variance	CPU Time
Parameters: $a = 0.01, b = 0.0004, \sigma = 0.3, r_0 = 0.02$			
$\hat{\theta}_{b,m,n,p}$	36.0561	0.001335	0.00953
$\hat{\theta}_{b,m,n,d}$	36.0553	0.000196	0.01018
Parameters: $a = 0.01, b = 0.0002, \sigma = 0.3, r_0 = 0.04$			
$\hat{\theta}_{b,m,n,p}$	35.9658	0.001416	0.01012
$\hat{\theta}_{b,m,n,d}$	35.9638	0.000206	0.01146
Parameters: $a = 0.01, b = 0.0003, \sigma = 0.4, r_0 = 0.03$			
$\hat{\theta}_{b,m,n,p}$	37.009	0.002546	0.00909
$\hat{\theta}_{b,m,n,d}$	37.012	0.000380	0.01025

Table 3: Monte Carlo EPE estimates for a single interest rate swap with different parameter settings

Now, consider the first term on the right side of (47). 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

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

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 (47) converges to zero a.s. based on (48). Analogous arguments led to (48) show that the second term on the right side of (47) 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.

E   Numerical Examples for Interest Rate Swaps

Table 3 on the preceding page above compares the variance and computing time of the DJS and PDS-based Monte Carlo estimators of EPE for a single interest rate swap contract in the Vasicek short rate framework (see, for instance, [2]). Specifically, the short rate, denoted by r, is modeled by $dr_t = (b - ar_t)dt + \sigma dB_t$, where B is a standard Brownian motion and $a > 0$. Recall that standard results from affine term structure modeling expresses the time-t value of a T-maturity zero coupon bond as $p(t, T) = e^{A(t,T) - B(t,T)r_t}$, where A and B are given deterministic functions of time, a, b, and σ. The interest rate swap in our examples is a forward swap settled in arrears (see, for instance, [2]) with quarterly payments, a principal value of 100, and maturity 1 year. We let $m = 10^4$ and $n = 12$. In particular, the 12 valuation points are equally spaced, i.e., monthly, within the one-year interval. Variances of the DJS and PDS-based estimators $\hat{\theta}_{b,m,n,d}$ and $\hat{\theta}_{b,m,n,p}$ (column 2 of Table 3 on the previous page) are estimated based on 1000 simulation runs. Table 3 presents instances of our numerical examples for interest rate swap where the DJS-based estimator of EPE outperforms the PDS-based estimator in terms of variance (by an order of at least 10) while the computing times are approximately equal.

F   Numerical Examples for EEPE$_{dst}$ estimation

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 5 and 6.

Similar to our numerical examples of Sections 3 and 5, we set $V \equiv S$ with S being 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 6.2. Each estimation procedure is replicated $10,000$ times to produce the estimates.

In Table 4 on the following page, in addition to presenting the estimator value, variance, MSE (which, unlike that in Section 6.2, 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 6. As this table shows, 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 5.

G   The DJS Method

This section gives an explanatory discussion of what we have referred to as the "DJS method" throughout our study. The underlying idea of the DJS method is rather simple. Suppose that the two given random variables V_i and V_j are dependent; Monte Carlo considers the joint distribution

	EEPE_{dst}	Variance	MSE	CPU Time	WrongOrderProb
Number of simulation runs $m = 50$					
$\hat{\theta}_{c,p}$	41.0514	20.6228	20.7011	0.000329	0.9613
$\hat{\theta}_{c,d}$	42.4019	7.901	10.5697	0.000431	1
Number of simulation runs $m = 500$					
$\hat{\theta}_{c,p}$	40.7481	2.1847	2.1849	0.00156	0.2789
$\hat{\theta}_{c,d}$	40.8833	0.77441	0.78761	0.00121	0.919
Number of simulation runs $m = 5000$					
$\hat{\theta}_{c,p}$	40.7756	0.21446	0.2146	0.0173	0
$\hat{\theta}_{c,d}$	40.7708	0.07379	0.07379	0.0090	0.1987
Number of simulation runs $m = 50000$					
$\hat{\theta}_{c,p}$	40.7708	0.02156	0.02156	0.1708	0
$\hat{\theta}_{c,d}$	40.7682	0.00711	0.00711	0.0932	0.0001

Table 4: Monte Carlo EEPE_{dst} estimates for based on increasing number of simulation runs

function of these two random variables to generate copies of V_i and V_j. This is what the paper refers to as the "PDS method". Since the functional form of the CCR measures considered in our study depends only on the marginal distributions of V_i and V_j as characterized by the notion of Marginal Matching, the dependence structure of V_i and V_j, specified by their joint distribution function, need not be preserved when Monte Carlo is used to estimate these classes of CCR measures. That is, V_i and V_j can be sampled from independently and so $cov(V_i, V_j) = 0$. This is what the paper refers to as the "DJS method". In the context of CCR, V_i, representing the exposure process evaluated that the i'th time point on the grid, could be a function of m dependent random variables, $S_1, ..., S_m$ and V_j could be a function of n dependent random variables $S_1, ..., S_m, ..., S_n$. In generating V_j, while the PDS method uses the same $S_1, ..., S_m$ that have been used in simulating V_i, the DJS method, starting anew, uses new copies of $S_1, ..., S_m$ that are independent from the previous ones. This is an additional computational burden, and so in cases where the DJS method proves to be statistically more efficient than the PDS method in terms of the variance of the Monte Carlo CCR estimators, (by using the analytical results of Proposition 1 and 2 or by conducting a small simulation study), computing time should also be taken in to account. The useful criterion that does this is,

$$\textit{variance per replication} \times \textit{expected computing time}.$$

This criterion, when used even as a rule of thumb, would specify whether to choose the DJS or the PDS method.

The DJS method in the CCR literature, to the best of our knowledge, often only refers to cases where the computing time of the DJS-based estimator is equivalent to that of the PDS-based estimator. The Path-Independent Case of Section 3.1.1 is the classical situation: V_i, in the above stylized example, is a function of a single random variable S_i, and V_j is a function of only S_j, where S_i and S_j are dependent with a given joint distribution function. What our study

refs to as the "DJS method", as illustrated above, would, then, be an extension of the DJS in its conventional sense. While we have not changed the well established CCR terminologies, "DJS and PDS", we invite the reader to be mindful of the distinction between the DJS in its traditional sense and in the more general sense of this paper which could require additional computing time.

H   Time-Discretized SDEs and Monte Carlo-Based Pricing

This section discusses cases that have not been formulated in the main body of the paper. We discuss Monte Carlo CCR measurement when the underlying risk factors are modeled by continuous time stochastic differential equations (SDEs) whose simulation requires time-discretization. We also discuss Monte Carlo CCR when in addition to generating the underlying risk factors up to each valuation point, Monte Carlo is also used for pricing the portfolio constituents. The potential applicability of the so called DJS method in this paper's setting is further clarified for these more general cases. Time-discretized SDEs and Monte Carlo-based pricing create additional sources of bias (for the estimators of the CCR measures) that have not been mathematically formulated in our study. This could be the subject of future research. We are to show and emphasize that the proposed framework does lead to basic Monte Carlo CCR efficiency improvements in these cases; the proposed framework assumes that the factors influencing these additional sources of bias are held constant and then specifies how the remaining factors determining the statistical efficiency and computing time of the CCR estimators should be dealt with. The simple stylized example below is used throughout this section to facilitate our discussion of the above-mentioned topics.

Example: Arithmetic Asian Options   Let $\{S_t; t \geq 0\}$ denote a geometric Brownian motion satisfying the following SDE,

$$dS_t = \mu S_t dt + \sigma S_t dB_t, \tag{50}$$

where $\{B_t; t \geq 0\}$ denotes a one-dimensional standard Brownian motion, and the drift μ and volatility σ parameters are given constants. Consider the equi-distant time grid $0 \equiv t_0 < t_1 < ... < t_n \equiv T$, $t_i - t_{i-1} \equiv h$, $i = 1, ..., n$. Let V_t denote the time-$t \geq 0$ value of an arithmetic Asian option with time-T payoff $(\bar{S} - K)^+$ for a given strike K, where $\bar{S} = (S_1 + ... + S_n)/n$ amd $S_i \equiv S_{t_i}$. For simplicity assume that risk-free rate is zero, and so

$$V_i = E[(\bar{S} - K)^+ | F_i],$$

where $F_i = F_{t_i}$ denote the filtration generated by the GBM up to time t_i. We are to use Monte Carlo to generate V_i's and estimate $\theta = \int_0^T E[V_t]dt$.

Time-Discretized SDEs   Suppose that we cannot analytically solve the SDE in (50). That is, suppose that the GBM could not have been simulated exactly (the joint distribution of the simulated values would not coincide with that of the continuous time model at the simulated

dates on the time grid), and an Euler time-discretization scheme would have to be used to sample from the SDE. Let $\hat{S}$ denote a time-discretized approximation to S on the above mentioned time grid,[17]

$$\hat{S}_{i+1} = \hat{S}_i + \mu\hat{S}_i h + \sigma\hat{S}_i \sqrt{h}Z_{i+1}, \tag{51}$$

with $Z_1, Z_2, ...$ independent standard normal random variables. Consider t_j and t_l on the time grid where $0 < t_j < t_l < T$. Generating $\hat{S}_j$ with the Euler scheme using (51) requires j i.i.d standard normals $Z_1, ..., Z_j$. Now, consider the simulation of $\hat{S}_l$. Consider two cases. If the set of normals $Z_1, ..., Z_j$ generated in calculating $\hat{S}_j$ are used again along with "new" i.i.d. standard normal random variables $Z_{j+1}, ...Z_l$ to generate $\hat{S}_l$, then $\hat{S}_j$ and $\hat{S}_l$ become positively correlated according to (51). If, however, l new i.i.d. standard normal random variables $Z_1, ..., Z_l$ are generated to simulate $\hat{S}_l$, then, clearly, $\hat{S}_j$ and $\hat{S}_l$ become independent. The price of achieving, $cov(\hat{S}_j, \hat{S}_l) = 0$, is the additional computational time to generate j new standard normal random variable instead of using the old ones that have already been generated when simulating $\hat{S}_j$.

Time-Discretized SDEs and the DJS Method Continuing to assume the Euler time-discretization is used to sample from the underlying GBM, we now return to the Asian option example. View V_i as a function of $(S_1, ..., S_i)$, i.e., $V_i = E[(\bar{S} - K)^+|F_i] = f(S_1, ..., S_i)$. Assume that given F_i, the conditional expectation $E[(\bar{S} - K)^+|F_i]$ can be calculated analytically, (later in this section, we discuss the implications of relaxing this assumption). Consider two Monte Carlo realizations of V_j and V_l. Similar to our analysis of the Path Dependent case of Section 3.1.2, $V_j = f(S_1, ..., S_j)$ and $V_l = f(S_1, ..., S_j, ..., S_l)$ become independent if we start anew in generating V_l. More specifically, assuming the Euler scheme in (51) is in place, if the set of $Z_1, Z_2, .., Z_l$ generated in calculating $\hat{V}_l = f(\hat{S}_1, ..., \hat{S}_l)$ are independent from the set of $Z_1, Z_2, ..., Z_j$ generated in sampling from $\hat{V}_j = f(\hat{S}_1, ..., \hat{S}_j)$, $\hat{V}_j$ and $\hat{V}_l$ become independent. We emphasize that it is only in this sense that what we continue to refer to as the DJS method is relevant to our study; that is, in sampling from the underlying risk factors in a way that V's evaluated at any two distinct time points on the grid become independent.

Recall the notion of Marginal Matching and note that in sampling from $V_j = f(S_1, ..., S_j)$, making $\hat{S}_1, ..., \hat{S}_j$ independent from each other is neither relevant to our objective - having $cov(\hat{V}_j, \hat{V}_l) = 0$ - nor possible in the sense that, depending on the functional form of f, the expected value of $\hat{V}_j$ under independent $\hat{S}_1, ..., \hat{S}_j$ may not become equal to the expected value of $\hat{V}_j = f(\hat{S}_1, ..., \hat{S}_j)$ when $\hat{S}_1, ..., \hat{S}_j$ are sequentially sampled from based on the finite dimensional distribution of the GBM time-discretized in (51). That is, our analysis and results are not affected by the inability to use a DJS type method in sampling from $V_j = f(S_1, ..., S_j)$ for a given fixed j. Again, what the DJS method, whose applicability is specified by the notion of Marginal Matching, is concerned with in our study is making Monte Carlo realizations of V_j and V_l independent - whether they have been simulated exactly or via an Euler time-discretization

---

[17]To simplify our discussion of time-discretized SDEs, we have assumed that the CCR time grid containing the valuations points and the time grid based on which the SDEs are time-discretized coincide. Relaxing this assumption does not change the results of this discussion section.

scheme.[18]

Time-Discretized SDEs and the Additional Source of Bias   Two types of bias have been formulated in the MSE minimization problems of our study: the time-discretization bias associated with the time integral of the exposure process in the definition of EPE, CVA, EEPE and also the bias induced by the maximum operator in the definition of the EEPE. The Euler-scheme-based time-discretized SDE, $\hat{S}$, which approximates S, creates an additional source of bias that has not been mathematically formulated in our study. A typical criterion based on which the quality of the SDEs' time discretization methods is measured is the convergence order of the form $|E[g(\hat{S}_i)] - E[g(S_i)]|$ for a given i with a given function g usually satisfying some smoothness conditions. See Chapter 6 of [13] and the references therein for typical weak and strong error criteria and for methods to improve the convergence order of Euler schemes based on these criteria. The simple Euler scheme (51) converges with respect to the above-mentioned criterion as the time step h decreases to zero. The proposed efficient CCR framework assumes that an Euler scheme (possibly an improved version of it by using the methods introduced in Chapter 6 of [13]) is already in place with a fixed time step h, and it then achieves an approximately optimal balance between the number of simulation runs and valuation points under the chosen Euler scheme.

Monte Carlo-Based Pricing   We now discuss the implications of Monte Carlo-based pricing in our setting. Suppose that given F_i, the conditional expectation $E[(\bar{S} - K)^+|F_i]$ could not be calculated analytically. Also, suppose that, motivated by the Longstaff and Schwartz American option pricing appraoch [21], American (or Least-Squares) Monte Carlo is to be used to estimate θ. In the m-simulation-run-based commonly-used version of American Monte Carlo for CCR, one often proceeds as follows (see [6]).[19]   First generate the m sample paths on the grid $0 \equiv t_0 < t_1 < ... < t_n \equiv T$, i.e., $\{S_{k1}, S_{k2}, ...S_{kn}\}$, $k = 1, ..., m$, based on the finite dimensional distribution of the process S. Then, all Monte Carlo estimators of $E[V_1], ..., E[V_n]$ are estimated based on this single set of sample path data.

To generate Monte Carlo realization of $V_i = E[(\bar{S} - K)^+|F_i] = f(S_1, ..., S_i)$, one first uses a regression-based method to approximate f with $\tilde{f}$ and uses all the simulated data $\{S_{k1}, S_{k2}, ...S_{kn}\}$, $k = 1, ..., m$ to estimate the regression coefficients. Then, using a given sampled path of $S_1, ..., S_i$, to generate a single approximate realization of V_i, which we denote by $\tilde{V}_i$,

---

[18]Important classes of models whose application requires time-discretization of the underlying SDEs are LIBOR and Swap Market models, (see, e.g., Chapter 25 of [2], Chapter 3 of [13], and the references therein). For instance, consider pricing Caplets in the LIBOR model where the LIBOR forward rates are GBM driven, (see Definition 25.5 of [2]), where the simulation of SDEs characterizing the underlying LIBOR forward rates (see Proposition 25.7 of [2]) require Euler time-discretization schemes, (see Section 25.5 of [2] and Section 3.7 of [13]). As illustrated above, at the cost of additional computing time compared with that of the PDS method, the DJS method can be applied in the LIBOR market models to make the simulated V's, representing the value of Caps or Caplets, independent at any two distinct time points on the CCR grid. That is, the so called DJS method is potentially applicable. However, the CCR modeler, considering both the computing time and the variance of the PDS and DJS-based estimators, may conclude that the PDS method should be chosen.

[19]We assume that the reader is familiar with American Monte Carlo option pricing and its application in counterparty credit risk.

one sets $\tilde{V}_i = \tilde{f}(S_1, ..., S_i)$. The two random variables $\tilde{V}_i = \tilde{f}(S_1, ..., S_i)$ and $V_i = f(S_1, ..., S_i)$ are not equal in distribution; it is in this sense that we refer to $\tilde{V}_i$ as an approximate realization of V_i. Note that this regression-based approximation of f with $\tilde{f}$ makes American Monte Carlo based estimators of θ biased. This bias has not been mathematically formulated in our framework.

As mentioned above, the commonly-used version of American Monte Carlo in CCR uses the same single set of simulated sample paths in estimation of $E[V_1], ..., E[V_n]$. To make American Monte Carlo estimators of $E[V_1], ..., E[V_n]$ uncorrelated, using the notion of Marginal Matching, each American Monte Carlo estimator of $E[V_i]$ should use a 'new' independent set of simulated sample paths. More specifically, fix i, and suppose that $E[V_i]$ has been estimated based on $\{S_{k1}, S_{k2}, ...S_{kn}\}$, $k = 1, ..., m$ as we summarized above. Now, using the notion of Marginal Matching (or what we continue to refer to as the DJS method), to estimate $E[V_{i+1}]$, instead of using the same set of sampled paths employed in the estimation of $E[V_i]$, a new independent set $\{S_{k1}, S_{k2}, ...S_{kn}\}$, $k = 1, ..., m$ is being generated to simulate $\tilde{V}_{i+1}$'s. So, under the DJS method, American Monte Carlo realizations of $\tilde{V}_i$ and $\tilde{V}_{i+1}$ become independent and $cov(\tilde{V}_i, \tilde{V}_{i+1}) = 0$. That is, at the cost of additional computational time, the DJS method can be used to make American Monte Carlo-based realizations of V_i's independent.[20]

Convergence Results of American Monte Carlo Algorithms It is well known that formulating and deriving precise convergence results on American option pricing algorithms, e.g., the Least Squares method of Longstaff and Schwartz, is quite difficult, (see Section 2.3 of [21] and Chapter 8 and Section 8.6 of [13]). Similar challenges exist when American Monte Carlo is used to estimate CCR measures since we need to consider limits as the number of valuation points (time points on the grid), the number of regression-based basis functions, and the number of simulation paths go to infinity. We would like to emphasize that when an American CCR "engine" is in place, even in the absence of the formulation and optimization of the MSE of the CCR estimators in the presence of all sources of bias, our framework would lead to basic efficiency improvements. It characterizes and differentiates the DJS and PDS methods by considering computational time in parallel with variance and then achieves an approximate optimal balance between the number of simulation runs (the number of sampled paths) and the number of valuation points assuming that the functional form and the number of basis functions used in the regression have been fixed.

I   Estimating CVA Sensitivities

We consider Monte Carlo estimation of the sensitivity of independent CVA with respect to the initial value of an underlying risk factor, (section 3 of [19] discusses this type of CVA sensitivity estimation with crude Monte Carlo; see Chapter 16 of [18] for a more complete overview of CVA 'Greeks' for hedging counterparty risk). We merely focus on comparing the DJS and PDS methods for CVA sensitivity estimation with respect to an underlying risk factor's initial value.

---

[20]Similar arguments hold for nested Monte Carlo. That is, the second layer of Monte Carlo used for pricing the contract at each valuation point makes the estimator of θ biased. And, the DJS method in the above mentioned sense can still be used to make V's independent at any two distinct time points on the grid.

Let S_0 denote an underlying risk factor's initial value and consider Monte Carlo estimation of $\frac{d\text{CVA}_I}{dS_0}$, where $\text{CVA}_I = \int_0^T E[V_t]dF_t$, using a finite-difference approximation method. The Monte Carlo estimator of CVA_I uses the equidistant discrete time grid, $0 \equiv t_0 < t_1 < ... < t_n \equiv T$. The central-difference estimator of $\frac{d\text{CVA}_I}{dS_0}$ based on m simulation runs at each valuation point is,

$$\hat{\theta}_k = \sum_{i=1}^{n} \hat{DV}_{i,m} \Delta F_i, \tag{52}$$

where

$$\hat{DV}_{i,m} = \frac{\bar{V}_i(S_0 + h) - \bar{V}_i(S_0 - h)}{2h}, \tag{53}$$

where $\bar{V}_i(S_0 + h)$ is the m-simulation-run average of $V_{i1}, ..., V_{im}$ with the underlying initial value at $S_0 + h$ for some $h > 0$. It is well-known that $\hat{DV}_{i,m}$ is a biased estimator of $\frac{dE[V_{t_i}]}{dS_0}$. It is also well-known that using *common random numbers* in estimating $\bar{V}_i(S_0 + h)$ and $\bar{V}_i(S_0 - h)$ will reduce the order of variance of the finite difference estimator. We assume that *common random numbers* are used. That is, suppose that V_i is a function of S_i, i.e., $V_i \equiv f(S_i)$. Then, the same set of generated $S_{i1}, ..., S_{im}$ used in the simulation of $\bar{V}_i(S_0 + h)$ are also being used in the simulation $\bar{V}_i(S_0 - h)$.[21] The subscript k in $\hat{\theta}_k$ takes d and p when the DJS and PDS methods are used, respectively.

Recall that under DJS and the notion of Marginal Matching, any two distinct pairs of $S_1, ..., S_n$ are uncorrelated, and so the variance of $\hat{\theta}_d$ is equal to the weighted sum of the variances of $\hat{DV}_{i,m} \equiv \hat{DV}_m(S_i)$, $i = 1, ..., n$, and the covariance terms, $cov(\hat{DV}_m(S_i), \hat{DV}_m(S_j))$, are zero for any $i \neq j$. So, to compare the variance of $\hat{\theta}_d$ and $\hat{\theta}_p$ it remains to consider the covariance terms under the PDS method. If using PDS results in positive covariance terms, then the DJS based estimator of $\frac{d\text{CVA}_I}{dS_0}$ outperform the PDS-baed estimator in terms of variance. Propositions 1 and 2 have identified conditions under which the covariance function $cov(V_u, V_t)$ is positive for contract level exposure for any $0 < u < t < T$. The same method of proof can be employed to show that when using PDS, under certain assumptions $cov(\hat{DV}_m(S_i), \hat{DV}_m(S_j)) > 0$ for any $i \neq j$. We give the proof here only for a simple stylized case.

Using the terminology and notation developed in Section 3, consider the contract level credit exposure for a unilateral transaction, i.e., $V_t = \max\{C_t, 0\} = C_t$ for all $0 < t \leq T$, where the time-t risk neutral value of the maturity-T derivatives contract is given by $C_t = n_t E[\frac{\Pi_T}{n_T}|\mathcal{F}_t]$ with n being the numeraire, $\mathcal{F}_t$ denoting the time-t filtration generated by the underlying risk factors, and Π being the payoff function. Consider the setting where the numeraire is the money market account, B, with deterministic short rate. Also, consider the single-simulation-run case, i.e., $m = 1$, where we set $\hat{DV}_{i,1} \equiv \hat{DV}_i$. We are to show that,

$$cov(\hat{DV}_i, \hat{DV}_j) > 0$$

---

[21]See Chapter 7 of [13] and the reference therein for finite-difference approximations. More specifically, see Section 7.1 on how the MSE of a finite-difference estimator can be approximately optimized.

for given $i < j$ on the discrete time grid. Using the conditional covariance formula as in the proofs of Propositions 1 and 2 gives,

$$cov(\hat{DV}_i, \hat{DV}_j) = cov(\hat{DV}_i, E[\hat{DV}_j|\mathcal{F}_i]).$$

Note that $V_j(S_0 + h) = B_j B_n^{-1} E_{h_+}[\Pi_T|F_j]$ and $V_j(S_0 - h) = B_j B_n^{-1} E_{h_-}[\Pi_T|F_j]$, where the subscript h_+ and h_- denote the cases where the underlying risk factor's initial prices are set at $S_0 + h$ and $S_0 - h$, respectively. Then, we have,

$$E[\hat{DV}_j|\mathcal{F}_i] = B_j B_n^{-1} \left(\frac{E_{h_+}[\Pi_T|F_i] - E_{h_-}[\Pi_T|F_i]}{2h} \right),$$

and so,

$$cov(\hat{DV}_i, \hat{DV}_j) = B_i B_j B_n^{-2} Var \left(\frac{E_{h_+}[\Pi_T|F_i] - E_{h_-}[\Pi_T|F_i]}{2h} \right) > 0.$$

Note that more general cases, e.g., stochastic short rate and bilateral transactions, require the application of changes of numeraire techniques and the Chebyshev's algebraic inequality and constraints on the number of risk factors and monotonicity of the payoff function as in the proofs of Proposition 1 and 2; we do not address these cases here as the proof methodology has already been developed in the paper. The mathematical proofs of the positivity of the covariance function under PDS are at the contract level exposure. For more general cases, comparing the variance of the PDS and DJS-based CVA sensitivity estimators requires small simulation studies before the CCR modeler makes a decision on whether to use DJS or PDS.

References

[1] "Capital requirements for bank exposures to central counterparties", *Basel Committee on Banking Supervision*, July (2012)

[2] T. Bjork, *"Arbitrage theory in continuous time"*, 3rd Edition, Oxford University Press, (2009)

[3] 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)

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

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

[6] Cesari, G., Aquilina, J., Charpillon, N., Filipovic, Z., Lee, G., Manda, I, "*Modelling, Pricing, and Hedging Counterparty Credit Exposure*", Springer Finance, (2010)

[7] Duffie, D., and Glynn, P., "Efficient Monte Carlo estimation of security prices", *Annals of Applied Porbability*, vol. 5, pp. 897-905, (1995)

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

[9] 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)

[10] Geman, H., El Karoui, N., Rochet., J. C., "Changes of Numeraire, changes of probability measure and option pricing", *Journal of Applied Probability*, vol. 32, pp. 443-458, (1995)

[11] Ghamami, S., "Book Review: Counterparty Credit Risk by Jon Gregory," *Quantitative Finance*, vol. 13, no. 12, pp. 1863-1865, (2013)

[12] Ghamami, S., Goldberg, L. R., "Stochastic intensity models of wrong way risk: wrong way CVA need not exceed independent CVA", *Journal of Derivatives*, vol. 21, no. 3, pp. 24-35, (2014)

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

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

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

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

[17] Gregory, J., "*Counterparty credit risk: the new challenge for global financial markets*", 1st edition, John Wiley and Sons, (2010)

[18] Gregory, J., "*Counterparty credit risk: a continuing challenge for global financial markets*", 2nd edition, John Wiley and Sons, (2012)

[19] Hull, J., and White, A., "CVA and wrong way risk", *Financial Analyst Journal*, vol. 68, no. 5, pp. 58-69, (2012)

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

[21] Longstaff, F. A., and Schwartz, E. S., "Valuing American options by simulation: A simple Least-Squares approach", *The Review of Financial Studies*, vol. 14, pp. 113-147, (2001)

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

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

[24] 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)

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

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

[27] Xing, J., Fu, M., Xiong, X., "Probabilistic error bounds for simulation quantile estimators", *Management Science*, vol. 14, no. 2, pp. 230-246, (2003)