TVA on American Derivatives

L. A. Abbas-Turki[†] and **M. A. Mikou**[‡]

[†]Laboratoire de Probabilités et Modéles Aléatoires, UMR 7599, UPMC,

4 place Jussieu, F-75252 Paris Cedex

[‡] EISTI, Laboratoire de Mathématiques,

avenue du Parc, 95011 Cergy-Pontoise Cedex, France

Abstract

This paper presents a full study of the Total Valuation Adjustment (TVA) simulation on American derivatives. It starts from the formulation of the problem under a general BSDE framework that includes the funding issue and the default of both parties. It finishes by giving a benchmark Nested Monte Carlo algorithm and discusses an appropriate implementation that provides accurate results within a one-minute simulation on Graphic Processing Units (GPUs). From a theoretical point of view, this paper can be considered as the extension to American derivatives of the work presented in Crépey (2012a,b). Regarding the algorithmic part, our study uses convergence rates developed in Newey (1997) as well as similar ideas to those presented in Gordy and Juneja (2010) and it goes beyond the square Monte Carlo algorithm detailed in Abbas-Turki et al. (2014) for European derivatives.

Keywords: TVA, American derivatives, BSDE, Nested Monte Carlo, GPU.

1 Introduction

After the 2007 economic crisis and the new Basel agreements that include the calculation of the CVA (Credit Valuation Adjustment) as an important part of the prudential rules, a large number of papers and books have been published on the CVA and the counterparty risk. For a comprehensive and detailed presentation of the subject, we refer the reader to Brigo et al. (2013) and to Crépey (2012a,b). The former reference provides an in-depth overview of the subject with a wide variety of compelling practical examples. The latter presents the mathematical intuition and details of the subject including a hedging framework. Other references that are not closely related to our work can be found in Brigo et al. (2013) and to Crépey (2012a,b).

Although there are a quite few of practitioner papers on the CVA as well as some important mathematical work that explains the problem, little research has been dedicated to developing a trustable numerical procedure that can be used to perform the computations. Cesari et al. (2009) is one of the first references that presents the industry practices in computing CVA. Among the research papers, maybe the most devoted to computing CVA are: P. Henry-Labordère (2012), Abbas-Turki et al. (2014a) and M. Fujii and A. Takahashi (2015). However, none of these papers develop a procedure that works for TVA for any portfolio of contracts and prove the convergence of the procedure with a reasonable error upper bound. The main reason for this is the mathematical complexity of the problem that makes the computational aspect very challenging.

Also due to both mathematical and computational complexity, to our knowledge, there is no paper that deals with the TVA problem when American contracts are involved. However, this point is capital especially in markets where American or Bermudan options are widely exchanged like in fixed income and equity markets. Thus, the first goal of this paper is to address this lack of theoretical development. We not only derive the TVA BSDE on one American option, we extend it to a portfolio of American derivatives.

The second goal of this paper is to propose a robust method to compute the TVA on

an exposure of American, European path-dependent and path-independent contracts. Consequently, this method not only provides accurate results when the exposure is explicit in terms of the underlying assets, it also works very well when the exposure must be simulated. This method is based on a nested Monte Carlo and it is studied in two situations: First, the funding constraints can be neglected and secondly they must be taken into account. For both situations, we express the upper bound of the Mean Square Error (MSE) in terms of the number of simulated trajectories. In particular, this allows us to establish an asymptotic relation between the number of trajectories M_0 simulated in the outer stage, the number of trajectories $\{M_j\}_{j=1,\dots,N-1}$ simulated in the inner stages and the number of time steps N.

The asymptotic relation between $\{M_j\}_{j=1,\dots,N-1}$, M_0 and N is useful to decrease the execution time of the simulation. Indeed, although the implementation is performed on a GPU with a high number of computing units that run the program in parallel, choosing appropriate values of M_j as a function of M_0 and N is necessary to perform the computations within a one-minute simulation. Thus, it is necessary to point out that the structure of the proposed method, based on nested Monte Carlo, allows us to compute an upper bound for the MSE and subsequently establish the desired relation between $\{M_j\}_{j=1,\dots,N-1}$, M_0 and N. Moreover, we will show that the resulting algorithm provides quite accurate values since the variance and the bias of the estimator are sufficiently small.

The rest of this paper is arranged as follows. In Section 2, we give the formulation of the TVA BSDE and the pre-default TVA BSDE when the contracts involved are American. In Section 3, we detail the simulation algorithms and we express an upper bound for the MSE according to the simulation parameters. Section 4 explains the implementation and contains some numerical results. Section 5 is dedicated to the proof of theorems 3.1 and 3.2.

2 TVA BSDE on American options

To simplify the presentation, we start by considering the TVA on only one American contract then extend it in Section 2.2 for a more general case.

2.1 TVA with an optimal stopping time

We consider two defaultable parties: The bank with a default time τ^b and the client with a default time τ^c. After selling to the client at time 0 an American contract that should generate a cumulative dividend D until the maturity $T > 0$, the bank sets-up a hedging strategy including collateralization and funding portfolio given by a price-and-hedge pair (Π, π). The promised dividend stream dD_t is effective only if none of the parties defaults till time t. We call "funder" of the bank a third party insuring the bank's funding strategy. Let $(\Omega, \mathcal{G}, (\mathcal{G}_t)_{t\in[0,T]}, P)$ be a filtrated probability space satisfying the usual conditions. We set $\mathcal{G}_t = \mathcal{G}_t^{\tau^b,\tau^c} \vee \mathcal{F}_t$ where $\mathcal{G}_t^{\tau^b,\tau^c} = \sigma(\tau^b \wedge t, \tau^c \wedge t)$ and $\mathcal{F}$ is generated by the underlying assets S such that D is $\mathcal{F}$-adapted. We define also $\bar{\mathcal{G}}_t = \sigma(\tau \wedge t) \vee \mathcal{F}_t$ with $\tau = \tau^b \wedge \tau^c$. We assume that all the considered processes are $\mathcal{G}$-adapted as well as integrable and P is a risk neutral probability. The price of the TVA contract is computed as the difference between a reference price and a price that takes into account the counterparty risk and the funding adjustment. We refer to Crépey (2012a) and Crépey (2012b) for more details on the market conventions and the meaning of a risk neutral probability in this context.

Assuming that the Azéma supermartingale associated with τ is a positive continuous and non-increasing process, we can admit Lemma 2.1 which will be useful in the sequel.

Lemma 2.1 *i. For any $\mathcal{G}$-measurable random variable Y and any $t \geq 0$*

$$E_t(Y\mathbf{1}_{\tau>t}) = \frac{\mathbf{1}_{\tau>t}}{\mathbb{E}_t(\mathbf{1}_{\tau>t})}\mathbb{E}_t(Y\mathbf{1}_{\tau>t}),$$

where E_t and $\mathbb{E}_t$ denote respectively the conditional P-expectation given $\mathcal{G}_t$ and $\mathcal{F}_t$.
ii. An $\mathcal{F}$-martingale stopped at τ is a $\bar{\mathcal{G}}$-martingale and a $\bar{\mathcal{G}}$-martingale is a $\mathcal{G}$ martingale.
iii. An $\mathcal{F}$-adapted càdlàg process cannot jump at τ.

The first point of this lemma can be found in Dellacherie-Meyer (1980) and a proof of the other two points is given in Crépey (2012b).

In the case of a European derivative, it is shown in Crépey (2012a) that the pair price-

and-hedge satisfies a BSDE on $[0, \tau \wedge T]$ under the risk neutral assumption. In the case of an American derivative, the situation is exactly the same until the optimal stopping time $\tau^* \in [0, T]$ after which the contract is exercised and the counterparty risk disappears. Denote $\bar{\tau} = \tau \wedge \tau^*$ the effective maturity and by analogy with Crépey (2012a) we characterize the bank portfolio as follows.

Definition 2.1 *We call a price-and-hedge the pair* (Π, π)*, comprising a* $\mathcal{G}$*-semimartingale* Π *and a hedge* π*, that satisfies the following BSDE on* $[0, \bar{\tau}]$*:*

$$\begin{aligned} d\Pi_t + \mathbf{1}_{t<\tau}\mathbf{1}_{t\leq\tau^*}dD_t - (r_t\Pi_t + g_t(\Pi_t, \pi_t))dt &= dm_t^\pi \\ \textit{with the final condition} \quad \Pi_{\bar{\tau}} &= \mathbf{1}_{\{\tau=\bar{\tau}\}}R \end{aligned}$$

where m^π *is a* $\mathcal{G}$*-martingale null at time* 0*,* g *is an* $\mathcal{F}$*-progressively measurable function and* R *is a* $\mathcal{G}_\tau$*-measurable recovery.*

Notice that, since the $\mathcal{F}$-adapted process D does not jump at time τ from Lemma 2.1 and by discounting, the previous BSDE becomes

$$d\beta_t\Pi_t + \beta_t dD_t - \beta_t g_t(\Pi_t, \pi_t)dt = \beta_t dm_t^\pi, \ \forall t \in [0, \bar{\tau}]. \tag{2.1}$$

where $\beta = e^{-\int_0 r_s ds}$ is the risk-free discounting asset with an interest rate r generally considered as the overnight indexed swap rate. In the integral form, the last BSDE is equivalent to the following

$$\beta_t\Pi_t = E_t\left(\int_t^{\bar{\tau}} \beta_s dD_s - \int_t^{\bar{\tau}} \beta_s g_s(\Pi_s, \pi_s)ds + \beta_\tau \mathbf{1}_{\{\tau=\bar{\tau}\}}R\right), \quad \forall t \in [0, \bar{\tau}]. \tag{2.2}$$

When ignoring counterparty risk and assuming a risk-free funding rate, we define p^a as the discounted cumulative clean price of an American contract with an optimal stopping time τ^*. Under the risk neutral measure P, it is known that $(\beta_{t\wedge\tau^*}p^a_{t\wedge\tau^*})_{0\leq t\leq T}$ is an $\mathcal{F}$-martingale

and p^a is given by

$$\begin{aligned}\beta_t p_t^a &= \mathbb{E}_t\left(\int_0^{\tau^*}\beta_s dD_s\right)\\ &= \int_0^t \beta_s dD_s + \mathbb{E}_t\left(\int_t^{\tau^*}\beta_s dD_s\right)\\ &=: \int_0^t \beta_s dD_s + \beta_t P_t^a, \quad \forall t\in[0,\tau^*],\end{aligned}$$

where P^a represents the price of the future cash flows of the contract which we call the clean price as in Crépey (2012b). Remark that the dividend D can be either seen as a native swap or as a virtual swap via a repo market. In the examples considered in this paper, we simplify D and make it only meaningful at τ^*. For instance, if the bank sells a put option on an asset S with a strike K then $D_s = (K - S_{\tau^*})_+ 1_{s\geq\tau^*}$ and thus P^a would be given by $\beta_t P_t^a = \mathbb{E}_t\left(\beta_{\tau^*}(D_{\tau^*} - D_{\tau^*-})\right) = \mathbb{E}_t\left(\beta_{\tau^*}(K - S_{\tau^*})_+\right)$ for each $t\in[0,\tau^*]$.

By substituting the time t by $\tau\wedge\tau^* = \bar\tau$ in the last equality and conditioning with respect to $\mathcal{G}_t$ we get

$$E_t\left(\beta_{\bar\tau} p_{\bar\tau}^a\right) = E_t\left(\int_0^{\bar\tau}\beta_s dD_s + \beta_{\bar\tau} P_{\bar\tau}^a\right), \quad \forall t\in[0,\bar\tau].$$

Using Lemma 2.1, the $\mathcal{F}$-martingale $(\beta_{t\wedge\tau^*} p^a_{t\wedge\tau^*})_{0\leq t\leq T}$ stopped at τ is a $\mathcal{G}$-martingale, we then deduce the following representation of P^a

$$\beta_t P_t^a = E_t\left(\int_t^{\bar\tau}\beta_s dD_s + \beta_{\bar\tau} P_{\bar\tau}^a\right), \quad \forall t\in[0,\bar\tau]. \tag{2.3}$$

The clean price can be seen also as the solution of the following BSDE

$$d\beta_t P_t^a + \beta_t dD_t = \beta_t dm_t^a, \quad \forall t\in[0,\bar\tau], \quad P_{\bar\tau}^a = 1_{\bar\tau=\tau} P_\tau^a \tag{2.4}$$

where, under usual assumptions on r, m^a is the $\mathcal{G}$-martingale null at time 0 defined by $m_t^a = \int_0^{t\wedge\bar\tau}\beta_s^{-1} d(\beta_s p_s^a)$ for $t\in[0,T]$.

The price of the TVA contract is defined by the process $\Theta = P^a - \Pi$. From (2.2) and (2.3), we deduce the integral form for the TVA

$$\beta_t\Theta_t = E_t\left(\beta_{\bar\tau}\Theta_{\bar\tau} + \int_t^{\bar\tau} \beta_s g_s(P^a_s - \Theta_s, \pi_s)ds\right), \quad \forall t \in [0, \bar\tau], \tag{2.5}$$
$$\text{where} \quad \Theta_{\bar\tau} = P^a_{\bar\tau} - \mathbf{1}_{\{\tau=\bar\tau\}}R.$$

Moreover, the TVA can be also seen as the solution of the following BSDE that combines both (2.1) and (2.4)

$$d\beta_t\Theta_t + \beta_t g_t(P^a_t - \Theta_t, \pi_t)dt = \beta_t dm_t, \quad \forall t \in [0, \bar\tau], \tag{2.6}$$
$$\text{with} \quad \Theta_{\bar\tau} = P^a_{\bar\tau} - \mathbf{1}_{\{\tau=\bar\tau\}}R,$$

where m is the $\mathcal{G}$-martingale null at time 0 defined by

$$m = m^a - m^\pi. \tag{2.7}$$

We assume now that the Azéma supermartingale of τ is time differentiable and we denote $\gamma_t = -\frac{d\ln(G_t)}{dt}$ the hazard intensity and $\alpha_t = e^{-\int_0^t \gamma_s ds}$. Denote $\xi_t := P^a_t - E_t(R\mathbf{1}_{\{t=\bar\tau\}})$, and $\tilde\xi_t := \frac{1}{P(\tau>t|\mathcal{F}_t)}\mathbb{E}_t(\xi_t\mathbf{1}_{\{\tau>t\}})$. Notice that $\xi_\tau = \Theta_{\bar\tau}$ and using Lemma 2.1 one can verify that $\tilde\xi_t\mathbf{1}_{\{\tau>t\}} = \xi_t\mathbf{1}_{\{\tau>t\}}$. Performing a filtration reduction, we introduce as in Crépey (2012b), a new process called the pre-default TVA.

Definition 2.2 *We call the pre-default TVA the solution of the following $\mathcal{F}$-BSDE.*

$$d\tilde\beta_t\tilde\Theta_t + \tilde\beta_t\tilde g_t(P^a_t - \tilde\Theta_t, \pi_t)dt = \tilde\beta_t d\tilde m_t, \ \forall t \in [0, \tau^*], \quad \textit{with} \quad \tilde\Theta_{\tau^*} = 0, \tag{2.8}$$

where $\tilde\beta = \alpha\beta$, $\tilde m$ is an $\mathcal{F}$-martingale and $\tilde g$ is the $\mathcal{F}$-progressively measurable function defined by

$$\tilde g_t(P^a_t - \tilde\Theta_t, \pi_t) = g_t(P^a_t - \tilde\Theta_t, \pi_t) + \gamma_t\tilde\xi_t, \quad \forall t \in [0, \tau^*].$$

In an integral form, the pre-default TVA is given by

$$\tilde{\beta}_t\tilde{\Theta}_t = \mathbb{E}_t\left(\int_t^{\tau^*} \tilde{\beta}_s\tilde{g}_s(P^a_s - \tilde{\Theta}_s, \pi_s) + \gamma_t\tilde{\xi}_s(P^a_s - \tilde{\Theta}_s, \pi_s)ds\right), \quad \forall t \in [0, \tau^*]. \tag{2.9}$$

Let M be the $\bar{\mathcal{G}}$-martingale defined by $M_t := \mathbf{1}_{\{\tau>t\}} + \int_0^{t\wedge\tau} \gamma_s ds$ and assume that the locale $\mathcal{G}$-martingale $\int_0^{\cdot\wedge\bar{\tau}}(\tilde{\Theta}_s - \xi_s)dM_s$ is a martingale.

Proposition 2.1 *Let the pair price-and-hedge (Π, π) be as in Definition 2.1 with a $\mathcal{G}$-martingale component m^π defined by $m^\pi := m^a_{\cdot\wedge\bar{\tau}} - \tilde{m}_{\cdot\wedge\bar{\tau}} - \int_0^{\cdot\wedge\bar{\tau}}(\tilde{\Theta}_s - \xi_s)dM_s$. Then, the TVA price is given by $\Theta = \tilde{\Theta}J + (1-J)\xi_\tau$, where $J_t = \mathbf{1}_{\{\tau>t\}}$.*

Proof. Let $\bar{\Theta} =: \tilde{\Theta}J + (1-J)\xi_\tau$. Using (2.8) and (2.7), we have for $t \in [0, \bar{\tau}]$

$$\begin{aligned}
d\beta_t\bar{\Theta}_t &= dJ_t\beta_t\tilde{\Theta}_t + d(1-J)\beta_t\xi_\tau = d\beta_{t\wedge\tau}\tilde{\Theta}_{t\wedge\tau} + \beta_t\tilde{\Theta}_t dJ_t - \beta_t\xi_t dJ_t \\
&= \frac{1}{\alpha_t}d\tilde{\beta}_t\tilde{\Theta}_t + \gamma_t\beta_t\tilde{\Theta}_t dt + \beta_t(\tilde{\Theta}_t - \xi_t)dJ_t \\
&= -\beta_t(g_t(P^a_t - \tilde{\Theta}_t, \pi_t) + \gamma_t\tilde{\xi}_t)dt + \beta_t d\tilde{m}_t + \beta_t[(\tilde{\Theta}_t - \xi_t)dJ_t + \gamma_t\tilde{\Theta}_t dt] \\
&= -\beta_t g_t(P^a_t - \tilde{\Theta}_t, \pi_t)dt + \beta_t d\tilde{m}_t + \beta_t[(\tilde{\Theta}_t - \xi_t)dJ_t + \gamma_t(\tilde{\Theta}_t - \tilde{\xi}_t)dt] \\
&= -\beta_t g_t(P^a_t - \tilde{\Theta}_t, \pi_t)dt + \beta_t\big(d\tilde{m}_t + (\tilde{\Theta}_t - \xi_t)dM_t\big) \\
&= -\beta_t g_t(P^a_t - \tilde{\Theta}_t, \pi_t)dt + \beta_t(dm^a_t - dm^\pi_t) \\
&= -\beta_t g_t(P^a_t - \bar{\Theta}_t, \pi_t)dt + \beta_t dm_t.
\end{aligned}$$

Then, $\bar{\Theta}$ satisfies the BSDE as Θ with the same limit condition, this ends the proof.

2.2 Multiple optimal stopping times

Let us consider now n American contracts with different maturities $T_i > 0$, different optimal stopping times $\tau^*_i \in [0, T_i]$ and different dividend streams dD^i_t. In that case, the counterparty risk vanishes after $\tau^* = max_{0\leq i\leq n}(\tau^*_i)$ and the effective maturity time becomes $\bar{\tau} = \tau \wedge \tau^*$.

From (2.3), the clean price $P^{a,i}$ of each contract is given by

$$\begin{aligned} \beta_t P^{a,i}_t &= E_t\left(\int_t^{\bar\tau^i} \beta_s dD^i_s + \beta_{\bar\tau^i} P^{a,i}_{\bar\tau^i}\right), \quad \forall t\in[0,\bar\tau^i] \\ \text{or equivalently} \quad \beta_{t\wedge\bar\tau^i} P^{a,i}_{t\wedge\bar\tau^i} &= E_t\left(\int_t^{\bar\tau} \beta_s dD^i_{s\wedge\bar\tau^i} + \beta_{\bar\tau^i} P^{a,i}_{\bar\tau^i}\right), \quad \forall t\in[0,\bar\tau], \end{aligned}$$

where $\bar\tau^i = \tau\wedge\tau^*_i$. We define the overall clean price of all contracts at time $t\in[0,\bar\tau]$ by

$$P^a_t =: \frac{1}{\beta_t}\sum_{i=1}^n \beta_{t\wedge\bar\tau^i} P^{a,i}_{t\wedge\bar\tau^i}.$$

This yields to the following BSDE for P^a

$$\begin{aligned} \beta_t P^a_t &= E_t\left(\int_t^{\bar\tau} \beta_s dD_s + \sum_{i=1}^n \beta_{\bar\tau^i} P^{a,i}_{\bar\tau^i}\right) \\ &= E_t\left(\int_t^{\bar\tau} \beta_s dD_s + \beta_{\bar\tau} P^a_{\bar\tau}\right), \quad \forall t\in[0,\bar\tau], \end{aligned}$$

where dD is the global dividend stream defined by $dD_t := \sum_{i=1}^n dD^i_{t\wedge\bar\tau^i}$. By analogy to the previous section, the price-and-hedge pair (Π,π) satisfies a BSDE given, in an integral form, by

$$\beta_t\Pi_t = E_t\left(\int_t^{\bar\tau}\beta_s dD_s - \int_t^{\bar\tau}\beta_s g_s(\Pi_s,\pi_s)ds + \beta_\tau \mathbf{1}_{\{\tau=\bar\tau\}}R\right), \quad \forall t\in[0,\bar\tau].$$

Thus, we deduce the integral form of the TVA BSDE

$$\begin{aligned} &\beta_t\Theta_t = E_t\left(\beta_{\bar\tau}\Theta_{\bar\tau} + \int_t^{\bar\tau}\beta_s g_s(P^a_s-\Theta_s,\pi_s)ds\right), \quad \forall t\in[0,\bar\tau], \\ &\text{where} \quad \Theta_{\bar\tau} = P^a_{\bar\tau} + \mathbf{1}_{\{\tau=\bar\tau\}}R. \end{aligned}$$

Consequently, Proposition 2.1 remains true in this setup with

$$m^a_t = \sum_{i=1}^n m^{a,i}_t = \sum_{i=1}^n \int_0^{t\wedge\bar\tau^i} \beta_s^{-1} d\beta_s p^{a,i}_s, \text{ where } \beta_t p^{a,i}_t = \mathbb{E}_t\left(\int_0^{\tau^*_i}\beta_s dD^i_s\right).$$

3 Benchmark simulation using nested Monte Carlo

We present here an overview of the overall algorithm. We also discuss the differences between Section 3.2 where we consider the funding issues and Section 3.1 where we study essentially an extension of the square Monte Carlo explained in Abbas-Turki et al. (2014). In the following, depending on the context, we use P_t either for the clean exposure of a portfolio or for the clean exposure of only one contract.

Since the paper by Brigo and Pallavicini (2008), the (CVA) Credit Valuation Adjustment can be viewed as an option on the clean exposure called Contingent Credit Default Swap (CCDS). When the clean exposure P_t is computed on a basket of contracts that are priced by closed expressions, the CVA and, more generally, the TVA can be calculated thanks to a one-stage simulation using either PDE discretization or Monte Carlo as in Crépey et al. (2014). However, when the underlying contracts have to be simulated as in the case of American options, it is more reasonable to perform a two-stage simulation: The outer stage for the TVA and the inner stages to compute the underlying contracts. Indeed, the TVA can be considered as a corrective value on P_t and mispricing the latter could produce significant errors on the former. So, using a one-stage simulation with the same set of trajectories for both TVA and the clean exposure would be a poor choice when implementing global methods like regressions or when the default is strongly dependent on the exposure P_t.

Our purpose is to develop a method that can be considered as a benchmark not only when the exposure involves American derivatives but also European derivatives that are not expressed by closed formulas. Although the proposed two-stage Monte Carlo is quite heavy to implement on a CPU, it will run much faster on a GPU. Moreover, in order to decrease the execution time and run a one-minute simulation, we propose in sections 3.1 and 3.2 a judicious procedure to choose the appropriate number of trajectories that have to be drawn in the inner stages.

For both sections 3.1 and 3.2, we fix the number of time steps used for the SDE discretization of the underlying asset $S = (S^1, ..., S^d)$ to be a multiple of N: $N_{sde} = q \times N$.

Using N_{sde} time steps, we simulate M_0 outer stage trajectories from $t_0 = 0$ to $t_{N_{sde}} = T$. From each value $S_{t_{qk}}$, with $k \in \{1, ..., N\}$, we simulate M_k inner stage trajectories that end at $t_{N_{sde}} = T$.

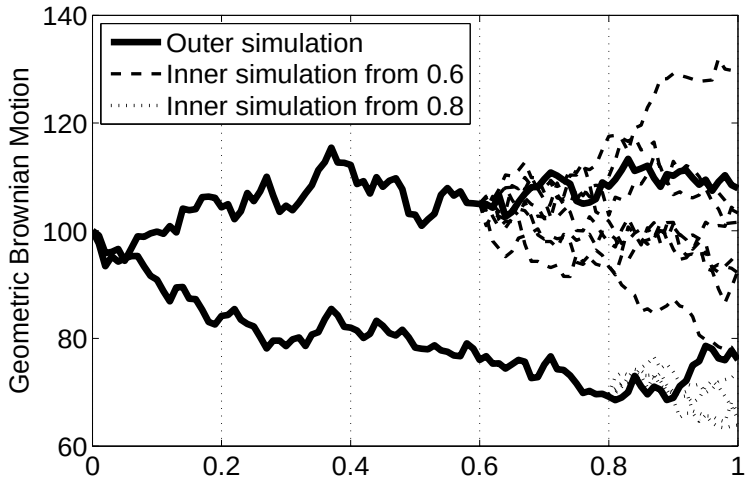


Figure 1: An example of a two-stage simulation with $M_0 = 2$, $M_6 = 8$ and $M_8 = 4$.

In Figure 1, we illustrate only two inner simulations starting at different times. Because the inner simulations are used to compute the clean exposure P_t, it is conceivable to draw fewer trajectories when t approaches T. Indeed, as $t \to T$ the simulation variance V_t associated to P_t decreases to 0. In addition, V_t produces a bias on the outer simulation of the TVA which gets smaller as $V_t \to 0$. We point out also that the bias of the estimator of P_t vanishes when the exposure involves only European contracts. For American contracts, the bias produced by Longstaff-Schwartz algorithm is generally small and will be neglected in this paper. We refer the reader to Glasserman (2003) for more details on the bias of American options estimators.

For fixed and sufficiently high values of M_0 and N, our purpose is to study the effect of V_t on the bias of the TVA estimator and thus on the choice of $M_1 ... M_{N-1}$. This study will

be first implemented, in Section 3.1, when funding constraints are ignored and thus when

$$\text{CVA}_{0,T} = \sum_{k=0}^{N-1} E\left(P^+_{t_{q(k+1)}} 1_{\tau\in(t_{qk},t_{q(k+1)}]}\right), \tag{3.1}$$

then for more general case, in Section 3.2,

$$\Theta_{t_{qk}} = E_{t_{qk}}\left(\Theta_{t_{q(k+1)}} + hg(t_{q(k+1)}, P_{t_{q(k+1)}}, \Theta_{t_{q(k+1)}})\right), \quad \Theta_{t_{qN}} = 0, \tag{3.2}$$

where Θ_t is the pre-default TVA process and $h = T/N$. In the sequel, P replaces P^a used in Section 2, $\text{CVA}_{0,T}$ represents Θ_0 given in (2.5) when we set g and R to zero and the Θ of (3.2) is used for the pre-default TVA $\tilde{\Theta}_t$ introduced in (2.8). Also to simplify notations, we assume that $T = 1$ then $h = 1/N$. In standard applications when $T \neq 1$, one should increase or decrease linearly the value of N depending on whether $T > 1$ or $T < 1$.

As for (3.2), we consider an intensity model for the default time τ of equation (3.1) and we assume that its hazard rate is a function of the exposition P_t. When compared to structural models or to intensity models with a hazard rate expressed as a function of the underlying assets, assuming that the hazard rate depends on P_t is more difficult to study as one must take into account the effect of V_t on simulating τ. Admitting that we are dealing with this complex setting, one can extend the obtained results bellow to simpler situations.

Unlike (3.1), (3.2) involves the computation of some conditional expectations during a backward induction. Consequently, although both (3.1) and (3.2) are based on an intensity model and on a two-stage simulation like the one illustrated in Figure 1, they are implemented differently. Below, we detail the implementation of each expression and calculate asymptotically its MSE. In the following, we replace the t_{q*k} index by k and we denote by $\widehat{P}_1, ..., \widehat{P}_{N-1}$ the simulated expositions using the inner trajectories. Then, we define $\Delta^P_k = \sqrt{M_k}\left(\widehat{P}_k - P_k\right)$, $\Delta^\Theta_k = \left(\widehat{\Theta}_k - \Theta_k\right)$ where $\widehat{\Theta}$ is the simulated value of Θ and we make the following assumption that is an extension of Assumption 1 in Gordy & Juneja (2010).

Assumption 3.1 *Defining* $\varphi_{M_0,...,M_{N-1}}(p_1,...,p_{N-1},\theta_1,...,\theta_{N-1},\delta_1^p,...,\delta_{N-1}^p,\delta_1^\theta,...,\delta_{N-1}^\theta)$ *as the density of the random vector* $(P_1,...,P_{N-1},\Theta_1,...,\Theta_{N-1},\Delta_1^P,...,\Delta_{N-1}^P,\Delta_1^\Theta,...,\Delta_{N-1}^\Theta)$ *with respect to the Lebesgue measure, we assume that its partial derivatives*

$$\partial_{u_k}\varphi_{M_0,...,M_{N-1}},\ \ \partial^2_{u_k,u_l}\varphi_{M_0,...,M_{N-1}},\ \ u_k=p_k \text{ or } \theta_k \text{ and } u_l=p_l \text{ or } \theta_l \text{ with } k,l=1,...,N-1$$

exist and are continuous for each $(M_0,...,M_{N-1})$. *With* u_k, u_l *as before and* $u_i=p_i$ *or* θ_i,

$$\text{for } i=1,...,N-1 \lim_{|u_i|\to\infty}\varphi_{M_0,...,M_{N-1}}=0,\ \lim_{|u_i|\to\infty}\partial_{u_k}\varphi_{M_0,...,M_{N-1}}=0,\ \lim_{|u_i|\to\infty}\partial^2_{u_k,u_l}\varphi_{M_0,...,M_{N-1}}=0$$

uniformly on all the variables except u_i *and uniformly on* $M_0,...,M_{N-1}$. *Moreover, for each* $(M_0,...,M_{N-1})$, *there exist nonnegative functions* $\varphi^0_{M_0,...,M_{N-1}}$, $\varphi^1_{M_0,...,M_{N-1}}$ *and* $\varphi^2_{M_0,...,M_{N-1}}$

$$\begin{aligned}
&\text{such that } \ \varphi_{M_0,...,M_{N-1}}\le\varphi^0_{M_0,...,M_{N-1}}(\delta_1^p,...,\delta_{N-1}^p,\delta_1^\theta,...,\delta_{N-1}^\theta),\\
&\left|\partial_{u_k}\varphi_{M_0,...,M_{N-1}}\right|\le\varphi^1_{M_0,...,M_{N-1}}(\delta_1^p,...,\delta_{N-1}^p,\delta_1^\theta,...,\delta_{N-1}^\theta),\\
&\left|\partial^2_{u_k,u_l}\varphi_{M_0,...,M_{N-1}}\right|\le\varphi^2_{M_0,...,M_{N-1}}(\delta_1^p,...,\delta_{N-1}^p,\delta_1^\theta,...,\delta_{N-1}^\theta)
\end{aligned}$$

for all $(p_1,...,p_{N-1},\theta_1,...,\theta_{N-1},\delta_1^p,...,\delta_{N-1}^p,\delta_1^\theta,...,\delta_{N-1}^\theta)$ *with*

$$\sup_{M_0,...,M_{N-1}}\int_{\mathbb{R}^{2N-2}}|\delta_k^p|^{r_1}|\delta_l^\theta|^{r_2}\varphi^i_{M_0,...,M_{N-1}}(\delta_1^p,...,\delta_{N-1}^p,\delta_1^\theta,...,\delta_{N-1}^\theta)d\delta_1^p...d\delta_{N-1}^p d\delta_1^\theta...d\delta_{N-1}^\theta<\infty$$

for $k,l=1,...,N-1$, $i=0,1,2$, $r_1\ge0$, $r_2\ge0$ *and* $0\le r_1+r_2\le4$.

This assumption is needed to justify the Taylor expansion performed in sections 3.1, 3.2 and ensures that one can ignore the higher order terms. In what follows, we assume also that the underlying asset S is a truncation of a positive Lévy process. The truncation should be performed such that the support of S is a Cartesian product of compact connected intervals on which the density of S is bounded away from zero.

3.1 TVA without funding constraint

We present an optimized version of the square Monte Carlo simulation MC2 taken as a benchmark algorithm in Abbas-Turki et al. (2014). Because MC2 was not the main subject of the latter paper, the authors implemented a simple version of MC2 with $M_0 = M_1 = ... = M_{N-1}$. However, here we would like to express M_1, ..., M_{N-1} as functions of (M_0, N) that have to be sufficiently big. The other difference with the MC2 presented in Abbas-Turki et al. (2014) is the possibility here to simulate American derivatives using the inner trajectories. This will be performed thanks to $N \times M_0$ local dynamic programming inductions that are explained at the end of this subsection and its implementation is detailed in Section 4.

We introduce new functions $F^1_{k+1}(x_1, ..., x_{k+1})$, $F^2_{k+1}(x_1, ..., x_{k+1})$, $F^3_{k+1}(x_1, ..., x_{k+1})$ and $F^4_{k+1}(x_1, ..., x_l, y)$ with

$$
\left\{
\begin{array}{c}
F^1_{k+1}(x_1, ..., x_{k+1}) = E\left(1_{\tau \in (kh,(k+1)h]} | P_1 = x_1, ..., P_{k+1} = x_{k+1}\right), \\
F^2_{k+1}(x_1, ..., x_{k+1}) = (x_{k+1})^+ F^1_{k+1}(x_1, ..., x_{k+1}), \\
F^3_{k+1}(x_1, ..., x_{k+1}) = (x_{k+1})^+ E\left(1_{\tau > (k+1)h} | P_1 = x_1, ..., P_{k+1} = x_{k+1}\right) \text{ and} \\
F^4_{k+1}(x_1, ..., x_j, y) = E\left(F^2_{k+1}(x_1, ..., x_j, P_{j+1}, ..., P_{k+1}) | S_j = y\right).
\end{array}
\right. \tag{3.3}
$$

Thus, the simulated value $\widehat{\mathrm{CVA}}_{0,T}$ of (3.1) is given by

$$
\widehat{\mathrm{CVA}}_{0,T} = \sum_{k=0}^{N-1} \frac{1}{M_0} \sum_{i=1}^{M_0} F^2_{k+1}\left(\widehat{P}_1(S^i_1), ..., \widehat{P}_{k+1}(S^i_{k+1})\right) \tag{3.4}
$$

where $\{S^i\}_{i \in \{1,...,M_0\}}$ are independent copies of the underling asset S that are generated in the outer simulation.

As said previously, we take the hazard rate of τ to be a function of the exposition P_t. This function is assumed constant by parts and can be decomposed using a family $\{f_k\}_{1 \leq k \leq N}$ of

twice differentiable functions such that

$$P\left(\tau > kh | P_1 = x_1, ..., P_k = x_k\right) = \exp\left(-\frac{1}{N}\sum_{i=1}^{k} f_i(x_i)\right). \tag{3.5}$$

Before announcing the main theorem of this section, we should consider an additional constraint on $M_1, ..., M_{N-1}$ that makes possible dealing with American contracts using regression methods. Indeed, an approximation of an American contract by a Bermudan leads to

$$P_k(x) = \sup_{\theta \in \mathcal{T}_{k,N}} \mathbb{E}\left(\Phi_{k,\theta}(S_\theta) | S_k = x\right), \tag{3.6}$$

where $\Phi_{s,t}(x) = \beta_t \Upsilon(x)/\beta_s$ with Υ is the payoff and β is the risk-free discounting asset. Besides, $\mathcal{T}_{k,N}$ represents the set of stopping times that take their values in $\{k, ..., N\}$. The computation of (3.6) is performed using the Longstaff-Schwartz algorithm introduced in Longstaff and Schwartz (2001) and well detailed in Clément & al. (2002). We set $P_k(x) = \mathbb{E}\left(\Phi_{k,\tau_k}(S_{\tau_k}) | S_k = x\right)$ with

$$\begin{aligned} &\tau_N = N, \\ &\forall k \in \{N-1, ..., 0\}, \quad \tau_k = k 1_{A_k} + \tau_{k+1} 1_{A_k^c}, \end{aligned} \tag{3.7}$$

where $A_k = \left\{\Phi_{k+1,k}(S_k) > E\left(P_{k+1}(S_{k+1}) \middle| S_k\right)\right\}$. The conditional expectation involved in A_k is approximated using a regression on a basis of monomial functions where K_k is its cardinal. This regression uses the inner trajectories and thus allows us to approximate $P_k(x)$ by

$$\widehat{P}_k(x) = \frac{1}{M_k}\sum_{i=1}^{M_k} \Phi_{k,\widehat{\tau}_k^i}\left(S^i_{\widehat{\tau}_k^i} \middle| S_k^i = x\right) \tag{3.8}$$

and the dependence on the inner trajectories can be seen from fixing S_k^i to be equal to x. In (3.8), $\widehat{\tau}_k$ is simulated thanks to a similar induction to (3.7) in which we replace the

conditional expectation involved in A_k by a regression.

Then, we are able to announce the following result.

Theorem 3.1 *As long as Assumption 3.1 is fulfilled and $K_j^3/M_j \to 0$ for each j when American contracts are involved, we get*

$$\begin{aligned}
MSE\left(\widehat{CVA}_{0,T} - CVA_{0,T}\right) = \quad & E\left[\left(\widehat{CVA}_{0,T} - CVA_{0,T}\right)^2\right] \\
\leq \quad & \frac{N^2}{M_0} \max_{k\in\{0,\dots,N-1\}} Var\left(F_{k+1}^2\left(\widehat{P}_1(S_1^i),...,\widehat{P}_{k+1}(S_{k+1}^i)\right)\right) \\
& + \sum_{j=1}^{N} \frac{1}{4NM_j^2}\left(E\left[V_j(S_j^i) f_j''(P_j(S_j^i)) F_j^3(P_1(S_1^i),...,P_j(S_j^i))\right]\right)^2 \\
& + \sum_{j=1}^{N} \frac{1}{4NM_j^2}\left(E\left[V_j(S_j^i) f_j''(P_j(S_j^i)) \sum_{k=j}^{N-1} F_{k+1}^4(P_1(S_1^i),...,P_j(S_j^i),S_j^i)\right]\right)^2 \\
& + \sum_{j=1}^{N} \frac{N}{4M_j^2}\left(E\left[V_j(S_j^i) F_j^1(P_1(S_1^i),...,P_j(S_j^i)) | P_j(S_j^i) = 0\right]\varphi_j(0)\right)^2 \\
& + N\sum_{j=1}^{N}(N-j+1)^2 O\left(\frac{1}{M_j^4}\right)
\end{aligned}$$

where φ_j is the density of $P_j(S_j^i)$ and $V_j(x) = Var\left(\sqrt{M_j}\left(\widehat{P}_j(x) - P_j(x)\right)\right)$.

Using this result, it is natural to take $M_j \sim \sqrt{M_0}/N$ when $\varphi_j(0)$ vanishes or generally when $\varphi_j(0)$ is small enough, otherwise, it is sufficient to take $M_j \sim \sqrt{M_0}$. However, we point out that with both choices, it is necessary to make sure that N is not too big in order to control the variance term in the previous inequality. Moreover, when American options are involved, K_j^3/M_j must be small enough.

Theorem 3.1 allows to establish a condition on the value of M_j for all $j \in \{1,...,N-1\}$. However, it is interesting to see how M_j should decrease with respect to j as $V_j(x)$ also decreases. It is easy to express a relationship between M_j and M_{j+1} when the exposition involves only European contracts and the underlying asset S is a truncation of a log-Normal process. In this case and assuming a sufficient regularity on the function $\Phi_{0,N}$, $V_j(x) =$

$\mathrm{Var}\left(\Phi_{0,N}\left(\sqrt{(N-j)/N}G\right)\right) \approx (N-j)/N\Phi'_{0,N}(0)+O((N-j)^2/N^2)$ where G is a truncation of a standard Gaussian variable. Thus, one can set

$$M_j = \frac{N-j}{N-1}M_1 \text{ with either } M_1 = \frac{\sqrt{M_0}}{N} \text{ or } M_1 = \sqrt{M_0}. \tag{3.9}$$

When American options are involved, we will see that the choice (3.9) is good numerically as long as M_0 is big enough. Indeed, this ensures that the value of M_{N-1}, which is the smallest among all the $M_{j>0}$, makes K_{N-1}^3/M_{N-1} small enough.

Even though we presented only asymptotic choices of M_j, we will make them quantitative in Section 4. In particular, we will see that $K_j^3/M_j \leq 1$ is quite sufficient because of the induction (3.7) robustness that is used for the dynamic programming.

3.2 Pre-default TVA BSDE

As said previously, although both the approximation of (3.1) and the approximation of (3.2) are based on an outer stage and on an inner stage simulation, the conditional expectation involved in (3.2) requires a more advanced implementation. In particular, we need to simulate two independent sets $\{S^i\}_{i\in\{1,\dots,M_0\}}$ and $\{\widetilde{S}^i\}_{i\in\{1,\dots,M_0\}}$ of the underlying asset S. Where $\{S^i\}_{i\in\{1,\dots,M_0\}}$ are used in the outer simulation and $\{\widetilde{S}^i\}_{i\in\{1,\dots,M_0\}}$ are used to compute the regression matrix Ψ_k that is given for each k by

$$\Psi_k = \mathfrak{T}\left(\frac{1}{M_0}\sum_{i=0}^{M_0}\psi(\widetilde{S}_k^i)^t\psi(\widetilde{S}_k^i)\right) \tag{3.10}$$

and ψ is a basis of monomial functions where K is its cardinal. $\mathfrak{T}$ is an operator that must satisfy assumption 3.2. Then, the simulated value $\widehat{\Theta}_k(S_k^i)$ of (3.2) is defined thanks to the

following induction

$$(3.11)\begin{cases}\text{For } k=1,...,N-1 \\ \widehat{\Theta}_k(x)={}^t\psi(x)\Psi_k^{-1}\left[\frac{1}{M_0}\sum_{j=1}^{M_0}\psi(S_k^j)\Big(\widehat{\Theta}_{k+1}(S_{k+1}^j)+\frac{1}{N}g\Big(k+1,\widehat{\Theta}_{k+1}(S_{k+1}^j),\widehat{P}_{k+1}(S_{k+1}^j)\Big)\Big)\right] \\ \text{and } \widehat{\Theta}_N(x)=0, \quad \widehat{\Theta}_0(S_0)=\frac{1}{M_0}\sum_{j=1}^{M_0}\left(\widehat{\Theta}_1(S_1^j)+\frac{1}{N}g\Big(1,\widehat{\Theta}_1(S_1^j),\widehat{P}_1(S_1^j)\Big)\right).\end{cases}$$

The proof of Theorem 3.2 involves an intermediary random function $\widetilde{\Theta}_k(x)$ that satisfies

$$(3.12)\begin{cases}\text{For } k=1,...,N-1 \\ \widetilde{\Theta}_k(x)={}^t\psi(x)\Psi_k^{-1}\left[\frac{1}{M_0}\sum_{j=1}^{M_0}\psi(S_k^j)\Big(\Theta_{k+1}(S_{k+1}^j)+\frac{1}{N}g\Big(k+1,\Theta_{k+1}(S_{k+1}^j),P_{k+1}(S_{k+1}^j)\Big)\Big)\right] \\ \text{and } \widetilde{\Theta}_N(x)=0, \quad \widetilde{\Theta}_0(S_0)=\frac{1}{M_0}\sum_{j=1}^{M_0}\left(\Theta_1(S_1^j)+\frac{1}{N}g\Big(1,\Theta_1(S_1^j),P_1(S_1^j)\Big)\right).\end{cases}$$

$\mathfrak{T}$ used in (3.10) denotes a set of transformations that makes Ψ_k satisfy:

Assumption 3.2 *$\Psi_k \to_{a.s.} E\left(\psi(\widetilde{S}_k^i){}^t\psi(\widetilde{S}_k^i)\right)$ as $M_0 \to \infty$ and Ψ_k remains symmetric. For a fixed M_0, denoting by $\{\chi_{k,i}\}_{i=1,...,K}$ the eigenvalues of Ψ_k we have also the property*

$$\max_{i=1,...,K} E\left(\left|\chi_{k,i}^{-1}\right|^4\right) < \infty \text{ and setting } \Sigma_k = E\left(\Psi_k^{-1}\right)$$

we assume that for any $K \times K$ deterministic matrix A there exists a positive function $h_k^A(M_0,K)$ that vanishes as $K^3/M_0 \to 0$ such that $\left|E(\Psi_k^{-1}A\Psi_k^{-1}) - \Sigma_k A \Sigma_k\right| \leq h_k^A(M_0,K)$.

Basically, Assumption 3.2 announces a fact that a programmer must check when implementing the regression. It is even natural to assume that $\mathfrak{T}$ allows to have $\{\chi_{k,i}^{-1}\}_{i=1,...,K} \in L^\infty$, when this fact is true one can compute the function h_k^A that fulfills the previous inequality. Now, we are ready to announce the theorem.

Theorem 3.2 *As long as assumptions 3.1 and 3.2 are fulfilled and $K_j^3/M_j \to 0$ for each*

j when American contracts are involved, if $\{\Theta_i(x)\}_{0\leq i\leq N-1}$ *is of class* $\mathcal{C}^s$ *on the support of* $S\in\mathbb{R}^d$ *then there exists a positive constant* C *such that for each* $0\leq k\leq N-1$

$$E\left[\left(\widehat{\Theta}_k(S_k^i)-\Theta_k(S_k^i)\right)^2\right]\leq\frac{CK}{N^2}\sum_{l=k}^{N-1}\left(E\left[\frac{V_{l+1}(S_{l+1}^j)\partial_P^2 g\left(l+1,\Theta_{l+1}(S_{l+1}^j),P_{l+1}(S_{l+1}^j)\right)}{2M_l}\right]\right)^2$$
$$+O\left(\frac{K}{M_0}+\frac{K}{N^4M_l^2}+\frac{K^2}{N^2M_0}+\frac{K^{1-2s/d}}{N^2}+K^{-2s/d}\right).$$

The s-continuous differentiability assumed in Theorem 3.2 can be gotten either from the regularity of g or from the regularity of the transition density of S using (3.2).

From this result, a reasonable choice of the number of inner trajectories is $M_j\sim\sqrt{M_0/N}$ for $j\in\{1,...,N-1\}$. Moreover, using the same arguments as the one presented in Section 3.1, when the exposition involves only European contracts and the underlying asset S is a truncation of a log-Normal process, we can set

$$M_j=\frac{N-j}{N-1}M_1 \text{ with } M_1=\sqrt{\frac{M_0}{N}}. \tag{3.13}$$

Also, when American options are involved, we will see that the choice (3.13) is good numerically as long as M_0 is big enough.

Although the proof of Theorem 3.2 is detailed in the last section, we should make few remarks on the key tools that are needed for the proof. The first point that is really essential is the independence between Ψ_k and $\{S^i\}_{i\in\{1,...,M_0\}}$, otherwise the computation of the square error would be much more difficult. This is why Ψ_k is computed thanks to a new set of outer trajectories $\{\widetilde{S}^i\}_{i\in\{1,...,M_0\}}$. The other important point is the use of Theorem 4 in Newey (1997) which announces the following.

Theorem 3.3 *We denote* $f(X)=E(Y|X)$ *and* $\widehat{f}$ *the approximation of* f *thanks to a regression on a basis of* K *monomials. We assume that the support of* $X\in\mathbb{R}^d$ *is a cartesian product of compact connected intervals on which the density of* X *is bounded away from zero. More-*

over, we compute $\widehat{f}$ *thanks to a regression that involves independent copies* $(X_i, Y_i)_{i\in\{1,\dots,M_0\}}$ *of the couple* (X, Y)*. If* f *is of class* $\mathcal{C}^s$ *on the support of* X *and if* $K^3/M_0 \to 0$*, then*

$$\int [f(x) - \widehat{f}(x)]^2 dF_0(x) = O_p(K/M_0 + K^{-2s/d})$$

where F_0 *is the cumulative distribution of* X*.*

Theorem 3.3 with a Ψ_k that satisfies Assumption 3.2 allows to conduct all the computations in order to get Theorem 3.2.

4 Simulation framework and results

This section starts by presenting additional details on how the simulations are implemented and computing the complexity induced by this implementation. Then, it finishes by some compelling numerical results that shows the robustness of the method presented in this paper and that illustrates the theoretical asymptotic result established in the previous section.

4.1 Implementation and complexity

As already explained in Section 3, the proposed algorithm is based on nested (square) Monte Carlo simulation. As Monte Carlo is well suited to parallel architecture, we perform the implementation of our algorithm on the GPU "NVIDIA geforce 980 GTX" which includes 2048 parallel processing units. This massive computing power allows to reduce the execution time of the overall solution to make it quite interesting to use for real applications in the industry. This fact is true despite the high complexity of our method that has however the property to be very accurate.

The parallelization of this method is performed according to both the outer and the inner trajectories. First, one has to simulate the outer trajectories and to store them on the GPU if enough memory is available, otherwise to store them on the machine RAM (Random Access

Memory). Then, at each time step $j \in \{1, ..., N\}$ of m outer trajectories with $m = M_0/m_0$, we simulate the M_j inner trajectories. The reason of considering the outer trajectories per a group of m realizations is due to the limitation of the memory space available on GPU which makes impossible dealing with $M_0 \times M_j$ memory space occupation when M_0 and M_j are sufficiently big. Consequently, for each $j \in \{1, ..., N\}$, we find ourselves obliged to repeat m_0 times the same operation but on different $m \times M_j$ data.

When the exposure involves only European contracts, the m_0 repeated operations compose a common Monte Carlo simulation for the $m \times M_j$ inner trajectories. However, when the exposure includes also American contracts, one has to perform a lot of regressions in addition to the standard Monte Carlo operations. Formally speaking, if the exposure includes also American contracts, one has to perform $N - j - 1$ regressions at each time step $j \in \{1, ..., N-1\}$ and for each outer trajectory using M_j inner trajectories involved in a projection on K_j monomials. The complexity of each regression is proportional to $M_j \times K_j^3$ and then the computations performed per each time step are of the order of $O((N-j-1)M_0M_j(d+K_j^3))$ where d is the number of underlying assets S. To explain exactly how the different operations are performed and optimized on the GPU, we are preparing a paper that will be submitted within a couple of months to a computing journal.

Let us now study the overall complexity of the algorithm assuming that $K_j = K_0$ does not change according to j with $M_j \gtrsim K_0^3$ and that M_0 is much bigger than N^2. The complexity induced by the computation of (3.4) and of (3.11) are almost the same and one can get the following orders:

- For $M_j = \dfrac{N-j}{N-1}M_1$ with $M_1 = \dfrac{\sqrt{M_0}}{N}$, then the complexity $\sim O\left((d+K_0)NM_0^{3/2}\right)$.
- For $M_j = \dfrac{N-j}{N-1}M_1$ with $M_1 = \dfrac{\sqrt{M_0}}{\sqrt{N}}$, then the complexity $\sim O\left((d+K_0)(NM_0)^{3/2}\right)$.
- For $M_j = \dfrac{N-j}{N-1}M_1$ with $M_1 = \sqrt{M_0}$, then the complexity $\sim O\left((d+K_0)N^2M_0^{3/2}\right)$.

These orders justify the necessity to employ the GPU computing power to reduce the execution time and make the overall solution usable by the banking industry.

4.2 Numerical results

In this section, we give some numerical results associated to three different expositions. For these examples we use the same three dimensional Black & Scholes model for the underlying assets $S = (S^1, S^2, S^3)$

$$dS_t^i = rdt + \sigma_i \sum_{j=1}^{i} \varrho_{ij} dW_t^j, \quad i = 1, 2, 3, \tag{4.1}$$

with r = the risk neutral interest rate = $\ln(1.1)$, $S_0^i = 100$, $\sigma_i = 0.2$ and $\varrho = \{\varrho_{ij}\}_{1 \leq i,j \leq d}$ comes from the Cholesky decomposition of the correlation matrix $(\delta_{i-j} + 0.5(1 - \delta_{i-j}))_{i,j=1,2,3}$ where δ is the Kronecker symbol. We point that our method is quite robust according to the dimension of the problem and when increasing the number of random factors one has to increase smoothly the number of outer trajectories M_0 to have the same accuracy order. Studying more quantitatively the affect of the dimension and the choice of the model will be done in a future work that deals with various practical examples.

For the following simulation, we take $T = 1$, $M_0 = 13 \times 10^4$ trajectories, $N = 10$ and $N_{sde} = 50$. Although one can simulate exactly (4.1) without discretization, we use here $N_{sde} > N$ because we need sufficient number of time steps in order to simulate $\overline{S}_T^3 = \sup_{0 \leq t \leq T} S_t^3$ by $\sup_{0 \leq k \leq N_{eds}} S_{t_k}^3$ that will be involved in two European path-dependent examples of tables 1 & 2.

Table 1: European exposition with a payoff $\Upsilon(S_T) = \left(\frac{S_T^1}{2} + \frac{S_T^2}{2} - \overline{S}_T^3\right)_+$ and $f_i(x_i) = 0.01 + 0.01(x_i)_+$, $g(k, p, \theta) = 0.01p(p - \theta)_+$.

M_1	Θ_0	Θ_0 std	$\text{CVA}_{0,T}$	$\text{CVA}_{0,T}$ std
$\frac{\sqrt{M_0}}{N}$	0.01364	4×10^{-5}	0.0296	2×10^{-4}
$\frac{\sqrt{M_0}}{\sqrt{N}}$	0.01307	4×10^{-5}	0.0294	2×10^{-4}
$\sqrt{M_0}$	0.01265	3×10^{-5}	0.0291	2×10^{-4}

In tables 1, 2 and 3, we study the changes in the values of the estimators of $\mathrm{CVA}_{0,T}$ and of the pre-default TVA Θ_0 when M_1 increases with $M_j = \dfrac{N-j}{N-1}M_1$. In all tables, the values postfixed by std represent the empirical standard deviation computed on 16 realizations of the estimator. From the values of these standard deviations, we conclude that $M_0 = 130K$ trajectories is sufficiently big as the variance of the simulations is quite small.

Table 2: European exposition with a payoff: $\Upsilon(S_T) = \left(\frac{3S_T^1}{10} + \frac{7S_T^2}{10} - \overline{S}_T^3\right)_+ - \left(\frac{7S_T^1}{10} + \frac{3S_T^2}{10} - \overline{S}_T^3\right)_+$ and $f_i(x_i) = 0.01 + 0.01(x_i)_+$, $g(k,p,\theta) = 0.05(p-\theta)_+$.

M_1	Θ_0	Θ_0 std	$\mathrm{CVA}_{0,T}$	$\mathrm{CVA}_{0,T}$ std
$\frac{\sqrt{M_0}}{N}$	2.72×10^{-3}	10^{-5}	0.0365	8×10^{-4}
$\frac{\sqrt{M_0}}{\sqrt{N}}$	2.44×10^{-3}	10^{-5}	0.0453	8×10^{-4}
$\sqrt{M_0}$	2.28×10^{-3}	10^{-5}	0.0520	8×10^{-4}
$\sqrt{N}\sqrt{M_0}$	2.24×10^{-3}	10^{-5}	0.0528	8×10^{-4}

Table 3: American exposition with a payoff: $\Upsilon(S_T) = \left(\kappa - \frac{S_T^1}{3} - \frac{S_T^2}{3} - \frac{S_T^3}{3}\right)_+$, $\kappa = 100$ and $f_i(x_i) = 0.01 + 0.01(x_i)_+$, $g(k,p,\theta) = 0.01p(p-\theta)_+$.

M_1	Θ_0	Θ_0 std	$\mathrm{CVA}_{0,T}$	$\mathrm{CVA}_{0,T}$ std
$\frac{\sqrt{M_0}}{\sqrt{N}}$	0.0242	10^{-4}	0.0356	2×10^{-4}
$\sqrt{M_0}$	0.0229	10^{-4}	0.0351	2×10^{-4}

Regarding the value of M_1 that affects the bias of the estimators, we see in Table 1 that $M_1 = \frac{\sqrt{M_0}}{N}$ is quite sufficient to compute $\mathrm{CVA}_{0,T}$ and taking $M_1 = \frac{\sqrt{M_0}}{\sqrt{N}}$ produces a very small bias on Θ_0 when compared to $\sqrt{M_0}$. We have also the same case for Θ_0 in Table 2, however $\mathrm{CVA}_{0,T}$ has a big bias if we take M_1 too small which is due to the fact that the density of the exposition has a significant value at zero. As far as Table 3 is concerned, we do not show the simulations for $M_1 = \frac{\sqrt{M_0}}{N}$ since this number is not sufficient to perform the regressions needed for the Longstaff-Schwartz algorithm required by our condition $M_j \gtrsim K_0^3$ (here $K_0 = 4$). Nevertheless, we see that $M_1 = \frac{\sqrt{M_0}}{\sqrt{N}}$ is quite sufficient to have very accurate results of Θ_0 and $\mathrm{CVA}_{0,T}$.

5 Proof of theorems 3.1 and 3.2

Proof of Theorem 3.1: The condition $K_j^3/M_j \to 0$ is related to the convergence of of the Longstaff-Schwartz algorithm studied in Stentoft (2004) which uses the results presented in Newey (1997) and Clément & al. (2002). Given (3.5), (3.3) becomes

$$(5.1)\quad \begin{cases} F_{k+1}^1(x_1,...,x_{k+1}) = e^{-\frac{1}{N}\sum_{i=1}^k f_i(x_i)}\left(1 - e^{-\frac{1}{N}f_{k+1}(x_{k+1})}\right), \\ F_{k+1}^2(x_1,...,x_{k+1}) = (x_{k+1})^+ F_{k+1}^1(x_1,...,x_{k+1}), \\ F_{k+1}^3(x_1,...,x_{k+1}) = (x_{k+1})^+ e^{-\frac{1}{N}\sum_{i=1}^{k+1} f_i(x_i)} \text{ and} \\ F_{k+1}^4(x_1,...,x_j,y) = e^{-\frac{1}{N}\sum_{i=1}^j f_i(x_i)} E\left(e^{-\frac{1}{N}\sum_{i=j+1}^{k+1} f_i(P_i(S_i))}\Big| S_j = y\right). \end{cases}$$

Besides, MSE $\left(\widehat{\text{CVA}}_{0,T} - \text{CVA}_{0,T}\right)$ can be decomposed into two terms: a variance term $E\left[\left(\widehat{\text{CVA}}_{0,T} - E\left[\widehat{\text{CVA}}_{0,T}\right]\right)^2\right]$ and a square bias term $\left(E\left[\widehat{\text{CVA}}_{0,T} - \text{CVA}_{0,T}\right]\right)^2$. Regarding the variance, one gets easily

$$E\left[\left(\widehat{\text{CVA}}_{0,T} - E\left[\widehat{\text{CVA}}_{0,T}\right]\right)^2\right] \leq \frac{N^2}{M_0} \max_{k\in\{0,...,N-1\}} \text{Var}\left(F_{k+1}^2\left(\widehat{P}_1(S_1^i),...,\widehat{P}_{k+1}(S_{k+1}^i)\right)\right).$$

As for the square bias term one has to perform a second order Taylor expansion according to P. Doing the computations, we obtain

$$\begin{aligned} E\left[\widehat{\text{CVA}}_{0,T} - \text{CVA}_{0,T}\right] = & \frac{1}{2}\sum_{j=1}^N \frac{1}{NM_j} E\left[(\Delta_j^p)^2 f_j''(P_j(S_j^i)) F_j^3(P_1(S_1^i),...,P_j(S_j^i))\right] \\ & + \frac{1}{2}\sum_{j=1}^N \frac{1}{M_j} E\left[(\Delta_j^p)^2 \varepsilon(P_j(S_j^i)) F_j^1(P_1(S_1^i),...,P_j(S_j^i))\right] \\ & - \frac{1}{2}\sum_{j=1}^N \frac{1}{NM_j} E\left[(\Delta_j^p)^2 f_j''(P_j(S_j^i)) \sum_{k=l}^{N-1} F_{k+1}^2(P_1(S_1^i),...,P_{k+1}(S_{k+1}^i))\right] \\ & + \frac{1}{2}\sum_{j=1}^N (N-j+1) O\left(\frac{1}{M_j^2}\right) \end{aligned}$$

and ε is the Dirac distribution at 0. Thanks to an integration by parts using Assumption 3.1:

$$E\Big[(\Delta_j^p)^2\varepsilon(P_j(S_j^i))F_j^1(P_1(S_1^i),...,P_j(S_j^i))\Big] = E\Big[(\Delta_j^p)^2F_j^1(P_1(S_1^i),...,P_j(S_j^i))|P_j(S_j^i)=0\Big]\varphi_j(0).$$

We point out that, conditionally to $\sigma(\{S_l^i\}_{l=1,...,j})$, (Δ_j^p) and $F_{k+1}^2(P_1(S_1^i),...,P_{k+1}(S_{k+1}^i))$ are independent as (Δ_j^p) involves the inner simulation which is conditionally to $\sigma(\{S_l^i\}_{l=1,...,j})$ independent from the outer simulation and thus independent from $\sigma(\{S_l^i\}_{l=j+1,...,k+1})$.

Finally, conditioning according to $\sigma(\{S_l^i\}_{l=1,...,j})$ in each expectation involved in the value of $E\Big[\widehat{\mathrm{CVA}}_{0,T} - \mathrm{CVA}_{0,T}\Big]$ and using the Markov property, then taking the square of it and using the inequality $(\sum_{i=1}^n a_i)^2 \leq n\sum_{i=1}^n a_i^2$, we get the required result.

■

Proof of Theorem 3.2: Like in the proof of Theorem 3.1, the condition $K_j^3/M_j \to 0$ is needed for the convergence of of the Longstaff-Schwartz algorithm. In this proof, we study separately the variance $E\left[\left(\widehat{\Theta}_k(x) - E\left[\widehat{\Theta}_k(x)\right]\right)^2\right]$ and the bias $E\left[\widehat{\Theta}_k(x) - \Theta_k(x)\right]$ involved in $E\left(\left[\widehat{\Theta}_k(x) - \Theta_k(x)\right]^2\right)$. It is quite heavy to develop each term involved in the variance. Because we are interested by the term that decreases the least as $M_0 \to \infty$, we assume that $\widehat{\Theta}_k(x)$ is independent from S_k^j as the terms $\{S_i^j\}_{k+1\leq i\leq N}$ are involved only once (weighted by $1/M_0$) when computing backwardly $\widehat{\Theta}_k(x)$.

Denoting

$$\overline{\Theta}_{k+1}(S_{k+1}^j) = \widehat{\Theta}_{k+1}(S_{k+1}^j) + \frac{1}{N}g\Big(k+1, \widehat{\Theta}_{k+1}(S_{k+1}^j), \widehat{P}_{k+1}(S_{k+1}^j)\Big),$$

then $\widehat{\Theta}_k(x) = {}^t\psi(x)\Psi_k^{-1}\left[\frac{1}{M_0}\sum_{j=1}^{M_0}\psi(S_k^j)\overline{\Theta}_{k+1}(S_{k+1}^j)\right]$ and using Assumption 3.2

$$E\Big[\Big(\widehat{\Theta}_k(x) - E\left[\widehat{\Theta}_k(x)\right]\Big)^2\Big] = E\left[tr\left(A_k(x)\left[\frac{1}{M_0^2}\sum_{i,j}^{M_0}\psi(S_{k+1}^i)^t\psi(S_{k+1}^j)\overline{\Theta}_{k+1}(S_{k+1}^i)\overline{\Theta}_{k+1}(S_{k+1}^j) \right.\right.\right.$$
$$\left.\left.\left. -m_{k+1}\,{}^t m_{k+1}\right]\right)\right] + h_k^{\psi(x)^t\psi(x)}(M_0,K)O\left(\frac{K}{M_0}\right)$$

with $A_k(x) = \Sigma_k\psi(x)^t\psi(x)\Sigma_k$, $m_{k+1} = E(\psi(S_{k+1}^i)\overline{\Theta}_{k+1}(S_{k+1}^i))$ and tr is the trace operator.

Thanks to the asymptotic independence of $\widehat{\Theta}_k(x)$ and S_k^i, we obtain

$$E\left[\left(\widehat{\Theta}_k(S_k^i) - E\left[\widehat{\Theta}_k(S_k^i)\right]\right)^2\right] = O\left(\frac{K}{M_0}\right).$$

Regarding the bias, if we denote $B_k(x) = E\left[\widehat{\Theta}_k(x) - \widetilde{\Theta}_k(x)\right]$ then

$$\begin{aligned}\left(E\left[\widehat{\Theta}_k(S_k^i) - \Theta_k(S_k^i)\right]\right)^2 &\leq 2\left(E\left[B_k(S_k^i)\right]\right)^2 + 2\left(E\left[\widetilde{\Theta}_k(S_k^i) - \Theta_k(S_k^i)\right]\right)^2 \\ &\leq 2\left(E\left[B_k(S_k^i)\right]\right)^2 + 2E\left(\left[\widetilde{\Theta}_k(S_k^i) - \Theta_k(S_k^i)\right]^2\right)\end{aligned}$$

and using Theorem 4 in Newey (1997) as well as the boundedness part in Assumption 3.2, we have

$$E\left(\left[\widetilde{\Theta}_k(S_k^i) - \Theta_k(S_k^i)\right]^2\right) = O\left(\frac{K}{M_0} + K^{-2s/d}\right).$$

Besides

$$B_k(x) = {}^t\psi(x)\Sigma_k E\left(\psi(S_k^j)\begin{bmatrix}\widehat{\Theta}_{k+1}(S_{k+1}^j) - \Theta_{k+1}(S_{k+1}^j) + \frac{1}{N}g\Big(k+1, \widehat{\Theta}_{k+1}(S_{k+1}^j), \widehat{P}_{k+1}(S_{k+1}^j)\Big) \\ -\frac{1}{N}g\Big(k+1, \Theta_{k+1}(S_{k+1}^j), P_{k+1}(S_{k+1}^j)\Big)\end{bmatrix}\right)$$

and performing a second order Taylor expansion according to $\Theta_{k+1}(S_{k+1}^j)$ and $P_{k+1}(S_{k+1}^j)$

$$B_k(x) = {}^t\psi(x)\Sigma_k E\left(\psi(S_k^j)\begin{bmatrix}\Delta_{k+1}^{\Theta}(S_{k+1}^j) + \frac{1}{N}\Delta_{k+1}^{\Theta}(S_{k+1}^j)\partial_{\Theta}g\Big(k+1, \Theta_{k+1}(S_{k+1}^j), P_{k+1}(S_{k+1}^j)\Big) \\ + \frac{1}{2NM_k}[\Delta_{k+1}^{P}(S_{k+1}^j)]^2\partial_P^2 g\Big(k+1, \Theta_{k+1}(S_{k+1}^j), P_{k+1}(S_{k+1}^j)\Big) \\ + \frac{1}{2N}[\Delta_{k+1}^{\Theta}(S_{k+1}^j)]^2\partial_{\Theta}^2 g\Big(k+1, \Theta_{k+1}(S_{k+1}^j), P_{k+1}(S_{k+1}^j)\Big) \\ + o\left(\left|\Delta_{k+1}(S_{k+1}^j)\right|_2^2\right)\end{bmatrix}\right)$$

where $\Delta_{k+1}(S_{k+1}^j) = (\Delta_{k+1}^{\Theta}(S_{k+1}^j), \Delta_{k+1}^{p}(S_{k+1}^j))$ and $|\cdot|_2$ is the Euclidean norm. In the previous equality, the first order term in $\Delta_{k+1}^{p}(S_{k+1}^j)$ vanishes as $E\left(\Delta_{k+1}^{p}(S_{k+1}^j)|S_{k+1}^j\right) = 0$ that removes also the term in $\Delta_{k+1}^{p}(S_{k+1}^j)\Delta_{k+1}^{\Theta}(S_{k+1}^j)$ as the random functions $\Delta_{k+1}^{p}(x)$ and $\Delta_{k+1}^{\Theta}(x)$ are independent because the first one is simulated using the inner trajectories and

the other one using the outer ones. Besides, one can only keep the first and the third term and ignore all the others and obtain

$$B_k(x)={}^t\psi(x)\Sigma_k E\left(\psi(S_k^j)\left[\begin{array}{c}\Delta_{k+1}^{\Theta}(S_{k+1}^j)+\frac{[\Delta_{k+1}^P(S_{k+1}^j)]^2}{2NM_k}\partial_P^2 g\Big(k+1,\Theta_{k+1}(S_{k+1}^j),P_{k+1}(S_{k+1}^j)\Big)\\ +O\left(\frac{1}{N^2M_l}+\frac{\sqrt{K}}{N\sqrt{M_0}}+\frac{K^{-s/d}}{N}\right)\end{array}\right]\right).$$

After conditioning according to S_{k+1}^j and ignoring $E\left(\widetilde{\Theta}_{k+1}(S_{k+1}^j)-\Theta_{k+1}(S_{k+1}^j)|S_{k+1}^j\right)$, we obtain the following induction

$$B_k(x)={}^t\psi(x)\Sigma_k E\left(\psi(S_k^j)\left[\begin{array}{c}B_{k+1}(S_{k+1}^j)+\frac{V_{k+1}(S_{k+1}^j)}{2NM_k}\partial_P^2 g\Big(k+1,\Theta_{k+1}(S_{k+1}^j),P_{k+1}(S_{k+1}^j)\Big)\\ +O\left(\frac{1}{N^2M_l}+\frac{\sqrt{K}}{N\sqrt{M_0}}+\frac{K^{-s/d}}{N}\right)\end{array}\right]\right)$$

and because $B_N(x)=0$, one can establish for each $0<k\leq N-1$

$$E\Big[B_k(S_k^j)\Big]=E\Big({}^t\psi(S_k^j)\Big)\,\Sigma_k\sum_{l=k}^{N-1}\left\{\prod_{i=k+1}^{l}\Phi_i\Sigma_i E\left(\psi(S_l^j)\left[\begin{array}{c}\frac{V_{l+1}(S_{l+1}^j)\partial_P^2 g\left(l+1,\Theta_{l+1}(S_{l+1}^j),P_{l+1}(S_{l+1}^j)\right)}{2NM_l}\\ +O\left(\frac{1}{N^2M_l}+\frac{\sqrt{K}}{N\sqrt{M_0}}+\frac{K^{-s/d}}{N}\right)\end{array}\right]\right)\right\}$$

where $\Phi_i=E\left(\psi(S_{i-1}^j)\;{}^t\psi(S_i^j)\right)$. For $k=0$, $E\Big[B_0(S_0^j)\Big]$ differs from the other $E\Big[B_k(S_k^j)\Big]\Big|_{k>0}$ by a last term ($l=0$) which does not involve $\psi(S_0^j)$ or Σ_0.

There exists then a positive constant C such that for each $0\leq k\leq N-1$

$$\begin{aligned}\Big(E\Big[B_k(S_k^j)\Big]\Big)^2 \;\leq\; \frac{CK}{N^2}\sum_{l=k}^{N-1}&\left(E\left[\frac{V_{l+1}(S_{l+1}^j)\partial_P^2 g\left(l+1,\Theta_{l+1}(S_{l+1}^j),P_{l+1}(S_{l+1}^j)\right)}{2M_l}\right]\right)^2\\ &+O\left(\frac{K}{N^4M_l^2}+\frac{K^2}{N^2M_0}+\frac{K^{1-2s/d}}{N^2}\right).\end{aligned}$$

Using this final expression as well as the asymptotic behavior of $E\left(\left[\widetilde{\Theta}_k(S_k^i)-\Theta_k(S_k^i)\right]^2\right)$ and of the variance term, we get the required result.

■

REFERENCES

L. A. Abbas-Turki, A. I. Bouselmi and M. A. Mikou (2014a): Toward a coherent Monte Carlo simulation of CVA. *Monte Carlo Methods and Applications,* 20(3), 195–216.

L. A. Abbas-Turki, S. Vialle, B. Lapeyre and P. Mercier (2014b): Pricing derivatives on graphics processing units using Monte Carlo simulation. *Concurrency and Computation: Practice and Experience*, 26(9), 1679–1697.

F. Black and J. C. Cox (1976): Valuing corporate securities: Some effects of bond indenture provisions. *Journal of Finance*, 31, 351–367.

D. Brigo, M. Morini and A. Pallavicini (2013): *Counterparty Credit Risk, Collateral and Funding: With Pricing Cases For All Asset Classes*. John Wiley and Sons.

D. Brigo and A. Pallavicini (2008): Counterparty Risk and Contingent CDS under correlation between interest-rates and default. *Risk Magazine*, February 84–88.

M. Broadie, Y. Du and C. C. Moallemi (2011): Efficient Risk Estimation via Nested Sequetial Simulation. *Management Science*, 57(6), 1172–1194.

G. Cesari & al (2009): *Modelling, Pricing and Hedging Counterparty Credit Exposure*. Springer Finance.

E. Clément, D. Lamberton, and P. Protter (2002): An analysis of a least squares regression algorithm for American option pricing, *Finance and Stochastics*, 17, 448–471.

S. Crépey (2012a): Bilateral Counterparty Risk under Funding Constraints–Part I: Pricing. Forthcoming in *Mathematical Finance*.

S. Crépey (2012b): Bilateral Counterparty Risk under Funding Constraints–Part II: CVA. Forthcoming in *Mathematical Finance*.

S. Crépey, Z. Grbac, N. Ngor and D. Skovmand (2014): A Lévy HJM multiple-curve model with application to CVA computation. Forthcoming in *Quantitative Finance*.

C. Dellacherie and P-A. Meyer (1980): Probabilités et Potentiel, chapitres V-VIII, Hermann. English translation : Probabilities and Potential, chapters V-VIII, North-Holland, (1982).

M. Fujii, A. Takahashi (2015): Perturbative expansion technique for non-linear FBSDEs

with interacting particle method. *Asia-Pacific Financial Markets*

P. Glasserman (2003):*Monte Carlo Methods in Financial Engineering*, Applications of Mathematics, Springer.

P. Glasserman and B. Yu (2004): Number of Paths Versus Number of Basis Functions in American Option Pricing. *The Annals of Applied Probability*, 14(4), 2090–2119.

M. B. Gordy and S. Juneja (2010): Nested Simulation in Portfolio Risk Measurement. *Management Science*, 56(10), 1833–1848.

P. Henry-Labordère (2012): Cutting CVA's complexity, Risk Magazine, July.

J. Hull and A. White (2012): CVA and Wrong Way Risk. *Financial Analysts Journal*, 68(5), 58–69.

F. A. Longstaff and E. S. Schwartz (2001): Valuing American options by simulation: A simple least-squares approach, *em The Review of Financial Studies*, 14(1), 113–147.

R. Merton (1974): On the pricing of corporate debt: The risk structure of interest rates. *Journal of Finance*, 3, 449–470.

W. K. Newey (1997): Convergence rates and asymptotic normality for series estimators. *Journal of Econometrics*, 79, 147–168.

W. H. Press, S. A. Teukolsky, W. T. Vetterling and B. P. Flannery (2002): *Numerical Recipes in C++: The Art of Scientific Computing*. Cambridge University Press.

L. Stentoft (2004): Convergence of the Least Squares Monte Carlo Approach to American Option Valuation. *Management Science*, 50(9), 1193–1203.