

Mémoire présenté devant l'ENSAE Paris
pour l'obtention du diplôme de la filière Actuariat
et l'admission à l'Institut des Actuaires
le 06/11/2024

Par : **Samy Mekkaoui**

Titre : **Pricing and hedging of XVA : from classic numerical methods to supervised learning algorithms with applications in finance and insurance**

Confidentialité : NON OUI (Durée : 1 an 2 ans)

Les signataires s'engagent à respecter la confidentialité indiquée ci-dessus

Membres présents du jury de la filière

Entreprise : Forvis Mazars

Nom : Caroline Hillairet

Signature :

*Membres présents du jury de l'Institut
des Actuaires*

Directeur du mémoire en entreprise :

Nom : David Dagane

Signature :

**Autorisation de publication et de
mise en ligne sur un site de
diffusion de documents actuariels
(après expiration de l'éventuel délai de
confidentialité)**

Signature du responsable entreprise

Secrétariat:

Signature du candidat

Bibliothèque:

Acknowledgements

I would like to thank all of the Mazars Financial Engineering and Actuarial teams for the great times shared during this end-of-study internship.

I would especially like to thank Fadi Sahbani for reviewing some parts of this actuarial dissertation and David Dagane for his trust throughout the internship allowing me to explore my ideas on this interesting subject related to XVA.s.

On the academic side, I want to thank the entire faculty, whether M2MO or ENSAE, for the high quality of the courses given.

I am particularly grateful to Caroline Hillairet for her exceptional availability and insightful advices during the writing of this actuarial dissertation.

Abstract

Following the financial crisis of 2008, banks and insurers became aware of the importance of taking into account counterparty risk in the valuation of transactions on over-the-counter markets resulting in the calculation of the credit valuation adjustment (*CVA*). Subsequently, other risks that were previously not taken into account gradually began to become a discussion of interest for banks and insurers, notably the question of the cost associated with liquidity : the funding valuation adjustment (*FVA*) but also the cost of capital with the capital valuation adjustment (*KVA*) and the cost linked to the deposit of an initial margin in collateralized contracts : the margin valuation adjustment (*MVA*). The computation of *XVA* which is a generic name for X-Valuation Adjustments is generally associated with fairly high computational costs and the banking and insurance industries are constantly looking for new numerical methods to reduce this cost in order to be able to estimate these quantities consistently with an acceptable computation time. In a risk management context, banks and insurers, after having evaluated the *XVAs*, must be able to proactively manage the risks associated with these value adjustments, whether through hedging instruments or by using strategies to mitigate these risks.

In the first part of this dissertation, a mathematical framework for the calculation of certain *XVAs* is presented, introducing the notion of *Exposure*. The *CVA* is calculated for different types of european or bermudan financial products across various asset classes. For this, the classical Monte-Carlo method or its american variant are used. The impact of the *Wrong Way Risk* (*WWR*) in *XVA* computations is also illustrated through multiple examples.

In a second part, different machine learning algorithms are introduced to overcome the principal weaknesses of the classical Monte-Carlo : the computational time due to the nested Monte-Carlo and the *curse of dimensionnality*. These algorithms will be based on the different representations of the *XVAs* whether from a probabilistic or a partial differential equation (*PDE*) point of view.

In the last part, a mechanism to mitigate counterparty credit risk is presented. For this, a dynamic hedging method called *Mean-Variance Minimization* is used. This hedging method is notably illustrated using an application on the counterparty risk on various financial instruments in the financial market but also on a stop-loss contract in the reinsurance market. Finally, a static hedging method based on the expected utility theory for hedging the counterparty risk of a reinsurer is also studied through the determination of optimal reinsurance and hedging contracts.

Keywords: *XVA, Wrong Way Risk, Monte-Carlo, Wrong Way Measure, Gaussian Process Regression, PDE, BSDE, Neural Networks, Deep XVA Solver, CDS, Quadratic Hedging*

⁰Some part of the code used in this dissertation are available on my GitHub page : <https://github.com/SamyMekk>.

Résumé

Suite à la crise financière de 2008, les banques et les assureurs ont pris conscience de l'importance de prendre en compte le risque de contrepartie dans l'évaluation des transactions sur les marchés de gré à gré, ce qui a conduit au calcul de la valeur d'ajustement de crédit (*CVA*). Par la suite, d'autres risques, qui n'étaient pas pris en compte auparavant, ont progressivement commencé à susciter l'intérêt des banques et des assureurs, notamment la question du coût associé à la liquidité : la valeur d'ajustement de liquidité (*FVA*), mais également celle liée au capital avec la valeur d'ajustement de capital (*KVA*) et celle liée au coût de dépôt d'une marge initiale dans les contrats collatéralisés (*MVA*). Le calcul des X-valeurs d'ajustement (*XVAs*) est généralement associé à des coûts de calcul assez élevés, et les industries bancaire et assurantiable cherchent constamment de nouvelles méthodes numériques pour réduire ces coûts afin de pouvoir estimer ces quantités de manière cohérente avec un temps de calcul acceptable. Dans un contexte de gestion des risques, les banques et les assureurs, après avoir évalué les *XVAs*, doivent être en mesure de gérer de manière proactive les risques associés à ces ajustements de valeur, que ce soit par le biais d'instruments de couverture ou en utilisant des stratégies visant à atténuer ces risques.

Dans une première partie de ce mémoire, un cadre mathématique pour le calcul de certaines *XVAs* est présenté, introduisant la notion d'*Exposure*. La *CVA* est calculée pour différents types de produits financiers européens ou bermudéens sur diverses classes d'actifs. Pour cela, la méthode classique de Monte-Carlo ou sa variante américaine sont utilisées. L'impact du *Wrong Way Risk* (*WWR*) dans le calcul de *XVAs* est également illustré à travers plusieurs exemples.

Dans une deuxième partie, différents algorithmes d'apprentissage supervisé sont introduits pour surmonter les principales faiblesses du Monte-Carlo classique : le temps de calcul dû au Monte-Carlo imbriqué et le problème de la dimensionnalité. Ces algorithmes seront basés sur différentes représentations des *XVAs*, que ce soit d'un point de vue probabiliste ou à partir d'équations aux dérivées partielles (*EDP*).

Dans la dernière partie, un mécanisme pour atténuer le risque de contrepartie sera présenté. Pour cela, une méthode de couverture dynamique appelée *Minimisation de la Variance Moyenne* est utilisée. Cette méthode de couverture est notamment illustrée à l'aide d'une application au risque de contrepartie sur divers instruments financiers mais également dans le cadre d'un contrat stop-loss sur le marché de la réassurance. Enfin, une méthode de couverture statique basée sur la théorie de l'espérance d'utilité pour la couverture du risque de contrepartie d'un réassureur est également étudiée à travers la détermination des contrats optimaux de réassurance et de couverture.

Mots-clés: *XVA*, *Wrong Way Risk*, *Monte-Carlo*, *Mesure Wrong Way*, *Régression par Processus Gaussiens*, *EDP*, *EDSR*, *Réseaux de Neurones*, *Deep XVA Solver*, *CDS*, *Couverture Quadratique*

⁰La plupart des codes utilisés dans ce mémoire sont disponibles sur ma page GitHub : <https://github.com/SamyMekk>.

Executive Summary

Context

Following the financial crisis of 2008, banks and insurers became aware of the importance of taking into account counterparty risk in the valuation of transactions on over-the-counter markets resulting in the computation of the Credit Valuation Adjustment (*CVA*). Subsequently, other risks that were previously not taken into account gradually began to become a discussion of interest for banks and insurers, notably the question of the cost associated with liquidity : the Funding Valuation Adjustment (*FVA*) but also the cost of capital with the Capital Valuation Adjustment (*KVA*) and the cost linked to the deposit of an initial margin in collateralized contracts : the Margin Valuation Adjustment (*MVA*). The calculation of X-Valuation Adjustments (*XVA*) is generally associated with fairly high computational costs and the banking and insurance industries are constantly looking for new numerical methods to reduce this cost in order to be able to estimate these quantities consistently with an acceptable calculation time. In a risk management context, banks and insurers, after having evaluated the *XVAs*, must be able to proactively manage the risks associated with these value adjustments, whether through hedging instruments or by using strategies to mitigate these risks.

Approach of this dissertation

This dissertation is structured around two major themes. The first focuses on the pricing of *XVAs*, transitioning from classic numerical methods based on Monte-Carlo approaches to supervised learning algorithms, namely deep neural networks and gaussian process regressions. Several numerical illustrations are provided to assess the relevance of these methods. Special attention is given to the modelling of *Wrong Way Risk* and to measuring its impact on the valuation of certain *XVAs* for common financial products. The second theme of this dissertation focuses on the hedging of counterparty exposure after having evaluated it. Following a brief definition of the characteristics of a Credit Default Swap (*CDS*), a dynamic hedging strategy based on this product will be analyzed to minimize the hedging error, with an application to reinsurance counterparty risk. Finally, a static hedging approach also for a reinsurer's counterparty risk, based on expected utility theory, will be examined. In this context, the optimal reinsurance and hedging contracts for an insurer will be determined.

XVA Pricing

In order to be able to value the *XVAs*, it is necessary to define a mathematical framework associated with these adjustment values. This dissertation ¹ will be based on the following

¹There is no absolute consensus in the literature (except for *CVA*) in the formulas which characterize the *XVAs*.

representation of the *XVAs* where we give the main notations : R^C the recovery rate, $\frac{1}{B_t}$ the discount factor at time t , V_t the portfolio value at t and τ^C the default time of C .²

$$\begin{aligned} CVA_0 &= (1 - R^C) \mathbb{E}^{\mathbb{Q}}[\mathbf{1}_{\tau^C \leq T} (V_{\tau}^C)^+ \frac{1}{B_{\tau^C}} | \mathcal{G}_0]. \\ FVA_0 &= \mathbb{E}^{\mathbb{Q}}[\mathbf{1}_{(\tau^C \wedge \tau^A) \leq T} \int_0^T \frac{1}{B_u} (s_b(u)(V_u)^+ - s_l(u)(V_u)^-) du | \mathcal{G}_0]. \\ MVA_0 &= \int_0^T f(s) \mathbb{E}^{\mathbb{Q}}[\frac{1}{B_s} IM(s) | \mathcal{G}_0] ds. \end{aligned}$$

The computation of CVA_0 which is the most fundamental *XVA* for banking and financial institutions is generally carried out using a reformulation of the previous equation of the *CVA* (noting $G(s) = \mathbb{Q}(\tau^C > s) = e^{-\int_0^s h(u)du}$ where h is called *hazard rate* and assuming the independence between the value of the portfolio and the default time of the counterparty) given by :

$$CVA_0 = -(1 - R^C) \int_0^T \mathbb{E}^{\mathbb{Q}}[\frac{1}{B_s} (V_s)^+] dG(s).$$

This representation incorporates the counterparty's probability of default through the term $dG(s)$, as well as the concept of exposure to default, captured by the exposure profile. The exposure profile is represented by the function *EPE*, defined as $EPE(t) = \mathbb{E}^{\mathbb{Q}}\left[\frac{1}{B_t}(V_t)^+\right]$. Discretizing over a temporal grid $0 = t_0 < t_1 < \dots < t_N = T$ allows for the following approximation of CVA_0 :

$$CVA_0 \approx - \sum_{i=0}^{N-1} (1 - R^C) EPE(t_i) (G(t_{i+1}) - G(t_i)).$$

This formula³ emphasizes the 3 main components of credit risk under Basel regulations :

- The loss given default represented by the term $1 - R^C$.
- The exposure at default at time t_i represented by the term $EPE(t_i)$.
- The default probability between t_i and t_{i+1} represented by the term $G(t_{i+1}) - G(t_i)$.

Classic methods for XVA pricing

Due to the mathematical structure of the *XVAs* we have introduced, Monte-Carlo methods appear to be the most suitable for evaluating the *XVAs*. However, calculating the exposure profile is particularly time-consuming because it often requires nested Monte-Carlo simulations, as in many cases, the price of the financial product or portfolio V_t at time t is itself evaluated using a Monte-Carlo method. When a closed formula is allowed for the calculation of V_t , we can limit ourselves to a simple Monte-Carlo procedure. This dissertation begins by illustrating this case with various financial products, ranging from european options and forward contracts in a Black-Scholes (*B-S*) model to interest rate swaps under Hull & White and *G2++* models. This

²The main notations used in this executive summary are introduced in the chapter 2 of this dissertation : *XVA Mathematical framework* and are recalled in the last section *Notations* of this executive summary.

³Other approximations could be considered like the trapezoidal rule for instance.

thesis then explores the case of options that can be exercised before maturity, where valuation becomes more complex and requires nested Monte-Carlo simulations. To address this, the dissertation examines the classic *Least Square Monte-Carlo (LSMC)* method for calculating efficient exposure profiles. This method, based on a dynamic programming approach, delivers particularly compelling results as it enables relatively fast computations of exposure profiles without the need for a computationally prohibitive nested Monte-Carlo procedure. *LSMC* remains the preferred method for banking institutions in their calculation of exposure profiles. Finally, this thesis illustrates the impact of *netting agreements* between two counterparties on the overall value of *CVA* and demonstrates how the netting principle can lead to a significant reduction in overall *CVA*.

Incorporating Wrong Way Risk

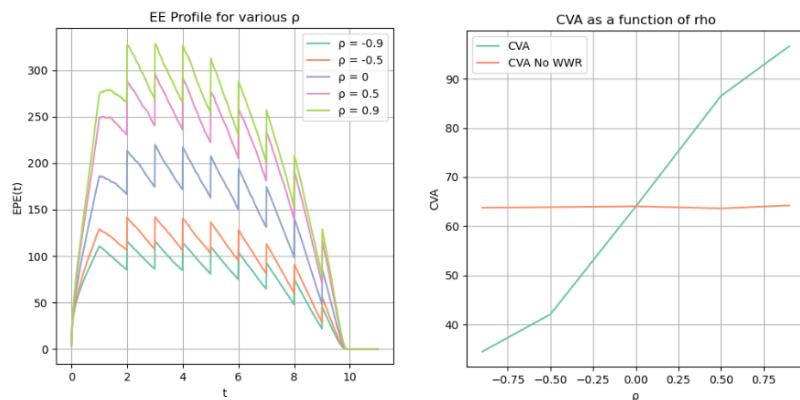
The assumption of independence between default exposure and counterparty default risk is generally far too crude and appropriate modelling of this dependence is generally necessary to better capture the true impact of *Wrong Way Risk (WWR)* on *XVAs*. This thesis proposes to study 2 modelling approaches, one based on a dynamic modelling of the default intensity denoted by $\lambda = (\lambda_t)_{t \in [0,T]}$ a stochastic process and the other based on a new probability measure called *Wrong Way Measure* which captures directly the *Wrong Way Risk* through a drift adjustment.

The *CVA* formula without the independence assumption is calculated as follows :

$$CVA_0 = -(1 - R^C) \int_0^T \mathbb{E}^{\mathbb{Q}} \left[\frac{1}{B_s} (V_s)^+ | \tau^C = s \right] dG(s).$$

Under the hypothesis of a Cox setup, that is to say that the process $S = (S_t)_{t \in [0,T]} = \mathbb{Q}(\tau^C > t | \mathcal{F}_t)$ can be written as $S_t = e^{-\int_0^t \lambda_s ds}$, we can then deduce a more general form of CVA_0 by taking into account the *Wrong Way Risk* given by :

$$CVA_0 = -(1 - R^C) \mathbb{E}^{\mathbb{Q}} \int_0^T \left[\frac{(V_s)^+}{B_s} \frac{\lambda_s S_s}{h(s) G(s)} \right] dG(s).$$



EPE Profile and corresponding CVA_0 of an IRS in the Hull & White model under *WWR*

Another approach based on the definition of a new probability measure is proposed by defining the measure $\mathbb{Q}^{WWR,t}$ whose Radon-Nikodym density M_s^t is given by:

$$M_s^t = \frac{d\mathbb{Q}^{WWR,t}}{d\mathbb{Q}}|_{\mathcal{F}_s} = \frac{\mathbb{E}^{\mathbb{Q}}[\frac{\lambda_t S_t}{B_t} | \mathcal{F}_s]}{\mathbb{E}^{\mathbb{Q}}[\frac{\lambda_t S_t}{B_t}]}.$$

Using this new probability measure, CVA_0 can be computed based on an appropriate exposure profile. Therefore, it is crucial to simulate $(V_t)^+$ under this new probability measure. In this thesis, this is achieved through a deterministic drift adjustment, which allows for analytical formulas under specific dynamics of V_t . The methodology is evaluated by using ground truth 2D full Monte-Carlo approximation for EPE and shows very good results.

Supervised Learning methods contribution for XVA pricing

This dissertation offers an in-depth study of supervised learning methods and their use in the context of the valuation of *XVAs*, particularly in relation to the limitations of classic Monte-Carlo methods.

Gaussian Process Regression

This thesis presents the Gaussian Process Regression (*GPR*) method for derivatives portfolio modelling, with applications for efficient expected exposure profile and CVA_0 computation. The main advantage of this method is its ability to efficiently learn price surfaces from a very small set of training data. However, it requires training data pairs (X, Y) , where the output labels Y may be noisy, for example, when prices are estimated using Monte-Carlo methods, which can make the approach less robust. This dissertation highlights the computational capabilities of *GPR*, both in the pricing of derivative products and in the valuation of insurance products such as Guaranteed Minimum Maturity Benefit (*GMMB*) contracts, while also addressing the method's limitations. An application of *GPR* for learning price surfaces at various times during the lifetime of a portfolio is presented, combined with a standard Monte-Carlo procedure to overcome the nested Monte-Carlo issue. This approach is applied to a portfolio of european derivatives under the $B - S$ model and a portfolio of swaps under the Hull & White model, showing great accuracy in the computation of the EE profile and the associated CVA_0 . Therefore, when combined with a simple Monte-Carlo loop, Gaussian Process Regression appears to be a promising candidate for overcoming the nested Monte-Carlo challenge, as long as accurate surface prices can be learned at each discretization point $0 = t_0 < t_1 < \dots < t_N = T$.

Neural Networks

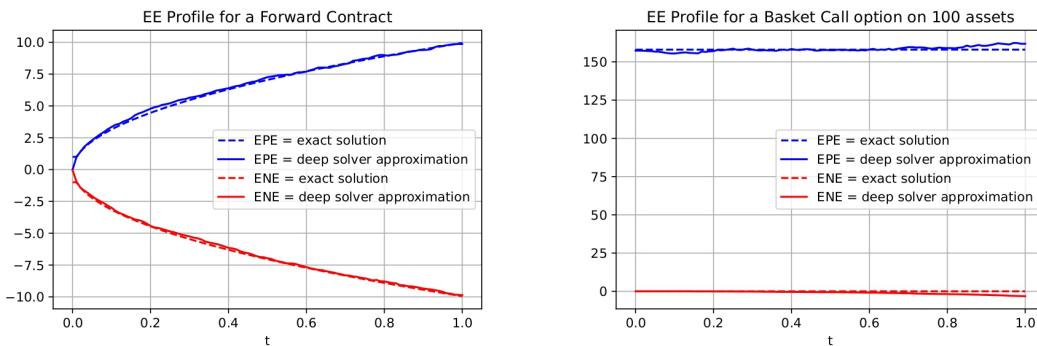
A probabilistic approach :

This dissertation examines the probabilistic formulation of *XVA* as conditional expectation, which, similar to Gaussian Process Regression, allows for efficient learning of price surfaces. Additionally, it leverages one of the key properties of conditional expectation—its role as a solution to a minimization problem. This property enables the avoidance of simulating prices to learn these surfaces, offering a computational advantage over Gaussian Process methods. Several examples of the *Deep Conditional Expectation* algorithm are presented in the thesis, particularly for an efficient calculation of the dynamic initial margin (*DIM*) and the associated MVA_0 for an interest rate swap. Normally, this task requires a nested Monte-Carlo procedure due to the complexity of the initial margin profile. Once the neural network is trained, the computational cost is almost negligible, whereas the computation of *DIM* using Monte-Carlo becomes

significantly more time-consuming, making nested Monte-Carlo impractical in an *XVA* engine where fast computations are crucial. The *Deep Conditional Expectation Solver* demonstrates excellent performance in this context, achieving relatively low error in the calculation of both the DIM profile and MVA_0 .

A PDE approach :

This dissertation also explores the formulation of *XVAs* as solutions of partial differential equations in a framework where default times are modelled by exponential variables with stochastic intensities. The solutions of these partial derivatives are then estimated using their intrinsic link with backward stochastic differential equations (*BSDEs*) which are very classic in finance where we give the payoff of an option at the terminal date. The idea of the method consists of rewriting the initial problem as a stochastic control problem where we seek to minimize a criterion allowing us to solve the *PDE* using *Deep Learning*. The main advantage of the *Deep XVA Solver* method lies in its ability to solve very high-dimensional problems compared to classic Monte-Carlo methods. Applications of this method are presented for the calculation of the expected exposure profile of a european option on a basket call option on $d = 100$ assets and show great accuracy on computing the *EE* profile and associated *XVAs*.



Exposure Calculation of a forward contract (left) and a basket option on $d = 100$ assets (right) both under $B - S$

Counterparty Exposure Hedging

This thesis addresses the problem of mean-variance minimization, specifically finding a self-financing portfolio strategy in *CDS* for the process $(CDS_t)_{t \in [0, T]}$, which minimizes the following quantity:

$$\min_{C_0, (\xi_t)_{t \in [0, T]}} \mathbb{E}^{\mathbb{Q}} \left[\left((1 - R^C)(V_{\tau^C})^+ \mathbf{1}_{\tau^C < T} - \left(C_0 + \int_0^T \xi_t d(CDS)_t \right) \right)^2 \right].$$

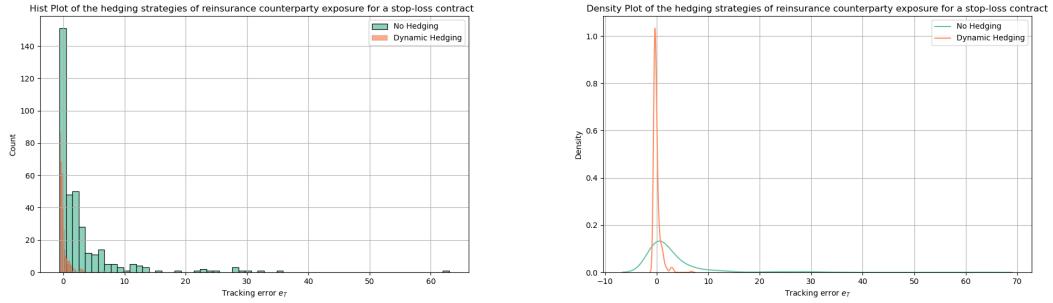
The insurer's loss is modelled using the process $L = (L_t)_{t \geq 0}$, defined as follows:⁴

$$L_t = \sum_{i=1}^{N_t} Z_i.$$

⁴The model considered is detailed in chapter 7 of the dissertation: *Towards Methods to Hedge CCR*.

where:

- N is a counting process such that $N_t = \sum_{n=1}^{+\infty} \mathbf{1}_{T_n \leq t}$, with $(T_n)_{n \in \mathbb{N}}$ representing the arrival times of the claims.
- $(Z_n)_{n \in \mathbb{N}}$ are strictly positive random variables modelling the amount of claims.
- $V_t = \mathbb{E}^{\mathbb{Q}} [e^{-r(T-t)}(L_T - K)^+ | \mathcal{G}_t]$.



Comparison of 2 hedging strategies in order to hedge the exposure on a stop-loss contract

This dissertation also proposes the study of a static hedging approach for an insurer subject to the counterparty risk of a reinsurer. The insurer seeks to find optimal reinsurance and hedging contracts that allow him to maximize his expected utility. Several cases are defined and a study is proposed in the case of an utility function of type *CARA*⁵.

Conclusion

Global conclusion

This dissertation provided an overview of numerical methods, transitioning from classical approaches to recent supervised learning methods, addressing the challenges associated with the computation of *XVAs*.

Firstly, a specific mathematical framework, inspired by academic literature, was proposed for the calculation of various *XVAs* and formed the foundation of this dissertation. From this framework, numerous illustrations were provided for the calculation of counterparty risk across different types of financial products, including european equity options and interest rate swaps under the *G2++* and Hull & White models, as well as more exotic products such as bermudan options, evaluated through a put and a swaption using the *Least Square Monte-Carlo* method. This dissertation also emphasized the significant impact of *Wrong Way Risk* in the valuation of *XVAs* by presenting two modelling approaches based on research articles. One of these approaches incorporates a change of measure, allowing the integration of *WWR* into the pricing of *CVA* under the *Wrong Way Measure*.

Secondly, this dissertation demonstrated the relevance of supervised learning algorithms for the calculation of *XVAs*, particularly in overcoming the classical issues of the nested Monte-Carlo approach. A detailed study was conducted on the contribution of Gaussian Process

⁵CARA means *Constant Absolute Risk Aversion* and is such that $u(x) = -e^{-\alpha x}$ where α is the absolute risk aversion parameter.

Regression, highlighting both the strengths and limitations of the algorithm. It was shown how combining *Gaussian Process Regression* with classical Monte-Carlo methods could efficiently learn the expected exposure profile of european derivatives or interest rate swap portfolios, thereby avoiding the need for nested Monte-Carlo procedures. An important part of this dissertation was also devoted to the study of fully connected deep neural networks and their usefulness in the calculation of *XVAs*. Two algorithms were studied : one which was based on the *PDE* representation of *XVAs* called *Deep XVA Solver* where we computed the expected profile of a very high dimensional european derivative which showed that it can overcome the curse of dimensionality and the other based on their probabilistic representation called *Deep Conditional Expectation Solver* where an efficient calculation of MVA_0 was computed avoiding also the nested Monte-Carlo procedure.

Finally, this dissertation focused on the hedging aspects of counterparty risk. A dynamic hedging approach was studied using quadratic hedging methods, with an investment in a benchmark hedging instrument for counterparty risk : Credit Default Swaps (*CDS*). An application to reinsurance counterparty risk in a stop-loss contract was presented, where we derived an analytical formula for the optimal investment strategy in the *CDS* to minimize the tracking error. A static approach, based on expected utility theory, was also proposed. In this approach, an insurer seeks to maximize the expected utility of its wealth by subscribing to a reinsurance contract and a hedging instrument to guard against the reinsurer's potential default. The insurer then determines the optimal contracts to solve his optimization problem.

In conclusion, this dissertation explored various quantitative aspects related to the management of *XVAs*, which have become a critical topic for banks and insurers since the 2008 financial crisis. The work highlighted the computational challenges of pricing *XVAs* and proposed new efficient numerical methods to address these challenges. Finally, it introduced key concepts in counterparty risk hedging, using both dynamic and static approaches.

Potential further research

- In addition to calculating the average exposure profile (*EE*), the financial industry is also interested in computing the exposure profile at a given percentile $\alpha \in [0, 1]$, defined as

$$PFE_t^\alpha = \inf\{y : \mathbb{Q}((V_t)^+ \leq y) \geq \alpha\}.$$

This complementary measure echoes the definition of *Value-at-Risk* and recently supervised learning methods have emerged for the calculation of these risk measures based on a dual representation of the *Value-at-Risk* and *Expected Shortfall* as minimization problems as introduced in the article *Learning Value-at-Risk and Expected Shortfall* by Barrera, Crépey, Gobet, Nguyen and Saadeddine.

- Supervised learning algorithms are also currently studied for the valuation of high-dimensional bermudan options as introduced in the article *Deep Optimal Stopping* by Becker, Cheridito, Jentzen where the optimal exercise time is learned from data samples.
- Neural networks can also be a great tool for hedging objectives as introduced in the article *Deep Quadratic Hedging* by Gnoatto, Lavagnini and Picarelli where quadratic hedging strategies are learned based on the *Deep BSDE Solver*. Recently, methods based on reinforcement learning have also emerged.

Notations

XVA	X Valuation Adjustment
CVA	Credit Valuation Adjustment
FVA	Funding Valuation Adjustment
MVA	Margin Valuation Adjustment
KVA	Capital Valuation Adjustment
V_t	Portfolio value at time t
$\frac{1}{B_t}$	Discount factor at time t
τ^C	Default time of counterparty C
R^C	Recovery Rate
$\mathbb{G} = (\mathcal{G}_t)_{t \geq 0}$	Augmented filtration with the information of default time τ^C
$\mathbb{F} = (\mathcal{F}_t)_{t \geq 0}$	Largest subfiltration of \mathbb{G} preventing the default time τ^C
s_B	Spread of borrowing
s_L	Spread of lending
f	funding spread between the collateral rate and the risk free rate
IM	Initial Margin
DIM	Dynamic Initial Margin
EE	Expected Exposure
EPE	Expected Positive Exposure
G	Survival probability function of τ^C
GMMB	Guaranteed Minimum Maturity Benefit
PDE	Partial Differential Equation
WWR	Wrong Way Risk
B-S	Black-Scholes
IRS	Interest Rate Swap
MC	Monte-Carlo
LSMC	Least Square Monte-Carlo
BSDE	Backward Stochastic Differential Equation
CDS	Credit Default Swap
PFE	Potential Future Exposure

Note de Synthèse

Contexte

Suite à la crise financière de 2008, les banques et les assureurs ont pris conscience de l'importance de prendre en compte le risque de contrepartie dans l'évaluation des transactions sur les marchés de gré à gré, ce qui a conduit au calcul de la valeur d'ajustement de crédit (*CVA*). Par la suite, d'autres risques, qui n'étaient pas pris en compte auparavant, ont progressivement commencé à susciter l'intérêt des banques et des assureurs, notamment la question du coût associé à la liquidité : la valeur d'ajustement de liquidité (*FVA*), mais également celle liée au capital avec la valeur d'ajustement de capital (*KVA*) et celle liée au coût de dépôt d'une marge initiale dans les contrats collatéralisés (*MVA*). Le calcul des X-valeurs d'ajustement (*XVAs*) est généralement associé à des coûts de calcul assez élevés, et les industries bancaire et assurantielle cherchent constamment de nouvelles méthodes numériques pour réduire ces coûts afin de pouvoir estimer ces quantités de manière cohérente avec un temps de calcul acceptable. Dans un contexte de gestion des risques, les banques et les assureurs, après avoir évalué les *XVAs*, doivent être en mesure de gérer de manière proactive les risques associés à ces ajustements de valeur, que ce soit par le biais d'instruments de couverture ou en utilisant des stratégies visant à atténuer ces risques.

Approche du mémoire

Ce mémoire s'articule autour de deux grands thèmes. Le premier thème concerne la valorisation des *XVAs*, en passant des méthodes numériques classiques basées sur des approches Monte-Carlo à des algorithmes d'apprentissage supervisé tels que les réseaux de neurones profonds et les régressions par processus gaussien. De multiples illustrations numériques sont proposées pour étudier la pertinence de ces méthodes. Une attention particulière est portée à la modélisation du *Wrong Way Risk* et à son impact sur la valorisation des *XVAs* pour des produits financiers courants. Le deuxième thème de ce mémoire se concentre sur la couverture de la *CVA*. Après une brève définition des caractéristiques d'un Credit Default Swap (*CDS*), une stratégie dynamique de couverture basée sur ce produit est étudiée afin de minimiser l'erreur de couverture, avec une application au risque de contrepartie sur le marché de la réassurance. Enfin, une approche statique de couverture du risque de contrepartie d'un réassureur, basée sur la théorie de l'espérance d'utilité, est également étudiée. Cette approche permet de déterminer les contrats de réassurance et de couverture optimaux pour un assureur.

Valorisation des *XVAs*

Afin de pouvoir valoriser les *XVAs*, il faut définir un cadre mathématique associé à ces valeurs d'ajustements. Ce mémoire⁶ va principalement se baser sur la représentation suivante des

⁶Il n'y a pas de consensus absolu (hors *CVA*) sur les formules qui caractérisent les *XVAs*.

XVAs dont on donne les principales notations : R^C le taux de récupération, $\frac{1}{B_t}$ le facteur d'actualisation en t , V_t la valeur du portefeuille en t et τ^C le temps du défaut de C .⁷

$$\begin{aligned} CVA_0 &= (1 - R^C) \mathbb{E}^{\mathbb{Q}}[\mathbf{1}_{\tau^C \leq T} (V_{\tau^C})^+ \frac{1}{B_{\tau^C}} | \mathcal{G}_0]. \\ FVA_0 &= \mathbb{E}^{\mathbb{Q}}[\mathbf{1}_{(\tau^C \wedge \tau^A) \leq T} \int_0^T \frac{1}{B_u} (s_b(u)(V_u)^+ - s_l(u)(V_u)^-) du | \mathcal{G}_0]. \\ MVA_0 &= \int_0^T f(s) \mathbb{E}^{\mathbb{Q}}[\frac{1}{B_s} IM(s) | \mathcal{G}_0] ds. \end{aligned}$$

Le calcul de la CVA_0 qui est la XVA la plus fondamentale pour les établissements bancaires et financiers est en général réalisé via une reformulation de l'équation de la CVA précédemment donnée (en notant $G(s) = \mathbb{Q}(\tau^C > s) = e^{-\int_0^s h(u)du}$ où h est appelé *taux de hasard* et en supposant l'indépendance entre la valeur du portefeuille et le défaut de la contrepartie) par :

$$CVA_0 = -(1 - R^C) \int_0^T \mathbb{E}^{\mathbb{Q}}[\frac{1}{B_s} (V_s)^+] dG(s).$$

Cette représentation fait intervenir les probabilités de défaut de la contrepartie à travers le terme $dG(s)$ ainsi que la notion d'exposition au défaut à travers la notion de profil d'exposition qui est la fonction EPE telle que $EPE(t) = \mathbb{E}^{\mathbb{Q}}[\frac{1}{B_t} (V_t)^+]$. La discrétisation sur une grille temporelle $0 = t_0 < t_1 < \dots < t_N = T$ permet alors d'approximer CVA_0 comme suit :

$$CVA_0 \approx - \sum_{i=0}^{N-1} (1 - R^C) EPE(t_i) (G(t_{i+1}) - G(t_i)).$$

Cette formule permet de mettre en lumière les 3 composantes principales du risque de crédit au titre de la réglementation bâloise :

- La perte en cas de défaut caractérisée par le terme $1 - R^C$.
- L'exposition au défaut en t_i caractérisée par le terme $EPE(t_i)$.
- La probabilité de défaut entre t_i et t_{i+1} caractérisée par le terme $G(t_{i+1}) - G(t_i)$.

Méthodes classiques de valorisation de XVAs

De par la structure mathématique des XVAs que nous avons introduite, les méthodes de Monte-Carlo apparaissent comme les plus adaptées pour évaluer les XVAs. Cependant, le calcul du profil d'exposition est particulièrement coûteux en terme de temps d'exécution car il nécessite de calculer des simulations de Monte-Carlo imbriqués étant donné que dans la plupart des cas le prix du produit financier/portefeuille V_t à l'instant t est lui-même évalué par une procédure Monte-Carlo. Lorsqu'une formule fermée est permise pour le calcul de V_t , on peut alors juste se restreindre à une simple procédure Monte-Carlo et ce mémoire commence par illustrer ce cas là pour des produits financiers variés allant d'options européennes et contrats à terme dans un modèle *Black – Scholes* à des swaps de taux d'intérêt dans des modèles Hull & White et *G2++*. Ce mémoire étudie ensuite le cas d'options dont l'exercice peut avoir lieu

⁷Les notations utilisées sont introduites dans le chapitre 2 du mémoire : *XVA Mathematical framework* et sont rappelées dans la section "Notations" de cette note de synthèse

avant la maturité là où la valorisation est plus complexe et nécessite l'utilisation de simulations Monte-Carlo imbriquées. Pour se faire, ce mémoire étudie la méthode classique de *Least Square Monte-Carlo* pour le calcul des profils d'exposition. Cette méthode basée sur une méthode de programmation dynamique apporte des résultats particulièrement intéressants et reste encore la méthode privilégiée par les institutions bancaires pour leur calcul des profils d'expositions. Enfin, ce mémoire illustre l'impact du *netting agreement* entre 2 contreparties sur la valeur globale de la *CVA*.

Incorporation du Wrong Way Risk

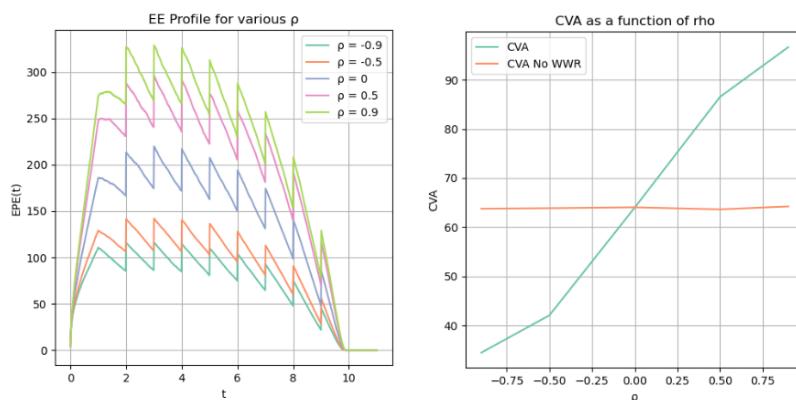
L'hypothèse d'indépendance entre l'exposition au défaut et le risque de défaut de la contrepartie est en général bien trop grossière et une modélisation appropriée de cette dépendance est en général nécessaire pour mieux capturer le vrai impact du *Wrong Way Risk* (*WWR*) sur les *XVAs*. Ce mémoire propose d'étudier 2 approches de modélisation, l'une basée sur une modélisation dynamique de l'intensité de défaut notée $\lambda = (\lambda_t)_{t \in [0, T]}$ par un processus stochastique et l'autre basée sur une évaluation de la *CVA* sans *Wrong Way Risk* apparent mais incorporé dans une nouvelle mesure de probabilité appelée *Wrong Way Measure* à travers un ajustement de dérive.

La formule de la *CVA* en omettant l'hypothèse d'indépendance peut alors être calculée de la manière suivante :

$$CVA_0 = -(1 - R^C) \int_0^T \mathbb{E}^{\mathbb{Q}} \left[\frac{1}{B_s} (V_s)^+ | \tau^C = s \right] dG(s).$$

Sous l'hypothèse d'un modèle Cox c'est à dire que le processus $S = (S_t)_{t \in [0, T]} = \mathbb{Q}(\tau^C > t | \mathcal{F}_t)$ peut être réécrit comme $S_t = e^{-\int_0^t \lambda_s ds}$, on peut alors en déduire une forme plus générale de la *CVA* en prenant en compte le *Wrong Way Risk* donnée par :

$$CVA_0 = -(1 - R^C) \mathbb{E}^{\mathbb{Q}} \int_0^T \left[\frac{(V_s)^+}{B_s} \frac{\lambda_s S_s}{h(s) G(s)} \right] dG(s).$$



Profil d'exposition EPE et *CVA* correspondante pour un swap de taux d'intérêt sous le modèle Hull & White avec *WWR*

Une autre approche basée sur la définition d'une nouvelle mesure de probabilité est proposé en définissant la mesure $\mathbb{Q}^{WWR,t}$ dont la densité de Radon-Nikodym M_s^t est donnée par :

$$M_s^t = \frac{d\mathbb{Q}^{WWR,t}}{d\mathbb{Q}}|_{\mathcal{F}_s} = \frac{\mathbb{E}^{\mathbb{Q}}[\frac{\lambda_t S_t}{B_t} | \mathcal{F}_s]}{\mathbb{E}^{\mathbb{Q}}[\frac{\lambda_t S_t}{B_t}]}.$$

A l'aide de cette nouvelle mesure de probabilité, la CVA_0 se calcule d'une manière similaire à la méthode utilisée par les établissement bancaires et financiers en spécifiant la loi de $(V_t)^+$ sous cette mesure ce qui est effectué via un ajustement de dérive déterministe. La méthode a été évaluée et a montré de très bons résultats dans la reproduction des profils d'expositions.

Contribution de l'apprentissage supervisé à la valorisation de dérivés et de XVA

Ce mémoire propose une étude approfondie des méthodes d'apprentissage supervisé et de leur utilisation dans le cadre de la valorisation des *XVAs* notamment par rapport aux limitations des méthodes classiques de Monte-Carlo.

Régression par processus gaussiens

Ce mémoire présente la méthode de régression par processus gaussiens (*GPR*) pour la modélisation de portefeuilles de produits dérivés, avec des applications pour le calcul efficace du profil d'exposition et de la CVA_0 associée. L'avantage principal de cette méthode est sa capacité à apprendre efficacement les surfaces de prix à partir d'un ensemble très réduit de données d'entraînement. Cependant, elle nécessite des paires de données d'entraînement (X, Y) , où les étiquettes de sortie Y peuvent être bruitées, par exemple lorsque les prix sont estimés à l'aide de méthodes de Monte-Carlo, ce qui peut rendre l'approche moins robuste. Ce mémoire met en évidence les capacités de calcul du *GPR*, tant dans la valorisation des produits dérivés que dans l'évaluation des produits d'assurance tels que les contrats de garantie minimale à maturité (*GMMB*), tout en abordant également les limites de la méthode. Une application du *GPR* pour apprendre les surfaces de prix à différents moments au cours de la durée de vie d'un portefeuille est présentée, combinée à une procédure standard de Monte-Carlo pour surmonter le problème des Monte-Carlo imbriqués. Cette approche est appliquée à un portefeuille de produits dérivés européens sous le modèle $B - S$ et à un portefeuille de swaps sous le modèle Hull & White, montrant une grande précision dans le calcul du profil d'exposition et de la CVA_0 associée.

Réseaux de Neurones

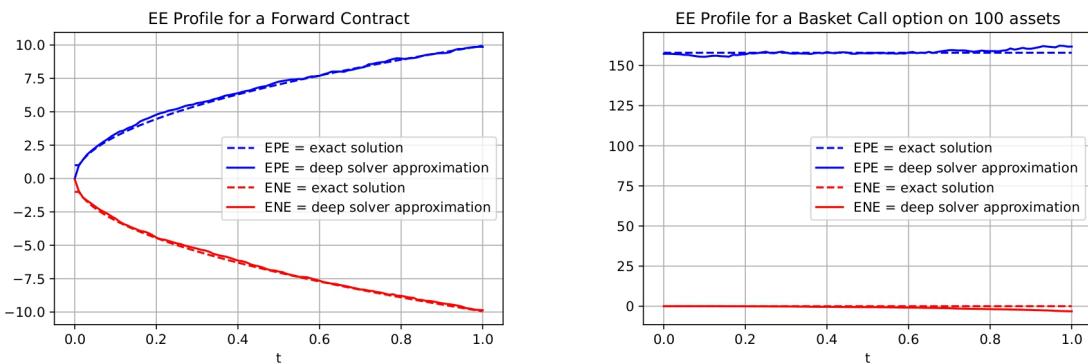
Une approche probabiliste :

Ce mémoire examine la formulation probabiliste des *XVA* en tant qu'espérance conditionnelle, ce qui, tout comme la régression des processus gaussiens, permet d'apprendre efficacement les surfaces de prix. De plus, elle exploite l'une des propriétés clés de l'espérance conditionnelle, à savoir son rôle en tant que solution d'un problème de minimisation. Cette propriété permet d'éviter la simulation des prix pour apprendre ces surfaces, offrant ainsi un avantage computationnel par rapport aux méthodes des processus gaussiens. Plusieurs exemples de la méthode d'apprentissage de l'espérance conditionnelle sont présentés dans le mémoire, notamment pour le calcul efficace de la marge initiale dynamique (*DIM*) et de la MVA_0 associée à un swap de taux d'intérêt. Normalement, cette tâche nécessite une procédure de Monte-Carlo imbriquée en raison du calcul lié au profil de la marge initiale. Une fois que le réseau de neurones est

entraîné, le coût computationnel est presque négligeable, tandis que le calcul de la *DIM* à partir du Monte-Carlo devient beaucoup plus chronophage, rendant la méthode de Monte-Carlo imbriquée inutilisable dans un moteur *XVA* où les calculs rapides sont cruciaux.

Une approche par EDP :

Ce mémoire explore également la formulation des *XVAs* comme solution d'équations aux dérivées partielles dans un cadre où les défauts sont modélisés par des variables exponentielles avec des intensités stochastiques. Les solutions de ces équations aux dérivées partielles sont alors estimées en utilisant leur lien intrinsèque avec les équations différentielles stochastiques rétrogrades qui sont très classiques en finance où on se donne le payoff d'un contrat financier à la date terminale. L'idée de la méthode consiste à réécrire le problème initial comme un problème de contrôle stochastique où l'on cherche à minimiser un critère nous permettant de résoudre l'*EDP*. L'avantage principal de la méthode du *Deep XVA Solver* réside dans sa capacité à résoudre des problèmes en très grande dimension par rapport à des méthodes de Monte-Carlo classique. Des applications de cette méthode sont présentées pour le calcul d'une option européenne en grande dimension afin de tester l'adaptabilité de la méthode.



Calcul du profil d'exposition pour un contrat forward (gauche) et d'une option basket sur $d = 100$ sous-jacents (droite) les 2 sous $B - S$

Couverture de l'exposition à la contrepartie

Ce mémoire se propose d'étudier le problème de la minimisation par variance moyenne c'est à dire de trouver une stratégie de portefeuille autofinançante dans un *CDS* dont on note la dynamique $(CDS_t)_{t \in [0, T]}$ qui minimise la quantité suivante :

$$\min_{C_0, (\xi_t)_{t \in [0, T]}} \mathbb{E}^{\mathbb{Q}} \left[\left((1 - R^C)(V_{\tau^C})^+ \mathbf{1}_{\tau^C < T} - \left(C_0 + \int_0^T \xi_t d(CDS)_t \right) \right)^2 \right].$$

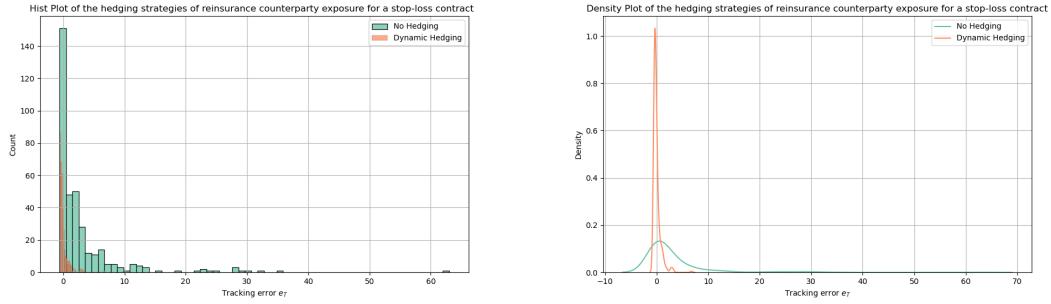
On modélise la perte pour l'assureur à travers le processus $L = (L_t)_{t \geq 0}$ défini ci-dessous : ⁸

$$L_t = \sum_{i=1}^{N_t} Z_i.$$

⁸Le modèle considéré est spécifié dans le mémoire dans le chapitre 7 : *Towards methods to hedge CCR*.

avec :

- Un processus de comptage N tel que $N_t = \sum_{n=1}^{+\infty} \mathbf{1}_{T_n \leq t}$ avec $(T_n)_{n \in \mathbb{N}}$ les dates d'arrivée des sinistres.
- $(Z_n)_{n \in \mathbb{N}}$ sont des variables aléatoires strictement positives qui modélisent les montant des sinistres.
- $V_t = \mathbb{E}^Q[e^{-r(T-t)}(L_T - K)^+ | \mathcal{G}_t]$.



Comparaison de 2 stratégies de couverture dans le but de couvrir l'exposition d'un contrat stop-loss

Ce mémoire propose également l'étude d'une approche statique de couverture pour un assureur soumis au risque de contrepartie d'un réassureur. L'assureur cherche alors à trouver les contrats optimaux de réassurance et de couverture qui lui permettent de maximiser son espérance d'utilité. Plusieurs cas sont définis et une étude est proposée dans le cas d'une fonction utilité de type *CARA*.⁹

Conclusion

Conclusion globale du mémoire

Ce mémoire a permis une vue d'ensemble des méthodes numériques en passant des méthodes classiques à des méthodes récentes d'apprentissage supervisé pour les challenges associés au calcul des *XVAs*.

Dans un premier temps, une revue succincte de la gestion des risques liée au défi des *XVAs* a été effectuée où l'importance d'ajustements pour les risque de contrepartie, liquidité, et de capital sont justifiés d'un point de vue réglementaire. Un cadre mathématique propre inspiré de la littérature académique a également été proposé pour le calcul des différentes *XVAs* et constitue le socle de ce mémoire. A partir de ce cadre mathématique, de nombreuses illustrations ont été proposées pour le calcul du risque de contrepartie sur différents types de produits : des options européennes de type equity ou des swaps de taux d'intérêt mais également des produits plus exotiques tels que les options bermudéennes à travers un put et un swaption en utilisant la méthode dite de *Least Square Monte-Carlo*. Ce mémoire a également mis en lumière l'impact conséquent du *Wrong Way Risk* sur la valorisation des *XVAs* en proposant deux approches de

⁹CARA signifie *Constant Absolute Risk Aversion* et est tel que $u(x) = -e^{-\alpha x}$ où α est le paramètre d'aversion absolu au risque.

modélisation basées sur des articles de recherche dont l'une basée sur un changement de mesure permettant d'incorporer le *WWR* dans le pricing de la *CVA* appelée *Wrong Way Measure*.

Dans un second temps, ce mémoire a permis d'illustrer la pertinence des algorithmes d'apprentissage supervisé pour le calcul des *XVAs*. Une étude particulière a été effectuée sur l'apport des *GPR* pour le calcul efficient de profil d'expositions. Une importante partie de ce mémoire est également consacrée à l'étude des réseaux de neurones profonds et à leur utilité dans le calcul des *XVAs*. Deux algorithmes ont été étudiés : l'un qui se base sur la représentation sous la forme d'*EDP* des *XVAs* appelée le *Deep XVA Solver* où un calcul efficient du profil d'exposition d'une option en très grande dimension a été effectué et l'autre basé sur leur représentation probabiliste appelé *Deep Conditional Expectation Solver* où un calcul efficient de la *MVA₀* d'un swap de taux d'intérêt a été présenté.

Enfin, ce mémoire s'est intéressé aux aspects de couverture du risque de contrepartie. Une approche dynamique de couverture a été étudiée basé sur les méthodes de couverture quadratique avec un investissement dans un instrument de couverture de référence pour le risque de contrepartie : les Credit Default Swaps (*CDS*). Une application au risque de contrepartie d'un réassureur basée sur cette approche dynamique a été proposée dans le mémoire. Une approche statistique basée sur la théorie de l'espérance d'utilité a également été proposée où un assureur cherche à maximiser l'espérance d'utilité de sa richesse en souscrivant à un contrat de réassurance et à un instrument de couverture sur le potentiel défaut du réassureur. Ce dernier cherche alors à trouver les contrats optimaux qui lui permettront de résoudre son problème d'optimisation.

En conclusion, ce mémoire a permis d'explorer différents aspects quantitatifs autour de la gestion des *XVAs* qui sont devenus depuis la crise financière de 2008 un sujet majeur pour les banques et les assureurs. Ce mémoire a permis de mettre en lumière les défis computationnels liés à la valorisation de ces *XVAs* ainsi que de proposer de nouvelles méthodes numériques efficientes pour répondre à ces défis. Enfin, il aura également permis une introduction aux notions de couverture du risque de contrepartie à l'aide d'une approche dynamique et d'une autre statique.

Potentiels axes de recherche

- L'industrie financière cherche également à calculer en plus du profil d'exposition moyen *EE* le profil d'exposition à un percentile donné $\alpha \in [0, 1]$ donné par :

$$PFE_t^\alpha = \inf\{y : \mathbb{Q}((V_t)^+ \leq y) \geq \alpha\}.$$

Cette mesure complémentaire fait écho à la définition de la *Value-at-Risk* et récemment des méthodes d'apprentissage supervisé ont émergé pour le calcul de ces mesures de risque en se basant sur une représentation duale de la *Value-at-Risk* et de l'*Expected Shortfall* comme des problèmes de minimisation comme introduit dans l'article *Learning Value-at-Risk and Expected Shortfall* de Barrera, Crépey, Gobet, Nguyen et Saadeddine.

- Les algorithmes d'apprentissage supervisé sont actuellement étudiés pour la valorisation d'options bermudéennes en grande dimension comme introduit dans l'article *Deep Optimal Stopping* de Becker, Cheridito, Jentzen où l'exercice optimal est appris sur un échantillon de données simulées.
- Les réseaux de neurones sont également étudiés pour des objectifs de couverture comme introduit dans *Deep Quadratic Hedging* de Gnoatto, Lavagnini et Picarelli où les stratégies de couverture quadratique sont apprises en se basant sur l'algorithme du *Deep BSDE Solver*. Récemment, des méthodes basées sur l'apprentissage par renforcement ont émergé.

Notations

XVA	Ajustement de la valorisation du coût de X
CVA	Ajustement de la valorisation du coût du crédit
FVA	Ajustement de la valorisation du coût de financement
MVA	Ajustement de la valorisation du coût la marge initiale
KVA	Ajustement de la valorisation du coût du capital
V_t	Valeur du portefeuille à la date t
$\frac{1}{B_t}$	Facteur d'actualisation au temps t
τ^C	Temps du défaut de la contrepartie C
R^C	Taux de récupération
$\mathbb{G} = (\mathcal{G}_t)_{t \geq 0}$	Filtration augmentée de l'information du marché avec τ^C
$\mathbb{F} = (\mathcal{F}_t)_{t \geq 0}$	Plus grande sous-filtration de \mathbb{G} qui ne contienne pas l'information de τ^C
s_b	Coût de l'emprunt
s_l	Coût du prêt
f	écart de financement entre le taux de garantie et le taux sans risque
IM	Marge initiale
DIM	Marge initiale dynamique
EE	Exposition moyenne
EPE	Exposition positive moyenne
G	Fonction de survie de τ^C
GMMB	Prestation minimale garantie à l'échéance
PDE	Equation aux dérivées partielles
WWR	Wrong Way Risk
B-S	Black-Scholes
IRS	Swap de taux d'intérêt
MC	Monte-Carlo
LSMC	Monte-Carlo par régression par moindres carrés
BSDE	Equations différentielles stochastiques rétrogrades
CDS	Couvertures de défaillance
PFE	Potentielle exposition future

Contents

Introduction	1
1 A risk management overview of XVA	2
1.1 An overview of the different risks associated to XVA	2
1.1.1 Counterparty credit risk	2
1.1.2 Funding liquidity risk	4
1.1.3 Cost of capital risk	5
1.2 An overview of the regulation around XVA	6
1.2.1 Capital requirements	6
1.2.2 Liquidity ratios	7
2 XVA Mathematical framework	9
2.1 CVA and BCVA modelling	9
2.1.1 Unilateral CVA	9
2.1.2 Bilateral CVA	13
2.1.3 Netting principle for CVA	14
2.1.4 Wrong and Right Way Risk for CVA	15
2.2 An FVA framework	15
2.2.1 FCA and FBA modelling	16
2.2.2 Wrong and Right Way Risk for FVA	18
2.3 An MVA framework	18
2.4 The KVA challenge	19
3 Some use cases for CVA computations	20
3.1 EE profile and CVA on some financial products under B-S model	20
3.1.1 EE profile of an european call	20
3.1.2 EE profile of a forward contract	23
3.2 EE profile and CVA of an IRS under Hull & White and G2++ models	26
3.2.1 Pricing of an IRS	26
3.2.2 EE profile of an IRS	27
3.3 EE profile of some bermudan options	39
3.3.1 The LSMC algorithm for pricing bermudan options	39
3.3.2 Some numerical results	41
3.4 EE profile of a portfolio	43
3.4.1 A quick overview of the model setup	43
3.4.2 Some numerical results	43
4 Taking into account the Wrong Way Risk	45
4.1 The Wrong Way Risk for CVA	45
4.1.1 Mathematical framework under the Cox setup	45
4.1.2 Some numerical results	47
4.2 A new way to tackle the WWR for CVA : a measure change	50

4.2.1	The Wrong Way Measure	50
4.2.2	Drift adjustment for exposure modelling	51
4.2.3	Some numerical results	53
4.3	The Wrong Way Risk for FVA	57
4.3.1	FVA decomposition under WWR	57
4.3.2	Some numerical results	58
5	Presentation of ML and DL methods for XVA computations	60
5.1	Deep Conditional Expectation Solver	61
5.1.1	Theoretical framework	61
5.1.2	An application of the method in a Markovian model	62
5.2	Gaussian Processes Regression	67
5.2.1	Theoretical framework	67
5.2.2	Some applications of GPR in finance and insurance	70
5.3	Deep BSDE Solver	75
5.3.1	Theoretical framework	75
5.3.2	Study of the algorithm	77
6	Application of ML and DL methods for XVA computations	80
6.1	Deep Conditional Expectation Solver for XVA computations	80
6.1.1	An application to CVA	81
6.1.2	An application to DIM and MVA	84
6.2	Gaussian Processes Regression for XVA computations	93
6.2.1	The GPR-MC for EE profile and CVA	93
6.3	Deep XVA Solver	102
6.3.1	EE profile computation of some derivatives	103
6.3.2	Direct computation of XVAs using PDE representation	106
7	Towards methods to hedge CCR	109
7.1	An introduction to Credit Default Swaps	109
7.2	Introduction to the Mean-Variance hedging framework	110
7.2.1	Mathematical framework	110
7.2.2	An application to the CCR in the financial market	111
7.2.3	An application to the CCR in the reinsurance market	117
7.3	Introduction to a static hedging approach for the CCR in the reinsurance market	123
7.3.1	Mathematical framework	123
7.3.2	Some numerical results	126
Conclusion		130
References		132
Glossary		136
Annexes		138
A	Key propositions used in this dissertation	138
B	A quick introduction to discretization schemes	143
C	An introduction to neural networks	144
D	Another deep PDE solver : The Deep Galerkin algorithm	150
E	Additional plots	158

List of Tables

1.1	An overview of the different <i>XVAs</i>	6
1.2	An overview of the regulatory changes since the 2008 crisis	8
3.1	Parameters of the swap used in the numerical experiments	31
3.2	Correlation matrix of the risk factors in the portfolio model	43
3.3	Parameters used in the numerical experiments for the global portfolio	43
4.1	Parameters used in the numerical experiments for the intensity of default process under <i>Wrong Way Risk</i>	48
4.2	Parameters used in the numerical experiments for the <i>FVA</i> pricing under <i>WWR</i>	58
5.1	Parameters used in the numerical experiments for the <i>B – S</i> pricing	62
5.2	Neural network architecture for derivatives pricing using <i>Deep Conditional Expectation Solver</i>	62
5.3	Correlation matrix of the underlying factors in the <i>B – S</i> model	65
5.4	Parameters used in the numerical experiments in the <i>B – S</i> setting	70
5.5	Parameters used in the numerical experiments for the GMMB contract pricing .	73
6.1	Parameters used in the numerical experiments for CVA_0 pricing using <i>Deep Conditional Expectation Solver</i>	81
6.2	Metric errors after the learning phase in <i>Deep Conditional Expectation Solver</i> for CVA_0 computation	81
6.3	Neural network architecture for the <i>EPE</i> profile computation of a call in the <i>B – S</i> model using the <i>Deep Conditional Expectation Solver</i>	82
6.4	Correlation matrix of the risk classes in the SIMM framework	85
6.5	Neural network architecture for the <i>DIM</i> computation in the <i>G2 ++</i> model . .	87
6.6	Lower and upper bounds for market state variable in the <i>G2 ++</i> model	87
6.7	Metric errors after the learning phase for the <i>DIM</i> computation	88
6.8	Set of parameters used to compare NN accuracy in <i>DIM</i> computation	89
6.9	Parameters of the 5 swaps used in the <i>DIM</i> computation	90
6.10	Parameters used in the <i>DIM</i> computation on a portfolio of 5 swaps	90
6.11	Metric errors after the learning phase on the portfolio of 5 swaps	91
6.12	Parameters used in the numerical experiments in the <i>GPR</i> methodology for <i>EE</i> profile computation of equity portfolios	94
6.13	Description of the composition of equity portfolios used in the numerical experiments	94
6.14	CVA_0 using the <i>GP – MC</i> methodology on the 1 st equity portfolio	95
6.15	CVA_0 using the <i>GP – MC</i> methodology on the 2 nd equity portfolio	96
6.16	CVA_0 using the <i>GP – MC</i> methodology on the 3 rd equity portfolio	97
6.17	Composition of the 1 st swap portfolio used in the numerical experiments	98
6.18	Composition of the 2 nd swap portfolio used in the numerical experiments	98
6.19	CVA_0 using the <i>GP – MC</i> methodology on the 1 st swap portfolio	99
6.20	CVA_0 using the <i>GP – MC</i> methodology on the 2 nd swap portfolio	100

6.21	Parameters used in the numerical experiments for the forward contract with the <i>Deep XVA Solver</i>	107
6.22	XVA_0 estimation from <i>Deep XVA Solver</i> for a forward contract with MC performed on 100000 samples	107
6.23	Parameters used in the numerical experiments for the basket call option with $d = 100$ assets with the <i>Deep XVA Solver</i>	108
6.24	XVA_0 estimation from <i>Deep XVA Solver</i> for a basket call option on $d = 100$ assets with MC performed on 100000 samples	108
7.1	Parameters used in the numerical experiments for the <i>Mean Variance Minimizing strategy</i> in the $B - S$ model	114
7.2	Norm 2 of e_T in case of an european call option in the $B - S$ model	115
7.3	Norm 2 of e_T in case of an european put option in the $B - S$ model	116
7.4	Norm 2 of e_T in case of a forward contract in the $B - S$ model	116
7.5	Parameter used in the numerical experiments in the <i>Mean Variance Minimizing strategy</i> for the stop loss contract	121
7.6	Norm 2 of e_T in case of a stop-loss contract	122
D.1	Parameters used in the <i>Deep Galerkin Method</i> illustrations	151
E.1	Parameters used in the numerical experiments in the <i>Mean Variance Minimizing strategy</i> in the Heston model	162
E.2	Norm 2 of e_T in case of an european call in the Heston model	163
E.3	Norm 2 of e_T in case of an european put in the Heston model	163

List of Figures

1.1	Contribution on <i>CVA</i> of each asset class (Figure from [1])	2
1.2	Illustration of funding and capital costs and their link with <i>XVAs</i> (Figure from [1])	5
1.3	An overview of the capital ratios requirements as part of Bale regulation (Figure from [1])	7
3.1	<i>EPE</i> and <i>ENE</i> profiles for a call option under a <i>B – S</i> model with the following parameters : ($S_0 = 100$, $K = 100$, $r = 0$ and $\sigma = 0.25$)	21
3.2	Evolution of CVA_0 for a call option as a function of intensity default λ with $R^C = 0.4$	22
3.3	<i>EPE</i> and <i>ENE</i> profiles for a forward contract under a <i>B – S</i> model with the following parameters : ($S_0 = 100$, $K = 100$, $r = 0$ and $\sigma = 0.25$)	24
3.4	Evolution of CVA_0 for a forward contract as a function of intensity default λ with $R^C = 0.4$	24
3.5	Simulation of r_t under Hull & White model and different choices of parameters	30
3.6	Sanity check implementation of Hull & White model	31
3.7	Value of a swap and <i>EPE</i> profile with $\sigma = 0.03$ and $\kappa = 0.5$ under Hull & White model with $N^{MC} = 20000$ simulations	31
3.8	Value of a swap and <i>EPE</i> profile with $\sigma = 0.06$ and $\kappa = 0.5$ under Hull & White model with $N^{MC} = 20000$ simulations	32
3.9	Value of a swap and <i>EPE</i> profile with $\sigma = 0.03$ and $\kappa = 1$ under Hull & White model with $N^{MC} = 20000$ simulations	32
3.10	Evolution of the CVA_0 profile of an IRS as a function of intensity default λ for $\sigma = 0.03$ (right) and $\sigma = 0.06$ (left) with $\kappa = 0.5$ and $R^C = 0.4$	33
3.11	Simulation of r_t under <i>G2++</i> model and different choices of parameters	34
3.12	Sanity check implementation of <i>G2++</i> Model	36
3.13	Value of a swap and <i>EPE</i> profile with $\rho = -0.99$ under the <i>G2++</i> model	37
3.14	Value of a swap and <i>EPE</i> profile with $\rho = 0$ under the <i>G2++</i> model	37
3.15	Value of a swap and <i>EPE</i> profile with $\rho = 0.99$ under the <i>G2++</i> model	37
3.16	Evolution of the CVA profile as a function of intensity default λ for $\rho = -0.99$ (right) and $\rho = 0.99$ (left) with $R^C = 0.4$	38
3.17	Calculation of the <i>EPE</i> profile of a bermudan put under <i>B – S</i> model with the following parameters : ($S_0 = 100$, $K = 100$, $r = 0.04$, $\sigma = 0.2$, $T = 1$ and $N = 13$)	41
3.18	Calculation of the <i>EPE</i> profile of a bermudan swaption under the Hull & White model with the following parameters : ($r_0 = 0.01$, $\kappa = 0.5$, $\sigma = 0.02$, $T = 10$) with payment dates and potential exercise dates given by $T_1 = 1$, $T_2 = 2$, . . . , $T_9 = 9$ with a nominal of $N = 10000$	42
3.19	Impact of netting in 2 simple portfolios with the following parameters : ($\rho_{12} = -0.8$, $\rho_{13} = -0.7$, $\rho_{23} = -0.6$ and $\rho_{xy} = 0$)	44
3.20	Impact of netting in 2 simple portfolios with the following parameters : ($\rho_{12} = 0.3$, $\rho_{13} = 0.4$, $\rho_{23} = 0.5$ and $\rho_{xy} = -0.99$)	44

4.1	<i>EPE</i> Profile of a call under <i>WWR</i> with the following parameters : ($S_0 = 100$, $K = 100$, $r = 0.03$, $\sigma = 0.2$, $T = 1$) and corresponding CVA_0 as a function of ρ .	48
4.2	<i>EPE</i> Profile of a forward under <i>WWR</i> with the following parameters : ($S_0 = 100$, $K = 100$, $r = 0.03$, $\sigma = 0.2$, $T = 1$) and corresponding CVA_0 as a function of ρ .	49
4.3	<i>EPE</i> Profile of an IRS under <i>WWR</i> with the following parameters : ($r_0 = 0.01$, $\kappa = 0.35$, $\sigma = 0.03$ with a fictitious initial <i>ZCB</i> curve given by $P(0, t) = e^{-r_0 t}$) and corresponding CVA_0 as a function of ρ .	49
4.4	Comparison of swap exposure profile between <i>2D Monte-Carlo</i> and the <i>Drift Adjustment</i> methods with the parameters : ($T = 5Y$, $y_0 = h = 0.15$, $\gamma = 0.001$, $\nu = 0.08$)	54
4.5	Comparison of swap exposure profile between <i>2D Monte-Carlo</i> and the <i>Drift Adjustment</i> methods with the parameters : ($T = 15Y$, $y_0 = h = 0.30$, $\gamma = 0.001$, $\nu = 0.08$)	54
4.6	Comparison of forward exposure profile between the <i>2D Monte-Carlo</i> and the <i>Drift Adjustment</i> with the parameters : ($T = 5Y$, $y_0 = h = 0.15$ and $\nu = 0.08$)	55
4.7	Comparison of forward exposure profile between <i>2D Monte-Carlo</i> and the <i>Drift Adjustment</i> with the parameters : ($T = 10Y$, $y_0 = h = 0.15$ and $\nu = 0.08$) . . .	55
4.8	Comparison of GBM exposure profile between the <i>2D Monte-Carlo</i> and the <i>Drift Adjustment</i> methods with the parameters : ($T = 5Y$, $y_0 = h = 0.15$, $V_0 = 0.02$, $r = 0.03$ and $\sigma^V = 0.20$)	56
4.9	comparison of GBM exposure profile between the <i>2D Monte-Carlo</i> and the <i>Drift Adjustment</i> methods with the parameters : ($T = 15Y$, $y_0 = h = 0.15$, $V_0 = 0.02$, $r = 0.03$ and $\sigma^V = 0.20$)	56
4.10	EPE_{FVA} profile for an interest rate swap (<i>WWR</i> (left) and <i>no WWR</i> (right)) .	58
4.11	Evolution of <i>FVA</i> as a function of ρ and ratio of $\frac{FVA}{FVA^{NoWWR}}$ to see the relative impact of <i>WWR</i> in <i>FVA</i> computation	59
5.1	Learning the price of a forward contract in the <i>B – S</i> model	63
5.2	Learning the price of an european put in the <i>B – S</i> model	64
5.3	Learning the price of a digital option in the <i>B – S</i> model	64
5.4	Learning the price of a max call option in the <i>B – S</i> model with $d = 6$ assets .	65
5.5	Learning the price of a min put option in the <i>B – S</i> model with $d = 6$ assets .	66
5.6	Pricing of a call option in <i>B – S</i> model and associated error using <i>GPR</i> . . .	70
5.7	Pricing of Delta on a call option in <i>B – S</i> Model and associated error using <i>GPR</i>	70
5.8	Pricing of a binary call option and associated error using <i>GPR</i> with $M = 10^6$ <i>M – C</i> simulations	71
5.9	Pricing of the delta of a binary call option in <i>B – S</i> model and associated error using <i>GPR</i> with $M = 10^6$ <i>M – C</i> simulations	71
5.10	Density and Survival Function of τ with the following set of parameters : ($c = 0.0750$, $\xi = 0.000597$, $\lambda_0 = 0.0087$)	72
5.11	1000 <i>MC</i> simulations to learn the price surface of a <i>GMMB</i> contract ($T = 10Y$)	73
5.12	10000 <i>MC</i> simulations to learn the price surface of a <i>GMMB</i> contract ($T = 10Y$)	73
5.13	100000 <i>MC</i> simulations to learn the price surface of a <i>GMMB</i> contract ($T = 10Y$)	73
5.14	1000 <i>MC</i> simulations to learn the price surface of a <i>GMMB</i> contract ($T = 20Y$)	74
5.15	10000 <i>MC</i> simulations to learn the price surface of a <i>GMMB</i> contract ($T = 20Y$)	74
5.16	100000 <i>MC</i> simulations to learn the price surface of a <i>GMMB</i> contract ($T = 20Y$)	74
6.1	CVA_0 learning in a <i>B – S</i> model by sampling τ^C for a forward (left) and an european put (right)	81

6.2	<i>EPE</i> profile learning using <i>Deep Conditional Expectation Solver</i> for a call option with ($S_0 = 100$ and $K = 100$ on left) and ($S_0 = 110$ and $K = 100$ on right)	83
6.3	Distribution of parameters r_0 and ρ	87
6.4	Noisy labels (with $M = 1$ inner path) for the <i>DIM</i> computation for the following set of parameters : ($\kappa_x = 0.10$, $\sigma_x = 0.02$, $\kappa_y = 0.12$, $\sigma_y = 0.02$, $\rho = -0.3$ and $r_0 = 0.01$)	88
6.5	Loss training process for the <i>DIM</i> computation with <i>Deep Conditional Expectation Solver</i> for a portfolio of 1 swap	88
6.6	Study of the <i>Deep Conditional Expectation Solver</i> algorithm with comparison with the nested Monte-Carlo for various set of parameters	89
6.7	Noisy labels (with $M=1$ inner path) for the following set of parameters ($\kappa_x = 0.10$, $\sigma_x = 0.015$, $\kappa_y = 0.12$, $\sigma_y = 0.015$, $\rho = -0.6$ and $r_0 = 0.015$)	91
6.8	Loss training process for the <i>DIM</i> computation with <i>Deep Conditional Expectation Solver</i> for a portfolio of 5 swaps	91
6.9	<i>DIM</i> computation with <i>Deep Conditional Expectation Solver</i> on a portfolio of 5 Swaps	92
6.10	<i>EE</i> profile of the 1 st equity portfolio using the <i>GP – MC</i> methodology and associated error in the calculation of the <i>EE</i> profile	95
6.11	<i>EE</i> profile of the 2 nd equity portfolio using the <i>GP – MC</i> methodology and associated error in the calculation of the <i>EE</i> profile	96
6.12	<i>EE</i> profile on the 3 rd portfolio using the <i>GP – MC</i> methodology and associated error in the calculation of the <i>EE</i> profile	97
6.13	<i>EE</i> profile of a single swap using the <i>GP – MC</i> methodology and associated error in the calculation of the <i>EE</i> profile	99
6.14	<i>EE</i> profile of a 5-swap Portfolio using the <i>GP – MC</i> methodology and associated error in the calculation of the <i>EE</i> profile	100
6.15	<i>EE</i> profile computation of a forward contract under B-S using the <i>Deep BSDE Solver</i>	104
6.16	<i>EE</i> profile computation of a basket option on $d = 100$ assets under B-S using the <i>Deep BSDE Solver</i>	105
7.1	Comparison of 2 hedging strategies in order to hedge the <i>CCR</i> on a call option with the <i>Mean Variance Minimizing</i> framework in Case 1	114
7.2	Comparison of 2 hedging strategies in order to hedge the <i>CCR</i> on a call option with the <i>Mean Variance Minimizing</i> framework in Case 2	115
7.3	Comparison of 2 hedging strategies in order to hedge the <i>CCR</i> on a put option with the <i>Mean Variance Minimizing</i> framework in Case 1	115
7.4	Comparison of 2 hedging strategies in order to hedge the <i>CCR</i> on a put option with the <i>Mean Variance Minimizing</i> framework in Case 2	115
7.5	Comparison of 2 hedging strategies in order to hedge the <i>CCR</i> on a forward contract with the <i>Mean Variance Minimizing</i> framework in Case 1	116
7.6	Comparison of 2 hedging strategies in order to hedge the <i>CCR</i> on a forward contract with the <i>Mean Variance Minimizing</i> framework in Case 2	116
7.7	Evolution of CVA_0 as a function of ρ for different values of γ (left) and as a function of γ for different values of ρ (right)	120
7.8	Comparison of 2 hedging strategies in order to hedge the <i>CCR</i> on a stop loss contract with the <i>Mean Variance Minimizing</i> framework in Case 1	122
7.9	Comparison of 2 hedging strategies in order to hedge the <i>CCR</i> on a stop loss contract with the <i>Mean Variance Minimizing</i> framework in Case 2	122
7.10	Evolution of d^* as a function of the aversion parameter α with $\rho_R = \rho_H = 0.1$. .	127

7.11 Evolution of d^* as a function of the aversion parameter α with $\rho_R = \rho_H = 0.2$	127
7.12 Evolution of d^* as a function of the aversion parameter ρ_R with $\alpha = 0.1$	128
7.13 Evolution of d^* as a function of the aversion parameter ρ_R with $\alpha = 0.2$	128
C.1 Neural network architecture with an input vector $\in \mathbb{R}^6$ with 5 hidden layers and with an output layer $\in \mathbb{R}^2$	145
C.2 Learning rate impact in the gradient descent algorithm	147
D.1 $CVA(t, S_t)$ surface for an european call with the following parameters : ($\lambda^C = 0.4$ and $R^C = 0$)	152
D.2 Projection of the CVA surface of an european call on $t = 0$ with in red line the $M - C$ estimation of $CVA(0, S_0)$ with the following parameters : ($\lambda^C = 0.4$ and $R^C = 0$)	152
D.3 $CVA(t, S_t)$ surface for an european call with the following parameters : ($\lambda^C = 0.1$ and $R^C = 0$)	153
D.4 Projection of the CVA surface of an european call on $t = 0$ with in red line the $M - C$ estimation of $CVA(0, S_0)$ with the following parameters : ($\lambda^C = 0.1$ and $R^C = 0$)	153
D.5 $FVA(t, S_t)$ surface for a forward contract with the following parameters : ($s^B = 0.02$, $s^L = 0$, $\lambda^C = 0.4$ and $\lambda^A = 0.1$)	154
D.6 Projection of the FVA surface of a forward contract on $t = 0$ with in red line the $M - C$ estimation of $FVA(0, S_0)$ with the following parameters : ($s^B = 0.02$, $s^L = 0$, $\lambda^C = 0.4$ and $\lambda^A = 0.1$)	154
D.7 $FVA(t, S_t)$ Surface for a forward contract with the following parameters : ($s^B = 0.02$, $s^L = 0.02$, $\lambda^C = 0.4$ and $\lambda^A = 0.1$)	155
D.8 Projection of the FVA surface of a forward contract on $t = 0$ with in red line the MC estimation of $FVA(0, S_0)$ with the following parameters : ($s^B = 0.02$, $s^L = 0.02$, $\lambda^C = 0.4$ and $\lambda^A = 0.1$)	155
D.9 $V(t, S_t)$ price surface for an european call	156
D.10 Projection of the price surface of an european call on $t = 0$ with in red line the MC estimation of $V(0, S_0)$	156
D.11 $V(t, S_t)$ price surface for a forward contract	157
D.12 Projection of the price surface of a forward contract on $t = 0$ with in red line the MC estimation of $V(0, S_0)$	157
E.1 Computation of the expected exposure profile of a bermudan call under $B - S$ model with the following parameters : ($S_0 = 100$, $K = 100$, $r = 0.04$, $q = 0$, $\sigma = 0.2$, $T = 1$ and $N = 13$)	158
E.2 Computation of the expected exposure profile of a bermudan call under $B - S$ model with dividends and with the following parameters : ($S_0 = 100$, $K = 100$, $r = 0.04$, $q = 0.10$, $\sigma = 0.2$, $T = 1$ and $N = 13$)	158
E.3 Computation of the expected exposure profile of a bermudan max call on 2 assets under $B - S$ model for different ρ with dividends and with the following parameters : ($S_0^1 = S_0^2 = 100$, $K = 100$, $r = 0.04$, $q^1 = q^2 = 0.10$, $\sigma^1 = \sigma^2 = 0.2$, $T = 1$ and $N = 13$)	159
E.4 Computation of the expected exposure profile of a bermudan min put option on 2 assets under $B - S$ model for different ρ with dividends and with the following parameters : ($S_0^1 = S_0^2 = 100$, $K = 100$, $r = 0.04$, $q^1 = q^2 = 0.10$, $\sigma^1 = \sigma^2 = 0.2$, $T = 1$ and $N = 13$)	160
E.5 Learning processes of neural networks for various european options from plots of chapter 5	161
E.6 Comparison of 2 hedging strategies in order to hedge the CCR on a call option under Heston model in Case 1	162

E.7	Comparison of 2 hedging strategies in order to hedge the CCR on a call option under Heston model in Case 2	162
E.8	Comparison of 2 hedging strategies in order to hedge the CCR on a put option under Heston model in Case 1	163
E.9	Comparison of 2 hedging strategies in order to hedge the CCR on a put option under Heston model in Case 2	163

List of Algorithms

3.1	Nested Monte-Carlo for exposure profile computation	25
3.2	Nested Monte-Carlo for CVA_0 computation	25
3.3	Sanity check implementation algorithm	30
3.4	Bermudan option pricing with dynamic programming approach	40
5.1	Deep Conditional Expectation Solver	61
6.1	Nested Monte-Carlo algorithm for DIM computation (Algorithm from [21]) . .	86
6.2	$\mathcal{GPR}-MC$ algorithm for exposure profile and CVA_0 computations	94
6.3	Deep algorithm for exposure simulation (Algorithm from [19])	103
6.4	Deep XVA Solver for non-recursive valuation adjustments (Algorithm from [19])	104
6.5	Deep XVA Solver (Algorithm from [19])	107
C.1	Gradient descent algorithm	147

Introduction

Following the financial crisis of 2008, banks and insurers became aware of the importance of taking into account counterparty risk in the valuation of transactions on over-the-counter markets resulting in the computation of the Credit Valuation Adjustment (*CVA*). Subsequently, other risks that were previously not taken into account gradually began to become a discussion of interest for banks and insurers, notably the question of the cost associated with liquidity : the Funding Valuation Adjustment (*FVA*) but also the cost of capital with the Capital Valuation Adjustment (*KVA*) and the cost linked to the deposit of an initial margin in collateralized contracts : the Margin Valuation Adjustment (*MVA*). The calculation of X-Valuation Adjustments (*XVA*) is generally associated with fairly high computational costs and the banking and insurance industries are constantly looking for new numerical methods to reduce this cost in order to be able to estimate these quantities consistently with an acceptable calculation time. In a risk management context, banks and insurers, after having evaluated the *XVAs*, must be able to proactively manage the risks associated with these value adjustments, whether through hedging instruments or by using strategies to mitigate these risks.

This dissertation is structured around two major themes. The first focuses on the pricing of *XVAs*, transitioning from classic numerical methods based on Monte-Carlo approaches to supervised learning algorithms, namely deep neural networks and Gaussian process regressions. Several numerical illustrations are provided to assess the relevance of these methods. Special attention is given to the modelling of *Wrong Way Risk* and to measuring its impact on the valuation of certain *XVAs* for common financial products.

The second theme of this dissertation focuses on the hedging of counterparty exposure after having evaluated it. Following a brief definition of the characteristics of a *CDS*, a dynamic hedging strategy based on this product will be analyzed to minimize the hedging error, with an application to reinsurance counterparty risk. Finally, a static hedging approach also for a reinsurer's counterparty risk, based on expected utility theory, will be examined. In this context, the optimal reinsurance and hedging contracts for an insurer will be determined.

Chapter 1

A risk management overview of XVA

This chapter will introduce the main risk management concepts introduced in [1] which is a book from Jon Gregory called *The XVA Challenge : Counterparty Risk, Funding, Collateral, Capital and Initial Margin*. In the first section of this chapter, we will describe the different risks which are associated with the main XVA's namely Credit, Funding, Capital and Initial Margin risks. In the second section, we will discuss the regulation around these XVA's by talking about the capital requirements and the liquidity ratios that need to be respected by banks in order to mitigate the risks they are facing as financial institutions.

1.1 An overview of the different risks associated to XVA

1.1.1 Counterparty credit risk

Counterparty Credit Risk (*CCR*) is often defined as the risk that an entity with whom one entered into a financial contract fail to honor her agreements. The *CCR* can happen in every situation but the most common way it occurs is on *OTC* derivatives market as they are less subject to regulations than exchanged markets. In the following figure, we show the contribution of each asset class in the global *CVA* in the *OTC* derivatives market.

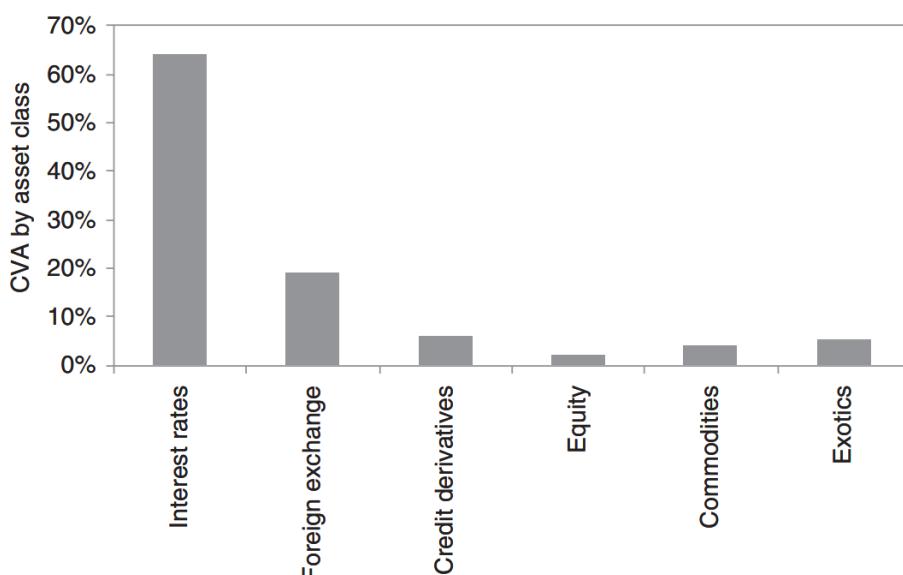


Figure 1.1: Contribution on *CVA* of each asset class (Figure from [1])

As we can see from the figure, the interest rate market has a significant proportion in the global counterparty risk which can be explained by the vast amount of transactions with important notional who are involved in this market. We have also an important amount of *CVA* from Foreign Exchange market (*FX*) which can be explained by the volatility of *FX* rates and long time maturity transactions.

Moreover, *CCR* is characterized by 2 aspects :

- The value of the contract in the future is uncertain. It will be the net value of all future cash flows under that contract and the value can be positive or negative depending of which type of contract we are considering. For example, a forward contract can have a negative or positive future value in the point of view of the buyer of the contract whereas a call option has always a positive value.
- Since the value of the contract can be positive or negative, counterparty risk can be bilateral and each counterparty can have a risk exposition to each other.

CVA Definition

The Counterparty Credit Risk can be calculated through the *Credit Valuation Adjustment (CVA)* which can be defined from one of the two following ways :

- The price is the cost of an associated hedging strategy which refers to the risk neutral pricing.
- The price represents the expected value of future cash flows, taking account the risk premium.

In the following of this dissertation and because it is the most dominant approach in the industry, we will use the risk-neutral approach. Moreover, this approach is justified by Basel III capital requirements where *CVA* is calculated through credit spreads which necessitates the use of risk-neutral default probabilities. Moreover, in *IFRS 13*, the *CVA* is defined from the default time price of the derivative who is priced under the risk neutral probability and so are the default probabilities.

Ways to Mitigate the *CCR* :

There exists quite a lot of methods to mitigate *CCR*. Some of them are based on contractual risk mitigants when others are more complex. We will quickly introduce some of these techniques in the following.

Netting

Netting is a technique which allows components like cashflows payments to be offset across a portfolio. This technique needs the validation of both parties involved in the transaction and is defined as *Netting Agreement*. There exists some different netting contracts like cash flow-netting or close-out netting and differ from how the counterparties aim to reduce their risk exposure.

Collateralisation

Collateralisation is a technique which involves the posting of cash or securities to cover the mark-to-market losses. However, there is still a residual risk named the *Margin Period of Risk (MPOR)* which is the time taken to receive the collateral amount. This technique is widely used in financial markets since the Global Crisis of 2008 as it is quite simple to setup with the help of central counterparties.

Hedging

Hedging is a technique that involves the use of instruments to protect against default events and credit movements. The main instruments used in the market to hedge the counterparty risk are based on *Credit Default Swap* which offer an insurance against the potential default of the counterparty. The Hedging technique will be introduced more in depth in the last chapter of this dissertation.

Central Counterparties

Central Counterparties act as intermediates for counterparties involved in a transaction by centralizing counterparty risk. Their use has grown exponentially since the global crisis of 2008 as they now become the norm in the major of financial transactions even in *OTC* derivatives markets.

An important fact about mitigation of counterparty risk is that it has both advantages and disadvantages. First, it helps to reduce counterparty risk and to improve financial market stability. However, due to the risk mitigants, markets concerned by counterparty risk have a big volume and it makes the financial system more exposed to a *systemic risk*.

1.1.2 Funding liquidity risk

Funding Liquidity Risk refers to the inability to fund contractual cashflows or collateral payments. Banks have 2 main types of financing : Debt and Equity and they can use different types of products as funding sources :

- *Debt Funding* with deposits and bonds mostly.
- *Equity Funding* with common equity and preferred shares.

Ways to Mitigate the Funding Liquidity Risk :

There exists quite a lot of methods to mitigate the funding liquidity risk. We will introduce quickly some of these techniques in the following.

Diversification by Debt Emission

The Debt emission can help to diversify the sources of funding and hence mitigate the funding liquidity risk. It can go from a large sample of financial actors such as hedge funds, insurers or portfolio managers for example.

Liquidity Reserve

In order to mitigate the funding liquidity risk, a bank or an insurer can use liquidity reserve as it will help if other sources of funding are falling down during a major global crisis. We can cite the following sources of liquidity reserve :

- Bank Central cash.
- Securities that can be sold as collateral at any time.

Note that since a bank gets funding from both Debt and Equity, the cost of each should be calculated based on the underlying costs, type of transaction and identity of counterparty. For instance, the cost of debt is known as the funding spread which is the spread from issuing a bond for example.

The quantification of the funding costs is done by the computation of the *Funding Valuation Adjustment (FVA)* and the *Margin Valuation Adjustment (MVA)*.

The *FVA* can be understood as a funding cost for uncollateralized or atleast partially collateralized transactions. Assets will provide funding costs whereas liabilities will provide benefit costs. Therefore, assets which will be in the money will be related to funding cost whereas out of the money assets will lead to benefit costs. *FVA* calculation has emerged after the global financial crisis because before transactions were achieved at the risk free rate. However, theses fundings have considerably increased and more on the uncollateralized transactions. *FVA* is usually related to *variation margin* which represents collateral variation due to variation of the portfolio value.

The *MVA*, unlike *FVA* can be understood as a cost of *overcollateralization* in the sense that *MVA* will represent the cost due to the posting of an *initial margin* over the lifetime of a transaction in the collateralized contracts. *MVA* is therefore related to the *initial margin* profile in opposition with *FVA* which is related to the *variation margin*.

1.1.3 Cost of capital risk

The cost of capital should represent the risk of a company's equity for investors and the bank will look to outperform the return by defining the *Return On Capital (ROC)*. The cost of capital is usually quantified by defining an appropriate percentage on *ROC* and is mostly a subjective parameter (10% being a classic assumption for banks). As profits will be taxed, the *ROC* on OTC derivatives is more around 15 – 20%. The quantification of the cost of capital is embedded into the *Capital Valuation Adjustment (KVA)* but this *XVA* is highly dependant of the capital cost profile of the bank which makes it hard to model.

We can summarize the following *XVAs* related to funding and capital costs in the following figure :

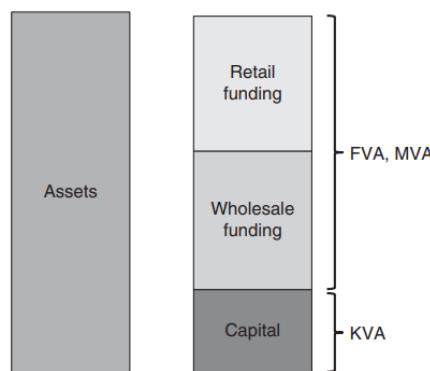


Figure 1.2: Illustration of funding and capital costs and their link with *XVAs* (Figure from [1])

We end up this section by giving the following table where we summarize all the major *XVAs* and their associated cost components.

Table 1.1: An overview of the different XVA

XVA	valuation adjustment	Expected Cost of the Bank
CVA	Credit Valuation Adjustment	Client Default Losses
DVA	Debt Valuation Adjustment	Bank Default Losses
FVA	Funding Valuation Adjustment	Funding cost for variation margin
MVA	Margin Valuation Adjustment	Funding cost for initial margin
KVA	Capital Valuation Adjustment	Remuneration of shareholder capital at risk

1.2 An overview of the regulation around XVA

1.2.1 Capital requirements

Under Bale Committee, there has been defined several capital requirements in order to prevent the different risks that banks can face. We will illustrate some of them as they are keys starting points of the mathematical definition of *XVAs* which we will introduce in the following chapter.

CCR Capital requirement

Counterparty Credit Risk Capital is defined as the product of three components :

- *Probability at default* (PD) : The default probability of the counterparty.
- *Loss Given Default* (LGD) : 100% minus the recovery rate.
- *Exposure at Default* (EAD) : The exposure to the counterparty.

Defining EAD for derivatives can be challenging in certain situations like when the value of a portfolio can either be negative or positive and being extremely volatile that's why there are specific methodologies on EAD for these situations to assess CCR Capital.

Capital Ratios requirements

- The capital ratio of the bank defined as the percentage of a bank's capital to its *Risk Weighted Assets (RWA)* which is supposed to be greater than 8% and composed of 2 types of capital : The *Common Equity Tier 1 (CET1)* which is mostly composed of common equity and the Tier 2 Capital composed mostly of preferred shares. Regulators have also imposed that *CET1* capital must hold to 4.5% in the 8% capital global ratio.
- The *Countercyclical capital buffer* which aims to ensure that banks take account the economic climate and is an additional percentage normally set between 0% and 2.5%.
- The *Capital conservation buffer* which is a ratio aiming to have capital buffers outside periods of stress and is set at 2.5% above the minimum capital requirement.
- *Globally-Systemically-important bank (G-SIB)* surcharge for systemic banks which can go from 1% to 3.5% additional costs.

It can be summarized in the following figure :

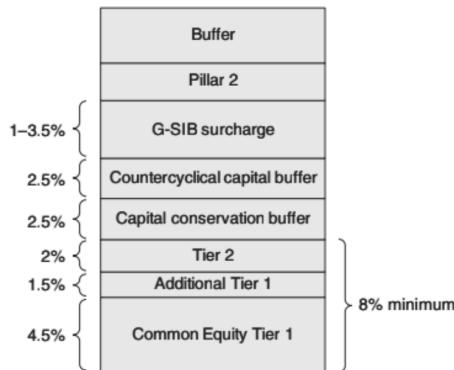


Figure 1.3: An overview of the capital ratios requirements as part of Bale regulation (Figure from [1])

Bilateral Margin requirements

- Variation Margin : It represents collateral variation due to variability in the portfolio transactions and is related to the *FVA*.
- Initial Margin : It represents extra collateral to cover costs in the case of a default in bilateral derivative markets and is related to the *MVA*.

1.2.2 Liquidity ratios

As banks use to fund long-dated illiquid assets with short-term liquid liabilities like deposits, there was a crucial need to manage the liquidity risk associated to these transactions. One of major problems for banks was to refinance short-term borrowing regularly, which created liquidity risk due to the fact that the funding costs would get more expensive. Therefore, the Bale Committee decided to impose international liquidity standards which are the following :

- The Liquidity Coverage Ratio (LCR) : This is a short term liquidity ratio to ensure that a company has the resources to survive to a negative liquidity scenario in the 30 following days.
- The Net Stable Funding Ratio (NSFR) : This is a long term liquidity ratio to ensure the proportion of long term assets funded by stable funding.

High-Quality Liquid Assets :

The concept of High-Quality Liquid Assets (HQLA) is particularly important in the Liquidity Ratios as it's a core concept in both LCR and NSFR. They are defined as the assets that can easily be converted into cash to ensure liquidity needs at any time. The HQLA can be decomposed into the following types :

- Level 1 : Assets of excellent liquidity and credit which may be : cash, central bank reserves, securities guaranteed by sovereign states.
- Level 2A : Less liquid assets than the Level 1 so they cannot represent more than 40% of the total of HQLA : It may be sovereign or corporate bonds.
- Level 2B : Most illiquid assets and they cannot represent more than 15% of the total of HQLA: It can concern equities, corporate bonds, mortgage backed securities.

The Liquidity Coverage Ratio :

As we said, the LCR is a short-term liquidity ratio to ensure that the institution can face an adversarial liquidity episode in the next 30 days. The LCR ratio can be defined as :

$$\frac{\text{Stock of HLQA}}{\text{Total net cash outflows over the next 30 days}} \geq 100\%. \quad (1.1)$$

Under the LCR, institutions must consider increased liquidity needs in relation to changes in the valuation of derivatives or in the collateral associated to theses transactions as we must hold the condition 1.1 which can have an impact both to *FVA* and *MVA*.

The Net Stable Funding Ratio :

The NSFR is a long-term liquidity ratio ensuring that banks hold a minimum amount of stable funding over one year to prevent liquidity transformation. The NSFR ratio can be defined as :

$$\frac{\text{Available amount of stable funding}}{\text{Required amount of stable funding}} \geq 100\%. \quad (1.2)$$

The NSFR is quite hard to handle for derivatives as they can be either assets or liabilities (depending on the sign associated in the transaction). Therefore, they are some features specific to derivatives which can receive a different treatment. For instance, the NSFR assumes that collateral generated from derivatives cannot fund another part in the institution balance sheet which can be an important factor in *FVA*. The NSFR also can have impact on the *MVA* and initial margin. Indeed, it's considering that receiving an initial margin doesn't afford available amount of stable funding which is normal because it is usually segregated.

In the table below, we give the main regulatory changes since the global crisis of 2008.

Table 1.2: An overview of the regulatory changes since the 2008 crisis

Strengthen Capital Bases	Strengthen Liquidity Standards	Clearing Mandate
Introduction of a <i>CVA</i> capital charge	Liquidity Coverage Ratio (<i>LCR</i>)	Mandatory clearing of standardised <i>OTC</i> derivatives
Strengthen Capital Ratio requirements	Net Stable Funding Ratio (<i>NSFR</i>)	Bilateral margin requirements for uncleared <i>OTC</i> derivatives

Chapter 2

XVA Mathematical framework

In this chapter, we are going to give the main mathematical framework we are going through in this dissertation. Except for *CVA*, there is not a clear mathematical definition of other *XVAs* and therefore we will refer to the existing litterature which is mostly based from the paper [2] of Green, Kenyon and Dennis where they derive expressions for major *XVAs*.

2.1 CVA and BCVA modelling

Credit Valuation Adjustment (*CVA*) is a financial metric used in order to measure the potential credit risk associated with financial instruments like derivatives. This metric allows to quantify the expected loss an agent will occur if the counterparty fails to meet his financial obligations. Mathematically, the *CVA* is defined as the difference between the risk free and the risky prices of a financial instrument. Debt Valuation Adjustment (*DVA*) is also a financial metric which concerns the measure of the own potential default of a counterparty. Combining both *CVA* and *DVA* will result in the *BCVA* modelling which will be the matter of the subsection 2.1.2.

In the following, we will refer to the letter C when it will be related to the counterparty C and to the letter A when it will be about the bank / insurer point of view.

2.1.1 Unilateral CVA

We will assume a probability filtered space $(\Omega, \mathcal{G}, \mathbb{G} = (\mathcal{G}_t)_{t \geq 0}, \mathbb{Q})$ with \mathbb{Q} assumed to be a risk neutral pricing probability measure. We will introduce firstly all the notations which will be necessary to compute the equations of unilateral *CVA*. In this order, we'll note :

- T : Contract Maturity.
- V_t : Value of the risk free contract at date t .
- V_t^D : Value of the risky contract at date t .
- τ^C : Random variable valued in $]0, +\infty[$ representing the default time of the counterparty C .
- H : Random process $H = (H_t)_{t \geq 0}$ such has H_t represents if the default has occurred before t : $H_t = \mathbf{1}_{\tau^C \leq t}$.
- R^C : The recovery rate in case of default of the counterparty C which will be assumed constant in this dissertation.

- LGD : Loss Given Default defined as $LGD = (1 - R^C)$.
- Z : The recovery process $Z = (Z_t)_{t \geq 0}$ which we assume has the form $Z_t = R^C \max(V_t, 0) = R^C(V_t)^+$.

We will now define the short risk free rate $r = (r_t)_{t \geq 0}$ and the bank account numeraire $B_t = e^{\int_0^t r_s ds}$ allowing us to write the following dynamics for B_t :

$$dB_t = B_t r_t dt.$$

As a main difference from the classic framework of *Option Pricing Theory* because of the potentiel default of the counterparty, we have to be careful about which information does the institution hold in the market. For this, we define the following filtration :

- $\mathbb{F} = (\mathcal{F}_t)_{t \in [0, T]}$: Filtration of the market without default.
- $\mathbb{H} = (\mathcal{H}_t)_{t \in [0, T]}$: Filtration associated with the default such that $\mathcal{H}_t = \sigma((H_s)_{s \leq t})$ where σ denotes the lowest σ -algebra on $((H_s)_{s \leq t})$.
- $\mathbb{G} = (\mathcal{G}_t)_{t \in [0, T]}$: Filtration containing all the market information such that $\mathcal{G}_t = \mathcal{F}_t \vee \mathcal{H}_t$.

It is crucial to use this new filtration \mathbb{G} as τ^C is not necessarily a \mathcal{F} -adapted stopping time but is a \mathcal{G} -adapted stopping time. This result of considering the filtration \mathcal{H} is called *Progressive Filtration Enlargement*.

One of the major issues of considering \mathbb{G} instead of \mathbb{F} is the properties of the stochastic processes under \mathbb{F} which may be different under \mathbb{G} .

Lemma 2.1. Passage Formula

Consider Y a \mathcal{G} -measurable function. Therefore we have the following identity $\forall t \geq 0$

$$\mathbb{E}^{\mathbb{Q}}[1_{\tau^C > t} Y | \mathcal{G}_t] = 1_{\tau^C > t} \frac{\mathbb{E}^{\mathbb{Q}}[Y 1_{\tau^C > t} | \mathcal{F}_t]}{\mathbb{Q}(\tau^C > t | \mathcal{F}_t)}. \quad (2.1)$$

Proof. The proof of this lemma is based on a technical result which allows to identify a \mathcal{G}_t measurable random variable as a \mathcal{F}_t measurable random variable on the event $\tau^C > t$. Assuming this result, we can therefore write for a random variable Y_t \mathcal{F}_t -measurable that we have :

$$1_{\tau^C > t} \mathbb{E}^{\mathbb{Q}}[Y | \mathcal{G}_t] = \mathbb{E}^{\mathbb{Q}}[1_{\tau^C > t} Y | \mathcal{G}_t] = 1_{\tau^C > t} Y_t. \quad (2.2)$$

Now by conditioning by \mathcal{F}_t , and using the fact that $\mathcal{F}_t \subset \mathcal{G}_t$, we therefore have that :

$$Y_t = \frac{\mathbb{E}^{\mathbb{Q}}[Y 1_{\tau^C > t} | \mathcal{F}_t]}{\mathbb{Q}(\tau^C > t | \mathcal{F}_t)}.$$

By remplacing now Y_t in (2.2) ends the proof. □

To derive an intuition of the *CVA* formula we will propose below, let's consider the case of an european derivative with final potential payoff $g(X_T)$ such as we have the following definitions for V_t and V_t^D .

$$\begin{aligned} V_t &= \mathbb{E}^{\mathbb{Q}}\left[\frac{B_t}{B_T} g(X_T) | \mathcal{F}_t\right] = \mathbb{E}^{\mathbb{Q}}\left[\frac{B_t}{B_T} g(X_T) | \mathcal{G}_t\right], \\ V_t^D &= 1_{\tau^C > t} \mathbb{E}^{\mathbb{Q}}\left[\frac{B_t}{B_T} g(X_T) 1_{\tau^C > T} + \frac{B_t}{B_{\tau^C}} Z_{\tau^C} 1_{t \leq \tau^C \leq T} | \mathcal{G}_t\right]. \end{aligned} \quad (2.3)$$

Now, we give the following lemma :

Lemma 2.2. $\forall t \in [0, T]$, we have :

$$\mathbb{E}^{\mathbb{Q}}[\mathbb{1}_{t \leq \tau^C \leq T} \frac{B_t}{B_{\tau^C}} V_{\tau^C} | \mathcal{G}_t] = \mathbb{E}^{\mathbb{Q}}[\int_t^T \frac{B_t}{B_s} V_s dH_s | \mathcal{G}_t] = \mathbb{E}^{\mathbb{Q}}[\mathbb{1}_{t \leq \tau^C \leq T} \frac{B_t}{B_T} g(X_T) | \mathcal{G}_t]. \quad (2.4)$$

Proof. The first equality comes from the definition of H_t . For the second equality, let's define the following stopping time $\sigma^C = (\tau^C \wedge T) \vee t$. It defines well a stopping time adapted to filtration \mathcal{G} by classic operations on stopping times. As σ^C is a bounded stopping time and because that the process $\frac{(V_t)}{B_t}$ is a $\mathcal{G} - \mathbb{Q}$ martingale, we therefore have using the Doob theorem:

$$V_{\sigma^C} = \mathbb{E}^{\mathbb{Q}}[\frac{B_{\sigma^C}}{B_T} \phi(X_T) | \mathcal{G}_{\sigma^C}]. \quad (2.5)$$

Or , on the event $\{t \leq \tau^C \leq T\}$, we have $\sigma^C = \tau^C$, and therefore using the equation 2.5, we have :

$$\mathbb{E}^{\mathbb{Q}}[\int_t^T \frac{B_t}{B_s} V_s dH_s | \mathcal{G}_t] = \mathbb{E}^{\mathbb{Q}}[\mathbb{1}_{t \leq \tau^C \leq T} \frac{B_t}{B_{\tau^C}} V_{\tau^C} | \mathcal{G}_t] = \mathbb{E}^{\mathbb{Q}}[\mathbb{E}^{\mathbb{Q}}[\mathbb{1}_{t \leq \tau^C \leq T} \frac{B_t}{B_T} \phi(X_T) | \mathcal{G}_{\sigma^C}] | \mathcal{G}_t].$$

The result follows from the tower law property as we have $\mathcal{G}_t \subset \mathcal{G}_{\sigma^C}$ \square

Now, let's consider the difference between V_t and V_t^D . We can rewrite V_t^D as the following:

$$V_t^D = \mathbb{1}_{\tau^C > t} \mathbb{E}^{\mathbb{Q}}[\frac{B_t}{B_T} g(X_T)(1 - \mathbb{1}_{\tau^C \leq T}) + \frac{B_t}{B_{\tau^C}} Z_{\tau^C} \mathbb{1}_{t \leq \tau^C \leq T} | \mathcal{G}_t].$$

Using now Lemma (2.2), we get assuming working on the event $\{\tau^C > t\}$ that :

$$V_t^D = V_t + \mathbb{E}^{\mathbb{Q}}[(R^C - 1)(\mathbb{1}_{\tau^C \leq T} \frac{B_t}{B_T} g(X_T)) | \mathcal{G}_t].$$

We then can define the *CVA* for an european derivative as follows :

$$\begin{aligned} CVA_t &= \mathbb{1}_{\tau^C \geq t} (V_t - V_t^D) = (1 - R^C) \mathbb{E}^{\mathbb{Q}}[\mathbb{1}_{t \leq \tau^C \leq T} \frac{B_t}{B_T} g(X_T) | \mathcal{G}_t], \\ CVA_t &= (1 - R^C) \mathbb{E}^{\mathbb{Q}}[\mathbb{1}_{t \leq \tau^C \leq T} \frac{B_t}{B_T} g(X_T) | \mathcal{G}_t] = (1 - R^C) \mathbb{E}^{\mathbb{Q}}[\int_t^T \frac{B_t}{B_s} V_s dH_s | \mathcal{G}_t]. \end{aligned}$$

From this definition of the *CVA* for an european derivative we clearly see that the Credit Valuation Adjustment is the difference between the risk-free derivative and the associated derivative defautable value. Now, we can give the global definition of the process *CVA*.

Definition 2.1.1. We define the *CVA* as the market consistent value of the future credit loss for a product with value at time t denoted by V_t and with a recovery process assumed to be of the form $Z_t = R^C(V_t)^+$.

$$CVA_t = (1 - R^C) \mathbb{E}^{\mathbb{Q}}[\int_t^T \frac{B_t}{B_s} (V_s)^+ dH_s | \mathcal{G}_t] = (1 - R^C) \mathbb{E}^{\mathbb{Q}}[\mathbb{1}_{t \leq \tau^C \leq T} (V_{\tau^C})^+ \frac{B_t}{B_{\tau^C}} | \mathcal{G}_t]. \quad (2.6)$$

A simplified version of the CVA in practice :

In practice, we are most interested in the computation of CVA_0 under common assumptions which will be discussed in the last section. Indeed, let's go back on equation (2.6) which we can rewrite as follows :

$$\begin{aligned} CVA_0 &= (1 - R^C) \mathbb{E}^Q[\mathbf{1}_{\tau^C \leq T} \frac{(V_{\tau^C})^+}{B_{\tau^C}}], \\ &= (1 - R^C) \mathbb{E}^Q[\mathbb{E}^Q[\mathbf{1}_{\tau^C \leq T} \mathbb{E}^Q[\frac{(V_{\tau^C})^+}{B_{\tau^C}} | \tau^C]]. \end{aligned}$$

Admitting that the random variable τ^C admits a density, we can therefore rewrite CVA_0 as follows by defining respectively the cumulative and survival probability function of τ^C denoted by $F(t) = P(\tau^C \leq t)$ and $G(t) = 1 - F(t) = P(\tau^C > t)$.

$$CVA_0 = -(1 - R^C) \int_0^T \mathbb{E}^Q[\frac{(V_t)^+}{B_t} | \tau^C = t] dG(t). \quad (2.7)$$

Under the assumption of independance between the credit exposure and the value of the derivate/ porfolio, we can rewrite CVA in the easiest following form :

$$CVA_0 = -(1 - R^C) \int_0^T \mathbb{E}^Q[\frac{(V_t)^+}{B_t}] dG(t). \quad (2.8)$$

The equation (2.8) is the most fondamental equation of CVA and is the most used by practitioners. Let's define some quantities why are key factors in the calculation of the CVA :

- The quantity $\frac{(V_t)^+}{B_t}$ is called **Positive Exposure** and is note $PE(t)$.
- The quantity $\frac{(V_t)^-}{B_t}$ is called **Negative Exposure** and is noted $NE(t)$.
- The quantity $\mathbb{E}^Q[\frac{(V_t)^+}{B_t}]$ is called **Expected Positive Exposure** and is noted $EPE(t)$.
- The quantity $\mathbb{E}^Q[\frac{(V_t)^-}{B_t}]$ is called **Expected Negative Exposure** and is noted $ENE(t)$.

The function defined by $t \in [0, T] \mapsto EPE(t)$ is often called positive exposure profile and has a lot of interest for practitioners.

The CVA from equation (2.8) can therefore be approximated considering on an homogenous timegrid $0 = t_0 < t_1 < \dots < t_N = T$ with $t_{i+1} - t_i = \frac{T}{N} \quad \forall i \in \llbracket 0; N-1 \rrbracket$ by the following :

$$CVA_0 \approx -(1 - R^C) \sum_{i=0}^{N-1} EPE(t_i)(G(t_{i+1}) - G(t_i)). \quad (2.9)$$

Some remarks on the CVA formula :

- In order to compute CVA_0 , we need to be able to derive survival probabilities G for the counterparty C which can be done using *CDS* bonds on the market or equivalent products.
- We need to be able to compute the exposition profile which can be computationally intense as we need to evaluate $\mathbb{E}^Q[\frac{(V_t)^+}{B_t}]$ which can lead to nested Monte-Carlo when V_t needs to be calculated himself using a Monte-Carlo procedure.

In general, if we consider products for which we don't get a closed formula for V_t , it will be computationally too intensive to perform the Nested Monte-Carlo. We will illustrate it in the following chapter.

An interest case study for the CVA : The Cox setup

In a lot of situations, the only observable quantity are the asset prices who generate the filtration \mathbb{F} so it's better to write V_t^D and after CVA_t in terms of \mathcal{F}_t instead of \mathcal{G}_t (in the case of an european derivative). For this, we can use the Lemma (2.1) and using the fact that $\mathbb{1}_{\tau^C > t}$ is \mathcal{G}_t measurable we have :

$$V_t^D = \mathbb{1}_{\tau^C > t} \frac{\mathbb{E}^{\mathbb{Q}}[\frac{B_t}{B_T} g(X_T) \mathbb{1}_{\tau^C > T} + \frac{B_t}{B_{\tau^C}} Z_{\tau^C} \mathbb{1}_{t \leq \tau^C \leq T} | \mathcal{F}_t]}{Q(\tau^C > t | \mathcal{F}_t)}.$$

$$CVA_t = \mathbb{1}_{\tau^C > t} (V_t - V_t^D).$$

We see from this equation that we need to give an appropriate form we denote $S = (S_t)_{t \in [0, T]}$ such as $S_t = \mathbb{Q}(\tau^C > t | \mathcal{F}_t)$.

The Cox setup is an intensity based approach where we assume the following representation for the process S_t :

$$S_t = e^{- \int_0^t \lambda_s ds}. \quad (2.10)$$

where $\lambda = (\lambda_t)_{t \in [0, T]}$ is assumed to be a non-negative \mathbb{F} -adapted process.

Moreover, and under proposition 5.1.1 from [3], we have the following representation for V_t^D :

$$V_t^D = \mathbb{1}_{\tau^C > t} \mathbb{E}^{\mathbb{Q}}[e^{- \int_t^T (r_s + \lambda_s) ds} g(X_T) | \mathcal{F}_t] + \mathbb{1}_{\tau^C > t} \mathbb{E}^{\mathbb{Q}}[\int_t^T Z_s \lambda_s e^{- \int_t^s (\lambda_u + r_u) du} | \mathcal{F}_t]. \quad (2.11)$$

Using now theorem 4.16 from [4] , we can show that V_t^D can be rewritten as :

$$V_t^D = \mathbb{1}_{\tau^C > t} (R^C \mathbb{E}^{\mathbb{Q}}[e^{- \int_t^T r_u du} g(X_T) | \mathcal{F}_t] + (1 - R^C) \mathbb{E}^{\mathbb{Q}}[e^{- \int_t^T (r_u + \lambda_u) du} g(X_T) | \mathcal{F}_t]). \quad (2.12)$$

We finally have a consequence the following formula for CVA of european derivatives in the Cox approach :

$$CVA_t = \mathbb{1}_{\tau^C > t} (V_t - V_t^D) = \mathbb{1}_{\tau^C > t} (1 - R^C) \mathbb{E}^{\mathbb{Q}}[e^{- \int_t^T r_u du} g(X_T) (1 - e^{- \int_t^T \lambda_u du}) | \mathcal{F}_t]. \quad (2.13)$$

Remark. In a more general framework, as how we defined S_t , we must verify that :

$$\mathbb{E}^{\mathbb{Q}}[S_t] = G(t), \quad \forall t \in [0, T].$$

It can be shown that CVA_t can be rewritten as :

$$CVA_t = \mathbb{1}_{\tau^C > t} (1 - R^C) \mathbb{E}^{\mathbb{Q}}[\int_t^T e^{- \int_t^s (r_u + \lambda_u) du} (V_s)^+ \lambda_s ds | \mathcal{F}_t]. \quad (2.14)$$

2.1.2 Bilateral CVA

Bilateral CVA is a modification of unilateral CVA, in the sense that we will allow the potential default of the buyer of the contract. As we notice in the beginning of this section, we denote by letter A the corresponding counterparty. Consequently, the value of the risky asset for A is calculated as the risk-free value of the derivative minus the CVA that incorporates the counterparty's default risk of C which is denoted by CVA^A plus the CVA computed by

considering the holder's own risk perceived by the counterparty C , denoted by CVA . For this, we will introduce the Debt Valuation Adjustment (DVA) which is equivalent to the CVA but represents the holder's counterparty risk.

We can therefore introduce the random time of default A , denote by τ^A and we can define as we did in the previous section.

We denote by $\tau = \tau^C \wedge \tau^A$ and \mathcal{G} the filtration augmented by the information with $(H_t^A)_{t \in [0, T]}$ such that $H_t^A = \mathbb{1}_{\tau^A < t}$.

Therefore, we can write the following equations for $V_t^{D,A}$ which is the risky value of the product as seen in the point of view of A , and similarly $V_t^{D,C}$ for the point view of the counterparty C :

$$BCVA_t^A = \mathbb{1}_{\tau > t} (V_t - V_t^{D,A}). \quad (2.15)$$

$$BCVA_t^C = \mathbb{1}_{\tau > t} (V_t - V_t^{D,C}). \quad (2.16)$$

where $BCVA$ means Bilateral CVA and is calculated as follows :

$$BCVA_t^A = CVA_t^A - DVA_t^A. \quad (2.17)$$

$$BCVA_t^C = CVA_t^C - DVA_t^C. \quad (2.18)$$

As we said, DVA is equivalent to the CVA but for the holder counterparty risk so the derivation of the formula is straightforward and can be calculated as.

$$DVA_t^A = CVA_t^C = (1 - R^A) \mathbb{1}_{\tau^C > t} \mathbb{E}^Q \left[\frac{B_t}{B_{\tau^A}} \mathbb{1}_{t \leq \tau^A \leq T} (-V_{\tau^A})^+ | \mathcal{G}_t \right]. \quad (2.19)$$

The sign $-$ comes from the fact that under A point of view, if he defaults, he is exposed to the negative part of V_t as he holds a long position in the product. Usually, as $(-x)^+ = \max(-x, 0) = x^-$, we usually rewrite 2.19 as the following :

$$DVA_t^A = (1 - R^A) \mathbb{1}_{\tau^C > t} \mathbb{E}^Q \left[\frac{B_t}{B_{\tau^A}} \mathbb{1}_{t \leq \tau^A \leq T} (V_{\tau^A})^- | \mathcal{G}_t \right]. \quad (2.20)$$

In the following, we won't really calculate $BCVA$ as when we will calculate CVA , we will assume that the the buyer of the contract is default-free so $\tau^A = +\infty$. However, when we will deal with FVA , we will have to model the creditworthiness of the counterparty A .

2.1.3 Netting principle for CVA

Let's consider that party A has many financial contracts with a counterparty B . When A computes the total CVA for the counterparty B , he can either consider the 2 following possibilities:

- Sum the CVA of all contracts with the counterparty.
- Compute a single CVA computed on the sum of valuations of all contracts considered as a single one.

Assume that counterparty A has 2 long positions on 2 differents contracts with C denoted respectively by value V_t^1 and V_t^2 .

If A computes the CVA under the first possibility, he will then charge the following amount to C :

$$CVA_t^{Case1} = (1 - R^C) \mathbb{E}^Q \left[\mathbb{1}_{t \leq \tau^C \leq T} \frac{B_t}{B_{\tau^C}} ((V_{\tau^C}^1)^+ + (V_{\tau^C}^2)^+) | \mathcal{G}_t \right].$$

However, if he decides to go under the second possibility, he will then charge the following amount to C :

$$CVA_t^{Case2} = (1 - R^C) \mathbb{E}^Q[\mathbf{1}_{t \leq \tau^C \leq T} \frac{B_t}{B_{\tau^C}} ((V_{\tau^C}^1 + V_{\tau^C}^2)^+ | \mathcal{G}_t)].$$

However, from the simple fact that $\forall (x_1, x_2) \in \mathbb{R}^2 \quad (x_1 + x_2)^+ \leq (x_1)^+ + (x_2)^+$, we have :

$$CVA_t^{Case2} \leq CVA_t^{Case1} \quad \forall t \in [0, T].$$

We see therefore that we can reduce the overall CVA by considering the second possibility which is called a *netting agreement* between counterparty A and C as the positive exposure of all contracts is lower than the sum of the exposures of the individual contracts.

2.1.4 Wrong and Right Way Risk for CVA

So far, we have considered for the computation of CVA the case of independence between credit exposure and the value of the portfolio. However, in reality, this assumption may not be realistic in a lot of situations and its manifestation can be potentially dramatic. *Wrong Way Risk (WWR)* occurs when the counterparty is more likely to default which results in a higher exposure profile and vice versa. *WWR* is often seen as an unavoidable consequence of financial markets. For instance, mortgage providers, who during an economic recession will face both falling of property prices and higher default rates. The Right Way Risk is the opposite of *WWR* meaning that there is a favourable dependence between credit exposure and exposure which can reduce the overall CVA .

Mathematically, the *Wrong Way Risk* appears from equation 2.7 where the conditioning exposure $\mathbb{E}^Q[\frac{V_t^+}{B_t} | \tau^C = t]$ cannot be simplified into $\mathbb{E}^Q[\frac{V_t^+}{B_t}]$. The modelling of *WWR* will be explored in details in chapter 4 where we will refer to intensity approaches.

2.2 An FVA framework

Funding Valuation Adjustment (*FVA*) is a financial metric used in order to measure the liquidity of funding associated with financial instruments. As the calculation of *FVA* is still under debate between practitioners and researchers, there exists multiple frameworks to handle the computation of *FVA*.

In this section, we will mainly refer to the *FVA* framework used in [5] as noticed in the preamble of the chapter. We can refer to [6] for another *XVA* framework where the *FVA* has a recursive form. We will consider the same notations as we used for the *CVA* and we will introduce the ones who need to be introduced in the *FVA* context.

2.2.1 FCA and FBA modelling

The FCA can be written as follows where $\tau = \tau^C \wedge \tau^A$ and \mathcal{G}_t is enlarged with the filtration associated with the potential default of A by denoting the random process $H_t^A = \mathbb{1}_{\tau^A < t}$.

Definition 2.2.1. *The FCA of a financial derivative V is defined as follows :*

$$FCA(t) = \mathbb{E}^{\mathbb{Q}} \left[\int_t^T \mathbb{1}_{u \leq \tau} \frac{B_t}{B_u} s_b(u) (V_u)^+ du \mid \mathcal{G}_t \right]. \quad (2.21)$$

where s_b denotes the spread of borrowing for A over the risk-free rate

The assumption of independance between the default of A and C is a common hypothesis as it has been shown that the impact is quite moderate so we will assume this hypothesis in the following proposition.

Proposition 2.1. *Under independance of defaults, FCA can be written as follows by supposing that $\mathbb{Q}(\tau^A > t \mid \mathcal{F}_t) = e^{- \int_0^t \lambda_s^A ds}$ and $\mathbb{Q}(\tau^C > t \mid \mathcal{F}_t) = e^{- \int_0^t \lambda_s^C ds}$.*

$$FCA(t) = \mathbb{1}_{t \leq \tau} \int_t^T \mathbb{E}^{\mathbb{Q}} \left[e^{- \int_t^u (\lambda_s^A + \lambda_s^C) ds} \frac{B_t}{B_u} s_b(u) (V_u)^+ \mid \mathcal{F}_t \right]. \quad (2.22)$$

Proof. The proof will be based on the Lemma 2.1 where we assume that the equality $\mathbb{E}^{\mathbb{Q}}[\mathbb{1}_{\tau^C > u} Y \mid \mathcal{G}_t] = \mathbb{1}_{\tau^C > t} \frac{\mathbb{E}^{\mathbb{Q}}[Y \mathbb{1}_{\tau^C > u} \mid \mathcal{F}_t]}{\mathbb{Q}(\tau^C > t \mid \mathcal{F}_t)}$ holds true $\forall 0 \leq t < u$. Going back into the equation 2.21, we can rewrite it :

$$\begin{aligned} FCA(t) &= \mathbb{E}^{\mathbb{Q}} \left[\int_t^T \mathbb{1}_{u \leq \tau} \frac{B_t}{B_u} s_b(u) (V_u)^+ du \mid \mathcal{G}_t \right], \\ FCA(t) &\stackrel{Fubini}{=} \int_t^T \mathbb{E}^{\mathbb{Q}} \left[\mathbb{1}_{u \leq \tau} \frac{B_t}{B_u} s_b(u) (V_u)^+ \mid \mathcal{G}_t \right] du, \\ FCA(t) &\stackrel{\text{Lemma 2.1}}{=} \mathbb{1}_{t \leq \tau} \int_t^T \frac{\mathbb{E}^{\mathbb{Q}} \left[\mathbb{1}_{u \leq \tau} \frac{B_t}{B_u} s_b(u) (V_u)^+ \mid \mathcal{F}_t \right]}{\mathbb{Q}(\tau > t \mid \mathcal{F}_t)} du, \\ FCA(t) &= \frac{\mathbb{1}_{t \leq \tau}}{e^{- \int_0^t (\lambda_s^A + \lambda_s^C) ds}} \int_t^T \mathbb{E}^{\mathbb{Q}} \left[\mathbb{1}_{u \leq \tau} \frac{B_t}{B_u} s_b(u) (V_u)^+ \mid \mathcal{F}_t \right] du. \end{aligned} \quad (2.23)$$

Now, let's handle the term $\mathbb{E}^{\mathbb{Q}}[\mathbb{1}_{u \leq \tau} \frac{B_t}{B_u} s_b(u) (V_u)^+ \mid \mathcal{F}_t]$. For this, let's condition on \mathcal{F}_u as $\mathcal{F}_t \subset \mathcal{F}_u$. We have therefore :

$$\begin{aligned} \mathbb{E}^{\mathbb{Q}}[\mathbb{1}_{u \leq \tau} \frac{B_t}{B_u} s_b(u) (V_u)^+ \mid \mathcal{F}_t] &= \mathbb{E}^{\mathbb{Q}}[\mathbb{E}^{\mathbb{Q}}[\mathbb{1}_{u \leq \tau} \frac{B_t}{B_u} s_b(u) (V_u)^+ \mid \mathcal{F}_u] \mid \mathcal{F}_t], \\ &= \mathbb{E}^{\mathbb{Q}}[\mathbb{E}^{\mathbb{Q}}[\mathbb{1}_{u \leq \tau} \mid \mathcal{F}_u] \frac{B_t}{B_u} s_b(u) (V_u)^+ \mid \mathcal{F}_t], \\ &= \mathbb{E}^{\mathbb{Q}}[e^{- \int_0^u (\lambda_s^A + \lambda_s^C) ds} \frac{B_t}{B_u} s_b(u) (V_u)^+ \mid \mathcal{F}_t]. \end{aligned}$$

The passage of the first line comes from the tower law for conditional expectation. The second line is justified by the \mathcal{F}_u -measurability of the process $e^{- \int_t^u r_s ds} s_b(u) (V_u)^+$. The third line obtained by the independance between defaults.

By reuniting this term in (2.23), we recover the expression in (2.25). \square

Definition 2.2.2. The FBA of a financial derivative V is defined as follows :

$$FBA(t) = \mathbb{E}^Q \left[\int_t^T \mathbf{1}_{u \leq \tau} \frac{B_t}{B_u} s_L(u)(V_u)^- du \mid \mathcal{G}_t \right]. \quad (2.24)$$

where s_L denotes the spread of lending for A over the risk free rate.

Proposition 2.2. Under independance of defaults, FBA can be written as follows by supposing that $Q(\tau^A \mid \mathcal{F}_t) = e^{-\int_0^t \lambda_s^A ds}$ and $Q(\tau^C \mid \mathcal{F}_t) = e^{-\int_0^t \lambda_s^C ds}$.

$$FCA(t) = \mathbf{1}_{t \leq \tau} \int_t^T \mathbb{E}^Q [e^{-\int_t^u (\lambda_s^A + \lambda_s^C) ds} \frac{B_t}{B_u} s_L(u)(V_u)^- \mid \mathcal{F}_t]. \quad (2.25)$$

Proof. The proof follows the identical approach as for the FCA. \square

Once theses quantites are computed, the process FVA is defined as follows :

Definition 2.2.3. The FVA on a financial derivative V is defined as :

$$FVA(t) = FCA(t) - FBA(t) \quad \forall t \in [0, T]. \quad (2.26)$$

We see immediatly that we can define :

$$\begin{aligned} EE_{FVA_t} &: [t, T] \rightarrow \mathbb{R}^+ \\ u &\mapsto \mathbb{E}^Q [e^{-\int_t^u (\lambda_s^A + \lambda_s^C) ds} \frac{B_t}{B_u} s_b(u)(V_u)^+ \mid \mathcal{F}_t] - \mathbb{E}^Q [e^{-\int_t^u (\lambda_s^A + \lambda_s^C) ds} \frac{B_t}{B_u} s_L(u)(V_u)^- \mid \mathcal{F}_t]. \\ FVA(t) &= \int_t^T EE_{FVA_t}(u) du. \end{aligned} \quad (2.27)$$

In the next chapters, as we will mostly focus on the cost of funding for A, we will assume that $FBA(t) = 0 \quad \forall t \in [0, T]$ so that FVA reduces to FCA.

A simplified version of the FVA in practice :

Moreover, the 2.25 can still be simplified by assuming again independance between exposure and credit default which is the approximation done in practice. The equation therefore simplifies for FCA_0 as :

$$FCA_0 = \int_0^T P(\tau^A > u) P(\tau^C > u) \mathbb{E}^Q \left[\frac{1}{B_u} s_b(u)(V_u)^+ \right] du.$$

which can be approximated over an homogenous timegrid $0 = t_0 < t_1 < \dots < t_N = T$ with $t_{i+1} - t_i = \frac{T}{N} \quad \forall i \in \llbracket 0; N-1 \rrbracket$ such that by supposing a deterministic spread of borrowing s_B , we have :

$$FCA_0 \approx \sum_{i=0}^{N-1} P(\tau^A > t_{i+1}) P(\tau^C > t_{i+1}) s_b(t_i) EPE(t_i)(t_{i+1} - t_i). \quad (2.28)$$

Therefore, we see that for both CVA and FVA, the exposure profile $EE(t)$ is a risk factor and the computation of this quantity is crucial for praticioners to calculate theses values adjustments.

2.2.2 Wrong and Right Way Risk for FVA

The *Wrong Way Risk* can have an impact on *FVA* and this measurement needs to be done in a risk management perspective. Indeed, during the Covid crisis, there was a global distress in financial markets which made the funding spreads explode and created unexpected losses. In the *FVA*, *Wrong Way Risk* means increased funding spreads due to increased market risk. In the case of a portfolio of receiver swaps, *Wrong Way Risk* will occur when there is a negative exposure between interest rates and funding spreads because when interest rates go down, exposure goes up and it increases the *FVA* and therefore the funding spread.

2.3 An MVA framework

The Margin Valuation Adjustment (*MVA*) is based on the calculation of the initial margin (*IM*) which is a margin introduced in order to cover the market risk of a portfolio during the necessary time to unwind the position. When an institution is involved in a bilateral transaction and the other counterparty defaults, there is close-out time for the position named as Margin Period of Risk (*MPOR*) which is typically of 10 business days. Therefore, the *IM* term should be able to capture the potential exposure that can happen during the *MPOR* and it has been required that the *IM* should correspond to a 99 % Value-At-Risk (*VaR*) change of the porfolio value under the *MPOR*.

As *IM* is a form of collateralization that cannot be netted and is segregated , it can lead to considerable costs for the entities involved in the exchange. The risk associated to theses funding costs is called *MVA* and it requires the knowledge of the *IM* profile during the life time of the transaction. Indeed, Dynamic Initial Margin (*DIM*) is understood as the expected *IM* that has to be posted at t and can be defined as :

$$DIM(t) = \mathbb{E}^Q[e^{-\int_0^t r_u du} IM(t)|\mathcal{F}_0]. \quad (2.29)$$

Once this quantity is computed and according to a funding spread between the collateral rate and the risk free rate denoted by f , we can then compute the *MVA* as follows :

$$MVA_0 = \int_0^T f(s) DIM(s) ds. \quad (2.30)$$

As we can see, the main feature of the computation of the *MVA* is the calculation of the *DIM* which himself requires the computation of the initial margin. This term is obviously the main factor in the *MVA* computation and is hard to handle which made a lot of debate in the industry. To manage this issue, the *Standard Initial Margin Model (SIMM)* proposed by *ISDA*¹ decided to promote a method for the computation of the *IM* and this is the one we will use in this dissertation. Other methods can exist for the computation of the *IM* profile which can be based on internal methods but need to be approved by regulatory authorities.

¹The *SIMM* framework will be presented in the chapter 6 when an *MVA* computation will be performed.

2.4 The KVA challenge

The Capital Valuation adjustment (*KVA*) is based on the fact that investors require a return on investment. The modelling of the *KVA* is quite challenging as it involves the computation of the capital profile of the banks and the associated cost of capital.

In [2], they propose a methodology to compute the KVA as follows :

$$KVA_0 = - \int_0^T e^{- \int_0^t r_u + \lambda_u^A + \lambda_u^C du} CC(t) \mathbb{E}^Q[K(t)] dt. \quad (2.31)$$

where :

- CC defines the cost of capital which is a really subjective term which will depend on the institution's policy.
- K defines the capital that is expected to be held by the institution.

As we can see from this definition of *KVA*, it is hard to define an appropriate CC cost profile and a capital K profile as they highly depend of the institution's policy.

In the litterature, there exists another way to model the *KVA* introduced in [?] :

$$KVA_0 = h \mathbb{E}^Q \left[\int_0^T e^{- \int_0^s (r_u + h) du} EC_s(L) \max(ES_s(L), KVA_s) ds \right]. \quad (2.32)$$

where :

- $EC_s(L)$ refers to the economic capital of the bank and is proposed to be $EC_s(L) = \max(ES_s^\alpha(L), KVA_s)$ where $ES_s^\alpha(L)$ is the *Expected Shortfall* at level α on the portfolio value L .
- h is the hurdle rate at which the investors of the bank should be remunerated.

From equation 2.31, the *KVA* is higly reliable on the institution's policy and from equation 2.32, it is computationally intensive as it requires the computation of a risk measure at any time $t < T$ on the portfolio of the bank.

In the following of this dissertation and for the reasons cited above, we won't focus on the computation of *KVA* and mainly focus on the others *XVAs*.

Chapter 3

Some use cases for CVA computations

In this chapter, we will illustrate some CVA_0 computations on various financial instruments. We will start with european options and forward contracts under Black-Scholes ($B - S$) model. Next, we will illustrate the computation of EE profile and associated CVA_0 for an interest rate swap under 2 models : Hull & White and $G2++$. We will also present some results for the case of bermudan options by considering a put and a swaption. Finally, we will end this chapter by showing the impact of a netting agreement in the CVA computation of a portfolio.

3.1 EE profile and CVA on some financial products under B-S model

For this section, we will consider a filtered probability space $(\Omega, \mathcal{F}, \mathbb{Q})$ with $\mathbb{F} = (\mathcal{F}_t)_{t \geq 0}$ the canonical filtration generated by a brownian motion $(W_t)_{t \geq 0}$. We assume that \mathbb{Q} defines a risk neutral probability measure. The $B - S$ model is given by the following dynamics for the underlying $S = (S_t)_{t \geq 0}$:

$$dS_t = S_t(rdt + \sigma dW_t), \quad S_0 \in \mathbb{R}_*^+. \quad (3.1)$$

We will consider either the case of an european call and a forward contract to see the core concepts and challenges of CVA computations. In the case of the equation (2.8) which we rewrite here for convenience, we know that CVA_0 can be computed as :

$$CVA_0 = (1 - R^C) \int_0^T \mathbb{E}^\mathbb{Q} \left[\frac{(V_t)^+}{B_t} \right] dF(t) = -(1 - R^C) \int_0^T \mathbb{E}^\mathbb{Q} \left[\frac{(V_t)^+}{B_t} \right] dG(t).$$

3.1.1 EE profile of an european call

We know that in the case of an european call with a strike K and a maturity T with final payoff given by $(S_T - K)^+$, we have :

$$V_t = \mathbb{E}^\mathbb{Q}[e^{-r(T-t)}(S_T - K)^+ | \mathcal{F}_t] \quad \forall t \in [0, T]. \quad (3.2)$$

This quantity can be calculated analytically and we recover the famous $B - S$ formula :

$$V_t = S_t N(d_1) - K e^{-r(T-t)} N(d_2) \quad \forall t \in [0, T]. \quad (3.3)$$

with :

- N is the cumulative distribution function of a $\mathcal{N}(0, 1)$ random variable defined by $N(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$.

- $d_1 = \frac{\ln(\frac{S_t}{K}) + (r + \frac{\sigma^2}{2})(T-t)}{\sigma\sqrt{T-t}}$.
- $d_2 = d_1 - \sigma\sqrt{T-t}$.

As a call is an option, it means that the buyer of the contract who holds a long position is always exposed to the counterparty credit risk and the exposure which is defined by $(V_t)^+$ reduces to V_t .

Therefore the exposition profile function EPE defined in our context is given by :

$$EPE(t) = \mathbb{E}^{\mathbb{Q}}[e^{-rt}V_t] \quad \forall t \in [0, T]. \quad (3.4)$$

But in our setting, the process $(e^{-rt}V_t)_{t \in [0, T]}$ defines a martingale and we therefore have that the exposition profile EPE for a call is given by : ¹

$$EPE(t) = V_0 = S_0N(d_1) - Ke^{-rT}N(d_2).$$

Therefore, the CVA_0 of an european call reduces to :

$$CVA_0 = (1 - R^C)V_0(1 - G(T)). \quad (3.5)$$

Moreover, we can also calculate the DVA for the call as it is related to the calculation of $(V_t)^-$ in the associated exposure. However, for an option, we have :

$$(V_t)^- = 0 \quad \forall t \in [0, T].$$

Therefore, we have that $ENE(t) = 0$ and the DVA_0 for the buyer of a call is therefore equal to 0.

We give below the calculation of the exposition profile EPE and ENE for a call of maturity $T = 1$:

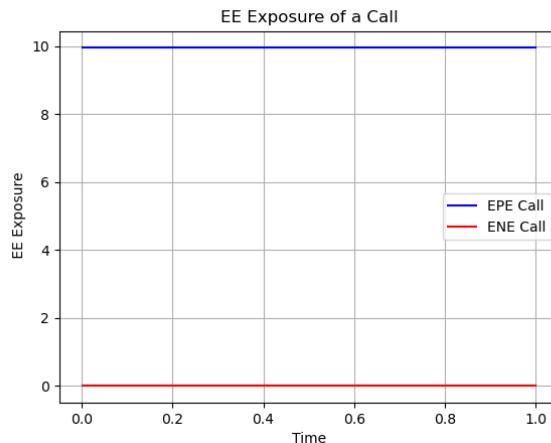


Figure 3.1: EPE and ENE profiles for a call option under a $B - S$ model with the following parameters : ($S_0 = 100$, $K = 100$, $r = 0$ and $\sigma = 0.25$)

¹Note that neither d_1 neither d_2 are functions of t here.

After calculating the exposition profile function, we are able to calculate the associated CVA_0 by assuming a cumulative distribution function for τ^C and using equation 3.5. For the illustration case, we will assume that $\tau^C \sim \mathcal{E}(\lambda)^2$.

Here is below an illustration of the evolution of CVA_0 for a call as a function of λ the default intensity parameter.

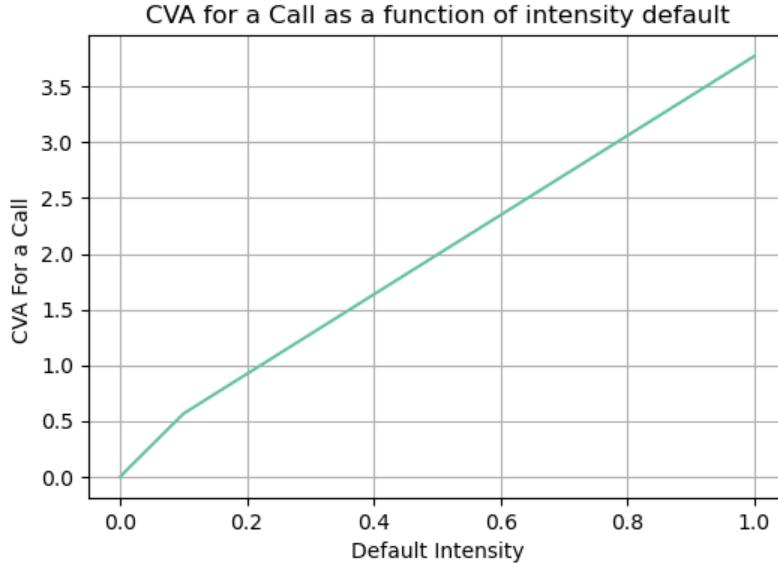


Figure 3.2: Evolution of CVA_0 for a call option as a function of intensity default λ with $R^C = 0.4$.

Some remarks on the computation of CVA_0 of a call option :

- As we can see, when λ increases, CVA_0 also increases which is an expected behavior because the counterparty is more likely to default.
- We assume that the default time of the counterparty is given by an exponential distribution as it is the most common choice but any other relevant choice could have been used. Moreover, we assume that the distribution parameter λ can be calibrated from market instruments like *CDS* or any other relevant credit derivatives.
- When we derive the equation for the CVA_0 of a call, we found out that it could be efficiently calculated by (3.5) but we can extend this formula to any european derivative as long as we have a final payoff given by $\phi(S_T)$. For instance, if we considered a put option with final payoff $\phi(S_T) = (K - S_T)^+$, we would have the same formula (3.5) with V_0 being in this case equal to :

$$V_0 = Ke^{-rT}N(-d_2) - S_0N(-d_1). \quad (3.6)$$

Anyways, by noting V_0^ϕ the price of an european option with payoff $\phi(S_T)$ and CVA_0^ϕ the associated CVA_0 , we have :

$$\begin{aligned} V_0^\phi &= \mathbb{E}^Q[e^{-rT}\phi(S_T)]. \\ CVA_0^\phi &= (1 - R^C)V_0^\phi(1 - G(T)). \end{aligned}$$

² $\mathcal{E}(\lambda)$ stands for the exponential distribution of parameter λ such that $F(t) = (1 - e^{-\lambda t})\mathbb{1}_{t \geq 0}$.

3.1.2 EE profile of a forward contract

We know that in the case of a forward contract with a strike K and a maturity T with final payoff given by $(S_T - K)$, we have :

$$V_t = \mathbb{E}^{\mathbb{Q}}[e^{-r(T-t)}(S_T - K)|\mathcal{F}_t] \quad \forall t \in [0, T]. \quad (3.7)$$

This quantity can be calculated analytically and we recover :

$$V_t = S_t - Ke^{-r(T-t)} \quad \forall t \in [0, T]. \quad (3.8)$$

A major difference between the consideration of the call option and the forward contract is the non simplification $(V_t)^+ = V_t$ as the price of the forward contract can be either positive or negative from the point of view of the institution A .

Therefore the exposition profile function EPE defined in our context by :

$$EPE(t) = \mathbb{E}^{\mathbb{Q}}[e^{-rt}(V_t)^+] = \mathbb{E}^{\mathbb{Q}}[e^{-rt}(S_t - Ke^{-r(T-t)})^+] \quad \forall t \in [0, T]. \quad (3.9)$$

We recognize the calculation of a call option with maturity t and strike given by $Ke^{-r(T-t)}$. It follows from the classical calculation for a call option in the $B - S$ model : ³

$$EPE(t) = V_0 = S_0N(d_1) - Ke^{-rT}N(d_2).$$

with :

- $d_1 = \frac{\ln(\frac{e^{-rt}S_0}{K}) + (r + \frac{\sigma^2}{2})t}{\sigma\sqrt{t}}$.
- $d_2 = d_1 - \sigma\sqrt{t}$.

Therefore, the CVA_0 of a forward contract has to be calculated following equation (2.9) as EPE is here a function of t .

Moreover, we can also calculate the DVA for the forward contract as it is related to the calculation of $(V_t)^-$ in the associated exposure.

$$(V_t)^- = (S_t - Ke^{-r(T-t)})^- = (Ke^{-r(T-t)} - S_t)^+ \quad \forall t \in [0, T]$$

Therefore, we have that the ENE at time t for the buyer of a forward contract is equivalent to the pricing a put option of maturity t with strike $Ke^{-r(T-t)}$ and we finally have

$$ENE(t) = Ke^{-r(T-t)}N(-d_2) - S_0N(-d_1). \quad (3.10)$$

³Note that d_1 and d_2 are functions of t here.

We give below the calculation of the exposition profile EPE and $-ENE$ for a forward contract of maturity $T = 1$:

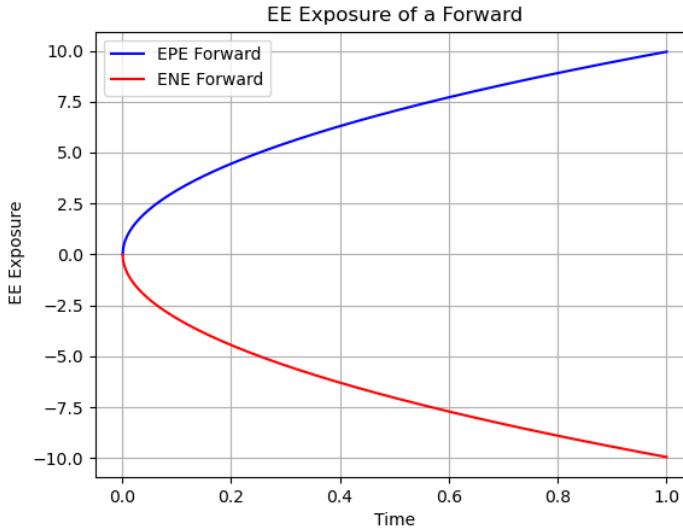


Figure 3.3: EPE and ENE profiles for a forward contract under a $B - S$ model with the following parameters : ($S_0 = 100$, $K = 100$, $r = 0$ and $\sigma = 0.25$)

After calculating the exposition profile function, we are able to calculate the associated CVA_0 by assuming a cumulative distribution function for τ^C and using 3.5. For the illustration case, we will assume also that $\tau^C \sim \mathcal{E}(\lambda)$.

Here is below an illustration of the evolution of CVA_0 for a forward contract as a function of λ the default intensity parameter.

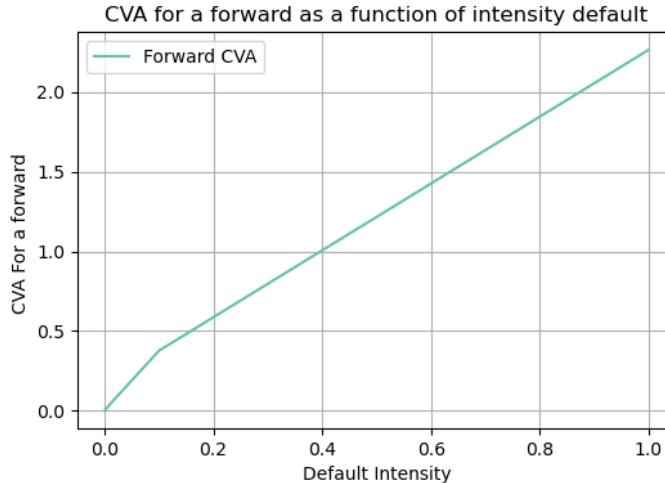


Figure 3.4: Evolution of CVA_0 for a forward contract as a function of intensity default λ with $R^C = 0.4$.

Some remarks on the computation of CVA_0 for a forward contract :

- As for the call option, when λ increases, CVA_0 also increases which is an expected behavior because the counterparty is more likely to default.
- When we derive the equation for the CVA_0 of a forward contract, we see that we cannot get rid of the sign + in the positive exposure $(V_t)^+$ as the value of the forward can become negative. In this case, we need to perform the numerical integration given by (2.9). which leads to additional computations over the life time of the contract. We cannot extend the methodology to similar products as forward as it will depend if we can derive a close formula for $EPE(t)$.

A global remark on the computation of CVA_0

- We illustrate the concept of calculation of the CVA for 2 products where we could have an analytic valuation of the financial instrument V and where we could also derive the EPE profile analytically but in a lot of cases we don't have closed formulas for the price of the financial instrument neither for the calculation of EPE which will result in a nested Monte-Carlo procedure. This procedure can then be computationally intense and that's why we didn't provide results for such a case.

We write below in Algorithm 3.1 the calculation of exposure based on a nested Monte-Carlo and the Algorithm 3.2 for CVA_0 computation based himself on Algorithm 3.1.

Algorithm 3.1 Nested Monte-Carlo for exposure profile computation

Input Parameters : M number of samples for the outer Monte-Carlo

N number of samples for the inner Monte-Carlo

ϕ the payoff function of the underlying risk factor S

for $i = 0, \dots, M - 1$ **do**

 Simulate the risk factor S_i up to date t_l denote by S_{i,t_l} using the forward dynamics of S

for $j = 0, \dots, N - 1$ **do**

 | Compute the risk factor dynamics until T starting from S_{i,t_l} at t_l for each j path S_{i,t_l}^j

end

 Compute the price $A_i = \frac{1}{N} \sum_{j=0}^{N-1} (\phi(S_{i,t_l}^j))^+$

end

Output Parameters : $EPE(t_l) = \frac{1}{M} \sum_{i=0}^{M-1} A_i$

Algorithm 3.2 Nested Monte-Carlo for CVA_0 computation

Input Parameters : T maturity contract, timegrid $t_0 < t_1 < \dots < t_K = T$, $CVA_0 = 0$

for $l = 0, \dots, K - 1$ **do**

 | Calculate EPE_{t_l} according to Algorithm 3.1

 | Compute $G(t_{l+1}) - G(t_l)$

 | $CVA_0 = LGD * EPE(t_l) * (G(t_{l+1}) - G(t_l))$

end

Output Parameters : CVA_0

3.2 EE profile and CVA of an IRS under Hull & White and G2++ models

We will now illustrate the calculation of the *EPE* profile and associated CVA_0 in the case of a vanilla interest rate swap under 2 models : Hull & White and $G2++$.

3.2.1 Pricing of an IRS

An interest rate swap is a financial derivative where two parties exchange interest rate cashflows, typically one party pays a fixed interest rate (*swap receiver*) and the other pays a floating interest rate (*swap payer*) which can be based on a reference rate. Let's define the notations we will used in the following :

- n : Number of Payments.
- $\tau_k = (T_{k+1} - T_k)$: Coupon frequency for the swap payments $\forall k \in \llbracket 0; n-1 \rrbracket$.
- $l_k(t) = l(t, T_k, T_{k+1})$: The forward rate at time t for an investment starting at T_{k-1} and ending at T_k .
- $B(t, T_k)$: the price of a zero-coupon-bond at t which gives 1 at T_k .
- $V(t)$: The price of the swap at t .
- K : The fixed strike payed by the receiver of the swap.
- N : The notional of the swap.

From the point of view of the receiver of the swap , we can therefore see the accumulated payments of the swap at the final date T_n assuming the same payment dates for the payer and the receiver of the swap and the same notional N .

$$V_{swap}(T_n) = N \sum_{k=0}^{n-1} \tau_k (l_k(T_k) - K) = N \sum_{k=0}^{n-1} (T_{k+1} - T_k) (l(T_k, T_k, T_{k+1}) - K).$$

By defining $\eta(t) = \min\{i \in \{0, \dots, n-1\} : T_{i+1} > t\}$, we have the following pricing formula for the value of the swap at t given by :

$$V_{Swap}(t) = N \sum_{k=\eta(t)}^{n-1} \tau_k B(t, T_{k+1}) (l(t, T_k, T_{k+1}) - K). \quad (3.11)$$

Moreover, we can rewrite the forward rate $l(t, T_k, T_{k+1})$ in terms of zero-coupon bonds as :

$$l(t, T_k, T_{k+1}) = \frac{1}{\tau_k} \left(\frac{B(t, T_k)}{B(t, T_{k+1})} - 1 \right).$$

With this reformulation, the formula for the swap reduces to :

$$V_{Swap}(t) = N (B(t, T_{\eta(t)}) - B(t, T_n) - K \sum_{k=\eta(t)}^{n-1} \tau_k B(t, T_{k+1})). \quad (3.12)$$

If we assume that the payment frequency is assumed to be constant denoted by $\tau_k = \delta T$, we can simplify further (3.12).

3.2.2 EE profile of an IRS

Because of the calculation of the exposure profile, we need to be able to calculate the quantities $(B(t, T_k))_{k \in \llbracket 1; N \rrbracket} \quad \forall t \geq 0$ which requires the use of an interest rate model. We will first introduce the Hull & White model and then the G2++.

The Hull & White Model :

The Hull & White model is a short rate model denoted by $(r_t)_{t \geq 0}$ which is an extension of the Vasicek model with the following dynamics ⁴:

$$dr_t = (\theta(t) - \kappa r(t))dt + \sigma dW_t, \quad r_0 \in \mathbb{R}. \quad (3.13)$$

with :

- r_0 is the initial short rate value.
- κ is the speed of the mean reversion.
- σ is the volatility associated with the diffusion process.
- θ is a deterministic function which helps to calibrate the model to the initial zero-coupon bond curve in the market $(B^{Market}(0, t))_{t > 0}$.

Proposition 3.1. *The solution of the EDS 3.13 is given by :*

$$r_t = r_0 e^{-kt} + \int_0^t \theta(u) e^{-k(t-u)} du + \sigma \int_0^t e^{-\kappa(t-u)} dW_u. \quad (3.14)$$

Proof. Let's consider the process $Y_t = r_t e^{\kappa t}$. By applying Itô to this process, we have :

$$\begin{aligned} dY_t &= e^{\kappa t} (\kappa r_t dt + dr_t) \\ dY_t &= e^{\kappa t} (\theta(t)dt + \sigma dW_t) \end{aligned}$$

The result follows by integrating over the interval $[0, t]$ and multiplying by $e^{-\kappa t}$ \square

Zero Coupon Bond Pricing under Hull & White Model :

Proposition 3.2. *The price of a zero-coupon bond with maturity T at time $t \leq T$ $B(t, T)$ is given by :*

$$B(t, T) = e^{m(t, T) - n(t, T)r_t}. \quad (3.15)$$

with :

$$\begin{aligned} n(t, T) &= \frac{1 - e^{-\kappa(T-t)}}{\kappa}, \\ m(t, T) &= - \int_t^T \theta(u) \eta(u, T) du + \frac{\sigma^2}{2} \int_t^T \eta(u, T)^2 du. \end{aligned}$$

⁴The discretisation scheme of the short rate process will be performed using the Euler method. See Annex B for more details on the methodology.

Proof. The results comes from the Markov property of the process r and from the Feymann-Kac formula which is presented in proposition A.1 in Annexes. Indeed, we have :

$$B(t, T) = \mathbb{E}^{\mathbb{Q}}[e^{-\int_t^T r_s ds} | \mathcal{F}_t].$$

$$B(t, T) \stackrel{\text{Markov}}{=} F^T(t, r_t).$$

Using Feymann-Kac formula, we have that F^T is solution of the following PDE :

$$\begin{cases} \frac{\partial F^T}{\partial t}(t, r) + (\theta(t) - \kappa r) \frac{\partial F^T}{\partial r} + \frac{\sigma^2}{2} \frac{\partial^2 F^T}{\partial r^2} - r F^T(t, r) = 0 & \forall (t, r) \in [0, T] \times \mathbb{R}, \\ F^T(T, r) = 1. \end{cases} \quad (3.16)$$

Supposing an affine structure for $B(t, T)$ and therefore for $F^T(t, r_t)$, we have the following equation : $\forall (t, r) \in [0, T] \times \mathbb{R}$:

$$F^T(t, r) \left(\frac{\partial m(t, T)}{\partial t}(t, T) - r \frac{\partial n(t, T)}{\partial t}(t, T) - (\theta(t) - \kappa r) \eta(t, T) + \frac{\sigma^2}{2} \eta(t, T)^2 - r \right) = 0. \quad (3.17)$$

From the fact that F^T is a stricly positive function and considering the terms in r or not, we arrive to the following system of ordinary differential equations :

$$\frac{\partial n}{\partial t}(t, T) - \kappa n(t, T) + 1 = 0. \quad (3.18)$$

$$\frac{\partial m}{\partial t}(t, T) = \theta(t)n(t, T) - \frac{\sigma^2}{2}n(t, T)^2. \quad (3.19)$$

The equation (3.18) can be solved explicitly and we recover the form from the proposition using the fact that $n(T, T) = 0$. Using the equation (3.19), we also recover the form of m using also that $m(T, T) = 0$. \square

Now, we need to characterize the form of θ to be able to compute prices of ZCB in the Hull & White Model. For this , we need to introduce the instantaneous forward rate denoted by $f_0(t)$ defined by :

$$f_0(t) = -\frac{\partial \ln(B(0, t))}{\partial t}. \quad (3.20)$$

Now, we have the following proposition :

Proposition 3.3. Assuming that $\forall t \geq 0$, $B^{\text{Market}}(0, t)$ are given by the market and θ takes the following form :

$$\theta(t) = \frac{\partial f_0(t)}{\partial t} + \kappa f_0(t) + \frac{\sigma^2}{2\kappa}(1 - e^{-2\kappa t}). \quad (3.21)$$

Then, the Hull & White model reproduces exactly the initial ZCB price curve $t \rightarrow B^{\text{Market}}(0, t)$.

Proof. Let's derive the form of $f_0(t)$ when we omit the notation $B^{Market}(0, t)$ and use instead for ease of notation $B(0, t)$.

$$\begin{aligned}
 f_0(t) &= -\frac{\partial \ln B(0, t)}{t}, \\
 &= -\frac{\partial m(0, t) - n(0, t)r_0}{\partial t}, \\
 &= \int_0^t \theta(u) \frac{\partial n(u, t)}{\partial t} du - \frac{\sigma^2}{2} \int_0^t \frac{\partial n(u, t)^2}{\partial t} + \frac{\partial n(0, t)}{\partial t} r_0, \\
 &= \int_0^t \theta(u) e^{-\kappa(t-u)} du - \frac{\sigma^2}{2} \int_0^t \frac{2}{k} (1 - e^{-\kappa(t-u)}) (e^{-\kappa(t-u)}) du + e^{-\kappa t} r_0, \\
 &= -\frac{\sigma^2}{2\kappa^2} (e^{-\kappa t} - 1)^2 + \int_0^t \theta(u) e^{-\kappa(t-u)} du + e^{-\kappa t} r_0, \\
 &= -g(t) + \phi(t).
 \end{aligned}$$

with :

$$\begin{aligned}
 g(t) &= \frac{\sigma^2}{2\kappa^2} (e^{-\kappa t} - 1)^2. \\
 \phi(t) &= \int_0^t \theta(u) e^{-\kappa(t-u)} du + e^{-\kappa t} r_0.
 \end{aligned}$$

Or, by deriving ϕ , we see that ϕ is solution to the following ordinary differential equation:

$$\phi'(t) = -\kappa\phi(t) + \theta(t), \quad \phi(0) = r_0.$$

Finally, we found that θ takes the following form :

$$\begin{aligned}
 \theta(t) &= \phi'(t) + \kappa\phi(t), \\
 &= \frac{\partial f_0(t)}{\partial t} + \frac{\partial g(t)}{\partial t} + \kappa(f_0(t) + g(t)), \\
 &= \frac{\partial f_0(t)}{\partial t} + \kappa f_0(t) + \left(\frac{\partial g(t)}{\partial t} + \kappa g(t)\right), \\
 &= \frac{\partial f_0(t)}{\partial t} + \kappa f_0(t) + \frac{\sigma^2}{2\kappa} (1 - e^{-2\kappa t}).
 \end{aligned}$$

Therefore, we recover the form of θ of the proposition (3.3) which ends the proof. \square

A Sanity Check Implementation

As we have now derived all the parameters for the simulation of the short rate process, we are going to do a sanity check implementation by comparing the groundtruth values of $B(0, T)$ from the market and the ones from the model as they should be close. We will consider that the market initial zero coupon curve is given by $B(0, t) = e^{-r_0 t}$.

First, we give some plots below of the simulation of the short rate process r_t under Hull & White model :

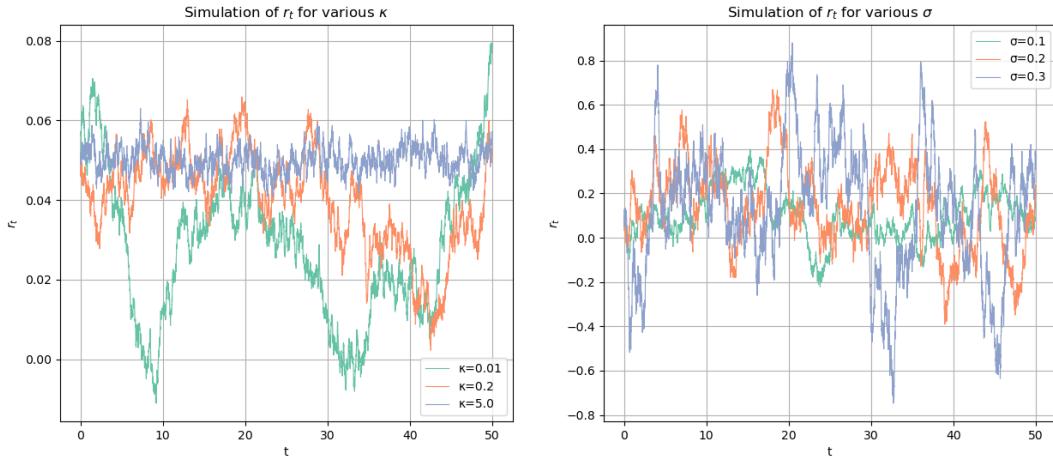


Figure 3.5: Simulation of r_t under Hull & White model and different choices of parameters

Some remarks on the impact of the parameters on r_t

- As we can see from the figures above, when κ increases, r_t become less volatile which is intuitive as κ controls the mean reversion speed of the short rate process r_t .
- Also, when the volatility parameter σ tends to increase, r_t become way more volatile but the impact of σ is more impactful compared to the speed parameter κ .

We will check the accuracy of the implementation by comparing the approximation of the zero-coupon bond price $B(0, t)$ defined as $\mathbb{E}^{\mathbb{Q}}[e^{-\int_0^t r_s ds}]$ following the algorithm 3.3 below based on the Law of Large Numbers.

Algorithm 3.3

Sanity check implementation algorithm

Input Parameters : M Number of Paths, \mathcal{T} timegrid $0 = t_0 < t_1 < \dots < t_N = t$, $((r_u^i)_{u \in \mathcal{T}})_{i \in [1, M]}$ dynamics of short rate over \mathcal{T} for each i

for $i = 1, \dots, M$ **do**

| Calculate $e^{-\int_0^t r_s^i ds}$ by approximating $\int_0^t r_s^i ds$ by trapezoidal rule $\sum_{j=0}^{N-1} \frac{1}{2}(r_{t_{j+1}}^i + r_{t_j}^i)(t_{j+1} - t_j)$

end

Output Parameters : Return $\hat{B}(0, t) = \frac{1}{M} \sum_{i=1}^M e^{-\sum_{j=0}^{N-1} \frac{1}{2}(r_{t_{j+1}}^i + r_{t_j}^i)(t_{j+1} - t_j)}$

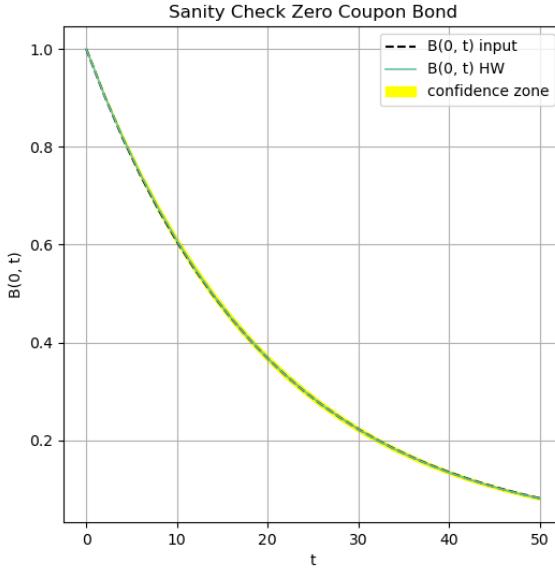


Figure 3.6: Sanity check implementation of Hull & White model

After checking the implementation , we can therefore calculate the exposure profile of a swap under this model.

We give in the table below the characteristics of the swap for which we performed the exposition profile computation :

Table 3.1: Parameters of the swap used in the numerical experiments

Parameters	Fixed Rate	T_{start}	T_{end}	Number of Payments
Value	0.01	1 Y	10 Y	10

Let's see in the following plots the impact of the parameters on the EPE profile of the swap.

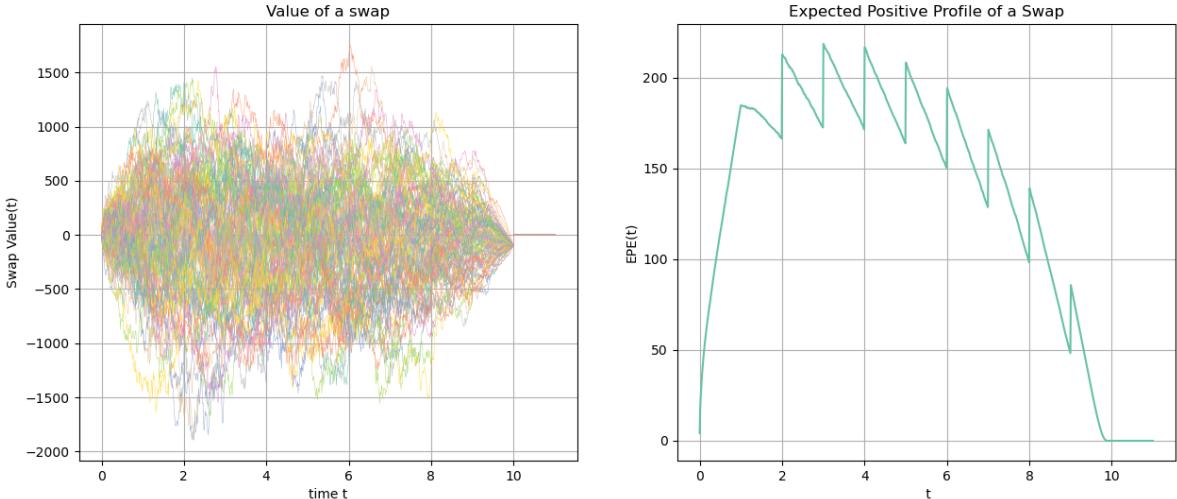


Figure 3.7: Value of a swap and EPE profile with $\sigma = 0.03$ and $\kappa = 0.5$ under Hull & White model with $N^{MC} = 20000$ simulations

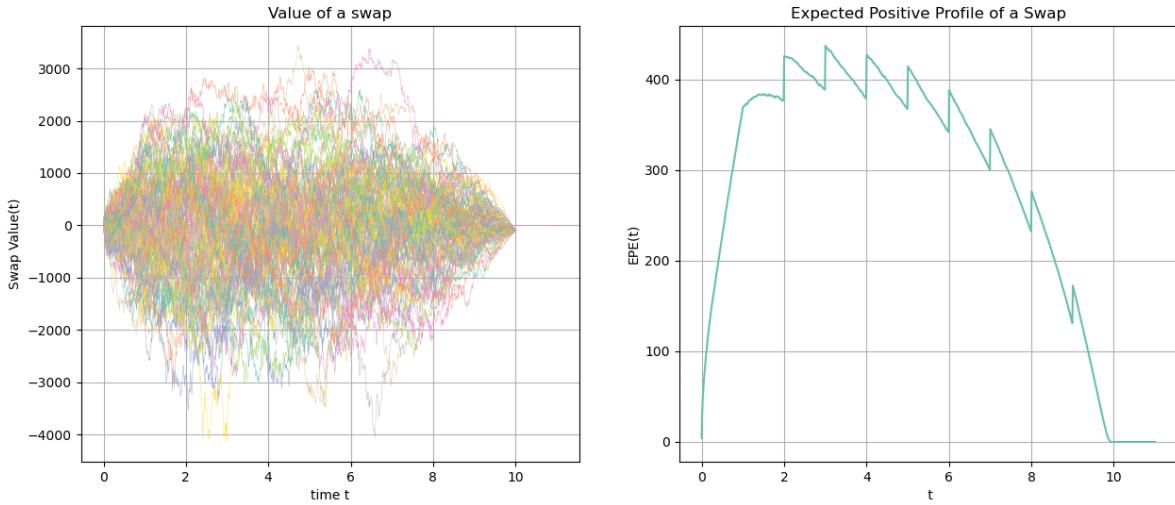


Figure 3.8: Value of a swap and EPE profile with $\sigma = 0.06$ and $\kappa = 0.5$ under Hull & White model with $N^{MC} = 20000$ simulations

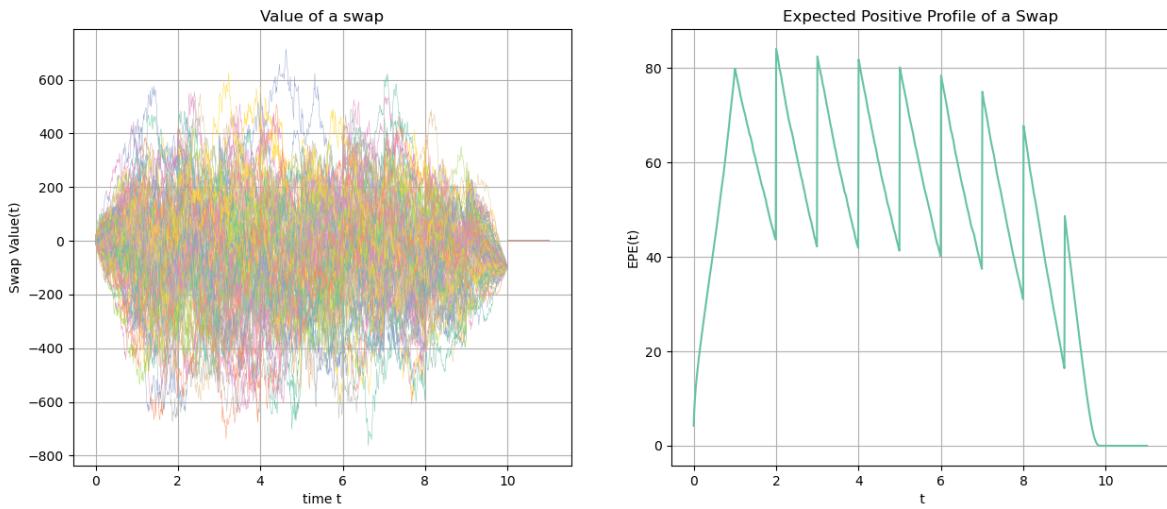


Figure 3.9: Value of a swap and EPE profile with $\sigma = 0.03$ and $\kappa = 1$ under Hull & White model with $N^{MC} = 20000$ simulations

Some remarks on the result :

- As we can see from the expected positive profile, we have several peaks which in fact represent the payment dates every year from $T = 1$ to $T = 10$.
- We also observe that when σ increases, the EPE profile also increases which is expected as the market price of the swap tends to become more volatile. Also, when κ increases, the EPE profile tends to decrease which is also intuitive as we can expect from the figure 3.5 as r_t becomes less volatile.
- At the beginning of the swap, his value is set to 0 and can become either positive or negative like for a forward contract. At the end of the swap, the value becomes 0 and so goes the EPE profile which is also intuitive.

Some CVA_0 profiles :

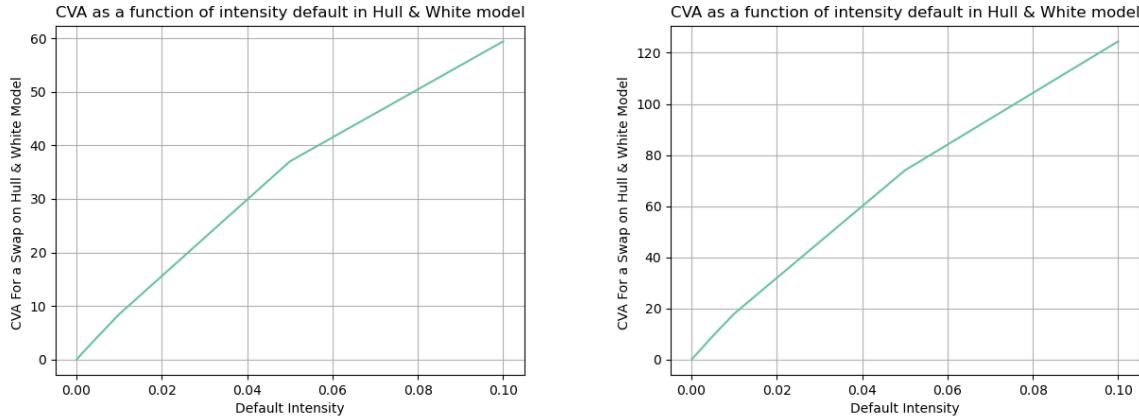


Figure 3.10: Evolution of the CVA_0 profile of an IRS as a function of intensity default λ for $\sigma = 0.03$ (right) and $\sigma = 0.06$ (left) with $\kappa = 0.5$ and $R^C = 0.4$

Some remarks on the computation of CVA_0 for an IRS under Hull & White model

- As we can see and from the expected exposure profile, when σ increases, the CVA_0 also increases which is expected.
- As we said previously, the value of CVA_0 also increases with the default intensity λ as the counterparty is more likely to default.

The G2 ++ Model :

The $G2++$ model is a short rate model where $r(t)$ takes the following form:

$$r(t) = x(t) + y(t) + \phi(t). \quad (3.22)$$

with :

- $r(t)$ the short interest rate at t .
- $dx(t) = -\kappa_x x(t)dt + \sigma_x dW_x(t)$ the 1st gaussian factor.
- $dy(t) = -\kappa_y y(t)dt + \sigma_y dW_y(t)$ the 2nd gaussian factor.
- $\phi(t) = f(0, t) + \frac{\sigma_x^2}{2\kappa_x^2}(1 - e^{-\kappa_x t})^2 + \frac{\sigma_y^2}{2\kappa_y^2}(1 - e^{-\kappa_y t})^2 + \frac{\rho\sigma_x\sigma_y}{\kappa_x\kappa_y}(1 - e^{-\kappa_x t})(1 - e^{-\kappa_y t})$.
- $d < W_x, W_y >_t = \rho dt$.
- $f(0, t)$ the instantaneous forward rate at t .

We omit the proof of the form of ϕ as it is similar to θ for the Hull & White model.⁵

⁵See [7] the book from Brigo and Mercurio where the main results on interest rate modelling are presented.

Some numerical results :

We give below some illustration of the evolution of r_t for some scenarios.

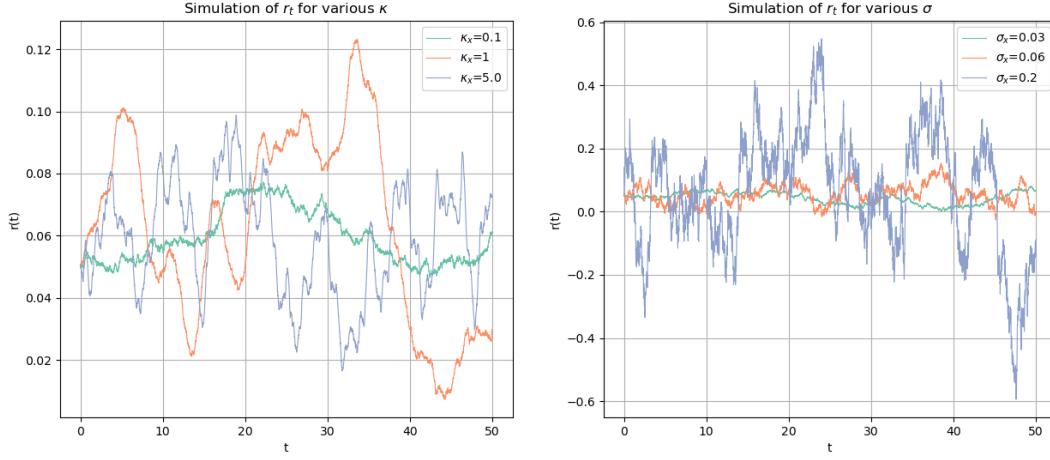


Figure 3.11: Simulation of r_t under $G2++$ model and different choices of parameters

Some remarks on the dynamic of r_t :

- As we can see from the plots above, the volatility of the first gaussian factor σ_x plays a crucial role in the range of values that the short rate process r_t can take.
- It's not as simple to analyze the impact of the parameter κ_x as it was in the Hull & White model as we are now considering a two factor model.

Zero Coupon Bond Pricing under G2++ model :

Proposition 3.4. *The price of a zero-Coupon Bond with maturity T at time $t \leq T$ $B(t, T)$ is given by :*

$$B(t, T) = \frac{B^M(0, T)}{B^M(0, t)} e^{A(t, T)}. \quad (3.23)$$

with :

$$\begin{aligned} A(t, T) &= \frac{1}{2}(V(t, T) - V(0, T) + V(0, t)) - \frac{1 - e^{-\kappa_x(T-t)}}{\kappa_x} x(t) - \frac{1 - e^{-\kappa_y(T-t)}}{\kappa_y} y(t). \\ V(t, T) &= -\frac{\sigma_x^2}{\kappa_x^2}(T-t + \frac{2}{\kappa_x}e^{-\kappa_x(T-t)} - \frac{1}{2\kappa_x}e^{-2\kappa_x(T-t)} - \frac{3}{2\kappa_x}) \\ &\quad + \frac{\sigma_y^2}{\kappa_y^2}(T-t + \frac{2}{\kappa_y}e^{-\kappa_y(T-t)} - \frac{1}{2\kappa_y}e^{-2\kappa_y(T-t)} - \frac{3}{2\kappa_y}) \\ &\quad + \frac{2\rho\sigma_x\sigma_y}{\kappa_x\kappa_y}(T-t + \frac{e^{-\kappa_x(T-t)} - 1}{\kappa_x} + \frac{e^{-\kappa_y(T-t)} - 1}{\kappa_y} - \frac{e^{-(\kappa_x+\kappa_y)(T-t)} - 1}{\kappa_x + \kappa_y}). \end{aligned}$$

Proof. We start from the definition of $B(t, T)$ as :

$$\begin{aligned} B(t, T) &= \mathbb{E}^{\mathbb{Q}}[e^{-\int_t^T r_s ds} | \mathcal{F}_t], \\ B(t, T) &= \mathbb{E}^{\mathbb{Q}}[e^{-\int_t^T (x_s + y_s + \phi_s) ds} | \mathcal{F}_t]. \end{aligned}$$

To compute, we need to characterize the law of the following random variable conditionally to \mathcal{F}_t :

$$I(t, T) = \int_t^T (x(u) + y(u)) du, \quad \forall 0 \leq t < T.$$

Following [7], we can show that $I(t, T)$ is a gaussian variable with mean $M(t, T)$ and variance $V(t, T)$ given by :

$$\begin{aligned} M(t, T) &= \frac{1 - e^{-\kappa_x(T-t)}}{\kappa_x} x(t) + \frac{1 - e^{-\kappa_y(T-t)}}{\kappa_y} y(t). \\ V(t, T) &= \frac{\sigma_x^2}{2\kappa_x^2} \left(T - t + \frac{2}{\kappa_x} e^{-\kappa_x(T-t)} - \frac{1}{2\kappa_x} e^{-2\kappa_x(T-t)} - \frac{3}{2\kappa_x} \right) \\ &\quad + \frac{\sigma_y^2}{2\kappa_y^2} \left(T - t + \frac{2}{\kappa_y} e^{-\kappa_y(T-t)} - \frac{1}{2\kappa_y} e^{-2\kappa_y(T-t)} - \frac{3}{2\kappa_y} \right) \\ &\quad + \frac{2\rho\sigma_x\sigma_y}{\kappa_{xy}} \left(T - t + \frac{e^{-\kappa_x(T-t)} - 1}{\kappa_x} + \frac{e^{-\kappa_y(T-t)} - 1}{\kappa_y} - \frac{e^{-(\kappa_x+\kappa_y)(T-t)} - 1}{\kappa_x + \kappa_y} \right). \end{aligned}$$

Now, from the lemma which says that if Z is a random variable with mean μ and variance σ^2 , we know that $\mathbb{E}[e^Z] = e^{\mu + \frac{\sigma^2}{2}}$, we therefore have :

$$B(t, T) = e^{M(t, T) + \frac{1}{2}V(t, T) - \int_t^T \phi(u) du}. \quad (3.24)$$

Now, we see that the model fits exactly the market curve of discount factors if and only if we have :

$$B^{Market}(0, T) = e^{-\int_0^T \phi(u) du + \frac{1}{2}V(0, T)}.$$

Therefore, we have that

$$\begin{aligned} e^{-\int_t^T \phi(u) du} &= e^{-\int_0^T \phi(u) du} e^{\int_0^t \phi(u) du}, \\ e^{-\int_t^T \phi(u) du} &= \frac{B(0, T)}{B(0, t)} e^{-\frac{1}{2}(V(0, T) + V(0, t))}. \end{aligned}$$

Putting this formula into (3.24), we recover the result of the proposition. □

Some numerical results :

We will provide some numerical results for an IRS in the $G2++$ model, starting from a sanity check implementation and then providing exposure profiles under different model parameters and associated CVA_0 .

Sanity Check Implementation :

We also provide a sanity check implementation by comparing the ZCB from the model and them from a fictitious market where $B(0, t) = e^{-r_0 t}$.

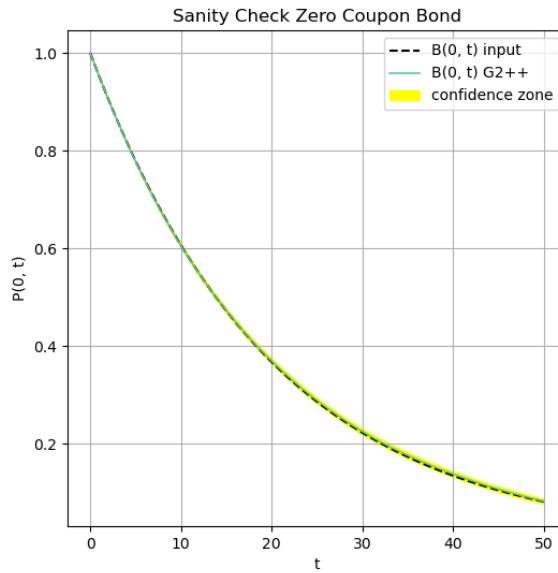


Figure 3.12: Sanity check implementation of $G2++$ Model

As we can see, there is a good adequation between the initial bond market price from the market and from the model which acts like a validation of our implementation. we can now derive some exposure and CVA_0 profiles for an interest rate swap under the $G2++$ model.

Some exposure profiles :

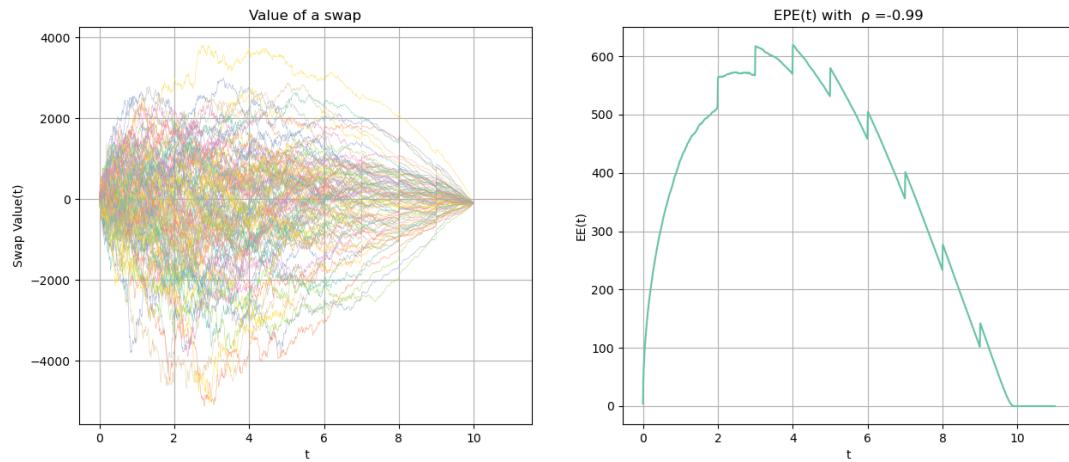


Figure 3.13: Value of a swap and EPE profile with $\rho = -0.99$ under the $G2++$ model

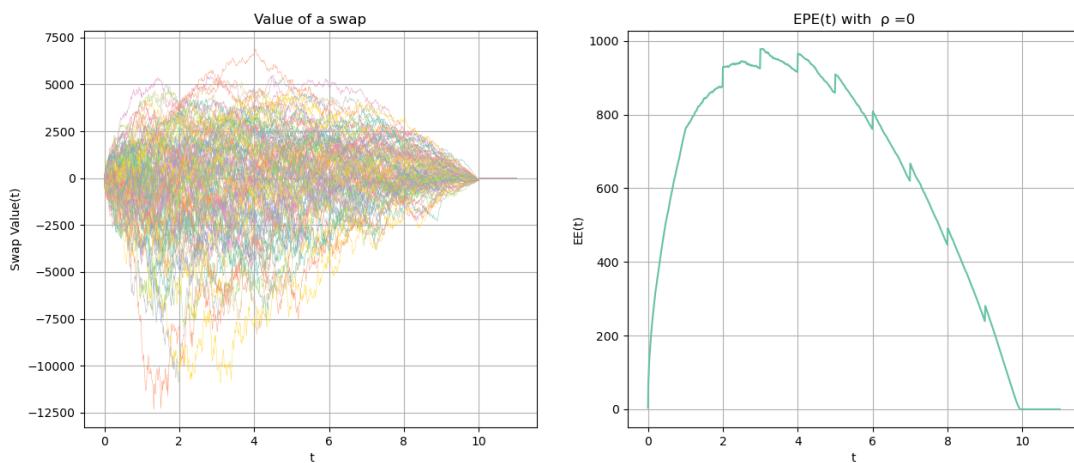


Figure 3.14: Value of a swap and EPE profile with $\rho = 0$ under the $G2++$ model

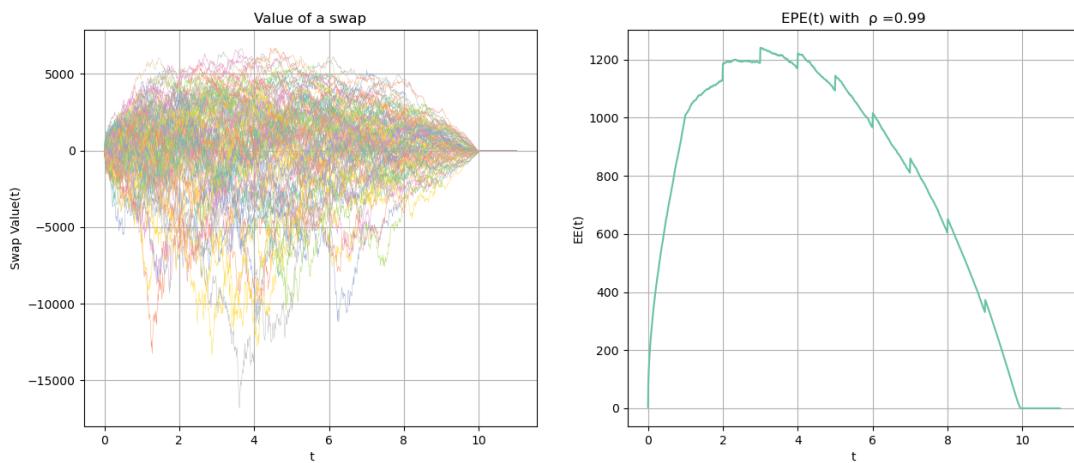


Figure 3.15: Value of a swap and EPE profile with $\rho = 0.99$ under the $G2++$ model

Some remarks on the computation of EE profile for an IRS under G2++ model

- First, we can see that the profile of exposure has some *jumps* which reflect the moment when a payment has been done. Moreover, the profile is really different from what we saw for equity products : call and forward.
- For the $G2++$ model, we provided some results when changing the ρ parameter and it ends up that when ρ increases, the exposition profile also increases, meaning that we have an higher exposition in a $G2++$ model with ρ next to 1.

Some CVA_0 profiles :

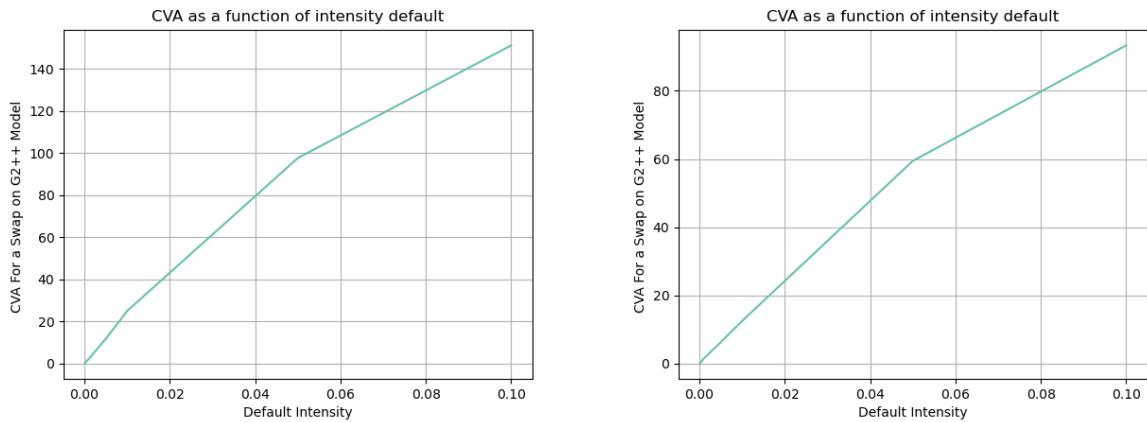


Figure 3.16: Evolution of the CVA profile as a function of intensity default λ for $\rho = -0.99$ (right) and $\rho = 0.99$ (left) with $R^C = 0.4$

Some remarks on the computation of CVA_0 for an IRS under G2++ model

- As we can see, as the default intensity increases, the CVA also increases which is expected but the behavior doesn't seem to be same if we used different ρ .
- Moreover, as we said before, if ρ increases, it seems to have an higher exposure which leads to an higher CVA which is also illustrated in the plots above. In practice, when practitioners are calibrating their $G2++$ model, an usual ρ parameter is around -0.9 which seems to lead to a lower CVA .⁶

⁶Note that the results could have been different if we choose a different set of parameters : κ_x , κ_y , σ_x and σ_y .

3.3 EE profile of some bermudan options

In this subsection, we will be interested in the computation of the EE profile of some bermudan Options as they are quite different from european options by the fact they can be exercised prior to maturity. We will mainly refer to the so called *Least Square Monte-Carlo Method* introduced in [8] by Tsitsiklis and Van Roy and in [9] by Longstaff and Schwartz to compute the price of such products. We will illustrate it with an interest rate swaption under the Hull & White model and for a put under the $B - S$ model.

3.3.1 The LSMC algorithm for pricing bermudan options

We will assume a probability space $(\Omega, \mathcal{F}, \mathbb{Q})$ and a finite interval $[0, T]$. We admit that \mathbb{Q} represents a risk neutral measure. We consider a stochastic process $(X_t)_{t \in [0, T]}$ which generates the filtration $(\mathbb{F} = \mathcal{F}_t)_{t \in [0, T]}$.

The pricing of an american option

As we said, unlike european options, american Options can be exercised prior to the maturity of the contract. For instance, if we consider a call option with maturity T on the stock price $(S_t)_{t \in [0, T]}$, with an american option, you can exercise your option at any time before maturity and receive the payoff $(S_t - K)^+$ at the time t .

To maximize their profits, the holder of the contract will try to exercise the option at the time that maximizes the discounted payoff. Therefore, we see that pricing an american option is equivalent to an optimal stopping problem on the space of $\mathcal{T}_{[t, T]}$ composed of all the stopping times adapted to \mathbb{F} with values in $[t, T]$. The price of such a product is therefore defined as :

$$P_t^{American} = \sup_{\tau \in \mathcal{T}_{[t, T]}} \mathbb{E}^{\mathbb{Q}}[B(t, \tau) \phi(S_\tau) | \mathcal{F}_t]. \quad (3.25)$$

where :

- ϕ is the payoff of the option.
- S is the underlying of the derivative.
- $B(t, \tau)$ is the discount factor between t and τ .

As we can see, the pricing of an american option is particularly hard as the option can be exercised at any time before the maturity T and it leads to a more challenging computational challenge than the pricing of european options.

The pricing of a bermudan option

We will consider an easier problem by considering bermudan options which are options that can only be exercised at discrete times prior to maturity. Let consider $N + 1$ potential different exercise times $0 = T_0 < T_1 < \dots < T_N = T$. The price of a bermudan option is similarly to the american option defined as the time over the timegrid $\mathcal{T}^{grid} = \{T_0, T_1, \dots, T_N\}$ that maximizes the discounted payoff. Pricing a Bermudan option is also an optimal stopping problem over $\mathcal{T}^{grid} \cap [t, T]$ denote by $\mathcal{T}^{Bermudan}$. The price of the Bermudan option is therefore given by :

$$P_t^{Bermudan} = \sup_{\tau \in \mathcal{T}^{Bermudan}} \mathbb{E}^{\mathbb{Q}}[B(t, \tau) \phi(S_\tau) | \mathcal{F}_t]. \quad (3.26)$$

Following the Snell Enveloppe Theory ⁷, it has been proven that the price of the bermudan option at each time of $\mathcal{T}^{Bermudan}$ is given by a dynamic programming approach. ⁸.

Algorithm 3.4 Bermudan option pricing with dynamic programming approach

Input Parameters : T maturity contract, timegrid $\mathcal{T} = \{0 = T_0 < T_1 < \dots < T_N = T\}$
 $P_{T_N}^{Bermudan} = \phi(S_{T_N})$

```

for  $i = N - 1, \dots, 0$  do
     $C_{T_i} = \mathbb{E}^{\mathbb{Q}}[B(T_i, T_{i+1})P_{T_{i+1}}^{Bermudan} | \mathcal{F}_{T_i}]$ 
     $P_{T_i}^{Bermudan} = \max(\phi(S_{T_i}), C_{T_i})$ 
end

```

Output Parameters : Price of Bermudan Option $(P_t^{Bermudan})_{t \in \mathcal{T}}$

The quantity $C_{T_i} = \mathbb{E}^{\mathbb{Q}}[B(T_i, T_{i+1})P_{T_{i+1}}^{Bermudan} | \mathcal{F}_{T_i}]$ is called continuation value at time T_i . The dynamic programming algorithm is a backward algorithm which can be understood as follows. At each time, the holder of the contract can either exercise the contract or hold it to the next date. If the current payoff $\phi(S_{T_i})$ exceeds the continuation value, then the holder will exercise his option. Otherwise, he will continue holding his option. As we can see, the challenging point in the Algorithm 3.4 is to compute the conditional expectation at each time. For this, we will assume that the process $(X_t)_{t \in [0, T]}$ is markovian so we can write the continuation value as follows.

$$C_{T_i} = \mathbb{E}^{\mathbb{Q}}[B(T_i, T_{i+1})P_{T_{i+1}}^{Bermudan} | X_{T_i}] \quad \forall i \in \llbracket 1, N-1 \rrbracket. \quad (3.27)$$

Therefore, we know that there exist ϕ_{T_i} a borelian measurable function such that $C_{T_i} = \phi_{T_i}(X_{T_i})$ and we will try to approximate this function ϕ_{T_i} .

The idea behind Longstaff-Schwartz Algorithm was to approximate ϕ_{T_i} using a suitable set of basis function denote by $\psi = (\psi_k^i)_{k \in [1, M]}$ and associated coefficients $\beta^i = (\beta_k^i)_{k \in [1, M]}$ such that we have:

$$\phi_{T_i} \approx \sum_{k=1}^M \beta_k^i \psi_k^i = \langle \beta^i, \psi^i \rangle. \quad (3.28)$$

where $\langle ., . \rangle$ is the usual scalar product on \mathbb{R}^M .

The choice of the set of basis function and of the number of factors M are crucial points as they influence the accuracy of the continuation value approximation. Classic basis functions are polynomials on X_{T_i} , Laguerre, Hermite or Legendre. The choice of the coefficients $(\alpha_k^i)_{k \in [1, M]}$ is done by using the conditional expectation representation as a minimization problem as presented in Annexes A.4 after choosing $(\psi_k^i)_{k \in [1, M]}$.

$$\hat{\beta}^i = \operatorname{argmin}_{\beta} \mathbb{E}^{\mathbb{Q}}[\|C_{T_i} - \langle \beta^i, \psi^i \rangle\|^2]. \quad (3.29)$$

Therefore, the continuation value is approximated by :

$$C_{T_i}(X_{T_i}) \approx \langle \hat{\beta}^i, \psi^i \rangle(X_{T_i}). \quad (3.30)$$

Using this dynamic programming approach, we can therefore evaluate the price of the bermudan at various nodes and discretization times and the exposure is calculated as follows :

$$EE^{Bermudan}(t) = (V_t^{Bermudan})^+ \mathbf{1}_{\tau > t}. \quad (3.31)$$

⁷See https://people.math.ethz.ch/~mschweiz/Files/bermuda_final.pdf for more theory.

⁸Note that this approach could be also performed with european options by taking $P_{T_i}^{European} = C_{T_i}$

3.3.2 Some numerical results

We will illustrate the calculation of the exposure in the case of a bermudan put under the $B - S$ model and a Bermudan swaption under the Hull & White model.

A bermudan put :

We consider the case of a bermudan put in a $B - S$ model such that the problem reformulation is as follows :

$$P_t^{Bermudan} = \sup_{\tau \in \mathcal{T}^{Bermudan}} \mathbb{E}^{\mathbb{Q}}[e^{-r(\tau-t)}(K - S_{\tau})^+ | \mathcal{F}_t]. \quad (3.32)$$

We consider at each time step T_i a polynomial basis function of S_{T_i} up to the degree 20 meaning that we approximate the continuation value at T_i as :

$$C_{T_i}(S_{T_i}) \approx \sum_{k=1}^{20} \beta_k^i S_{T_i}^k. \quad (3.33)$$

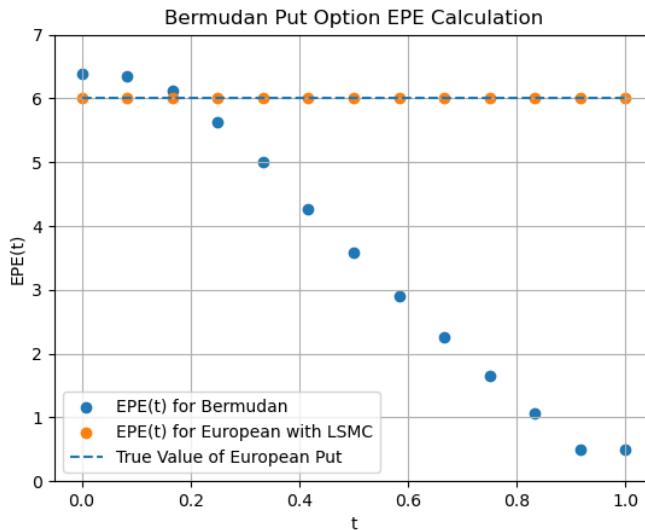


Figure 3.17: Calculation of the EPE profile of a bermudan put under $B - S$ model with the following parameters : ($S_0 = 100$, $K = 100$, $r = 0.04$, $\sigma = 0.2$, $T = 1$ and $N = 13$)

Some remarks :

- We see that at time $t = 0$ the exposure or equivalently the price of the bermudan option is higher than the european one which is something expected as the bermudan offers the potential exercise before maturity.
- We choose a polynomial basis function and we show that for the EPE profile in the case of the european put is quite similar with the true value.
- One interesting fact is during the last timestep between T_{n-1} and T_n , the option becomes an european style one and we recover the type of exposure of the european option (constant during the last period).

A Bermudan swaption

We also consider the case of a bermudan swaption in the Hull & White model. A swaption is an option which allows the holder to enter on a swap and the payoff assume that the exercise date of the option is at the beginning of the swap denoted by T_0 . Therefore, the payoff of the option is given by :

$$(V_{swap}(T_0))^+.$$

The price of the bermudan swaption can therefore be written as :

$$V_t^{BermudanSwaption} = \sup_{\tau \in \mathcal{T}^{Bermudan}} \mathbb{E}^Q[B(t, \tau) V_{swap}(\tau)^+ | \mathcal{F}_t]. \quad (3.34)$$

We give below the figure for the bermudan swaption ⁹ where we suppose that the early-exercise dates are defined by the payment dates of the swap. We also exclude the final date T_N from being a payment date so there are only only 9 exercise dates here.

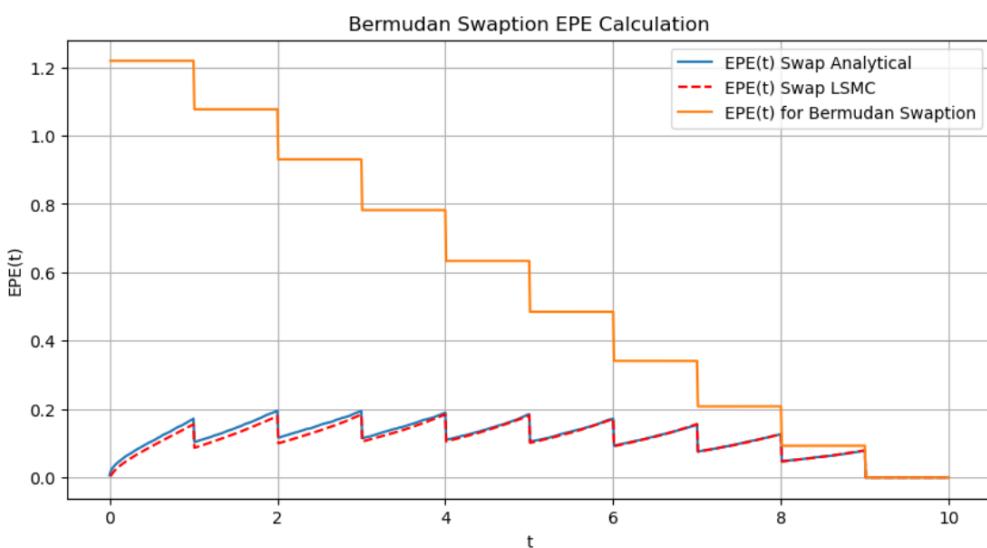


Figure 3.18: Calculation of the *EPE* profile of a bermudan swaption under the Hull & White model with the following parameters : ($r_0 = 0.01$, $\kappa = 0.5$, $\sigma = 0.02$, $T = 10$) with payment dates and potential exercise dates given by $T_1 = 1$, $T_2 = 2$, ..., $T_9 = 9$ with a nominal of $N = 10000$.

Some remarks :

- We observe that the expected exposure of the swaption is higher than the exposure of the corresponding underlying swap which is expected as it permits to his holder to decide if he wants to enter to the swap. Moreover, a swap exposure can either be positive or negative for the counterparty whereas a swaption is an option which will implies an higher exposure.
- As we said, after the payment T_9 , both value of the swap and bermudan swaption become zero because T_{10} is excluded.

⁹Additional Plots are available on the Annex E a Bermudan Call, Max-Call and Min Put on $d = 2$ assets options

3.4 EE profile of a portfolio

On this last section, we will illustrate the impact of netting on the exposition profile of a portfolio.

3.4.1 A quick overview of the model setup

We will consider the following model with the short rate process $r = (r_t)_{t \geq 0}$ assumed to follow a $G2++$ model with a zero coupon bond market given by $B(0, t) = e^{-r_0 t}$.

$$\begin{aligned} dS_t^1 &= S_t^1(r(t)dt + \sigma^1 dW_t^1), & S_0^1 &= S_0 \in \mathbb{R}_*^+, \\ dS_t^2 &= S_t^2(r(t)dt + \sigma^2 dW_t^2), & S_0^2 &= S_0 \in \mathbb{R}_*^+, \\ dS_t^3 &= S_t^3(r(t)dt + \sigma^3 dW_t^3), & S_0^3 &= S_0 \in \mathbb{R}_*^+, \\ r(t) &= x(t) + y(t) + \phi(t), & r_0 &= r_0 \in \mathbb{R}. \end{aligned}$$

We will assume the following correlation matrix for the risk factors by denoting by W^x and W^y respectively the risk factors for x and y in the $G2++$ model.

Table 3.2: Correlation matrix of the risk factors in the portfolio model

Risk Factors	W_t^1	W_t^2	W_t^3	W_t^x	W_t^y
W_t^1	1	$\rho_{1,2}$	$\rho_{1,3}$	0	0
W_t^2	$\rho_{1,2}$	1	$\rho_{2,3}$	0	0
W_t^3	$\rho_{1,3}$	$\rho_{2,3}$	1	0	0
W_t^x	0	0	0	1	$\rho_{x,y}$
W_t^y	0	0	0	$\rho_{x,y}$	1

We also give in the table below the parameters used in the modelling of the portfolios exposure :

Table 3.3: Parameters used in the numerical experiments for the global portfolio

Parameters	κ_x	σ_x	κ_y	σ_y	σ_1	σ_2	σ_3	S_0	r_0
Value	0.6	0.06	0.15	0.04	0.2	0.1	0.1	100	0.01

3.4.2 Some numerical results

We will consider 2 cases in the calculation of the *EE* to see the impact of netting :

- A simple portfolio with 2 swaps : one payer and one receiver with same final payment date $T_n = 10$ with strike rate $K = 0.01$ but with different payment dates and starting date.
- A simple portfolio with 3 long positions in a forward contract written on each underlying $(S_t^i)_{i \in [1;3]}$ with strike set at the money forward ¹⁰ and maturity $T = 10$.

¹⁰It means that we set K such that $S_0 - KB(0, T) = 0$ that is to say $K = \frac{S_0}{B(0, T)}$.

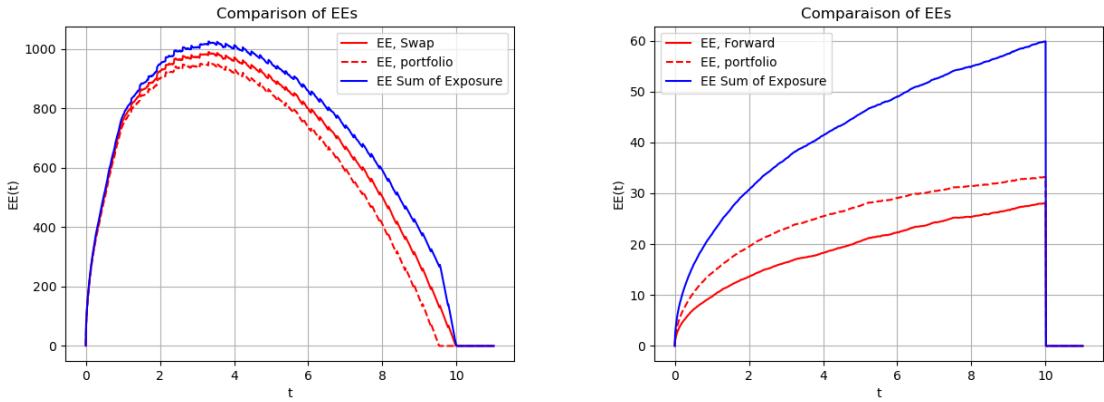


Figure 3.19: Impact of netting in 2 simple portfolios with the following parameters : ($\rho_{12} = -0.8$, $\rho_{13} = -0.7$, $\rho_{23} = -0.6$ and $\rho_{xy} = 0$)

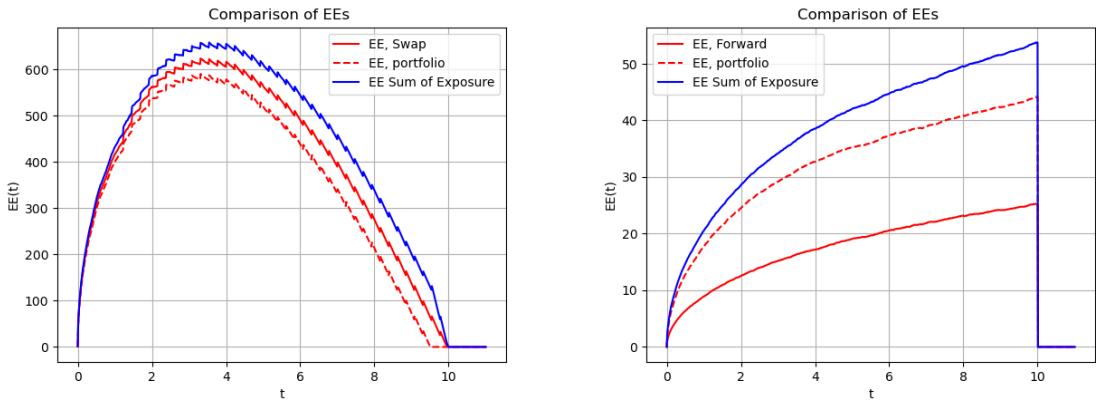


Figure 3.20: Impact of netting in 2 simple portfolios with the following parameters : ($\rho_{12} = 0.3$, $\rho_{13} = 0.4$, $\rho_{23} = 0.5$ and $\rho_{xy} = -0.99$)

Some remarks on the result :

- As we can see, the global *EE* of the portfolio is reduced and this will lead to a lower *CVA* as we mentioned in the previous chapter. The impact of netting for portfolios can then be seen as an important way to reduce the *CVA* cost.
- We can also see that under the fact the stocks price are uncorrelated with $\rho_{13} = 0.8$, the exposure of the portfolio with netting is really way lower from the exposure of the portfolio without netting and therefore according to model parameters, the impact of netting can be really important.

Chapter 4

Taking into account the Wrong Way Risk

In this chapter, a discussion about the impact of the *Wrong Way Risk* (*WWR*) in *XVAs* computation will be proposed. In a first section, the Cox setup will be introduced and will be the key point of the modelling of the *WWR*. Some numerical results on basic financial products will be proposed to see the relevance of the *WWR* in the CVA_0 computation. In the second section, a new approach introduced in [11] by Brigo and Vrins will be discussed. This approach consists in embedding *WWR* in a new probability measure using a deterministic drift adjustment. Some numerical results will also be proposed to see the relevance of the method. Finally, the impact of *WWR* on the computation of *FVA* will be studied with some numerical results on an interest rate swap.

4.1 The Wrong Way Risk for CVA

4.1.1 Mathematical framework under the Cox setup

As a starting point, we need to come back to the expression of the CVA_0 we derived in (2.7). We assumed that τ admits a density towards Lebesgue with $G(t) = P(\tau^C \geq t)$ that :

$$CVA_0 = -(1 - R^C) \int_0^T \mathbb{E}^{\mathbb{Q}}\left[\frac{(V_t)^+}{B_t} \mid \tau^C = t\right] dG(t).$$

Now and instead of what we did previously, we won't suppose the independance between τ^C and $(V_t)_{t \in [0, T]}$ and therefore we need an appropriate modelling of the quantity $\mathbb{E}^{\mathbb{Q}}\left[\frac{(V_t)^+}{B_t} \mid \tau^C = t\right]$. For this, we need to introduce an important tool which is called *\mathbb{F} -Azéma supermartingale* $S = (S_t)_{t \in [0, T]}$ defined as follows:

$$S_t = \mathbb{E}^{\mathbb{Q}}[1_{\tau^C > t} \mid \mathcal{F}_t] = \mathbb{Q}[\tau^C > t \mid \mathcal{F}_t].$$

Remark. We can observe that $\mathbb{E}^{\mathbb{Q}}[S_t] = G(t)$ which is referred to the calibration equation as in practice G will be given by the market and therefore the dynamics of S needs to be calibrated using this equation.

We can observe that S is indeed a \mathbb{F} -supermartingale as for $0 < s < t$

$$\begin{aligned} \mathbb{E}^{\mathbb{Q}}[S_t \mid \mathcal{F}_s] &= \mathbb{E}^{\mathbb{Q}}[\mathbb{E}^{\mathbb{Q}}[1_{\tau^C > t} \mid \mathcal{F}_t] \mid \mathcal{F}_s] \\ &= \mathbb{E}^{\mathbb{Q}}[1_{\tau^C > t} \mid \mathcal{F}_s] \\ &\leq \mathbb{E}^{\mathbb{Q}}[1_{\tau^C > s} \mid \mathcal{F}_s] = S_s. \end{aligned}$$

The supermartingale S has been introduced because there is an important result usually referred as *Key-Lemma* in the litterature which allows to write CVA_0 as follows : ¹

¹The proof of this lemma can be found in Lemma 3.1.3 in [12].

$$CVA_0 = (1 - R^C) \mathbb{E}^{\mathbb{Q}} \left[\frac{(V_{\tau^C})^+}{B_{\tau^C}} \mathbf{1}_{\tau^C \leq T} \right] = -(1 - R^C) \mathbb{E}^{\mathbb{Q}} \left[\int_0^T \frac{(V_t)^+}{B_t} dS_t \right]. \quad (4.1)$$

Now, we see that we are able to derive an equivalent formula to (2.8) but we are considering S instead of G . Even if we have another theoretical representation of the CVA with the equation (4.1), we have to specify a form for S to make this formula useful in practice. An important setup is the case when S is assumed to follows a *Cox Process* that is to say $S = (S_t)_{t \geq 0}$ is a positive decreasing process starting from $S_0 = 0$ with zero quadratic variation defined by : ²

$$S_t = e^{- \int_0^t \lambda_s ds}. \quad (4.2)$$

with $\lambda = (\lambda_t)_{t \in [0, T]}$ a positive \mathbb{F} -adapted process. This definition of S is highly appreciable as it allows an easy simulation of the default event τ^C according to the following proposition :

Proposition 4.1. *By defining τ^C as the following in the Cox setup :*

$$\tau = \inf \{t \geq 0 : \int_0^t \lambda_s ds \geq v\}. \quad (4.3)$$

where v denotes an exponential variable under \mathbb{Q} with parameter 1 and assumed to be independant from \mathbb{G} . Therefore, the process S is well defined with $S_t = e^{- \int_0^t \lambda_s ds}$.

Proof. Let's consider v an exponential variable with parameter 1 assumed to be independant from \mathbb{G} . We then have that considering that λ is a stochastic process \mathbb{F} -adapted : $\mathbb{Q}(\tau^C > t | \mathcal{F}_t) = \mathbb{Q}(v > \int_0^t \lambda_s ds | \mathcal{F}_t) == e^{- \int_0^t \lambda_s ds} = S_t$ which ends the proof \square

Coming back to the expression (4.1), we have the dynamics of S_t given by $dS_t = -\lambda_t S_t dt$. If we now considerer the following process $\xi = (\xi_t)_{t \in [0, T]}$ defined as :

$$\xi_t = \frac{\lambda_t S_t}{h(t)G(t)}. \quad (4.4)$$

We can therefore write the CVA_0 including WWR as follows :

$$CVA_0 = -(1 - R^C) \int_0^T \mathbb{E}^{\mathbb{Q}} \left[\frac{(V_t)^+}{B_t} \xi_t \right] dG(t). \quad (4.5)$$

As we saw in the first section, we can now derive an EPE function which now contains the WWR as follows :

$$EPE(t) = \mathbb{E}^{\mathbb{Q}} \left[\frac{(V_t)^+}{B_t} \xi_t \right].$$

Remark. *We can also verify the consistency of the approach by verifying that under independance of the processes V and λ , we have :*

$$CVA_0 = -(1 - R^C) \int_0^T \mathbb{E}^{\mathbb{Q}} \left[\frac{(V_t)^+}{B_t} \right] \mathbb{E}^Q[\xi_t] dG(t).$$

Or, we have assuming suitable integrability properties :

$$\mathbb{E}^{\mathbb{Q}}[\lambda_t S_t] = -\mathbb{E}^{\mathbb{Q}} \left[\frac{d}{dt} S_t \right] = -\frac{d}{dt} \mathbb{E}^{\mathbb{Q}}[S_t] = -G'(t) = h(t)G(t).$$

So we have the following property for ξ :

$$\mathbb{E}^{\mathbb{Q}}[\xi_t] = 1, \quad \forall t \in [0, T].$$

Therefore, it shows that the approach is consistent in the case of independance.

²The Cox process setup is widely used in quantitative finance in the modelling of credit risk.

4.1.2 Some numerical results

As we see the CVA_0 under WWR needs the computation of the quantity $\mathbb{E}^Q[\frac{(V_t)^+}{B_t}\xi_t]$. We will assume a CIR model for the process λ as it is the most common choice in the litterature and as it provided closed formulas. For this we will rely to the following proposition which gives an analytic formula for $G(t)$.

Proposition 4.2. *Assuming a process $\lambda = (\lambda_t)_{t \geq 0}$ following a CIR process³ with the following dynamic under a probability measure \mathbb{Q}*

$$d\lambda_t = \kappa^\lambda(\theta^\lambda - \lambda_t)dt + \sigma^\lambda \sqrt{\lambda_t} dW_t, \quad \lambda_0 \in \mathbb{R}_*^+. \quad (4.6)$$

Therefore, the process λ is said to be affine as the quantity $\mathbb{E}^Q[e^{-\int_t^s \lambda_u du} | \mathcal{F}_t]$ has the following form :

$$B^\lambda(t, s) = \mathbb{E}^Q[e^{-\int_t^s \lambda_u du} | \mathcal{F}_t] = A^\lambda(t, s)e^{-D^\lambda(t, s)\lambda_t}.$$

with :

- $D^\lambda(t, s) = \frac{e^{h\tau} - 1}{h + \frac{\kappa+h}{2}(e^{h\tau} - 1)}$.
- $A^\lambda(t, s) = \left(\frac{h \exp(\frac{\kappa+h}{2}\tau)}{e^{h\tau} - 1} D^\lambda(t, s) \right)^{\frac{2\kappa^\lambda \theta^\lambda}{(\sigma^\lambda)^2}}$.
- $\tau = s - t$ and $h = \sqrt{(\kappa^\lambda)^2 + 2(\sigma^\lambda)^2}$.

and the quantity $\mathbb{Q}(\tau > t)$ can then be calculated as :

$$G(t) = \mathbb{Q}(\tau > t) = \mathbb{E}^Q[e^{-\int_0^t \lambda_s ds}].$$

Proof. The proof of this result is based on solving the Riccati equations associated with the process. Starting from the fact that $P^\lambda(t, s)$ has an affine structure which will be assumed and using as we did in the previous chapter the PDE representation of P^λ , we can show that A^λ and D^λ are solution to the following system of differential equations⁴ :

$$\begin{aligned} \partial_t D^\lambda(t, s) - \kappa D^\lambda(t, s) - \frac{1}{2}\sigma^2 D^\lambda(t, s)^2 + 1 &= 0, \quad D(s, s) = 0. \\ \partial_t(\ln(A^\lambda(t, s))) - \kappa\theta D^\lambda(t, s) &= 0, \quad A(s, s) = 1. \end{aligned}$$

Solving this system leads to the form of the theorem. \square

From now, we are now able to derive the analytic formula for $G(t)$ and we will also compute $h(t)$ as follows :⁵

$$h(t) = -\frac{\partial \ln(G(t))}{\partial t}. \quad (4.7)$$

We are now able to compute the process ξ and derive the exposure profile under *Wrong Way Risk*. We will illustrate this concept with 3 different financial instruments, namely an european call and a forward contract under the $B - S$ model and an interest rate swap under the Hull & White model.

³Under the Feller condition which is $(2\kappa^\lambda \theta^\lambda > (\sigma^\lambda)^2)$, λ is a strictly positive process.

⁴The proof of this result can be found in [7] the book of Brigo and Mercurio about Interest Rate Modelling.

⁵Numerically, we will calculate h by approximating the derivative.

We give below the intensity default λ parameters we used for the numerical results :

Table 4.1: Parameters used in the numerical experiments for the intensity of default process under *Wrong Way Risk*

Parameters	λ_0	κ^λ	θ^λ	σ^λ
Value	0.12	0.35	0.12	0.12

To model the *WWR*, we will according to the model dynamics suppose that the randomneses are correlated using a parameter ρ .

An european call : In this case, we consider the following model :

$$\begin{aligned} dS_t &= S_t(rdt + \sigma dW_t^1), \quad S_0 \in \mathbb{R}_*^+ \\ d\lambda_t &= \kappa^\lambda(\theta^\lambda - \lambda_t)dt + \sigma^\lambda \sqrt{\lambda_t} dW_t^2. \\ d < W_t^1, W_t^2 >_t &= \rho dt. \end{aligned}$$

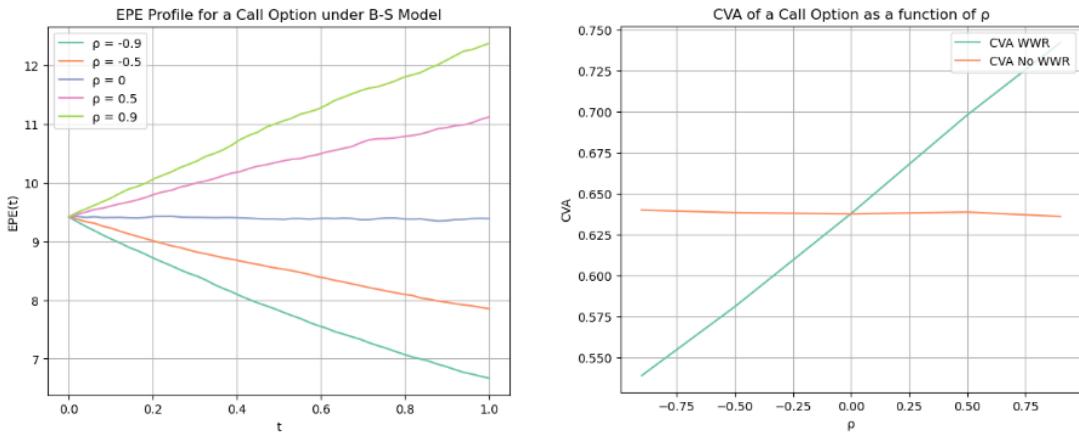


Figure 4.1: *EPE* Profile of a call under *WWR* with the following parameters : ($S_0 = 100$, $K = 100$, $r = 0.03$, $\sigma = 0.2$, $T = 1$) and corresponding CVA_0 as a function of ρ .

A forward contract : Under the same model dynamics, we also provide the results obtained when considering a forward contract.

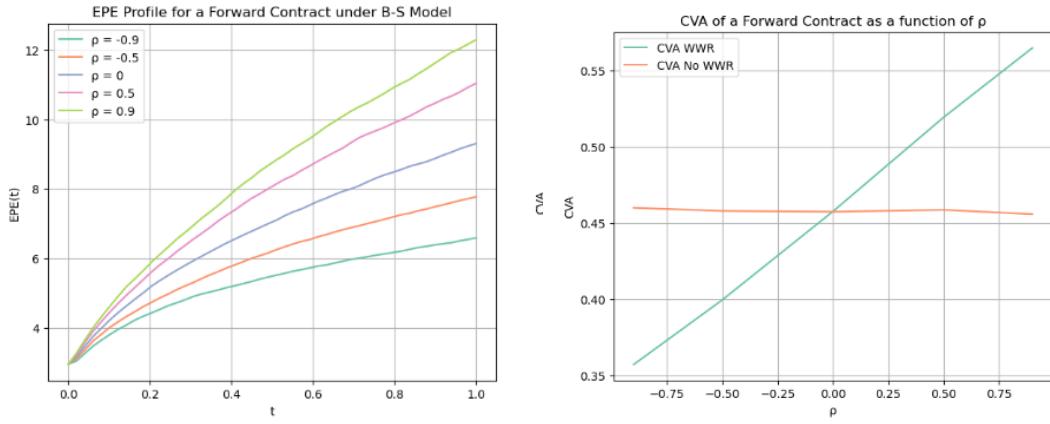


Figure 4.2: *EPE* Profile of a forward under *WWR* with the following parameters : ($S_0 = 100$, $K = 100$, $r = 0.03$, $\sigma = 0.2$, $T = 1$) and corresponding CVA_0 as a function of ρ .

An interest rate swap in the Hull & White model : In this case, we will consider the following model dynamics :

$$\begin{aligned} dr_t &= (\theta(t) - \kappa r_t)dt + \sigma dW_t^1, \quad r_0 \in \mathbb{R}. \\ d\lambda_t &= \kappa^\lambda (\theta^\lambda - \lambda_t)dt + \sigma^\lambda \sqrt{\lambda_t} dW_t^2. \\ d < W_t^1, W_t^2 >_t &= \rho dt. \end{aligned}$$

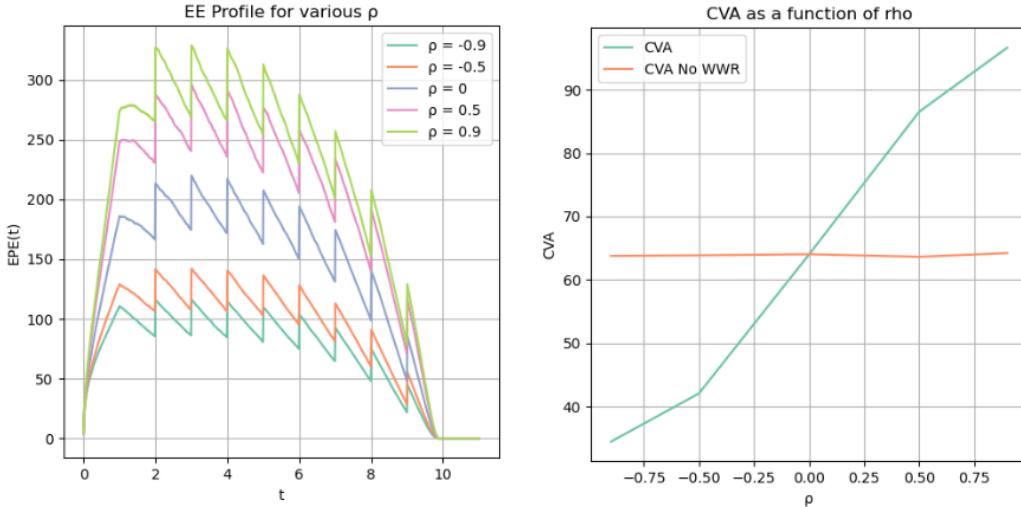


Figure 4.3: *EPE* Profile of an IRS under *WWR* with the following parameters : ($r_0 = 0.01$, $\kappa = 0.35$, $\sigma = 0.03$ with a fictitious initial ZCB curve given by $P(0, t) = e^{-r_0 t}$) and corresponding CVA_0 as a function of ρ .

Some remarks on the results :

- As we can see from the plots above, the impact of *WWR* is not negligible and has to be taken account when we are valuing *EE* of various financial products.
- In these cases, we are in situation of *Wrong Way Risk* as when ρ increases, the exposition to the default increases also.

4.2 A new way to tackle the WWR for CVA : a measure change

We will now introduce the new approach introduced in the paper [11] of Brigo and Vrins to tackle the *WWR* issue with a probability change measure and a drift adjustment.

4.2.1 The Wrong Way Measure

Let's consider the following process $(C_s^{\mathcal{F},t})_{s \in [0,t]}$ defined by :

$$C_s^{\mathcal{F},t} = \mathbb{E}^{\mathbb{Q}}\left[\frac{B_s}{B_t}\lambda_t S_t | \mathcal{F}_s\right] = B_s \mathbb{E}^{\mathbb{Q}}\left[\frac{1}{B_t}\lambda_t S_t | \mathcal{F}_s\right]. \quad (4.8)$$

As the process $(M_s^t)_{s \in [0,t]}$ defined by $M_s^t = \mathbb{E}^{\mathbb{Q}}\left[\frac{1}{B_t}\lambda_t S_t | \mathcal{F}_s\right]$ clearly defines a positive \mathbb{F} -martingale on the filtration generated by the brownian motion W driving the exposure, we can therefore using the *martingale representation theorem*⁶ say that there exists $(\gamma_s)_{s \in [0,t]}$ an \mathbb{F} -adapted process such that :

$$dM_s^t = M_s^t \gamma_s dW_s.$$

We can therefore derive the dynamics of $C^{\mathcal{F},t}$ as follows :

$$dC_s^{\mathcal{F},t} = dB_s M_s^t + B_s dM_s^t = r_s B_s M_s^t ds + B_s M_s^t \gamma_s dW_s = r_s C_s^{\mathcal{F},t} ds + C_s^{\mathcal{F},t} \gamma_s dW_s.$$

We see that the process $C^{\mathcal{F},t}$ grows as the risk free rate. As it's a strictly positive process, we can define the following Radon-Nikodym process⁷ :

$$Z_s^t = \frac{C_s^{\mathcal{F},t} B_0}{C_0^{\mathcal{F},t} B_s} = \frac{M_s^t}{M_0^t} = \frac{\mathbb{E}^{\mathbb{Q}}\left[\frac{\lambda_t S_t}{B_t} | \mathcal{F}_s\right]}{\mathbb{E}^{\mathbb{Q}}\left[\frac{\lambda_t S_t}{B_t}\right]}. \quad (4.9)$$

Then defining the probability measure $Q^{C^{\mathcal{F},t}}$ such as $\frac{dQ^{C^{\mathcal{F},t}}}{d\mathbb{Q}} \Big|_{\mathcal{F}_s} = Z_s^t$ and going back to formula 4.5, we can calculate the quantity $\mathbb{E}^Q\left[\frac{(V_t)^+}{B_t} \xi_t\right]$ as follows using the *Bayes Lemma* :

$$\mathbb{E}^Q\left[\frac{(V_t)^+}{B_t} \xi_t\right] = \mathbb{E}^{C^{\mathcal{F},t}}\left[\frac{C_0^{\mathcal{F},t}}{C_t^{\mathcal{F},t}} \xi_t (V_t)^+\right] = \frac{C_0^{\mathcal{F},t}}{h(t)G(t)} \mathbb{E}^{C^{\mathcal{F},t}}[(V_t)^+] = \mathbb{E}^{C^{\mathcal{F},t}}[(V_t)^+] \mathbb{E}^Q\left[\frac{\xi_t}{B_t}\right]. \quad (4.10)$$

Finally, we are able to derive a new formula for the *CVA* under *WWR* using the new measure $Q^{C^{\mathcal{F},t}}$. As we can see the formula is very similar to (4.5) except that we have to evaluate the exposure under the new probability measure and the bank account numeraire appears in the expectation embedding the credit risk. We can thereby derive a new formula for the *EPE* function under *WWR* :

$$EPE(t) = \mathbb{E}^{C^{\mathcal{F},t}}[(V_t)^+] \mathbb{E}^Q\left[\frac{\xi_t}{B_t}\right]. \quad (4.11)$$

⁶See Annex A.5 for a proof of this result.

⁷See Annex A.2 for an introduction to the *Bayes Lemma* and the definition of the Radon-Nikodym process in such a setting.

Remark. If we assume independance between the risk-free rate and the credit which is a common assumption in the industry, then we can write $\mathbb{E}^{\mathbb{Q}}[\frac{\xi_t}{B_t}] = \mathbb{E}^{\mathbb{Q}}[\frac{1}{B_t}] = P^r(0, t) = \mathbb{E}^{\mathbb{Q}}[e^{-\int_0^t r_s ds}]$. We can write the CVA taking account the WWR under this assumption as follows :

$$CVA_0 = -(1 - R^C) \int_0^T \mathbb{E}^{C^{\mathcal{F},t}}[(V_t)^+] P^r(0, t) dG(t). \quad (4.12)$$

4.2.2 Drift adjustment for exposure modelling

In this subsection, we will assume that the portfolio value V and the credit risk intensity λ are driven by a one dimensional \mathbb{Q} brownian motion namely W^V and W^λ and we will keep the assumption that W^λ is uncorrelated with the short-rate risk driver.

We assume the following dynamics for the portfolio V where $\mu = (\mu_s)_{s \geq 0}$ and $\sigma = (\sigma_s)_{s \geq 0}$ are assumed to be continous and \mathbb{F} adapted processes.

$$dV_s = \mu_s ds + \sigma_s dW_s^V. \quad (4.13)$$

From the *Girsanov Theorem*⁸ and by the fact that the process $C^{\mathcal{F},t}$ is the numeraire associated to the probability measure $Q^{C^{\mathcal{F},t}}$, we have that the process $\tilde{W}^V = (\tilde{W}_s^V)_{s \in [0,t]}$ defined by :

$$W_s^V - \int_0^s d\langle W^V, \ln(C^{\mathcal{F},t}) \rangle_u$$

is a $Q^{C^{\mathcal{F},t}}$ brownian motion adapted to \mathbb{F} . Following the dynamics of $C^{\mathcal{F},t}$, it is clear that we have the following dynamic for the portfolio V under $Q^{C^{\mathcal{F},t}}$:

$$dV_s = (\mu_s + \theta_s^t) ds + \sigma_s d\tilde{W}_s^V. \quad (4.14)$$

where θ_s^t is the *drift adjustment*. We also know that we have the following relation for the process θ :

$$\theta_s^t ds = \sigma_s d\langle W^V, \ln(C^{\mathcal{F},t}) \rangle_s.$$

We see that we now need to evaluate the covariation between the 2 processes W^V and $\ln(C^{\mathcal{F},t})$. For this, we need to go back to the definition of the process $C^{\mathcal{F},t}$ in (4.8). Assuming like we said independance between risk free rate and credit risk and the fact that λ is a \mathbb{F} -adapted process, we have :

$$C_s^{\mathcal{F},t} = -B^r(s, t) S_s \mathbb{E}^{\mathbb{Q}}[\lambda_t e^{-\int_s^t \lambda_u du} | \mathcal{F}_s] = -B^r(s, t) S_s \frac{\partial B^\lambda(s, t)}{\partial t}. \quad (4.15)$$

where :

- $B^\lambda(s, t) = \mathbb{E}^{\mathbb{Q}}[e^{-\int_s^t \lambda_u du} | \mathcal{F}_s]$.
- $B^r(s, t) = \mathbb{E}^{\mathbb{Q}}[e^{-\int_s^t r_u du} | \mathcal{F}_s]$.

⁸See Annex A.3 for a proof of this result.

As we see, we know need to continue the computations to specify a form for $P^\lambda(s, t)$. We will considerer the classic framework of affine stochastic processes for both the risk free rate r and the the default intensity process λ . We therefore choose the following modelling :

$$\begin{aligned} dr_s &= \mu_s^r ds + \sigma_s^r dW_s^r \\ d\lambda_s &= \mu_s^\lambda ds + \sigma_s^\lambda dW_s^\lambda. \end{aligned}$$

such as we have $d\langle W^r, W^\lambda \rangle = 0$ and the drift and diffusion coefficients have the following forms for $x \in \{r, \lambda\}$:

$$\begin{aligned} \mu_s^x &= a^x(s) + b^x(s)x_s \\ \sigma_s^x &= \sqrt{c^x(s) + d^x(s)x_s}. \end{aligned}$$

with a^x, b^x, c^x and d^x being deterministic functions which will be specify later. In the classical theory of interest rate pricing with affine models, we know how to evaluate a zero coupon bond assuming such dynamics and it ends up that :

$$\begin{aligned} B^r(s, t) &= A^r(s, t)e^{-D^r(s, t)r_s} \\ B^\lambda(s, t) &= A^\lambda(s, t)e^{-D^\lambda(s, t)\lambda_s}. \end{aligned} \tag{4.16}$$

with A^x and B^x being solutions to Riccati ordinary differentials equations that we don't need to specify here. As we know have the form of B^λ , we can therefore write :

$$\frac{\partial B^\lambda(s, t)}{\partial t} = \frac{\partial A^\lambda(s, t)}{\partial t}e^{-D^\lambda(s, t)\lambda_s} - \frac{\partial D^\lambda(s, t)}{\partial t}\lambda_s D^\lambda(s, t) = B^\lambda(s, t)\left(\frac{\frac{\partial A^\lambda(s, t)}{\partial t}}{A^\lambda(s, t)} - \lambda_s D^\lambda(s, t)\right).$$

Therefore, we have the following dynamic for $\ln(C^{\mathcal{F}, t})$ using the affine structure of both B^r and B^λ . We refer to [11] for the complete proof.

$$\begin{aligned} d\ln C_s^{\mathcal{F}, t} &= d(-\ln(S_s) + \ln(B^r(s, t)) + \ln(B^\lambda(s, t))) + \ln\left(\frac{\frac{\partial A^\lambda(s, t)}{\partial t}}{A^\lambda(s, t)} - \lambda_s D^\lambda(s, t)\right) \\ d\ln C_s^{\mathcal{F}, t} &= (\dots)ds - D^\lambda(s, t)d\lambda_s - D^r(s, t)dr_s + \frac{1}{\frac{\frac{\partial A^\lambda(s, t)}{\partial t}}{A^\lambda(s, t)} - \lambda_s D^\lambda(s, t)}d\left(\frac{\frac{\partial A^\lambda(s, t)}{\partial t}}{A^\lambda(s, t)} - \lambda_s D^\lambda(s, t)\right) \\ d\ln C_s^{\mathcal{F}, t} &= (\dots)ds + \left(\frac{A^\lambda(s, t)\frac{\partial D^\lambda(s, t)}{\partial t}}{A^\lambda(s, t)\frac{\partial D^\lambda(s, t)}{\partial t}\lambda_s - \frac{\partial A^\lambda(s, t)}{\partial t}} - D^\lambda(s, t)\right)\sigma_s^\lambda dW_s^\lambda - D^r(s, t)\sigma_s^r dW_s^r. \end{aligned} \tag{4.17}$$

From the dynamic and introducing the instant correlation between risk factors ρ_s^λ and ρ_s^r such as we have $d\langle W^V, W^\lambda \rangle_s = \rho_s^\lambda ds$ and $d\langle W^V, W^r \rangle_s = \rho_s^r ds$, we have the following expression for the drift adjustment :

$$\theta_s^t = \rho_s^\lambda \sigma_s \sigma_s^\lambda \left(\frac{A^\lambda(s, t)\frac{\partial D^\lambda(s, t)}{\partial t}}{A^\lambda(s, t)\frac{\partial D^\lambda(s, t)}{\partial t}\lambda_s - \frac{\partial A^\lambda(s, t)}{\partial t}} - D^\lambda(s, t) \right) - \rho_s^r \sigma_s \sigma_s^r D^r(s, t). \tag{4.18}$$

Deterministic approximation of the drift adjustment : As we want to focus on the impact of the WWR , we will assume deterministic risk free rates and correlation so the drift adjustment simplifies to :

$$\theta_s^t = \rho_s^\lambda \sigma_s \sigma_s^\lambda \left(\frac{A^\lambda(s, t) \frac{\partial D^\lambda(s, t)}{\partial t}}{A^\lambda(s, t) \frac{\partial D^\lambda(s, t)}{\partial t} \lambda_s - \frac{\partial A^\lambda(s, t)}{\partial t}} - D^\lambda(s, t) \right). \quad (4.19)$$

However, at this point even if we found an expression of the drift adjustment θ^t , it still has a stochastic behavior and in order to reduce the dimensionality of the problem , we can look for deterministic approximations of θ_s^t noted $\theta(s, t)$. In [11], they propose 2 possibilities :

- Replace λ_s in 4.19 by his expected value $\bar{\lambda}(s) = \mathbb{E}^Q[\lambda_s]$.
- Replace λ_s in 4.19 by the implied hazard rate $h(s)$.

Remark. They justify the connections between the two potential approximations by the fact that assuming that $Cov^Q[\lambda_t, S_t] = o(\mathbb{E}^Q[S_t])$, we then have :

$$h(s) = -\frac{d}{ds} \ln(G(s)) = -\frac{G'(s)}{G(s)} = \frac{\mathbb{E}^Q[\lambda_s S_s]}{\mathbb{E}^Q[S_s]} = \bar{\lambda}(s) + \frac{Cov^Q[\lambda_s, S_s]}{\mathbb{E}^Q[S_s]} \approx \bar{\lambda}(s). \quad (4.20)$$

Calibration step of the model : As the calibration equation $G(t) = \mathbb{E}^Q[S_t]$ needs to be verified, it means that we cannot decide to use only a *CIR* model for the process λ . We will need a deterministic shift ϕ which allows the calibration equation to be verified. Therefore, λ can be written as :

$$\lambda_t = y_t + \phi(t). \quad (4.21)$$

where :

- y is a *CIR* process with the following parameters : $y_0 = h(0)$, $\kappa^y = 0.35$, $\theta^y = 0.12$ and $\sigma^y = 0.12$.
- ϕ is such that we have $G(t) = \mathbb{E}^Q[S_t]$.

Therefore, we will have using the affine property of y :

$$\begin{aligned} B^\lambda(s, t) &= \mathbb{E}^Q[e^{-\int_s^t y_u du} | \mathcal{F}_s] e^{-\int_s^t \phi(u) du} \\ &= A^y(s, t) e^{-D^y(s, t)y_s} e^{-\int_s^t \phi(u) du}. \end{aligned} \quad (4.22)$$

Therefore the shifted process λ is also affine and by noting $\Phi(s, t) = e^{-\int_s^t \phi(u) du}$, we have :

- $A^\lambda(s, t) = A^y(s, t) e^{D^y(s, t)\phi(s) - \Phi(s, t)} e^{-D^y(s, t)\lambda_s}$.
- $D^\lambda(s, t) = D^y(s, t)$.

As λ is an affine process, we can therefore use the deterministic drift adjustment we derived in the previous paragraph. When being defined by 4.21, λ is said to be a *CIR++* process.

4.2.3 Some numerical results

In the following, we will illustrate the calculation of the EPE profile under both the *2D Monte-Carlo* diffusion setting with $M = 10000$ samples and with the drift adjustment method. For this, we will focus on a *CIR++* stochastic intensity model and will considerer different gaussian exposures that can actually reproduce the *EPE* profile of different financial derivatives.

A swap profile : Let's consider the following dynamic for the portfolio process V under Q with $\mu_s = \gamma(T - s) - \frac{V_s}{T-s}$ and $\sigma_s = \nu$:

$$dV_s = \left(\gamma(T - s) - \frac{V_s}{T - s} \right) ds + \nu dW_s^V$$

Under this dynamics, it can be shown :

$$EPE^{WWR}(s) = \sigma(s)\phi\left(\frac{\mu(s)}{\sigma(s)}\right) + \mu(s)\Phi\frac{\mu(s)}{\sigma(s)}. \quad (4.23)$$

with :

- ϕ the density of a $\mathcal{N}(0, 1)$ and Φ her CDF
- $\theta(u, s)$ the deterministic proxy obtained by replacing λ_s by $h(s)$.
- $\mu(s) = \gamma s(T - s) + (s - T) \int_0^s \frac{\theta(u, s)}{u - T} du$
- $\sigma(s) = \nu \sqrt{s(1 - \frac{s}{T})}$

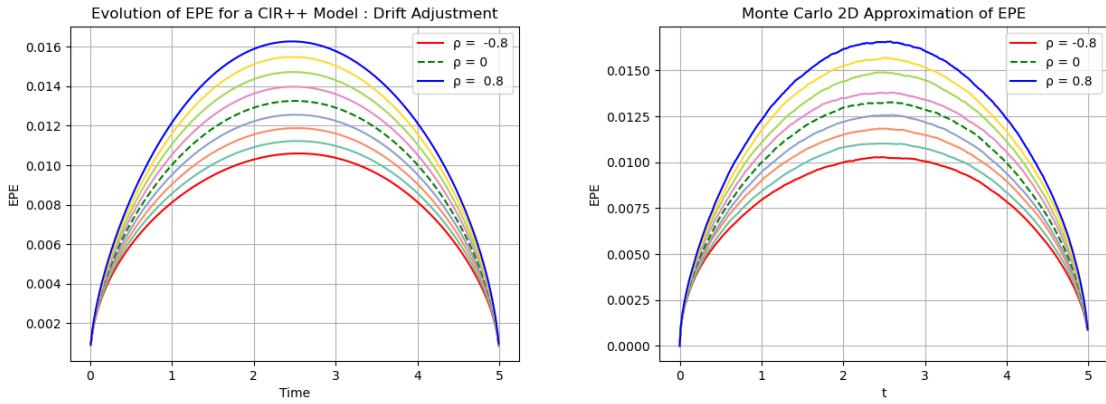


Figure 4.4: Comparison of swap exposure profile between 2D Monte-Carlo and the Drift Adjustment methods with the parameters : ($T = 5Y$, $y_0 = h = 0.15$, $\gamma = 0.001$, $\nu = 0.08$)

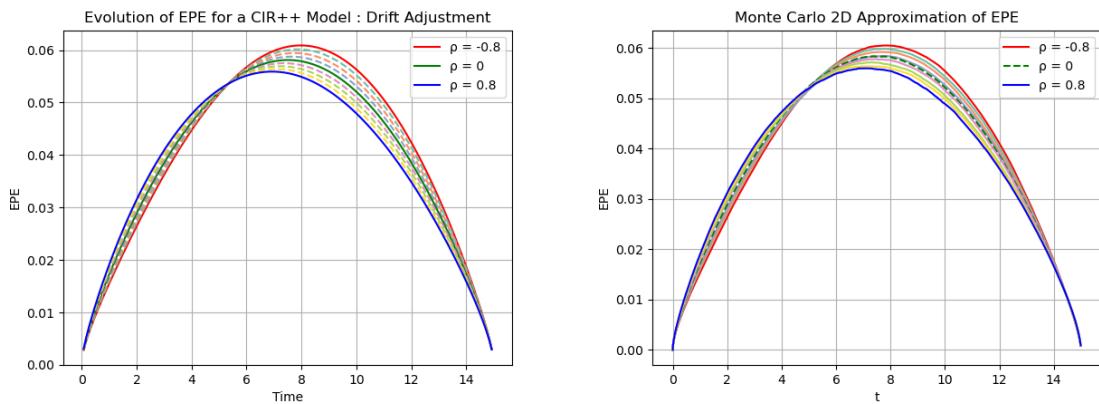


Figure 4.5: Comparison of swap exposure profile between 2D Monte-Carlo and the Drift Adjustment methods with the parameters : ($T = 15Y$, $y_0 = h = 0.30$, $\gamma = 0.001$, $\nu = 0.08$)

A forward profile : Let's consider now the following dynamic for the portfolio process V under Q with $\mu_s = 0$ and $\sigma_s = \nu$:

$$dV_s = \nu dW_s^V$$

Under this dynamics, it can be shown that :

$$EPE^{WWR}(s) = \nu \sqrt{s} \phi\left(\frac{\Theta(s)}{\nu \sqrt{t}}\right) + \Theta(s) \Phi\left(\frac{\Theta(s)}{\nu \sqrt{s}}\right) \quad (4.24)$$

with :

- ϕ the density of a $\mathcal{N}(0, 1)$ and Φ her CDF
- $\Theta(s) = \int_0^s \theta(u, s) du$

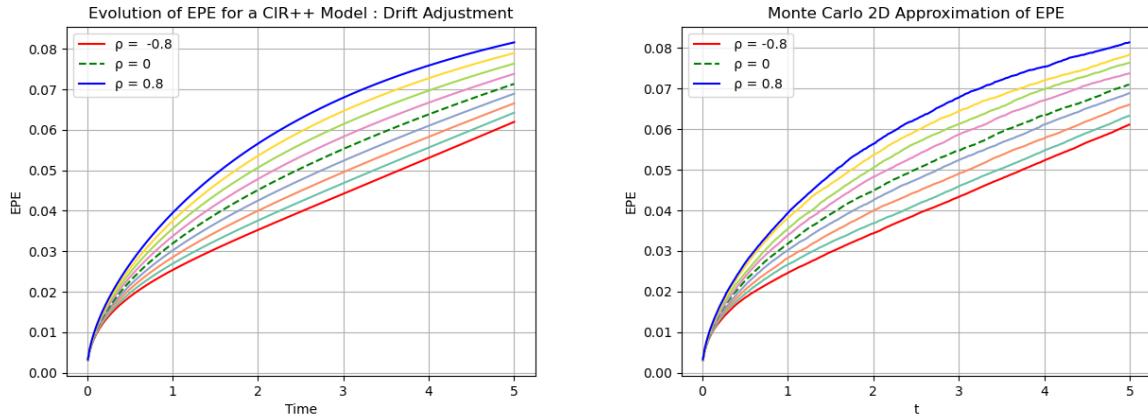


Figure 4.6: Comparison of forward exposure profile between the 2D Monte-Carlo and the Drift Adjustment with the parameters : ($T = 5Y$, $y_0 = h = 0.15$ and $\nu = 0.08$)

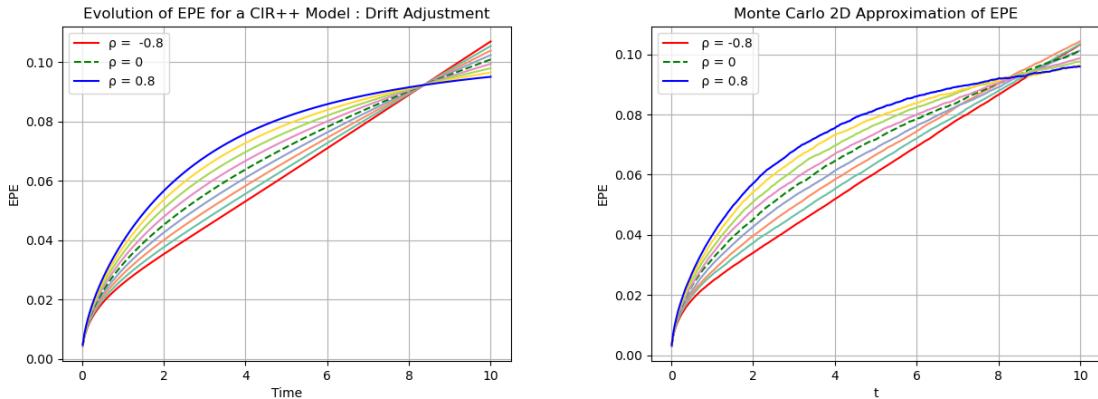


Figure 4.7: Comparison of forward exposure profile between 2D Monte-Carlo and the Drift Adjustment with the parameters : ($T = 10Y$, $y_0 = h = 0.15$ and $\nu = 0.08$)

A geometric brownian motion profile : Let's consider finally the following dynamic for the portfolio process V under Q with $\mu_s = \mu(s)V_s$ and $\sigma_s = \sigma(s)V_s$:

$$dV_s = V_s(\mu(s)ds + \sigma(s)dW_s^V)$$

Under this dynamics, it can be shown that :

$$EPE^{WWR}(s) = V_0 e^{\int_0^s \mu(u)du} e^{\int_0^s \theta(u,s)du} \quad (4.25)$$

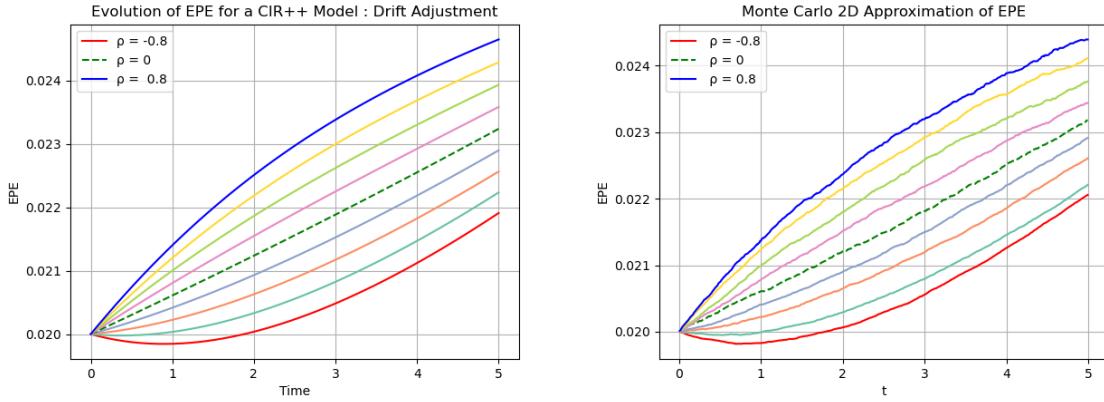


Figure 4.8: Comparison of GBM exposure profile between the *2D Monte-Carlo* and the *Drift Adjustment* methods with the parameters : ($T = 5Y$, $y_0 = h = 0.15$, $V_0 = 0.02$, $r = 0.03$ and $\sigma^V = 0.20$)

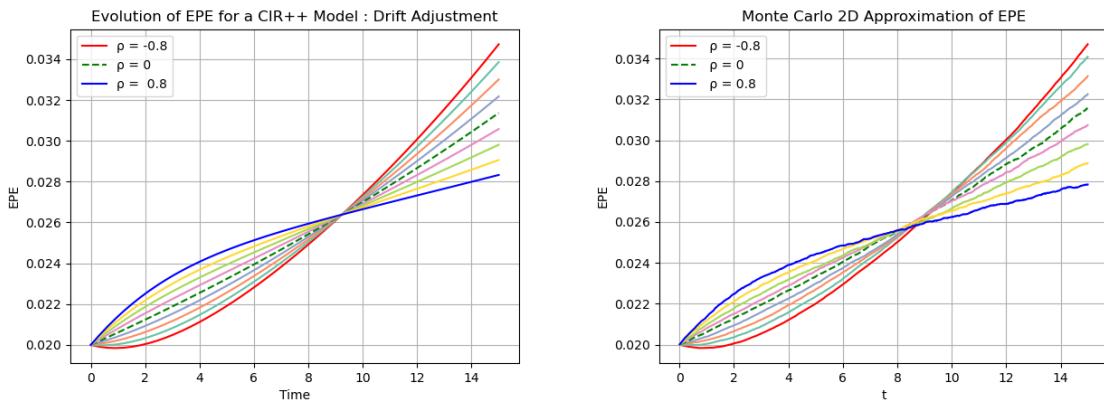


Figure 4.9: comparison of GBM exposure profile between the *2D Monte-Carlo* and the *Drift Adjustment* methods with the parameters : ($T = 15Y$, $y_0 = h = 0.15$, $V_0 = 0.02$, $r = 0.03$ and $\sigma^V = 0.20$)

Some remarks on the results :

- As we can see from the plots above, the approximation with the drift adjustment seems to perform well when we are comparing it with the actual *2D Monte-Carlo*. Note that however, this approximation is sensitive to model parameters as it has been shown in [11] when the Feller condition is violated.
- Using the drift adjustment technique allows to reduce the computational cost as in theses cases, the *EPE* profile under *Wrong Way Risk* could be computed analytically

4.3 The Wrong Way Risk for FVA

In this section, we will refer to the work introduced in [5] by Zwaard, Grzelak and Oosterlee. We will denote by $\mathbb{E}_t^{\mathbb{Q}}[\cdot] = \mathbb{E}^{\mathbb{Q}}[\cdot | \mathcal{F}_t]$.

4.3.1 FVA decomposition under WWR

We will give the decomposition of the FVA under WWR

$$FVA(t) = \mathbb{E}_t^{\mathbb{Q}} \left[\int_t^T e^{-\int_t^u \lambda_s^C + r_s ds} (s_u^B(V_u)^+ - s_u^L(V_u)^-) du \right] = \int_t^T EPE_{FVA}(t, u) du$$

$$EPE_{FVA}(t, u) = \mathbb{E}_t^{\mathbb{Q}} [e^{-\int_t^u \lambda_s^C + r_s ds} (s_u^B(V_u)^+ - s_u^L(V_u)^-)]$$

We will make the following assumptions for the computation of FVA .

- We will suppose that $s^L = 0$ so $FBA = 0$ and FVA reduces to FCA so we just take in consideration the cost of borrow and therefore the positive exposure $(V_t)^+$ profile.
- We will assume a deterministic funding spread s^B which can be either be constant or can be time dependant.
- We assume independance between defaults of A and C .

Proposition 4.3. FVA Exposure Decomposition

Under the Assumptions we defined above, the FVA can be derived as follows :

$$EPE_{FVA}(t, s) = EPE_{FVA}^{NoWWR}(t, s) + EPE_{FVA}^{WWR}(t, s)$$

$$EPE_{FVA}^{NoWWR}(t, s) = s^B(t, s) \mathbb{E}_t^{\mathbb{Q}} [e^{-\int_t^s \lambda_A(u) du}] \mathbb{E}_t^{\mathbb{Q}} [e^{-\int_t^s \lambda_C(u) du}] \mathbb{E}_t^{\mathbb{Q}} [e^{-\int_t^s r(u) du} (V_s)^+]$$

$$EPE_{FVA}^{WWR}(t, s) = s^B(t, s) \mathbb{E}_t^{\mathbb{Q}} [(e^{-\int_t^s r_u du} (V_s)^+ - \mathbb{E}^{\mathbb{Q}} [e^{-\int_t^s r_u du} (V_s)^+]) e^{-\int_t^s \lambda_A(u) + \lambda_C(u) du}]$$

Proof. The idea is to decompose $EPE_{FVA}(t, s)$ in the generic form :

$$EPE_{FVA}(t, s) = \mathbb{E}_t^{\mathbb{Q}} [f(t, s, \lambda_A, \lambda_C) h(t, s, r, V)]$$

with :

- $f(t, s, \lambda_A, \lambda_C) = e^{-\int_t^s \lambda_A(s) + \lambda_C(s) ds} s^B(s)$
- $h(t, s, r, V) = e^{-\int_t^s r_u du} (V_s)^+$

Therefore, $EPE_{FVA}(t, s)$ can be rewritten denoting $f = f(t, s, \lambda_A, \lambda_C)$ and $h = h(t, s, r, V)$ for ease of notation as follows :

$$EPE_{FVA}(t, s) = \mathbb{E}_t^{\mathbb{Q}} [f] \mathbb{E}_t^{\mathbb{Q}} [h] + \mathbb{E}_t^{\mathbb{Q}} [(h - \mathbb{E}_t^{\mathbb{Q}} [h]) f]$$

Or, under the assumption of independance between A and C , we recover the expressions for EPE_{FVA}^{NoWWR} and EPE_{FVA}^{WWR} which ends the proof. \square

As we see, we can therefore define the FVA as follows :

$$\begin{aligned} FVA^{NoWWR}(t) &= \int_t^T EPE_{FVA}^{NoWWR}(t, u) du \\ FVA^{WWR}(t) &= \int_t^T EPE_{FVA}^{WWR}(t, u) du \\ FVA(t) &= FVA^{NoWWR}(t) + FVA^{WWR}(t) \end{aligned} \tag{4.26}$$

4.3.2 Some numerical results

We will illustrate the impact of WWR for the FVA for the case of an interest rate swap under the Hull & White model.

As the aim of the FVA is to capture the funding ability of an institution in the market, we will focus on a deterministic spread capturing the institution credit risk. We will also suppose that there is no default for the counterparty C . According to [5], we define the funding spread as follows :

$$s_B(t) = (1 - R^A)\mathbb{E}^Q[\lambda_t^A] \quad (4.27)$$

We aim to evaluate the impact of FVA^{WWR} by omitting the value in comparison with FVA . We give in the table below the parameters used in the numerical results :

Table 4.2: Parameters used in the numerical experiments for the FVA pricing under WWR

Parameters	λ_0^A	κ_A	σ_A	θ_A	r_0	R^A	κ_r	σ_r
Value	0.12	0.35	0.12	0.12	0.01	0.4	0.5	0.03

FVA for an interest rate swap in the Hull & White model

We will considerer the following model :

$$\begin{aligned} dr_t &= \kappa_r(\theta(t) - r_t)dt + \sigma_r dW_t^r \\ d\lambda_t^A &= \kappa_A(\theta_A - \lambda_t^A)dt + \sigma_A \sqrt{\lambda_t^A} dW_t^A \\ d < W^A, W^r >_t &= \rho^{r,A} dt \end{aligned}$$

For the computation of s_b , we need to compute $\mathbb{E}^Q[\lambda_A(t)]$ for a CIR process which is given by the following :

$$\mathbb{E}^Q[\lambda_t^A] = \lambda_0 e^{-\kappa_A t} + \theta_A (1 - e^{-\kappa_A t})$$

We give below the plots we obtained

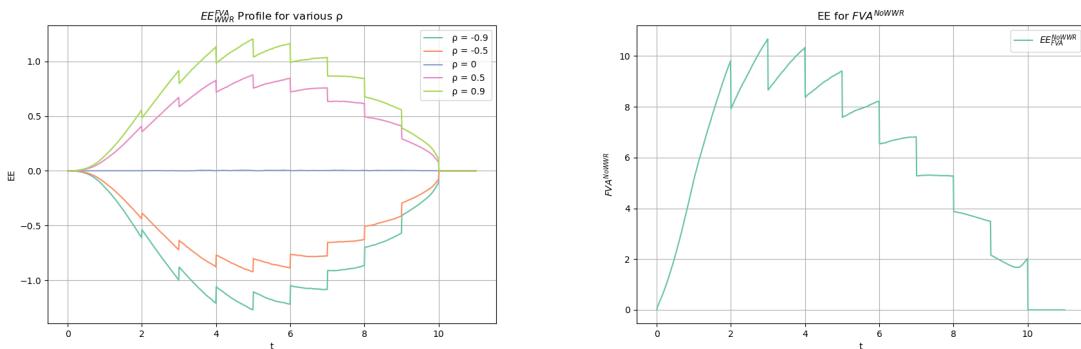


Figure 4.10: EPE_{FVA} profile for an interest rate swap (WWR (left) and $no WWR$ (right))

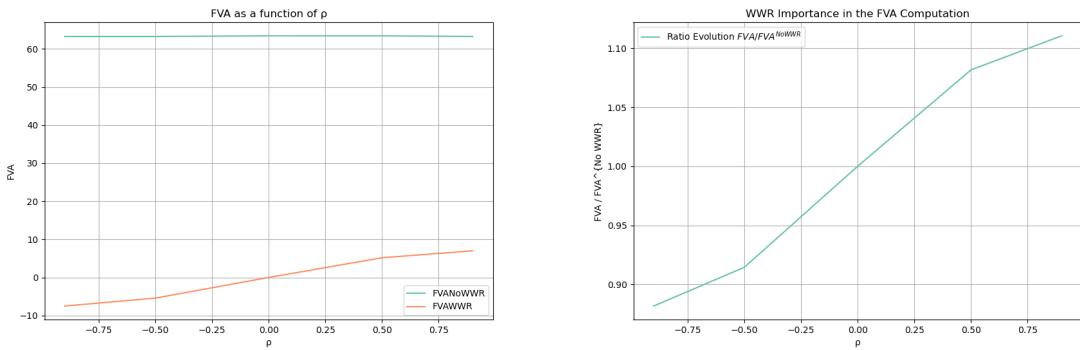


Figure 4.11: Evolution of FVA as a function of ρ and ratio of $\frac{FVA}{FVA^{No WWR}}$ to see the relative impact of WWR in FVA computation

Global remarks on the results :

- As we can see from the plots above, the impact of the WWR on the FVA can be relatively significant in the case of really highly positive correlation leading to +10% of misevaluation of the FVA . Therefore, as for the CVA computation, the WWR is an important factor that needs to be taken account by financial institutions in their FVA calculation.
- The results we obtained are of course highly dependant of the choice of parameters we took in the model. Therefore, the calibration of the parameters in the default intensity process λ^A are really important as they are the key starting points of the potential impact of the WWR .

Global remarks on the chapter :

- As we demonstrate throughout this chapter, the *Wrong Way Risk* is a crucial parameter in the quantification of the potential dependance between exposure and default. We saw through multiple numerical experiments the impact it can have on the EE profile and so on CVA_0 and FVA_0 computation.
- Throughout this chapter, we adopt some choices for our modelling purposes by defining a default intensity process λ and by considering the WWR impact by correlating the risk factors of the exposure and default intensity. We mainly refer to the CIR process which is a very common choice in the litterature as it provides a strictly positive process under the Feller condition with analytic formulas. Other methodologies to model the WWR could have been used. For example, copulas are another important tool in the modelling of the dependance between exposure and the default time of the counterparty. This tool consists of making an assumption on the joint distribution of the risk factors without considering the marginal distributions.

Chapter 5

Presentation of ML and DL methods for XVA computations

In this chapter, we will present some numerical methods based on supervised learning algorithms which will help us to overcome the main difficulties associated with the classic Monte-Carlo framework. Assuming that the price of a derivative or a *XVA* at time t is given by a function $\pi(t, X_t)$ where $t \in [0, T]$ and $X_t \in D$ where X_t denotes the underlying risk factor, we know that when we use a Monte-Carlo approach, we are able to derive the value $\pi(t, X_t)$ at a current market state $X_t = x \in D$ whereas with suitable machine or deep learning algorithms, it can help us to retrieve the function $\pi(t, .)$ on all the domain D which can be way more convenient as once the algorithm is trained, it doesn't lead to further computations. Moreover, Monte-Carlo methods suffer from the so called *curse of dimensionality* as it becomes almost unfeasible to compute prices when the dimension of the underlying d is greater than 5. Deep Learning algorithms¹ can help us to overcome this issue and make high computational challenges feasible.

We will introduce 3 supervised machine or deep learning based methods :

- Conditional Expectation Learning introduced in [13] by Huge and Savine in 2020.
- Gaussian Processes Regression introduced in [14] by Rasmussen and Williams in 2006.
- Deep BSDE Solver introduced in [15] by Weinan, Han and Jentzen in 2017.

Assume, we have N *i.i.d* samples $(X_i, Y_i)_{i \in \llbracket 1; N \rrbracket}$ defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and we want to find f^* such as $f^*(X) \approx Y$. For this, we need to minimize a *loss / criterion* l on a space of functions $\mathcal{C}(X, Y)$ which means that we have to solve the following minimization problem :

$$\min_{f \in \mathcal{C}(X, Y)} \mathbb{E}^{\mathbb{P}}[l(f(X), Y)]. \quad (5.1)$$

The algorithms above will specify the form of $\mathcal{C}(X, Y)$ and the criterion l form will depends of the problem we are tackling with. As we don't know the true law of (X, Y) , the minimisation problem 5.1 is solved using the data samples (which are assumed to be *i.i.d*)² from the distribution under \mathbb{P} as the law of large numbers holds. Therefore, the problem becomes :

$$\min_{f \in \mathcal{C}(X, Y)} \frac{1}{N} \sum_{i=1}^N l(f(X_i), Y_i). \quad (5.2)$$

¹An introduction to neural networks is presented in Annex C.

²*i.i.d* refers to independant and identically distributed.

5.1 Deep Conditional Expectation Solver

We will first introduce a deep learning based method to learn conditional expectation. As in finance and insurance, the pricing of derivatives relies to the computation of such quantities, it can therefore be interesting to design an algorithm which can perform well for these type of computations.

5.1.1 Theoretical framework

In this section, we will introduce a numerical method based on a representation of the conditional expectation we already used when we introduced the Longstaff-Schwartz Algorithm. From proposition A.4 in Annexes , we know that for 2 random variables Y and X such that $E[Y|X]$ is in $L^2(X)$, we have:

$$\operatorname{argmin}_{f \in L^2(X)} \mathbb{E}[(Y - f(X))^2].$$

As the space $L^2(X)$ leads to an infinite dimension problem, we will replace this space by the space of functions generated by neural networks parametrized by a vector θ of finite dimension denoted by f^θ . The problem can therefore be rewritten by

$$\operatorname{argmin}_\theta \mathbb{E}[(Y - f^\theta(X))^2].$$

From the definition of the problem, we see that the appropriate loss to consider is the *MSE* loss and then we can train the neural network by sampling $((X_i, Y_i))_{i \in [1;N]}$. The associated algorithm is described in 5.1.1.

Benefits of Conditional Expectation Solver :

- Assuming we want to price derivatives or *XVAs*, we don't need to compute prices as labels but only what is inside the conditional expectation which can help to reduce the computational cost of the method in case we have to compute prices using a classic numerical method such as Monte-Carlo or PDEs.
- Instead of a classical method where we price at a single current state ($X_t = x$), the neural network can help to learn the entire price function on a domain of interest D .

Algorithm 5.1 Deep Conditional Expectation Solver

Input Parameters : B batch size

Fix Architecture of NN which defines the number of parameters in the vector θ

Deep Conditional Expectation Solver (B) :

Assume N iid samples (X, Y)

Define the neural network f^θ

Minimize over θ by considering a mini-batch gradient descent with batch B

$$\frac{1}{N} \sum_{i=1}^N (Y_i - f^\theta(X_i))^2$$

end

Output Parameters : θ^* the optimized parameter of the NN

5.1.2 An application of the method in a Markovian model

In this subsection, we will illustrate the efficiency of the algorithm 5.1.1 for various derivatives in potentially high dimensions for Markovian models.

We will assume the $B - S$ for an underlying S and we will consider the case of an european option with final payoff given by $g(S_T)$ at a maturity T . The price at date t is given under the risk neutral probability measure \mathbb{Q} by :

$$C_t = \mathbb{E}^{\mathbb{Q}}[e^{-r(T-t)}g(S_T)|\mathcal{F}_t].$$

The important trick here is under the markovian property of the process $S = (S_t)_{t \in [0,T]}$, the process price $C = (C_t)_{t > 0}$ can be written as $C_t = F(t, S_t)$ for an unknown measurable function F . Therefore, we expect the neural network denoted by F^θ with θ representing the NN to approximate well the unknown form F . According to the algorithm 5.1.1 , we can look for such a candidate by putting our training data as follows :³

- X as S_t .
- Y as $e^{-r(T-t)}g(S_T)$.

In order to generate our data, we generated $N = 10000$ samples of data $(X, Y)_{i \in [1:N]}$ assuming that $X \sim \mathcal{U}(20, 200)$ ⁴. Therefore, our goal is to be able to learn the function price F on the domain $D = [20, 200]$. As we assumed a $B - S$ model, for each sample S^i , we have :

$$S_T^i = S_t^i e^{(r - \frac{\sigma^2}{2})(T-t) + \sigma \sqrt{T-t} Z^i}. \quad (5.3)$$

with r the interest risk free rate, σ the volatility and $Z^i \sim \mathcal{N}(0, 1)$.

Table 5.1: Parameters used in the numerical experiments for the $B - S$ pricing

Parameters	r	σ	t	T	K
Value	0.025	0.3	0	1.5	102.5

Table 5.2: Neural network architecture for derivatives pricing using *Deep Conditional Expectation Solver*

Number of Inputs	1
Number of Outputs	1
Number of Hidden Layers	3
Number of Neurons per Layer	15
Activation Function	Sigmoid
Weight Initialization	Xavier/Goriot
Gradient Descent Algorithm	Adam Optimizer (learning rate = 0.001)

³We will standardize our data because it shows better results by considering the standard scaling of our data such as $\tilde{X}^i = \frac{X^i - \mu_X}{\sigma_X}$ and $\tilde{Y}^i = \frac{Y^i - \mu_Y}{\sigma_Y}$ where μ and σ respectively stand for the mean and the standard deviation on the data.

⁴ $\mathcal{U}([a, b])$ refers to the uniform distribution on $[a, b]$ with density function given by $f_X(x) = \frac{1}{b-a} \mathbf{1}_{x \in [a, b]}$

Some remarks on the parametrization :

- The neural network architecture is subjective and any other decent architecture could lead to similar or better results.
- As an improvement to the algorithm we could train a neural network with more potential inputs like r , σ , T and K . This is of course possible but as we just aim to show the accuracy of the method here, we do not focus on this aspect.
- Theoretically, we can achieve the same thing for any more complex model as long as we have the Markov property and know which are the risk factors such that we can write $C_t = F(t, X_t)$ with X_t the relative risk factor which can potentially be multidimensionnal. For example, in the Heston model which is known to be markovian we will have $X_t = (S_t, v_t) \in (\mathbb{R}_*^+)^2$ where v_t stands for the variance of the underlying S_t .

Some numerical results :

We will illustrate the results for several european options starting from options for which closed formulas are known to check the accuracy of the NN approximation. We will also test the algorithm in a high multidimensionnal setting to see the scalability of the approach. For this, we will plot the true price when available, the network output and also the discounted payoffs which are the output labels in our setting.⁵

A forward contract :

We consider the case of a forward contract where they payoff is given by :

$$g(S_T) = S_T - K.$$

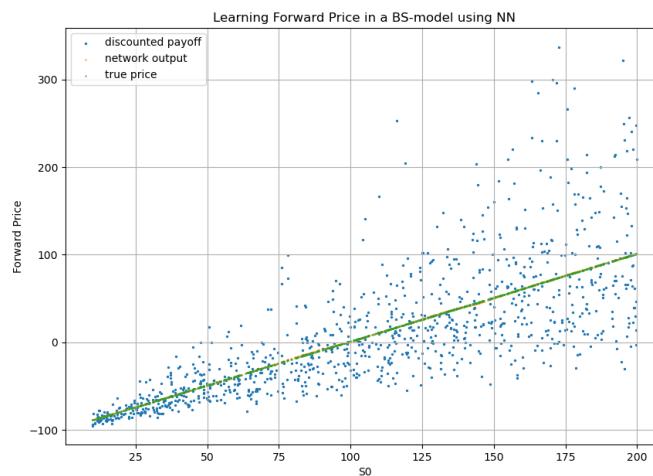


Figure 5.1: Learning the price of a forward contract in the $B - S$ model

⁵In the Annex E.5, we show how the learning process works out which shows the evolution of the MSE on the training set and the test/validation set as the MSE is the appropriate loss in our context

An european put option :

We consider the case of a put option where the payoff is given by :

$$g(S_T) = (K - S_T)^+$$

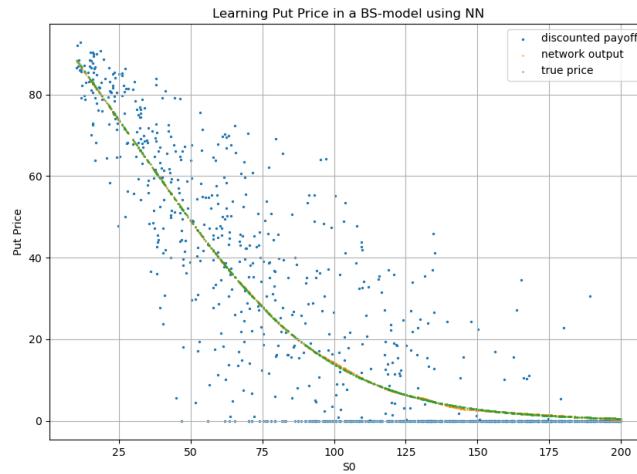


Figure 5.2: Learning the price of an european put in the $B - S$ model

A digital option :

We consider the case of a digital option with a non-continuous payoff given by :

$$g(S_T) = \mathbb{1}_{S_T > K}$$

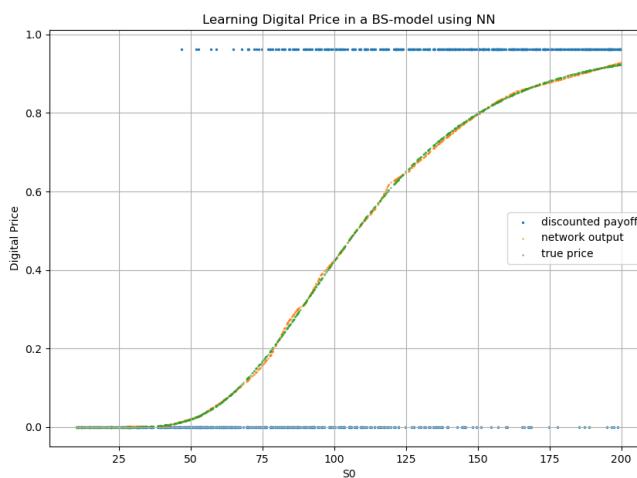


Figure 5.3: Learning the price of a digital option in the $B - S$ model

A multidimensional case :

In the following, we will illustrate the algorithm in a multidimensional case in the $B - S$ model. We will suppose 2 options on $d = 6$ assets with the following correlation matrix :

Table 5.3: Correlation matrix of the underlying factors in the $B - S$ model

1	ρ	ρ	ρ	ρ	ρ
ρ	1	ρ	ρ	ρ	ρ
ρ	ρ	1	ρ	ρ	ρ
ρ	ρ	ρ	1	ρ	ρ
ρ	ρ	ρ	ρ	1	ρ
ρ	ρ	ρ	ρ	ρ	1

We assume for the following numerical results that $\forall i \in \llbracket 1; 6 \rrbracket$, $\sigma_i = 0.3$ and $\rho = 0.3$

A max call option :

We consider the case of a Max Call Option with the following payoff given by :

$$g(S_T^1, \dots, S_T^6) = (\max_{i \in \llbracket 1; 6 \rrbracket} S_T^i - K)^+ \quad (5.4)$$

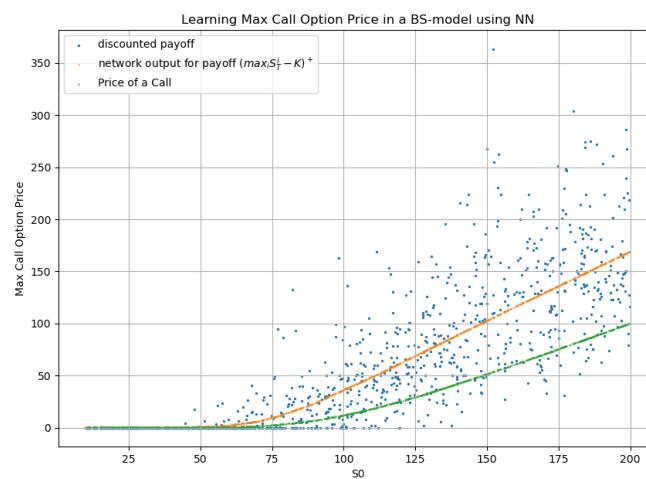


Figure 5.4: Learning the price of a max call option in the $B - S$ model with $d = 6$ assets

A min put option :

We consider the case of a min put option with the following payoff given by :

$$g(S_T^1, \dots, S_T^6) = (K - \min_{i \in [1:6]} S_T^i)^+ \quad (5.5)$$

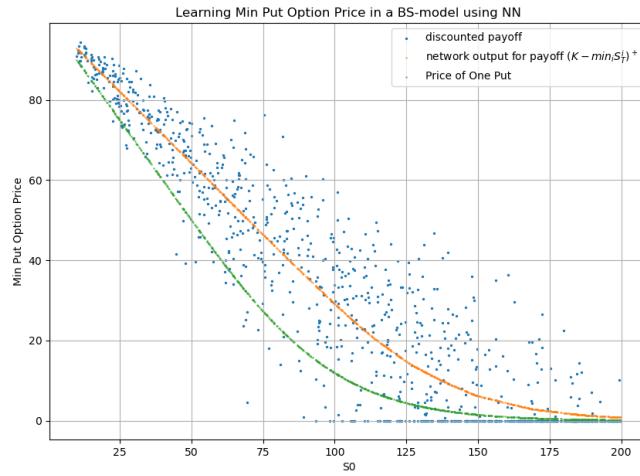


Figure 5.5: Learning the price of a min put option in the $B - S$ model with $d = 6$ assets

Some remarks on the results :

- As we can observe on each plots, there is a true convergence of the value of the NN towards a function whereas the outputs Y are still random. It is particularly significant in the case of the binary option, where the outputs labels are either 0 or 1 up to a discount factor. As the payoff option is not continuous, this is interesting to see that the *Deep Conditional Learning* algorithm still works in this case.
- For the multidimensional case, we don't provide the analytic formula but we recover the fact that the max call option is higher than the version with $d = 1$ asset. We have similar results for the min put option where we expect the price of the option to be higher.

Global remarks on the *Deep Conditional Learning Algorithm*

- This method is interesting in the sense that we don't have to compute prices for the output labels which reduce the overall computational cost when these prices needs to be computed using an appropriate numerical method like Monte-Carlo or by solving a PDE.
- This method seems to be well suited for *XVAs* as we need to compute conditional expectations. However, in the *CVA* or *FVA* cases, our labels Y will need to be directly options price at any time t during the lifetime of the financial product so if we want to use this technique to reduce the nested Monte-Carlo challenge of *XVAs*, we need to be able to compute these quantities with an efficient manner. By combining various numerical methods, it may be possible to build an efficient algorithm which benefits from the advantage of each method.
- However, in the case of the *MVA* computation, this method is particularly well suited with the *ISDA* approach as we will illustrate it in the next chapter.

5.2 Gaussian Processes Regression

We now introduce a new technique widely used in finance recently for derivatives pricing and calibration named *Gaussian Processes Regression* (\mathcal{GPR}). This technique allows for really fast computation of price surfaces. As we mention in the previous technique, we need to be able for the computations of the *CVA* and *FVA* to calculate the value of the product at any time t during the lifetime of the transaction and we will analyze \mathcal{GPR} performance to do this.

5.2.1 Theoretical framework

As we said in the introduction of this section, a machine learning implies the definition of a space function $\mathcal{C}(X, Y)$ and we will specify the form of $\mathcal{C}(X, Y)$ in the case of \mathcal{GPR} . Let's assume that our input variable is $\in \mathbb{R}^d$ with $d \in \mathbb{N}^*$ and our output variable is in \mathbb{R}

Definition 5.2.1. We say that a function $f : \mathbb{R}^d \rightarrow \mathbb{R}$ is distributed by a $\mathcal{GPR}(\mu, K_{X,X})$ if $\forall n \in \mathbb{N}^* \forall x_1, x_2, \dots, x_n \in \mathbb{R}^d$, we have that :

$$[f(x_1), f(x_2), \dots, f(x_n)] \sim \mathcal{N}(\mu_X, K_{X,X}).$$

with $\mu \in \mathbb{R}^n$ and $K_{X,X} \in \mathcal{M}_n(\mathbb{R})$ symmetric semi-definite positive matrix with general term defined by :

$$\begin{aligned} \mu_i &= \mu(x_i). \\ K_{X,X}(i, j) &= K(x_i, x_j). \end{aligned}$$

The function K is called kernel and has the following property which permits to $K_{X,X}$ to be a symmetric semi-definite positive matrix.

$$\sum_{i=1}^n \sum_{j=1}^n k(x_i)k(x_j)a_i a_j \geq 0 \text{ for any } a_k \in \mathbb{R}.$$

There exists some kernels which can be used according to the kind of problem you are dealing with.

- The *Squared Exponential Radial basis Kernel* which depends only of the norm $\|x - x'\|$ defined by $K(x, x') = \sigma^2 e^{-\frac{1}{2l^2}\|x-x'\|^2}$ where l is the length-scale parameter and σ the amplitude which characterizes the maximum of the correlation. The intuition with this kernel is as close as inputs are in term of the norm 2 distance, as more they are correlated and if they are "far", then they are uncorrelated.
- Periodic Kernels defined by $K(x, x') = \sigma^2 e^{-\frac{1}{2l^2} \sum_{j=1}^d \sin^2(\frac{\pi}{\lambda_j}(x_j - x'_j))}$ where each dimension of the input space has a period λ_j . These type of kernels are useful when dealing with time series with clear periodic effects.

In practice, an additive noise is added to the data to add one more hyperparameter to fit the model as it is done in the linear regression model. Therefore, in a \mathcal{GPR} , we assume that the data is distributed as follows :

$$y_i = f(x_i) + \epsilon_i, \quad \forall i \in \llbracket 1; N \rrbracket. \quad (5.6)$$

where $\epsilon_i \sim \mathcal{N}(0, \sigma_\epsilon^2)$ and $\sigma_\epsilon^2 > 0$ supposed to be *i.i.d* and f is a $\mathcal{GPR}(0, K_{X,X})$ ⁶. Therefore, assuming the training sample $\mathbf{x} = (x_1, x_2, \dots, x_n)$ and x^* the new input that we want to predict.

⁶In practice, f is assumed to be a \mathcal{GPR} with mean $\mu = 0$ as it can be recalibrated later.

Proposition 5.1. Under the model specification given by (5.6), we have :

$$\begin{bmatrix} \mathbf{y} \\ y^* \end{bmatrix} \sim \mathcal{N}(0, \begin{bmatrix} K(\mathbf{x}, \mathbf{x}) + \sigma_\epsilon^2 I_n & K(\mathbf{x}^*, \mathbf{x})^T \\ K(\mathbf{x}^*, \mathbf{x}) & K^* + \sigma_\epsilon^2 \end{bmatrix}) \quad (5.7)$$

Therefore, the law of $y^*|\mathbf{x}, \mathbf{y}, x^*$ is given by a $\mathcal{N}(\mathbb{E}[y^*|\mathbf{x}, \mathbf{y}, x^*], \mathbb{V}[y^*|\mathbf{x}, \mathbf{y}, x^*])$ with :

$$\begin{aligned} \mathbb{E}[y^*|\mathbf{x}, \mathbf{y}, x^*] &= \mu_{X^*} + K_{X^*, X} [K_{X, X} + \sigma_\epsilon^2 I_n]^{-1} \mathbf{y} \\ \mathbb{V}[y^*|\mathbf{x}, \mathbf{y}, x^*] &= K_{X^*, X^*} - K_{X^*, X} [K_{X, X} + \sigma_\epsilon^2 I_n]^{-1} K_{X, X^*} \end{aligned}$$

Proof. The model specification comes immediatly from the definition in (5.6). The proof is based on the following lemma :

Lemma 5.1. Assume $(X_1, X_2) \in \mathbb{R}^n \times \mathbb{R}^m$ a gaussian vector such as

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} m_1 \\ m_2 \end{bmatrix}, \begin{bmatrix} \Sigma_1 & \Sigma_{21}^\top \\ \Sigma_{21} & \Sigma_2 \end{bmatrix}\right) \quad (5.8)$$

Therefore, $X_2|X_1 = x_1 \sim \mathcal{N}(m_{2|1}, \Sigma_{2|1})$ where

- $m_{2|1} = \mathbb{E}[X_2|X_1 = x_1] = m_2 + \Sigma_{21}\Sigma_1^{-1}(x_1 - m_1)$
- $\Sigma_{2|1} = \mathbb{V}[X_2|X_1 = x_1] = \Sigma_2 - \Sigma_{21}\Sigma_1^{-1}\Sigma_{21}^T$

Proof. As $(X_1, X_2) \in \mathbb{R}^n \times \mathbb{R}^m$ is a gaussian vector, we know that for $\lambda \in \mathcal{M}_{m \times n}(\mathbb{R})$ $(X_2 - \lambda X_1)$ is independant from X_1 if and only if $Cov(X_2 - \lambda X_1, X_1) = 0$ which simplifies to $\Sigma_{21} - \lambda\Sigma_1 = 0$ meaning that by choosing $\lambda = \Sigma_{21}\Sigma_1^{-1} \in \mathcal{M}_{m \times n}(\mathbb{R})$, we have that $X_2 - \lambda X_1$ independant from X_1 . Therefore, we have that :

$$\begin{aligned} \mathbb{E}[X_2|X_1 = x_1] &= \mathbb{E}[X_2 - \lambda X_1 + \lambda X_1|X_1 = x_1] = \mathbb{E}[X_2 - \lambda X_1] + \lambda x_1 = m_2 + \lambda(x_1 - m_1) \\ \mathbb{V}[X_2|X_1 = x_1] &= \mathbb{V}[X_2 - \lambda X_1 + \lambda X_1|X_1 = x_1] = V[X_2 - \lambda X_1] = \Sigma_2 - 2\lambda\Sigma_{21}^\top + \lambda\Sigma_1\lambda^\top \\ &= \Sigma_2 - 2\lambda\Sigma_{21}^\top + \lambda\Sigma_1\Sigma_1^{-1}\Sigma_{21}^\top = \Sigma_2 - \Sigma_{21}\Sigma_1^{-1}\Sigma_{21}^\top \end{aligned} \quad (5.9)$$

which ends the proof by setting the appropriate terms in the case of the \mathcal{GPR} . \square

Therefore, using a \mathcal{GPR} prior modelling on the data allows easy computations of the posterior distribution using properties of gaussian vectors. Moreover and in comparison with the deep learning and neural network approach, we can quantify the uncertainty in the \mathcal{GPR} interpolation which is something particularly important.

Hyperparameter tuning The hyperparameter tuning is done by maximizing the likelihood of the data. It can be show from [16] that the logarithm of the marginal distribution $\log(p(Y|X, \theta))$ parametrized by the vector parameter θ is given by :

$$\log(p(Y|X, \theta)) = -[Y^\top(K_{X, X} + \sigma^2 I)^{-1}Y + \log(\det(K_{X, X} + \sigma^2 I))] - \frac{N}{2}\log(2\pi) \quad (5.10)$$

where I is the identity matrix of $\mathcal{M}_N(\mathbb{R})$.

Therefore, we aim to find the parameter θ (which depends of the type of kernel K and the number of hyperparameters used in the definition of K) which solves the following problem :

$$\operatorname{argmax}_{\theta} p(Y|X, \theta).$$

As there is not an analytic solution for θ , the maximization is performed using a gradient descent algorithm as the gradient of the log likelihood with respect to θ is given by :

$$\frac{\partial p(Y|X, \theta)}{\partial \theta} = \operatorname{Tr}(\alpha \alpha^\top - (K_{X,X} + \sigma^2 I_N)^{-1}) \frac{\partial (K_{X,X} + \sigma^2 I_N)^{-1}}{\partial \theta}.$$

where $\alpha = (K_{X,X} + \sigma^2 I_N)^{-1}Y$.

Benefits and disadvantages of GPR :

- The GPR does scale poorly with the number of data N as it involves the computation of the Cholesky decomposition for the matrix K for the training set. For the test set, it only involves a matrix multiplication which is performed in $O(N^2)$. However, it does require quite few data for the GPR to fit the actual model that's why they bring an important computational efficiency.
- GPR are used in practice because they benefit from convergence properties like neural networks. They are also linked to neural networks (with some hypothesis on the biais and weights) because considering a NN having only one hidden layer with the number of neurons M which tends to $+\infty$, then the NN tends to a GPR. The form of the GPR depends on the choice of the activation function in the neural network architecture.
- Once trained and similarly to neural networks, we can use the model to derive the map function and not only the value at a current state ($X_t = x$) unlike Monte-Carlo.
- In the case of finance and insurance, the input vector X will correspond to our risk factors depending of the model and Y will be the price of the financial instruments unlike the *Deep Conditional Learning Algorithm* where the labels were the payoffs. Therefore, using the GPR, we will need to compute noisy labels Y in case if they are estimated using appropriate numerical methods like Monte-Carlo or PDE.

5.2.2 Some applications of GPR in finance and insurance

We will show numerical applications of \mathcal{GPR} in the case of pricing some derivative contracts either finance or insurance one to illustrate the major benefits and disadvantages of the method.

A B-S setting :

Table 5.4: Parameters used in the numerical experiments in the $B - S$ setting

Parameters	r	σ	T	K
Value	0.03	0.3	1	100

In this case, we fitted the model with 40 grid points on a set $\Omega \subset [0, 1]$ which will represent the normalized value of the underlying. We used the *Radial Basis Kernel* and we test the model on 40 test points. The training process is almost immediate and the fitting is particularly accurate.

Pricing of a Call Option and the Delta under B-S :

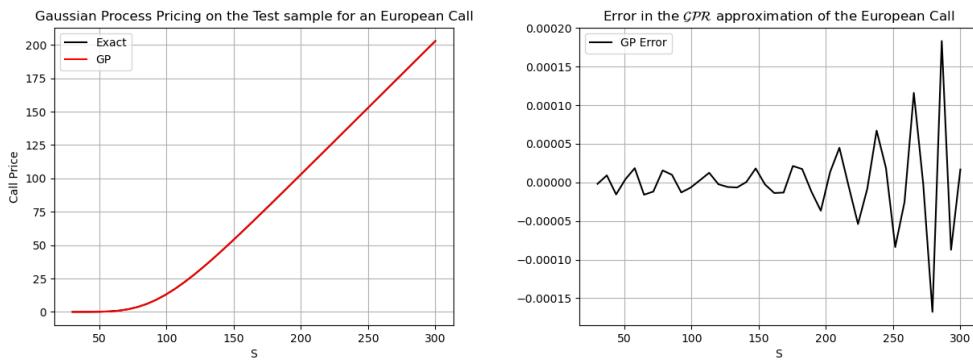


Figure 5.6: Pricing of a call option in $B - S$ model and associated error using \mathcal{GPR}

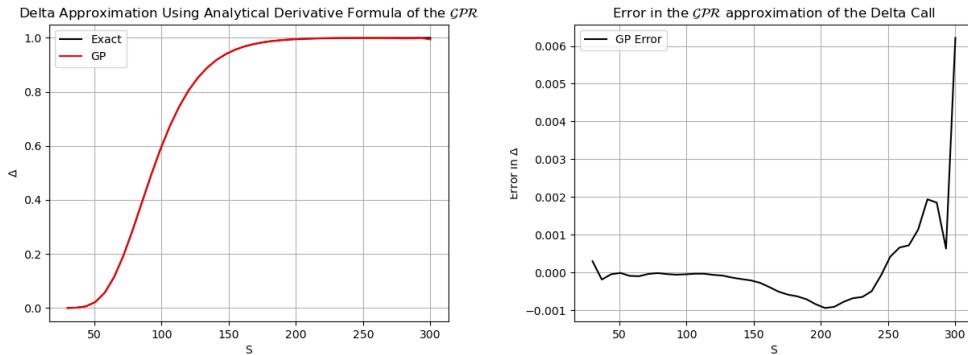


Figure 5.7: Pricing of Delta on a call option in $B - S$ Model and associated error using \mathcal{GPR}

Some remarks on the result :

- The \mathcal{GPR} approximation is really accurate as we can observe on both plots.
- The Δ is calculated through the derivative of the *Kernel* as it allows an analytic formula showing really good results.
- The model has learned the price function by showing the true labels. In the next example, we will show the performance when the output labels are noisy.

Pricing of a binary option and his delta using M-C samples :

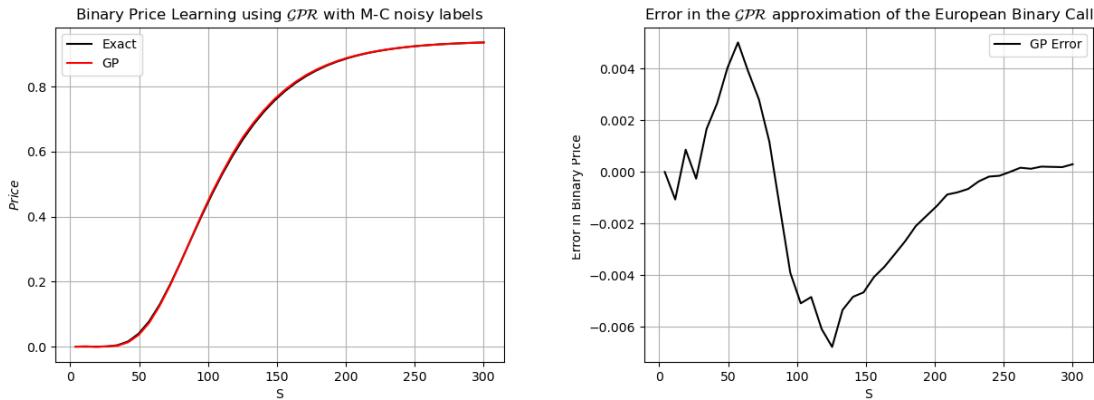


Figure 5.8: Pricing of a binary call option and associated error using \mathcal{GPR} with $M = 10^6$ $M - C$ simulations

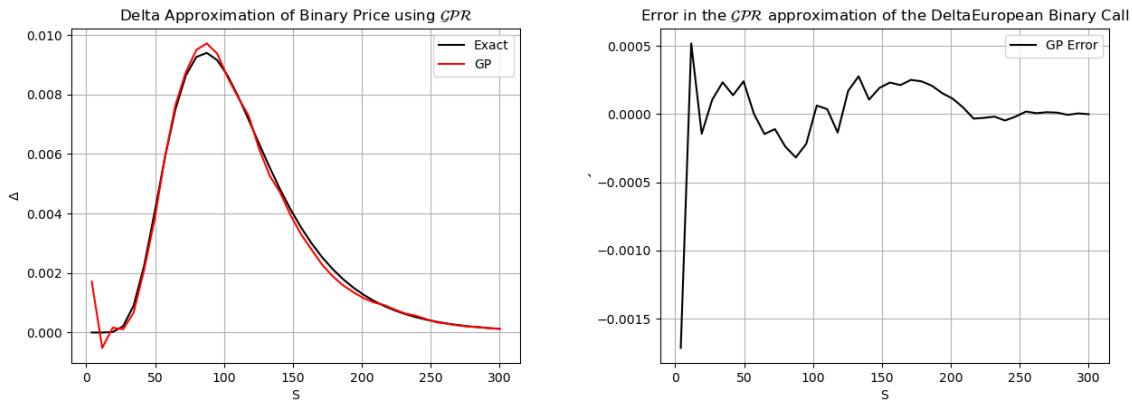


Figure 5.9: Pricing of the delta of a binary call option in $B - S$ model and associated error using \mathcal{GPR} with $M = 10^6$ $M - C$ simulations

Some remarks on the results :

- The \mathcal{GPR} approximation is not as much accurate as it was for the european call with true labels. In our case, the learning of the binary price is still decent but for the derivative *Delta*, the performance is way worst.
- The impact of the labels is therefore significant in the sense the \mathcal{GPR} will have more difficulties to learn the true pricing surface and the derivatives can have significant differences. This feature is particularly important and it highlights the fact that the data which is provided to the \mathcal{GPR} needs to be the most accurate as possible.

Pricing of a GMMB Contract :

We now present the case of a *GMMB* Contract which is defined by the following payoff at maturity T for someone who is aged x at $t=0$.⁷

$$\mathbb{1}_{\tau>T} \max(S_T, K).$$

where :

- τ denotes the mortality date of the insured starting from 0 at age x .
- S_T is the value of the underlying stock at time T with $S_0 \in \mathbb{R}_+^*$
- K is a minimum guarantee for the insured.

We assume the following dynamics for the underlying stock and the mortality rate λ for someone aged of x at $t = 0$:

$$\begin{aligned} dS_t &= S_t(rdt + \sigma dW_t^1), \\ d\lambda_t &= c\lambda_t dt + \xi \sqrt{\lambda_t} dW_t^2, \\ d\langle W^1, W^2 \rangle_t &= \rho dt. \end{aligned} \tag{5.11}$$

As we saw in a previous chapter, we define $\tau = \inf\{t \geq 0 : \int_0^t \lambda_s ds \geq \nu\}$ such that $\mathbb{Q}(\tau > t) = \mathbb{E}^\mathbb{Q}[e^{-\int_0^t \lambda_s ds}]$.

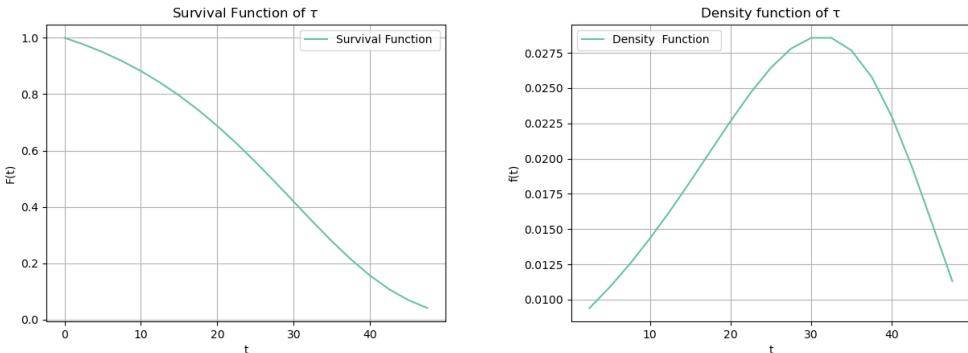


Figure 5.10: Density and Survival Function of τ with the following set of parameters : ($c = 0.0750$, $\xi = 0.000597$, $\lambda_0 = 0.0087$)

The fair value of the *GMMB* contract is defined as $t = 0$ by :

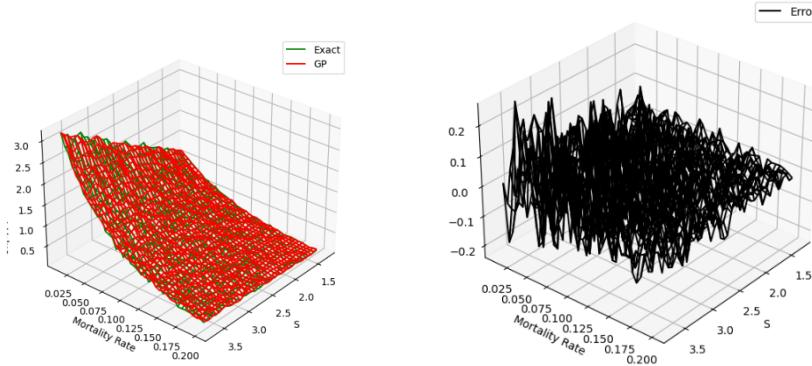
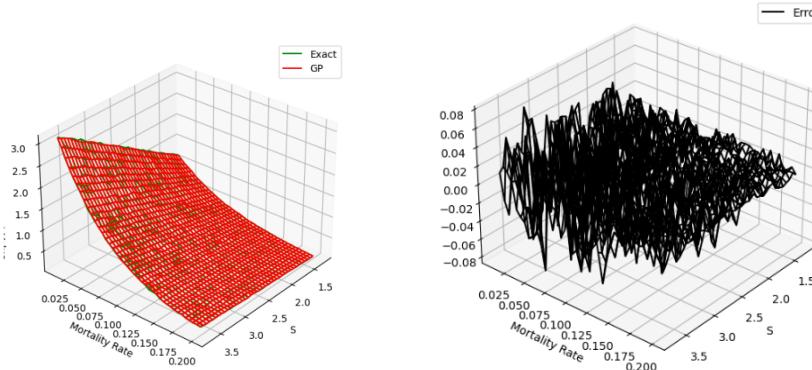
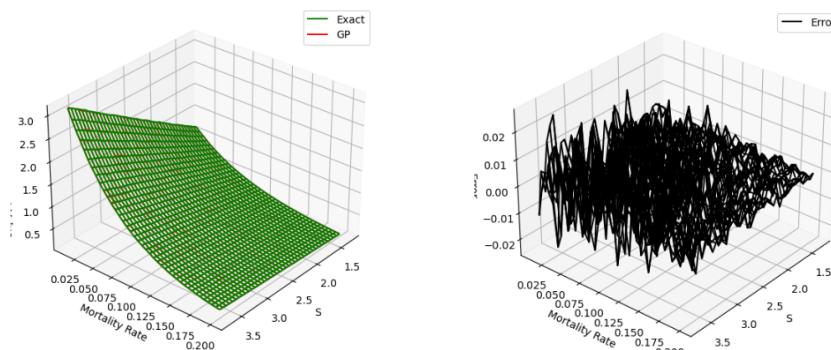
$$P_0^{GMMB} = \mathbb{E}^\mathbb{Q}[e^{-rT} \mathbb{1}_{\tau>T} \max(S_T, K)] \tag{5.12}$$

As we know that the model (5.11) is markovian, we know that P_0^{GMMB} is a function of (S_0, λ_0) so it defines our input vector \mathbf{x} . We therefore used 40×40 grid points on a set $\Omega \subset [0, 1] \times [0, 1]$. We will now provide some numerical results by showing how the number of *MC* simulations for the noisy price labels can influence the accuracy of the *GPR* in reproducing the price surface. We give below the parameters we used in the numerical experiments for the *GMMB* contract pricing.

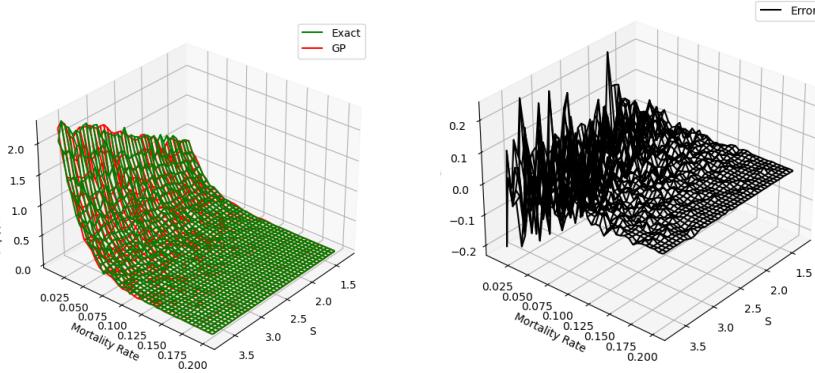
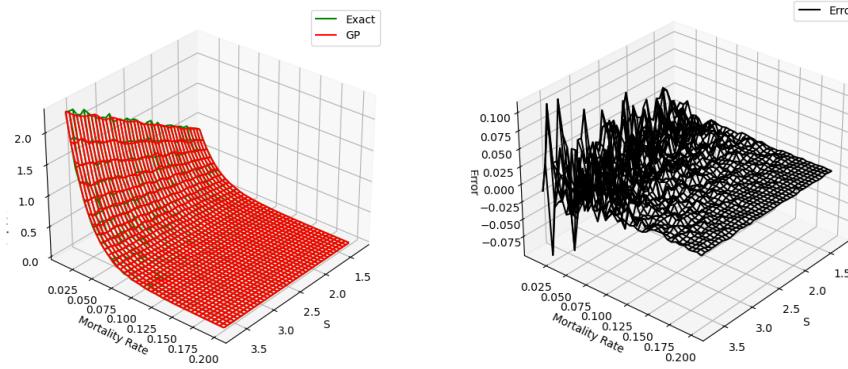
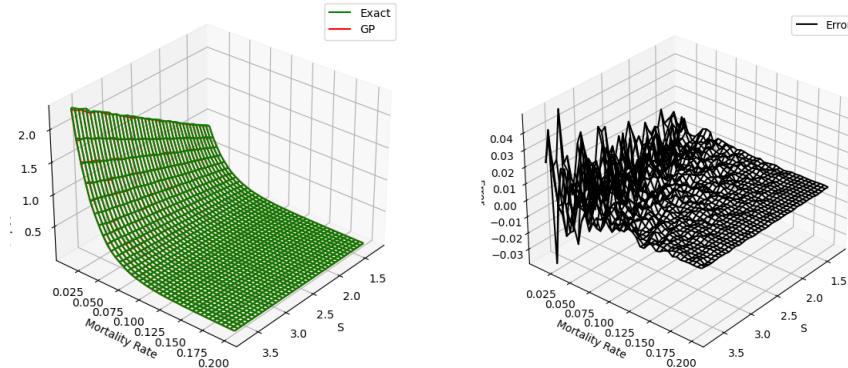
⁷The parameters are taken from the article *Actuarial-consistency and two-step actuarial valuations* of Karim Barigou.

Table 5.5: Parameters used in the numerical experiments for the GMMB contract pricing

Parameters	c	ξ	r	σ	ρ	K
Value	0.0750	0.000597	0.02	0.2	-0.7	1


 Figure 5.11: 1000 MC simulations to learn the price surface of a GMMB contract ($T = 10Y$)

 Figure 5.12: 10000 MC simulations to learn the price surface of a GMMB contract ($T = 10Y$)

 Figure 5.13: 100000 MC simulations to learn the price surface of a GMMB contract ($T = 10Y$)

- As we can see from the price evolution of the GMMB contract, it's an increasing function of both λ_0 and S_0 as we could expect from the definition of the payoff.
- However, the efficiency of the method is highly dependant of the quality of the data we are given in inputs. When the prices computed by $M - C$ are less noisy, then the learning process is more accurate as the error becomes less important.


 Figure 5.14: 1000 MC simulations to learn the price surface of a *GMMB* contract ($T = 20Y$)

 Figure 5.15: 10000 MC simulations to learn the price surface of a *GMMB* contract ($T = 20Y$)

 Figure 5.16: 100000 MC simulations to learn the price surface of a *GMMB* contract ($T = 20Y$)

Global remarks on the \mathcal{GPR} methodology

- As we show, the \mathcal{GPR} methodology can be a great tool to deal with really fast computations of full surfaces. However, it needs to learn from accurate labels as the pricing of the *GMMB* contract shows.
- In the next chapter, we will illustrate how we can use the \mathcal{GPR} algorithm to compute accurate exposure profile by learning as much \mathcal{GPR} as discretization times for the CVA_0 computation.

5.3 Deep BSDE Solver

In this subsection, we will introduce a recent methodology introduced in [15] by Weinan, Han and Jentzen to solve parabolic partial differential equations. As in the financial world, pricing problems can be related to the solution of PDE, this method has gained a lot of interest since it can handle the so called *curse of dimensionality* from which the major of classic numerical algorithms suffer. The Deep BSDE Algorithm is an algorithm that leverages deep learning algorithms and Backward Stochastic Differential Equations (*BSDEs*).

5.3.1 Theoretical framework

We assume $(\Omega, \mathcal{F}, \mathbb{Q})$ a probability space and $(\mathcal{F}_t)_{t \in [0, T]}$ a filtration adapted to a d -dimensional Brownian motion $W = (W_t)_{t \in [0, T]}$ with $T > 0$ and we introduce the following spaces :

- $L^2(\mathbb{R}^d)$ is the space of \mathcal{F}_T measurable variables in \mathbb{R}^d such that $\mathbb{E}^\mathbb{Q}[\|X\|^2] < \infty$.
- $\mathbb{S}^2(0, T)^d$ the space of progressively measurable process $Y = (Y_t)_{t \in [0, T]}$ valued in \mathbb{R}^d such that $\mathbb{E}^\mathbb{Q}[\sup_{t \in [0, T]} \|Y_t\|^2] < \infty$.
- $H_T^2(\mathbb{R}^d)$ is the space of predictable processes $Y = (Y_t)_{t \in [0, T]}$ valued in \mathbb{R}^d such that $\|Y\|^2 = \mathbb{E}^\mathbb{Q}(\int_0^T |Y_t|^2 dt) < \infty$.

The BSDE method : A generalization of Feynmann-Kac

BSDEs are an important tool and are particularly well suited to financial problems as they are stochastic differential equations with a terminal condition, very used in finance and in stochastic control theory.

Let's consider the following *BSDE* equation given by :

$$dY_t = -f(t, ., Y_t, Z_t)dt + Z_t^T dW_t^\mathbb{Q}, \quad Y_T = \xi. \quad (5.13)$$

where :

- $\xi \in L^2(\mathbb{R}^d)$.
- $Y_t \in \mathbb{R}^d$ and $Z_t \in \mathbb{R}^{n \times d}$.
- $f : [0, T] \times \Omega \times \mathbb{R}^d \times \mathbb{R}^{n \times d} \mapsto \mathbb{R}^d$ is such that the function $f(t, ., y, z)$ is a progressively measurable function called driver $\forall (y, z) \in \mathbb{R}^d \times \mathbb{R}^{n \times d}$.

However, there is no guarantee that there exists 2 process Y and Z which solve (5.13) as the solution may not be adapted to the filtration. To handle this issue, the following theorem has been derived to ensure existence and unicity to the BSDE equation.

Proposition 5.2. *If f and ξ satisfy the following properties :*

- $\xi \in L^2(\mathbb{R}^d)$.
- $f(., ., 0, 0) \in H_T^2(\mathbb{R}^d)$.
- f is uniformly Lipschitz in y and z meaning that $dt \otimes d\mathbb{Q}$ almost surely, we have for 2 constants L_y and L_z and $\forall (y_1, y_2, z_1, z_2)$

$$|f(t, w, y_1, z_1) - f(t, w, y_2, z_2)| \leq L_y(|y_1 - y_2|) + L_z(|z_1 - z_2|). \quad (5.14)$$

Then the BSDE has an unique adapted solution $(Y, Z) \in \mathbb{S}^2(0, T)^d \times H_T^2(\mathbb{R}^{n \times d})$.

Proof. The proof can be found in the book [17] from Huyêñ Pham with a proof based on a fixed point argument. \square

In finance, ξ will correspond to the terminal value and will be of the form $\xi = g(X_T)$ for a process $X = (X_t)_{t \in [0, T]}$ and under some assumptions, Y will correspond to the associated price ξ that's why it's called a generalisation of the Feymann-Kac formula. We now assume that Y is valued in \mathbb{R} .

The FBSDE representation

Let's considerer the following Forward Backard Stochastic Differential Equation (FBSDE) :

$$\begin{aligned} X_t &= x + \int_0^t b(s, X_s) ds + \int_0^t \sigma(s, X_s)^T dW_s^{\mathbb{Q}}. \\ Y_t &= g(X_T) + \int_t^T f(s, X_s, Y_s, Z_s) ds - \int_t^T Z_s^T dW_s^{\mathbb{Q}}. \end{aligned} \quad (5.15)$$

where we assume that $b : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$, $\sigma : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^{d \times d}$, $f : [0, T] \times \mathbb{R} \times \mathbb{R} \times \mathbb{R}^d \rightarrow \mathbb{R}$ and $g : \mathbb{R}^d \rightarrow \mathbb{R}$ follow the classical assumptions ensuring existence and unicity.

We denote by $(X_t^x)_{t \in [0, T]}$ and $(Y_t^y, Z_t)_{t \in [0, T]}$ the unique adapted solution to (5.15) where Y_t^y means that Y starts from $Y_0 = y$.

Let's now consider the semilinear parabolic PDE of which $u : [0, T] \times \mathbb{R}^d \mapsto \mathbb{R}$ is solution :

$$\begin{aligned} (\partial_t + \mathcal{L})u(t, x) + f(t, x, u(t, x), \sigma^T(t, x)D_x u(t, x)) &= 0, \quad \forall (t, x) \in [0, T] \times \mathbb{R}^d. \\ u(T, x) &= g(x), \quad \forall x \in \mathbb{R}^d. \end{aligned} \quad (5.16)$$

where the operator \mathcal{L} is the one of the diffusion process X that is to say :

$$\mathcal{L}(u)(t, x) = \frac{1}{2} \text{Tr}(\sigma \sigma^T(t, x) D_x^2 u(t, x)) + \langle b(t, x), D_x u(t, x) \rangle. \quad (5.17)$$

This class of PDEs is relatively large and a lot of financial pricing problems can be written as a semilinear parabolic PDE equation.

Now, let's formulate the main proposition which will be the key motivation for the design of the Deep BSDE Solver.

Proposition 5.3. *The processes $(Y_t = u(t, X_t))_{t \in [0, T]}$ and $(Z_t = \sigma^T(t, X_t)D_x u(t, X_t))_{t \in [0, T]}$ are solution to 5.15*

Proof. Let's apply the Itô formula to the process $u(t, X_t)$ assumed to be $C^{1,2}$. Therefore, we have :

$$\begin{aligned} du(t, X_t) &= (\partial_t + \mathcal{L})u(t, X_t)dt + \sigma^T D_x u(t, X_t) dW_t^{\mathbb{Q}} \\ du(t, X_t) &= -f(t, X_t, u(t, X_t), \sigma^T D_x u(t, X_t))dt + \sigma^T D_x u(t, X_t) dW_t^{\mathbb{Q}} \end{aligned}$$

By integrating and using the terminal condition, we have :

$$u(t, X_t) = g(X_T) + \int_t^T f(s, X_s, u(s, X_s), \sigma^T D_x u(s, X_s))ds - \int_t^T \sigma^T D_x u(s, X_s) dW_s^{\mathbb{Q}} \quad (5.18)$$

Therefore, we recover the form of Y_t and by unicity of the solution, it follows that Y_t and Z_t are solutions of the BSDE and are adapted to the filtration generated by $W^{\mathbb{Q}}$ \square

5.3.2 Study of the algorithm

The reformulation of the problem in terms of $FBSDE$ is related to the following stochastic optimal control problem :⁸

$$\min_{y, (Z_t)_{t \in [0, T]}} \mathbb{E}[|g(X_T) - Y_T^{y, Z}|^2]. \quad (5.19)$$

where :

- $X_t = x + \int_0^t b(s, X_s) ds + \int_0^t \sigma(s, X_s)^T dW_s.$
- $Y_t^{y, Z} = y - \int_0^t f(s, X_s, Y_s^{y, Z}, Z_s) ds + \int_0^t Z_s dW_s.$

A solution (Y, Z) to the BSDE is a minimisor of 5.19 and that's the key starting point of the Deep BSDE Algorithm. Indeed, as we have formulated the problem into a minimisation problem, we can now make use of machine learning methods. Let's consider $0 = t_0 < t_1 < \dots < t_N = T$ a uniform meshgrid with step size $\Delta t = \frac{T}{N}$ and $N \in \mathbb{N}^*$ such that $t_n = n\Delta t$. Using the Euler scheme discretization, we can approximate X_t and $Y_t^{y, z}$ as the following such that we have $\forall n \in \llbracket 0; N-1 \rrbracket$:

$$X_{t_{n+1}} = X_{t_n} + b(t_n, X_{t_n})\Delta t + \sigma(t_n, X_{t_n})(W_{t_{n+1}} - W_{t_n}), \quad X_0 = x. \quad (5.20)$$

$$Y_{t_{n+1}}^{y, Z} = Y_{t_n}^{y, Z} - h(t_n, X_{t_n}, Y_{t_n}^{y, Z}, Z_{t_n})\Delta t + Z_{t_n}^T(W_{t_{n+1}} - W_{t_n}), \quad Y_0^{y, Z} = y. \quad (5.21)$$

The idea of the Deep BSDE is to approximate at each time step t_n the control process Z_{t_n} in (5.21) by using a $FFNN$. Indeed, in a markovian setting Z_{t_n} is assumed to be in the form $\phi_n(X_{t_n})$. As we also aim to learn the optimal parameter y from the stochastic control problem, we will set y approximated by ξ as a trainable parameter of the neural network which will be optimised during the learning procedure.

Let's denote by θ a vector associated to a specified architecture of a neural network. For sake of simplicity, we will assume that each neural network at each time step has the same structure. Therefore, we can introduce a family of neural networks $(\phi_n^\theta)_{n \in \llbracket 0; N \rrbracket}$ valued from \mathbb{R}^d to \mathbb{R}^d such as by defining $Z_{t_n}^\theta = \phi_n^\theta(X_{t_n})$, we can define the following discretisation scheme :

$$Y_{t_{n+1}}^{\xi, \theta} = Y_{t_n}^{\xi, \theta} - h(t_n, X_{t_n}, Y_{t_n}^{\xi, \theta}, Z_{t_n}^\theta)\Delta t + (Z_{t_n}^\theta)^\top(W_{t_{n+1}} - W_{t_n}), \quad Y_0^{\xi, \theta} = \xi. \quad (5.22)$$

Therefore, the global minimization problem becomes :

$$\min_{\xi, \theta} \mathbb{E}[(g(X_T) - Y_T^{\xi, \theta})^2]. \quad (5.23)$$

By sampling L Monte-Carlo paths $(X_{t_n}^{(l)}, Y_{t_n}^{(l)})_{n \in \llbracket 0; N-1 \rrbracket}$ for $l \in \llbracket 1; L \rrbracket$, we will look to minimize the approximated expected cost :

$$\min_{\xi, \theta} \frac{1}{L} \sum_{l=1}^L (g(X_T^{(l)}) - Y_T^{\xi, \theta, (l)})^2. \quad (5.24)$$

⁸Other algorithms have been recently developed exploiting the same idea of the BSDE representation (see [18]).

Some remarks on the algorithm :

- The *Deep BSDE Solver* can be used to solve a lot of financial problems as long as they are semilinear which is the case of *XVAs* under suitable hypothesis.
- The *Deep BSDE solver* can overcome the *curse of dimensionality* problem that we mentionned and we will illustrate it in the next chapter where we will learn the exposure profile of a high dimensionnal european derivative.

In the following, we are giving some informations of the control error of processes Y and Z to understand better the algorithm.

Proposition 5.4. Error Bound for Y in term of error of Z ⁹

There exists a constant C independant of ξ and θ but dependant of other model parameters such that we have :

$$\sup_{t \in [0, T]} \mathbb{E}[|Y_t - Y_t^{\xi, \theta}|^2] \leq C(\mathbb{E}[(Y_0 - \xi)^2] + \int_0^T \mathbb{E}[Z_t - Z_t^\theta|^2] dt). \quad (5.25)$$

Proof. We will study the error between the true solution of the *BSDE* (Y, Z) with her neural network approximation $(Y^{\xi, \theta}, Z^\theta)$. Let's define first the continuous form of the process $Y_t^{\xi, \theta}$

$$\begin{aligned} Y_t &= Y_0 - \int_0^t f(s, X_s, Y_s, Z_s) ds + \int_0^t Z_s^T dW_s^Q. \\ Y_t^{\xi, \theta} &= \xi - \int_0^t f(s, X_s, Y_s^{\xi, \theta}, Z_s^\theta) ds + \int_0^t (Z_s^\theta)^T dW_s^Q. \end{aligned} \quad (5.26)$$

We now use the following convexity inequality for $a_1, \dots, a_N \in \mathbb{R}$ and $N \in \mathbb{N}^*$ which comes the Jensen inequality applied to the squared function.

$$(\sum_{i=1}^N a_i)^2 \leq N \sum_{i=1}^N a_i^2.$$

Therefore, we can write :

$$|Y_t - Y_t^{\xi, \theta}|^2 \leq 3(|Y_0 - \xi|^2 + |\int_0^t f(s, X_s, Y_s, Z_s) - f(s, X_s, Y_s^\theta, Z_s^\theta) ds|^2 + |\int_0^t (Z_s - Z_s^\theta)^T dW_s^Q|^2). \quad (5.27)$$

Before taking the expectation, we will use the following convexity inequality for ϕ convex and $(a, b) \in \mathbb{R}^2$ and f continuous function.

$$\phi\left(\frac{1}{b-a} \int_a^b f(x) dx\right) \leq \frac{1}{b-a} \int_a^b \phi(f(x)) dx. \quad (5.28)$$

Applying the previous inequality to the central term in 5.27 with $\phi(x) = x^2$ and using the Lipschitz property of f in y and z (noting L the associated Lipschitz constant) and the *Burkholder – Davis – Gundy* inequality, we have :

$$\mathbb{E}|Y_t - Y_t^{\xi, \theta}|^2 \leq 3(\mathbb{E}|Y_0 - \xi|^2 + 2tL_y^2 \int_0^t \mathbb{E}|Y_s - Y_s^{\xi, \theta}|^2 ds + (2tL_z^2 + 1) \int_0^t \mathbb{E}|Z_s - Z_s^\theta|^2 ds). \quad (5.29)$$

⁹The proof of this result is based from [19].

Using Grownwall lemma ¹⁰, we can write :

$$\mathbb{E}|Y_t - Y_t^{\xi,\theta}|^2 \leq 3(\mathbb{E}|Y_0 - \xi|^2 + (2tL_z^2 + 1) \int_0^t \mathbb{E}|Z_s - Z_s^\theta|^2 ds)e^{6tL_y^2}$$

From the inequality 5.30, we can by considering a constant C which only depends on L and T write :

$$\mathbb{E}[|Y_t - Y_t^{\xi,\theta}|^2] \leq C(\mathbb{E}[(Y_0 - \xi)^2] + \int_0^T \mathbb{E}[Z_s - Z_s^\theta]^2 ds) \quad \forall t \in [0, T] \quad (5.30)$$

As the equality holds $\forall t \in [0, T]$, it ends the proof. \square

From proposition 5.4, we see that the error control on Y is done by the approximation of ξ by the process $Y_t^{\xi,\theta}$ and by the control process Z^θ . In the proof, we used the continuous form of $Y_t^{\xi,\theta}$ which is theoretical and we have to consider in the global error term the error associated to the discretisation.

Other bounds on the algorithm can be found under suitable assumptions on the coefficients of the *FBSDE* in [19].

Use cases of the Deep BSDE Solver in derivatives pricing :

Let's consider a Black-Scholes model defined by the following dynamics of the underlying S :

$$dS_t = S_t(rdt + \sigma dW_t), \quad S_0 \in \mathbb{R}_+^*. \quad (5.31)$$

Let's consider the pricing of an European derivative with payoff $g(S_T)$. It is known that the price u of the option is given by solution of the following PDE :

$$\begin{aligned} \partial_t u + \frac{1}{2}\sigma^2 S^2 \partial_{S^2} u + rS \partial_S u - ru &= 0 \quad \forall(t, x) \in [0, T] \times \mathbb{R}_+^*. \\ u(T, x) &= g(x) \quad \forall x \in \mathbb{R}_+^*. \end{aligned}$$

In this setting, we have that $f(t, x, y, z) = -ry$ which verifies the classical assumptions and therefore the Deep BSDE Solver Algorithm can be used in the case of an european derivative and we will use this representation for the *XVA* computation also in higher dimensions. The case of pricing using the *Deep BSDE Solver* is highly currently studied in the litterature as it has opened new doors to overcome the limits of the classical numerical methods. ¹¹

¹⁰See <https://www.ceremade.dauphine.fr/~mischler/Enseignements/M2evo12018/chap0.pdf> for an introduction to the lemma.

¹¹A study of barrier options with the *Deep BSDE Solver* has been done in the article *Deep-learning based numerical BSDE method for barrier options* from Yu, Xing and Sudjianto where they rewrite integrate these type of products in the *Deep BSDE* framework.

Chapter 6

Application of ML and DL methods for XVA computations

6.1 Deep Conditional Expectation Solver for XVA computations

To illustrate the *Deep Conditional Expectation Solver*, we will come back to the definition of CVA_0 and FVA_0 ¹ and we need to distinct 2 cases :

- The first case when we suppose that we know how to sample the default time τ^C .
- The second case when we don't know how to sample the default time τ^C .

Using a sampling of τ^C :

In the case where we are able to sample τ^C , we know that we can evaluate CVA_0 by using the following representation :

$$CVA_0 = \mathbb{E}^{\mathbb{Q}}[\mathbf{1}_{\tau^C \leq T} \frac{(V_{\tau^C})^+}{B_{\tau^C}} | \mathcal{G}_0].$$

Not using a sampling of τ^C :

In the case where we are not able to sample τ^C , we saw that we could under no *WWR* formulate an approximation of the *CVA* given by equation 2.9 :

$$CVA_0 \approx -(1 - R^C) \sum_{i=0}^{N-1} EPE(t_i)(G(t_{i+1}) - G(t_i)).$$

$$FVA_0 \approx \sum_{i=0}^{N-1} P(\tau^A > t_{i+1})P(\tau^C > t_{i+1})(\mathbb{E}^{\mathbb{Q}}[\frac{1}{B_{t_{i+1}}} (s_b(t_{i+1})(V_{i+1})^+ - s_L(t_{i+1})(V_{i+1})^-)](t_{i+1} - t_i)).$$

$$EPE(t_i) = \mathbb{E}^{\mathbb{Q}}[\frac{1}{B_{t_i}}(V_{t_i})^+ | \mathcal{F}_0].$$

These 2 forms of the *XVA* will help us to illustrate the *Conditional Expectation Learning Algorithm* 5.1.1 as we have for both representations to calculate conditional expectations.

¹We won't focus on the version of *CVA* or *FVA* under *Wrong Way Risk* in the further computations.

6.1.1 An application to CVA

Case of sampling τ^C :

In this case, according to algorithm 5.1.1, we will need to simulate N i.i.d samples of the form $(X_i, Y_i)_{i \in [1; N]}$ where in our case we have :

- X_i is the Markov state of the $i - th$ sample in the considered model.

$$\bullet Y_i = \mathbb{1}_{\tau_i^C < T} \frac{(V_{\tau_i^C})^+}{B_{\tau_i^C}}.$$

As we did in the previous experiments, we assumed that $S_0 \sim \mathcal{U}(20, 200)$ and $\tau \sim \mathcal{E}(\lambda^C)$.

For the training process of the neural network, we train it using $N = 10000$ samples of (X, Y) . As we can see, it requires to be able to price the derivative on which the XVA is calculated at any potential time between 0 and T so it doesn't reduce the computational cost of a nested Monte-Carlo but it can atleast help to learn the XVA function on the domain of definition of the markovian state of the process.

We provide again some numerical results in the case of a $B - S$ model for a forward contract and an european put option as we have closed formulas to evaluate $V_{\tau^C}^+$ in this context.

We give below the parameters used in the numerical results.

Table 6.1: Parameters used in the numerical experiments for CVA_0 pricing using *Deep Conditional Expectation Solver*

Parameters	r	σ	T	K	R^C	λ^C
Value	0.02	0.2	1	100	0	0.5

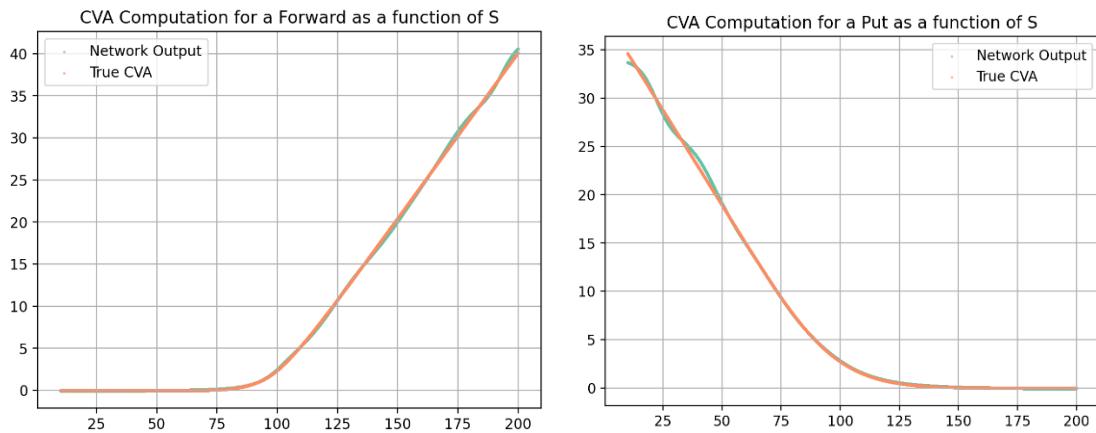


Figure 6.1: CVA_0 learning in a $B - S$ model by sampling τ^C for a forward (left) and an european put (right)

Table 6.2: Metric errors after the learning phase in *Deep Conditional Expectation Solver* for CVA_0 computation

	MSE	Forward	European Put
Training Set	0.757	0.719	
Test Set	0.745	0.720	

Some potential improvements on the methodology :

As we said, the main disadvantage of the method is the fact that we need to compute $(V_t)^+$ at any time $t \in [0, T]$. In this case the algorithm doesn't help to reduce the computational cost. However, we can imagine combining an other neural network method to calculate efficiently the surface price as we know that in our context that $V_t = C(t, S_t)$.

Case of not sampling τ^C :

Following the representation of CVA_0 under (2.9), we see that knowing the exposure profile at a timegrid $0 = t_0 < \dots < t_N = T$ is sufficient to characterize the CVA . Therefore, we can consider to compute the following vector **EPE** defined as follows :

$$\begin{aligned}\mathbf{EPE} &= (EPE(t_0), \dots, EPE(t_N)). \\ \mathbf{EPE} &= (\mathbb{E}^\mathbb{Q}[\frac{1}{B_{t_0}}(V_{t_0})^+ | \mathcal{F}_0], \dots, \mathbb{E}^\mathbb{Q}[\frac{1}{B_{t_N}}(V_{t_N})^+ | \mathcal{F}_0]). \\ \mathbf{EPE} &= \mathbb{E}^\mathbb{Q}[(\frac{1}{B_{t_0}}(V_{t_0})^+, \dots, \frac{1}{B_{t_N}}(V_{t_N})^+) | \mathcal{F}_0].\end{aligned}$$

Following this definition, we can look to approximate the vector **EPE** using the *Deep Conditional Expectation Solver* by assuming that the initial market state \mathcal{F}_0 is characterized by a vector X . Thereby, we can write $\mathbf{EPE} = (F_{t_0}(X), \dots, F_{t_N}(X)) = \mathbf{F}(X)$ for unknown deterministic functions F_{t_0}, \dots, F_{t_N} .²

A numerical result with a call option :

We suppose a $B - S$ model as it allows computation of closed formulas for a call option. We assume that the model is calibrated and we use $r = 0.03$, $\sigma = 0.2$ and $T = 1$ in the computations.

The inputs of our neural network will be both the current value of the option S_0 and the associated strike K assumed to be sampling using an uniform distribution such that $S_0 \sim \mathcal{U}(70, 130)$ and $K \sim \mathcal{U}(70, 130)$ and we will considered the output vector to be of dimension 100 meaning that we consider 100 time steps for the computation of the exposure profile. As we said previously, the main issue of the method is that we still need to compute the price of the product at discrete times $0 = t_0 < \dots < t_N = T$ for the output labels. However, assuming that it can be done in an efficient manner, the method can be interesting as it will immediately provide the exposure profile leading to direct computation of CVA_0 and FVA_0 .

We give below the architecture of the neural network we used :

Table 6.3: Neural network architecture for the *EPE* profile computation of a call in the $B - S$ model using the *Deep Conditional Expectation Solver*

Number of Inputs	2
Number of Outputs	100
Number of Hidden Layers	3
Number of Neurons per Layer	256
Activation Function	$\phi(x) = x^+$ (ReLU)
Weight Initialization	Xavier/Goriot
Gradient Descent Algorithm	Adam Optimizer (learning rate = 0.001)

²In this case, we will consider a neural network with a multidimensional output.

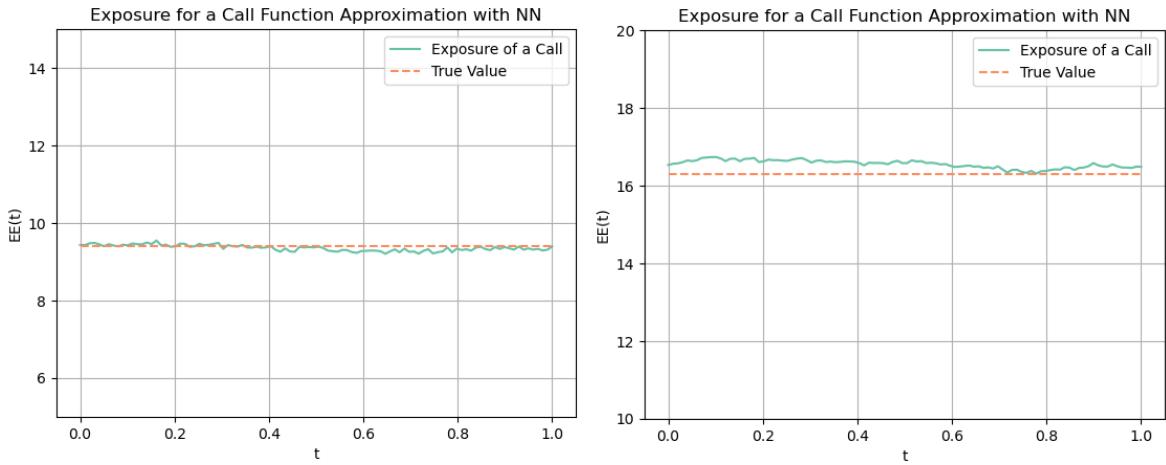


Figure 6.2: *EPE* profile learning using *Deep Conditional Expectation Solver* for a call option with ($S_0 = 100$ and $K = 100$ on left) and ($S_0 = 110$ and $K = 100$ on right)

As we can see from the figure 6.2 above, the neural network seems to fail learning decently the exposure profile. We can see that in the case where we put as inputs of the NN ($S_0 = 100$ and $K = 100$), the output profile is decent but in the case of ($S_0 = 110$ and $K = 100$), it does make quite a lot of error.

Some remarks on the *Deep Conditional Expectation Solver* for CVA_0 and FVA_0 computations

- Assuming we can't sample τ^C , the computation of CVA and FVA both need the knowledge of the exposure profile. In the case we don't have closed formulas for prices at any time t during the lifetime of the transaction which is the most common case in more complicated models, therefore the algorithm doesn't seem to be a good alternative to the classical nested Monte-Carlo if it is not combined with other numerical methods to make it useful in practice.
- Moreover, from the numerical results we did on an european call, we see that the learning of the exposure is not really efficient and there is still a lot of error in the *NN* approximation which could be explained by the *NN* architecture.

The algorithm therefore needs to be improved to be useful in practice but look promising as we can expect to learn a function that can learn the expected profile on discrete times during the life time of the transaction. Working on the neural network architecture by maybe using other architectures than simple *FFNN* could maybe improve the learning process. An appropriate hyperparameter tuning of the neural networks can also be relevant as it could improve globally the performance of the neural networks.

6.1.2 An application to DIM and MVA

As we mentioned in the first chapters, initial margin named as IM aims to cover the potential future exposure that could arise between the default time of a counterparty and the time that the surviving party has closed its position after default. This is the key component of the MVA computation which we will discuss in this subsection. We will refer to the $SIMM$ methodology which we will introduce below for the computation of this XVA .

The SIMM framework : International Swaps and Derivatives Association (ISDA) has introduced the Standard Initial Margin Model (SIMM) which is a sensitivity approach based on the risk profile of a position and which is defined for the 4 following product classes :

- RatesFX
- Credit
- Equity
- Commodity

Each trade is assigned to an individual product class and SIMM is calculated separately for each product class. For each product class, we can define the risk class that are part of the product class and each risk class is kept separated between product classes even if they share a same risk class. The SIMM considers the 6 following risk classes :

- Interest Rate
- Credit (Qualifying)
- Credit (Non Qualifying)
- Equity
- Commodity
- FX

From that, they define the IM for each risk class considering delta, vega and curvature sensitivities such as we have for each risk class R :

$$IM_R = DeltaMargin_R + VegaMargin_R + CurvatureMargin_R. \quad (6.1)$$

They then define $SIMM_X$ the initial margin associated with the product class X as follows :

$$SIMM_X = \sqrt{\sum_r IM_r^2 + \sum_{r,s} \psi_{r,s} IM_r IM_s}. \quad (6.2)$$

where IM_r refers to the margin associated with the risk class r and $\psi_{r,s}$ the correlation between the risk classes r and s with a correlation matrix given in Table 6.4 from [20]. The total SIMM is then the sum over each product class resulting in :

$$SIMM = SIMM_{RatesFX} + SIMM_{Credit} + SIMM_{Equity} + SIMM_{Commodity}. \quad (6.3)$$

In [20], they give details of how to calculate the sensitivities in different scenarios.

Table 6.4: Correlation matrix of the risk classes in the SIMM framework

Risk Class	Interest Rate	Credit Qualifying	Credit Non Qualifying	Equity	Commodity	FX
Interest Rate		4%	4%	7%	37%	14%
Credit Qualifying	4%		54%	70%	27%	37%
Credit Non Qualifying	4%	54%		46%	24%	15%
Equity	7%	70%	46%		35%	39%
Commodity	37%	27%	24%	35%		35%
FX	14%	37%	15%	39%	35%	

As we wrote in equation (2.29), the dynamic initial margin profile can be calculated as :

$$DIM(t) = \mathbb{E}^Q[e^{-\int_0^t r_u du} IM(t)|\mathcal{F}_0].$$

Once this quantity is computed and according to a funding spread between the collateral rate and the risk free rate denoted by f , we can then compute the MVA_0 as follows :

$$MVA_0 = \int_0^T f(s) DIM(s) ds.$$

As we can see, the main feature of the computation of the MVA is the calculation of the DIM as it will involve the calculation of the IM profile.

We will considerer the following approach introduced in [21] by Villarino and Leitao where they used the *Deep Conditional Expectation Solver* to compute in an efficient manner the DIM profile and associated MVA_0 . The methodology is similar to the one we proposed in the previous subsection so we just formulate it briefly.

As we can approximate the equation (6.4) by the following approximation considering $t_0 = 0 < t_1 < \dots < t_N = T$ a timegrid of $[0, T]$.

$$MVA_0 \approx \sum_{i=0}^{N-1} \frac{1}{2}(f(t_i)DIM(t_i) + f(t_{i+1})DIM(t_{i+1}))(t_{i+1} - t_i).$$

We consider simlarly to the vector **EPE** the vector **DIM** $\in \mathbb{R}^{N+1}$ defined as times $0 = t_0 < \dots < t_N = T$. According to the equation (2.29), we can therefore write the vector **DIM** as the following :

$$\mathbf{DIM} = (\mathbb{E}^Q[IM(t_0)|\mathcal{F}_0], \dots, \mathbb{E}^Q[e^{-\int_0^{t_N} r_s ds} IM(t_N)|\mathcal{F}_0]).$$

Now, assuming that \mathcal{F}_0 is characterized by a a vector X of initial market state variable assumed to be markovian , we then know that we can rewrite the vector **DIM** using deterministic functions $(F_{t_i})_{i \in [0, N]}$. If we note $\mathbf{F} = (F_{t_0}, \dots, F_{t_N})$, we then have :

$$\mathbf{DIM} = (\mathbb{E}^Q[IM(t_0)|X], \dots, \mathbb{E}^Q[e^{-\int_0^{t_N} r_s ds} IM(t_N)|X]) = (F_{t_0}(X), \dots, F_{t_N}(X)) = \mathbf{F}(X). \quad (6.4)$$

We then know aim to approximate **F** by using the subspace of neural networks. Writing down the vector $\mathbf{IM} = (IM(t_0), \dots, e^{-\int_0^{t_N} r_s ds} IM(t_N))$, we have also the following representation for **DIM** :

$$\mathbf{DIM} = \mathbb{E}^Q[\mathbf{IM}|X]. \quad (6.5)$$

We write below the following algorithm which will be used to assess the accuracy of the *Deep Conditional Expectation Solver*³ by calculating the full nested Monte-Carlo procedure.

Algorithm 6.1 Nested Monte-Carlo algorithm for *DIM* computation (Algorithm from [21])

Inputs : Simulation times $0 = t_0 < t_1 < \dots < t_N = T$ with h_t time step.

Yield curve points at the ISDA-SIMM tenors for time t_0 , $\{Y_k^0\}_{k=1}^{12}$.

Model parameters Θ .

Number of MC paths, M .

Samples of the standard normal distribution i.i.d., $(W)_j^i$, $j = 0 \dots, M - 1$, $i = 0 \dots N - 1$.

Output : Compute DIM_i , the dynamical IM at time t_i .

$x^0 \leftarrow Y^0$

for $i = 1, i < N + 1, i++$ **do**

for $j = 0, j < M, j++$ **do**

$x_j^i \leftarrow (x_j^{i-1}, W_j^i, \Theta)$

$\{Y_{k,j}^i\}_{k=1}^{12} \leftarrow (x_j^i, \Theta)$

$V_j^i \leftarrow \{Y_{k,j}^i\}_{k=1}^{12}$

for $k = 1, k < 13, k++$ **do**

$Y_{k,j}^i \leftarrow Y_{k,j}^i + 1\text{bp}$

$V_{k,j}^i \leftarrow \{Y_{k,j}^i\}_{k=1}^{12}$

$S_{k,j}^i \leftarrow V_{k,j}^i - V_i^j$

$Y_{k,j}^i \leftarrow Y_{k,j}^i - 1\text{bp}$

end

$\text{IM}_j^i \leftarrow \{S_{k,j}^i\}_{k=1}^{12}$

$\text{IM}_j^i \leftarrow \text{IM}_j^i e^{-\sum_{k=0}^i r_k^j(t_{i+1} - t_i)}$

end

$\text{DIM}_i \leftarrow \frac{1}{M} \sum_{j=0}^M \text{IM}_j^i$ Approximate DIM from its MC estimator.

end

Some numerical results :

We provided here some numerical results after applying Algorithm 6.1 and comparing it with the classical nested Monte-Carlo simulation

An interest rate swap :

Let's go back on the *G2++* model that we already introduced in the previous section with the following dynamics :

$$\begin{aligned} dx(t) &= -\kappa_x dt + \sigma_x dW_t^x, \quad x(0) = 0. \\ dy(t) &= -\kappa_y dt + \sigma_y dW_t^y, \quad y(0) = 0. \\ r(t) &= x(t) + y(t) + \phi(t), \quad r(0) = r_0. \end{aligned} \tag{6.6}$$

with :

³The Algorithm is taken directly from [21] as we exactly perform the same case instead that we use a *G2++* model.

- $\phi(t) = f^M(0, t) + \frac{\sigma_x^2}{2\kappa_x}(1 - e^{-\kappa_x t})^2 + \frac{\sigma_y^2}{2\kappa_y}(1 - e^{-\kappa_y t})^2 + \rho \frac{\sigma_x \sigma_y}{\kappa_x \kappa_y} (1 - e^{-\kappa_x t})(1 - e^{-\kappa_y t}).$
- $d < W^x, W^y >_t = \rho dt.$

As we know, the $G2++$ model is fully characterized by the following set of parameters $X = (\kappa_x, \sigma_x, \kappa_y, \sigma_y, \rho, r_0)$. We will considerer $N = 100$ discretization timesteps.

As we will consider an interest rate swap which is not subject to optionality , we won't take care of the vega of such derivatives and we will work only with the interest rate risk factor such as :

$$IM_{IR} = DeltaMargin_{IR}. \quad (6.7)$$

We will consider in the following numerical result an interest swap starting at $t_0 = 1Y$, ending at $t_N = 6Y$ with 20 payments dates and strike $K = 0.01$.

We give the actual architecture of the neural network in the table below :

Table 6.5: Neural network architecture for the DIM computation in the $G2++$ model

Number of Inputs	6
Number of Outputs	101
Number of Hidden Layers	3
Number of Neurons per Layer	256
Activation Function	$\phi(x) = x^+$ (ReLU)
Weight Initialization	Xavier/Goriot
Gradient Descent Algorithm	Adam Optimizer (learning rate = 0.001)

We also sample $K = 10000$ vectors of initial parameters X with the lower and upper bounds parameters considered in the table below. We consider for each variable an uniform sampling of the form $\mathcal{U}(\min(X), \max(X))$

Table 6.6: Lower and upper bounds for market state variable in the $G2++$ model

X	κ_x	σ_x	κ_y	σ_y	ρ	r_0
$\min(X)$	2.4%	0.5%	3%	0.5%	-0.999	-3%
$\max(X)$	12%	2.5%	15%	2.5%	0.999	6%

We provide some plots of the distribution of some parameters according to the uniform sampling with the bounds from Table 6.6.

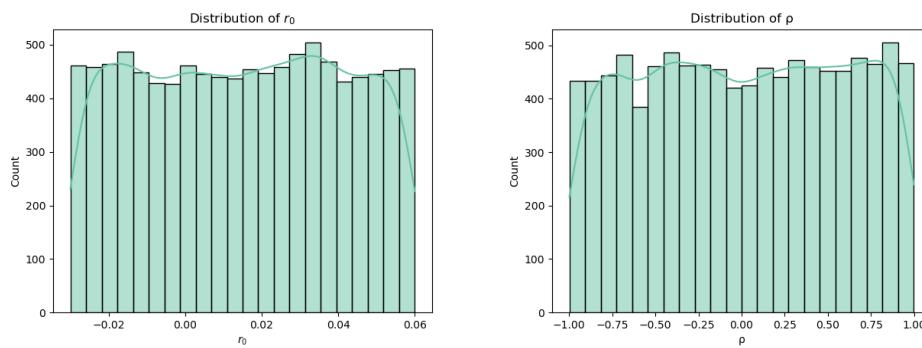


Figure 6.3: Distribution of parameters r_0 and ρ

As we said, we aim to learn from noisy labels of DIM in the fact that only one Monte-Carlo path will generate our output dataset. We provided below some noisy labels from the same set of parameters :

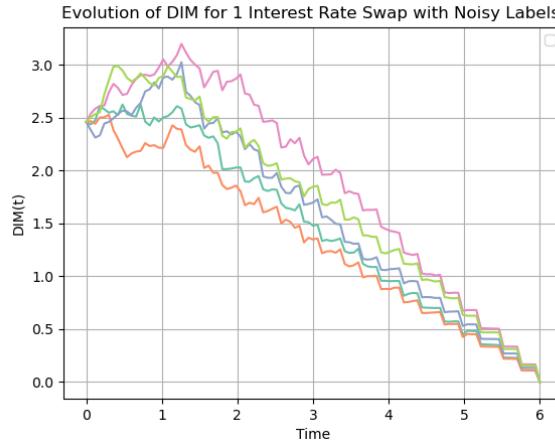


Figure 6.4: Noisy labels (with $M = 1$ inner path) for the DIM computation for the following set of parameters : $(\kappa_x = 0.10, \sigma_x = 0.02, \kappa_y = 0.12, \sigma_y = 0.02, \rho = -0.3$ and $r_0 = 0.01)$

Once we have built our dataset, we can then train our neural network according to the *Deep Conditional Expectation Solver* algorithm. We provide below some metrics on the training phase.

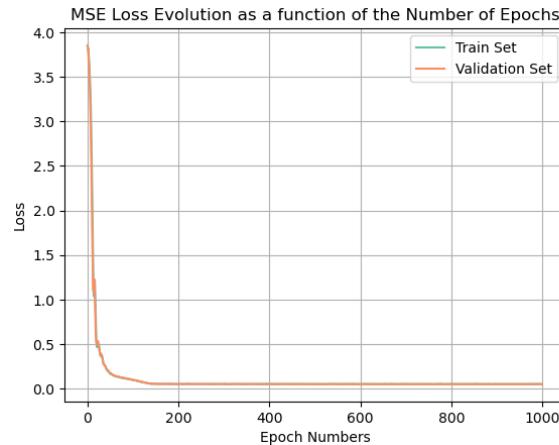


Figure 6.5: Loss training process for the DIM computation with *Deep Conditional Expectation Solver* for a portfolio of 1 swap

Table 6.7: Metric errors after the learning phase for the DIM computation

	MSE	MAE
Training Set	0.051206	0.150324
Test Set	0.048375	0.146877

We will illustrate the accuracy of the learning phase by providing several plots which compares the actual nested Monte-Carlo approximation which will be considered as the actual groundtruth true value and the NN approximation. However, as the nested Monte-Carlo is computationally intense, we had to consider only $M = 1000$ inner paths in Algorithm 6.1. We give in the following table the parameters that were used for each case of the figure 6.6.

Table 6.8: Set of parameters used to compare NN accuracy in DIM computation

Parameters	κ_x	σ_x	κ_y	σ_y	ρ	r_0
Case 1	0.10	0.015	0.12	0.015	-0.8	0.04
Case 2	0.10	0.02	0.12	0.02	-0.3	0.03
Case 3	0.10	0.015	0.12	0.015	-0.6	0.015
Case 4	0.10	0.02	0.12	0.02	0	0.01

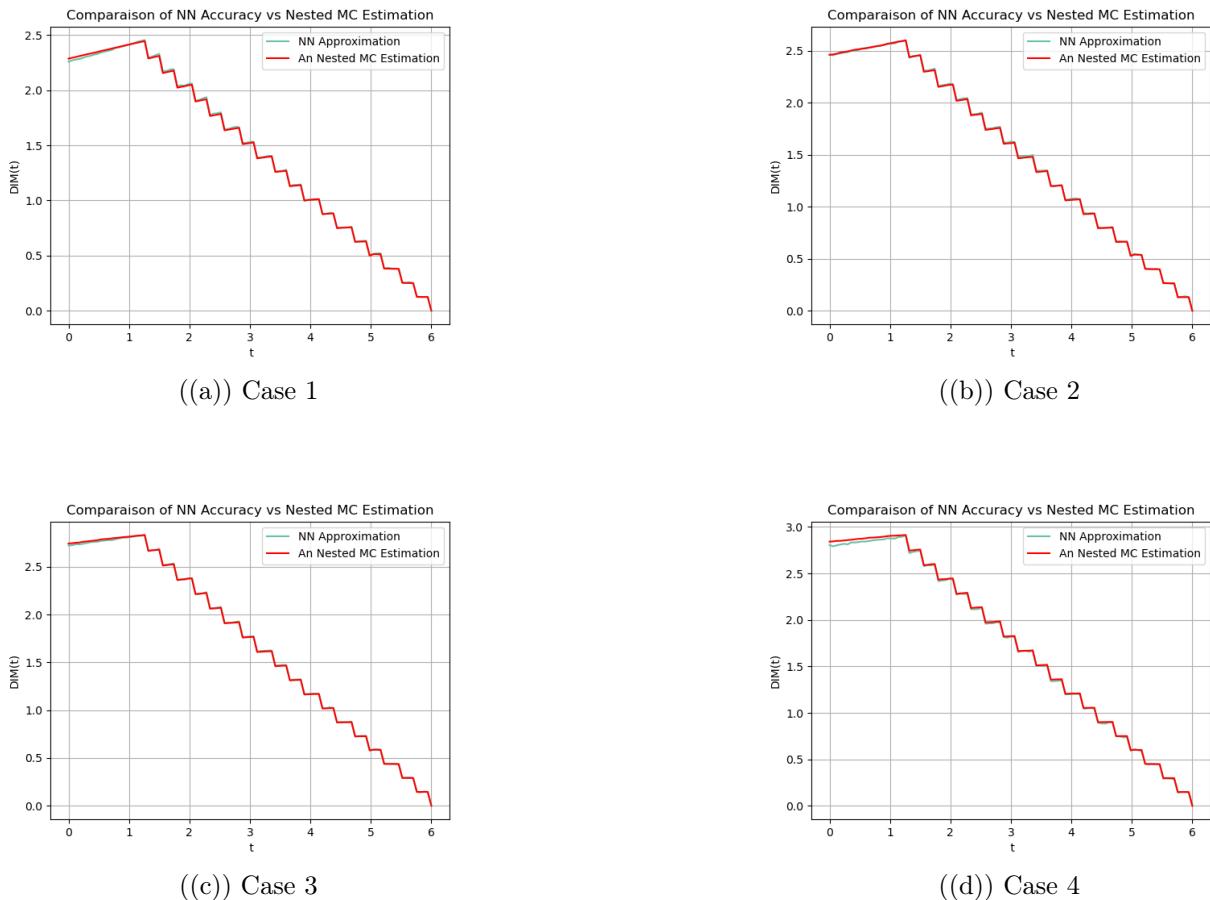


Figure 6.6: Study of the *Deep Conditional Expectation Solver* algorithm with comparison with the nested Monte-Carlo for various set of parameters

Some remarks on the methodology :

As we can see from figure 6.6, the NN seems to approximate quite well the groundtruth value from the nested Monte-Carlo except for the figure 6.6(d) where it looks to be less accurate than for the others.

As we said, the algorithm has been trained for only 10000 values and we could of course use more scenarios to feed the neural network.

Moreover, as we said, we only used $M = 1000$ paths for the Nested-Monte-Carlo as it was way too computationally intense leading potentially to inaccurate results.

In our case, we illustrate the algorithm for a portfolio consisting of a single interest swap with defined parameters but we can imagine that we could extend this methodology to other securities and to calculate the DIM and the resulting MVA from an efficient manner in a real XVA engine by taking profit of the methodology we used in this section.

For the IM profile on itself, we can observe a decreasing step function which is intuitive for an interest rate swap because at each exchange of payment, the residual value of the swap is reduced and therefore the initial margin which represents a value at risk is also reduced. We can also observe that the decay takes place until 0 which is also expected because at the end of the swap no more exchanges are done and therefore there is no more risk.

A case study on a portfolio of 5 swaps

To assess the methodology on a more complex setting, we will illustrate it on a portfolio of 5 interest swaps which we will assume to have the following parameters :

Table 6.9: Parameters of the 5 swaps used in the DIM computation

Parameters	Start Date	Terminal Date	Number of Payments	Fixed Rate
Swap 1	1Y	6Y	20	0.01
Swap 2	0Y	6Y	10	0.01
Swap 3	1.5Y	6Y	10	0.01
Swap 4	0.5Y	6Y	5	0.01
Swap 5	3Y	6Y	20	0.01

We only show 2 plots of portfolios for the DIM profile with the following parameters configuration :

Table 6.10: Parameters used in the DIM computation on a portfolio of 5 swaps

Parameters	κ_x	σ_x	κ_y	σ_y	ρ	r_0
Case 1	0.10	0.12	0.015	0.015	-0.6	0.015
Case 2	0.10	0.12	0.02	0.02	-0.8	0.01

We give below the loss training process on a dataset composed of 10000 noisy labels of DIM for 5 swaps and the comparison of the neural network with the nested Monte-Carlo.

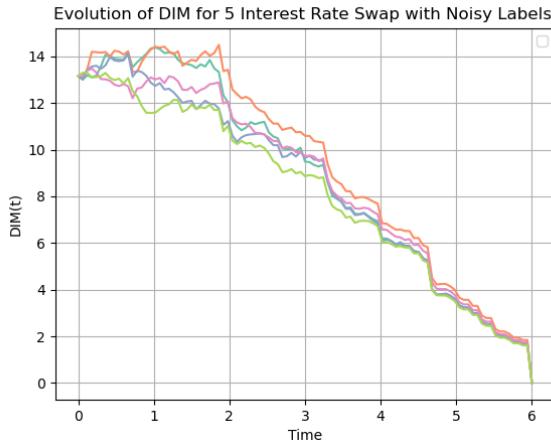


Figure 6.7: Noisy labels (with $M=1$ inner path) for the following set of parameters ($\kappa_x = 0.10$, $\sigma_x = 0.015$, $\kappa_y = 0.12$, $\sigma_y = 0.015$, $\rho = -0.6$ and $r_0 = 0.015$)

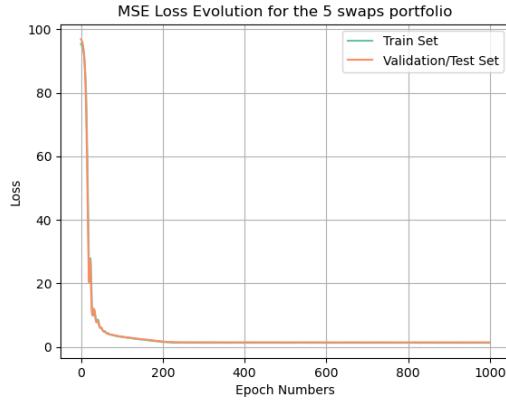


Figure 6.8: Loss training process for the DIM computation with *Deep Conditional Expectation Solver* for a portfolio of 5 swaps

Table 6.11: Metric errors after the learning phase on the portfolio of 5 swaps

	<i>MSE</i>	<i>MAE</i>
Training Set	1.220055	0.752931
Test Set	1.180061	0.740112

As we can observe with comparison with the case of a single interest rate swap, the *MSE* loss after the learning phase is way higher but it can be explained as the values taken by the *DIM* are bigger and we didn't rescale the data at the beginning so it could maybe lead to improvements on this aspect.

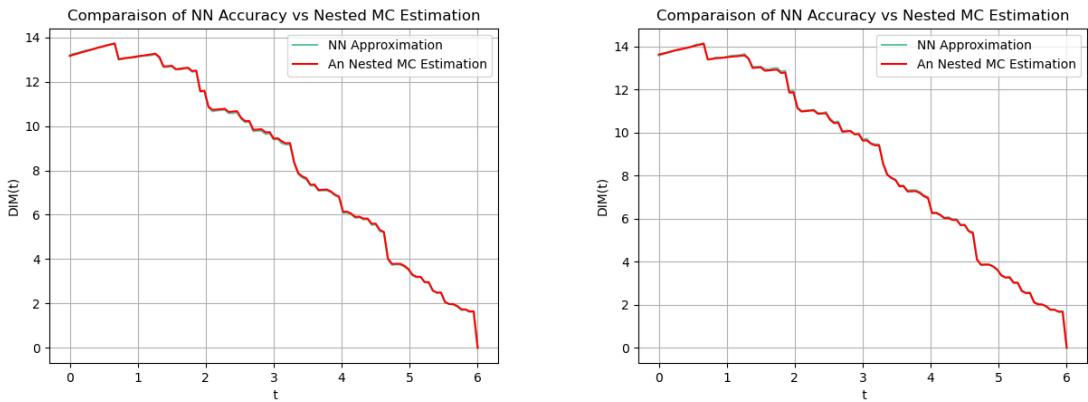


Figure 6.9: DIM computation with *Deep Conditional Expectation Solver* on a portfolio of 5 Swaps

Some remarks on the results

- As we can see from the plots above, the *NN* tends to learn accurately the form of the *DIM* in a more complex setup with a portfolio composed of 5 swaps in 2 different parameters settings. It highlights the fact that the *Deep Conditional Expectation Solver* can therefore be used in an efficient *XVA* engine.
- The form of the *DIM* is way more hard to understand in this context in comparison to the previous one when we only considered the case of a single interest rate swap but as we can see at the final date T , the initial margin still goes to 0 as there are no more exchange flows.

Global remarks on the *Deep Conditional Expectation Solver* for computing MVA_0

- The results on this more complex portfolio highlights the fact that the *Deep Conditional Expectation Solver* can be used in an efficient *XVA* engine for fast computations of *DIM* profile once the neural network is trained compared to single computations of *DIM* which took in our case more than half an hour and is therefore unfeasible in a production setting.
- The model can of course be improved by the choice of a better architecture of the neural network and hyperparameters settings but the core idea of the *NN* to approximate efficiently the *DIM* profile is illustrated. The computation of MVA_0 follows from the integration of the *DIM* profile multiplied by the function f over the lifetime of the portfolio.

6.2 Gaussian Processes Regression for XVA computations

In this section, we will follow the methodology from [22] by Crépey and Dixon⁴ where they use both \mathcal{GPR} and $M - C$ properties to build an efficient computation of expected exposure profile and associated CVA_0 .

6.2.1 The GPR-MC for EE profile and CVA

We will focus on the following representation of CVA_0 :

$$CVA_0 \approx -(1 - R^C) \sum_{i=0}^{N-1} EPE(t_i)(G(t_{i+1}) - G(t_i)).$$

where N is the number of exposure dates. We approximate the expression of $EPE(t_i)$ using $M - C$ which gives :

$$CVA_0 \approx -\frac{(1 - R^C)}{M} \sum_{j=1}^M \sum_{i=0}^{N-1} \frac{V(t_i, X_{t_i}^j)^+}{B_{t_i}^j} (G(t_{i+1}) - G(t_i)). \quad (6.8)$$

where M is the number of samples. As we said in the previous chapter, our goal will be to approximate the quantity $EPE(t_i)$ by using a \mathcal{GPR} such as we have the following estimator :

$$\hat{CVA}_0 = -\frac{(1 - R^C)}{M} \sum_{j=1}^M \sum_{i=0}^{N-1} \frac{(\mathbb{E}[V_*|X, Y, x^* = X_{t_i}^j])^+}{B_{t_i}^j} (G(t_{i+1}) - G(t_i)). \quad (6.9)$$

We also have a measurement of the MC -sampling error⁵ defined by :

$$\frac{1}{M-1} \sum_{j=1}^M [-(1 - R^C) \sum_{i=0}^{N-1} \frac{1}{B_{t_i}^j} (\mathbb{E}[V_*|X, y, x^* = X_{t_i}^j])^+ (G(t_{i+1}) - G(t_i)) - \hat{CVA}_0]^2. \quad (6.10)$$

Remark. Hence, we use machine learning to learn the component derivative exposures as a function of the underlying and other parameters, including (by slicing in time) time to maturity. We will therefore train as much \mathcal{GPR} as much as discretization times N we set. The ensuing CVA computations are then done by Monte-Carlo simulation based on (6.9). The authors of [22] named this method $MC-GP$. On the computational aspect, it saves one level of nested Monte-Carlo as in the normal case the price $V(t_i, X_{t_i}^j)$ would need to be computed using a Monte-Carlo procedure.

We will illustrate the methodology on the case of portfolios of european derivatives under the $B - S$ model and interest rate swaps under the Hull & White model as they provide benchmark solutions.

The algorithm can be written as follows :

⁴Their base code is available on <https://github.com/mfrdixon/GP-CVA/tree/master>.

⁵See [22] for further informations.

Algorithm 6.2 $\mathcal{GPR} - MC$ algorithm for exposure profile and CVA_0 computations

Input Parameters : T maturity contract, timegrid $\mathcal{T} = \{0 = t_0 < t_1 < \dots < t_N = T\}$, $CVA_0 = 0$, $(\mathcal{GPR}_i)_{i \in [0; N-1]}$ a family of \mathcal{GPR} trained at each time t_i

for $i = 0, \dots, N - 1$ **do**

$$CVA+ = (1 - R^C) \sum_{j=1}^M \frac{\mathcal{GPR}_i(X_{t_i}^j)}{B_{t_i}^j} (G(t_{i+1}) - G(t_i))$$

end

Output Parameters : CVA_0 computed with the $\mathcal{GPR} - MC$ algorithm

Some numerical results on portfolios of European derivatives

We will illustrate the $\mathcal{GPR} - MC$ methodology on the case of 3 portfolios of european derivatives. We will start from simple portfolio and we will complexify it to see the scalability of the algorithm.

We give in the tables below the parameters used in the numerical experiments below and the composition of equity portfolios.

Table 6.12: Parameters used in the numerical experiments in the \mathcal{GPR} methodology for EE profile computation of equity portfolios

Parameter Description	Symbol	Value
Strike of the Call Option	K^C	110
Strike of the Put Option	K^P	90
Initial Stock Price	S_0	100
Interest Risk Free Rate	r	0
Volatility	σ	0.3
Default Intensity	λ	0.01
Number of simulations	M	1000
Number of time steps in the Euler Scheme	N_{MC}	100
Number of Exposure Time	N	10
Number of Training points	$N_{Training}$	30
Number of Testing points	$N_{Testing}$	40

Table 6.13: Description of the composition of equity portfolios used in the numerical experiments

Portfolio Composition	Number of calls	Number of puts
Portfolio 1	1	0
Portfolio 2	10	5
Portfolio 3	5	-5

For each portfolio , we will compute the following quantities :

- The expected profile of the portfolio.
- The error on the learning profile of the exposure by the \mathcal{GPR} with a confidence interval using the properties of the \mathcal{GPR} .
- The computation of CVA_0 using the Monte-Carlo procedure which allows the computation of a confidence interval.

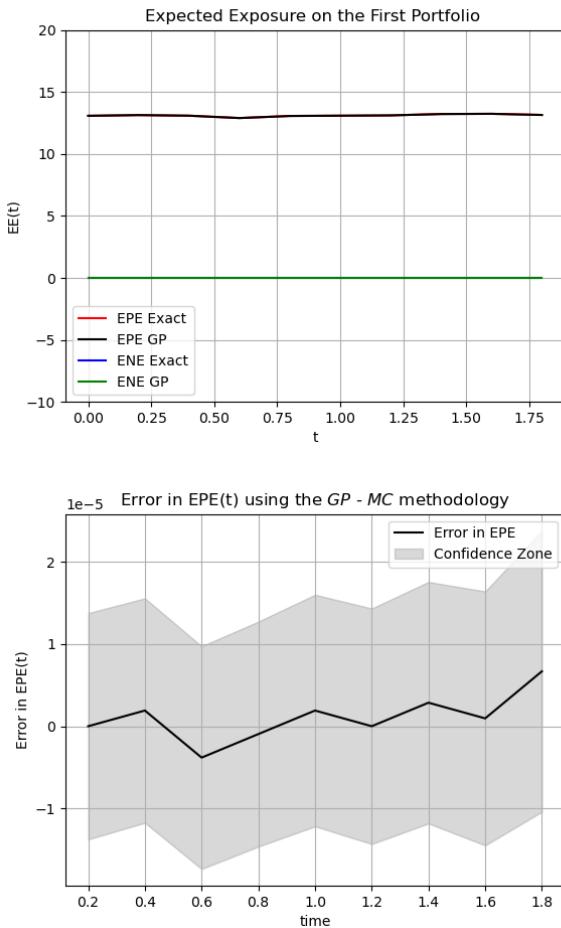
Portfolio 1 :


Figure 6.10: EE profile of the 1^{st} equity portfolio using the $GP - MC$ methodology and associated error in the calculation of the EE profile

We give also in the table below the results of CVA_0 with confidence interval.

Table 6.14: CVA_0 using the $GP - MC$ methodology on the 1^{st} equity portfolio

	True Value	$GP - MC$ estimation	Upper Bound	Lower Bound
CVA_0 on Portfolio 1	0.15802482	0.15802473	0.16180155	0.15424791

Remarks on the results on the 1^{st} portfolio :

- As we can see from the plots above, we recover the profile we had in the first chapters of this dissertation. The profile is not perfectly flat as it comes from the relatively low MC simulations M . However, when comparing it with the exact exposure calculated with the true $B - S$ prices, we see on the error plot that the error on the EE is really low which means that the performance of the \mathcal{GPR} is particularly high.
- The computation of CVA_0 shows the remarkable accuracy of the \mathcal{GPR} as the difference between the "true value" and the $GP - MC$ estimation is almost similar up to 10^{-6} .

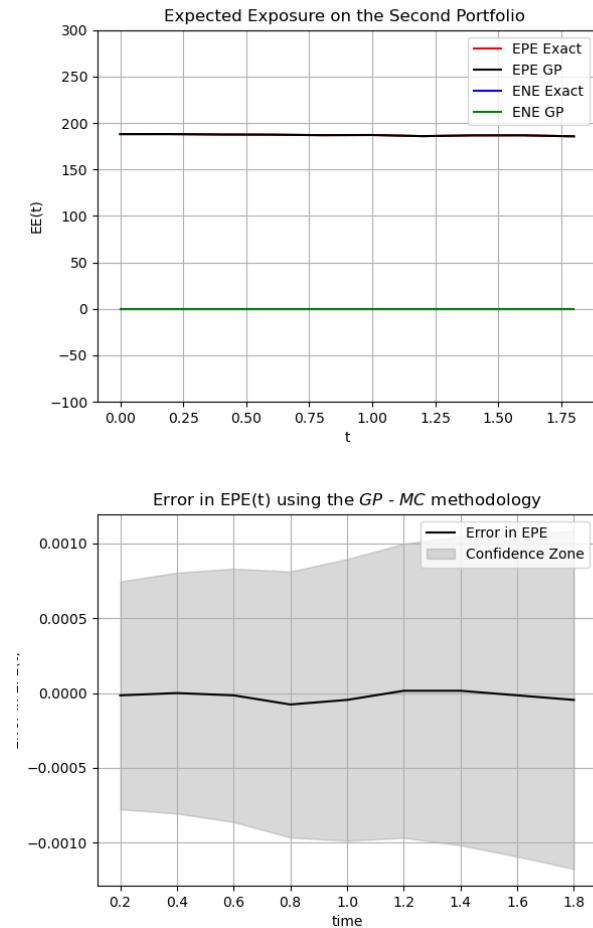
Portfolio 2 :


Figure 6.11: EE profile of the 2^{nd} equity portfolio using the $GP - MC$ methodology and associated error in the calculation of the EE profile

We give also in the table below the results of CVA_0 with confidence interval.

Table 6.15: CVA_0 using the $GP - MC$ methodology on the 2^{nd} equity portfolio

	True Value	$GP - MC$ estimation	Upper Bound	Lower Bound
CVA_0 on Portfolio 2	2.2333603	2.2333624	2.2654195	2.2013054

Remarks on the results on the 2^{nd} portfolio

- As we can see from the profile of the exposure, we see again the flat profile of the exposure which is intuitive as the investor holds only long positions on the derivatives. Similarly to the first portfolio, the results are quite similar expect that the error is a bit higher but up to a 10^{-4} factor.
- The computation of CVA_0 shows also the remarkable accuracy of the \mathcal{GPR} as the difference between the "True Value" and the $GP - MC$ estimation is almost similar up to 10^{-6} .

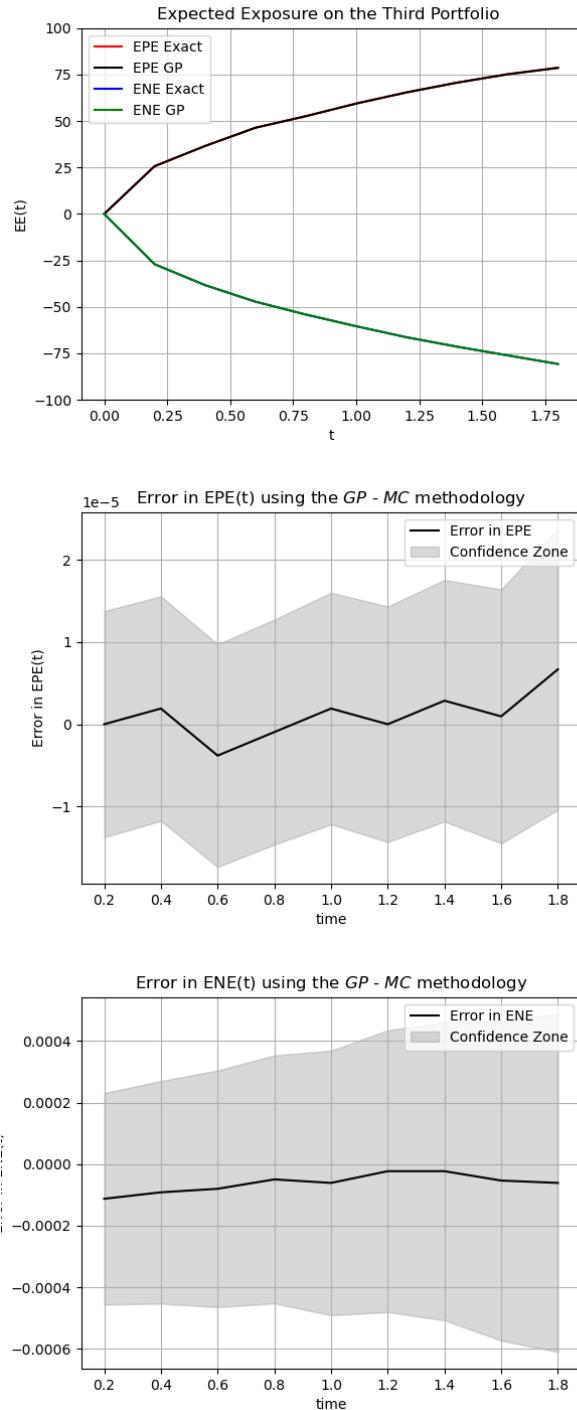
Portfolio 3 : ⁶


Figure 6.12: EE profile on the 3^{rd} portfolio using the $GP - MC$ methodology and associated error in the calculation of the EE profile

Table 6.16: CVA_0 using the $GP - MC$ methodology on the 3^{rd} equity portfolio

	True Value	$GP - MC$ estimation	Upper Bound	Lower Bound
CVA_0 on Portfolio 3	0.6092085	0.6092076	0.61602855	0.6023867

⁶For this portfolio, we choose the same strike $K^C = K^P = 100$ and we plot also the ENE as the investor has short positions in put options. By choosing these strikes, we should reproduce the exposure of a forward according to Call-Put parity.

Remarks on the results on the 3rd portfolio

- As we can see from the exposure profile, we recover the exposure of the forward we had in the previous sections. As we noticed, the portfolio composition holds to lead to this result so it's a way to compare the accuracy of the $\mathcal{GPR} - MC$ methodology in computation the exposure profile. Similarly to the previous results, the errors done on learning the profile exposures are really low up to 10^{-5} in the EPE computation and 10^{-4} in the ENE computation
- Similarly to the previous results, the CVA_0 computation is really accurate up to 10^{-6} .

Some numerical results on swap portfolios

We also consider the following swap portfolios to assess the methodology. We decided to work under the Hull & White model for the computation of the exposure profile and associated CVA_0 .

Table 6.17: Composition of the 1st swap portfolio used in the numerical experiments

Parameters	Start Date	Terminal Date	Number of Payments	Fixed Rate
Swap 1	1Y	10Y	20	0.01

Table 6.18: Composition of the 2nd swap portfolio used in the numerical experiments

Parameters	Start Date	Terminal Date	Number of Payments	Fixed Rate
Swap 1	1Y	10Y	20	0.01
Swap 2	0Y	10Y	10	0.01
Swap 3	1.5Y	10Y	10	0.01
Swap 4	0.5Y	10Y	5	0.01
Swap 5	3Y	10Y	20	0.01

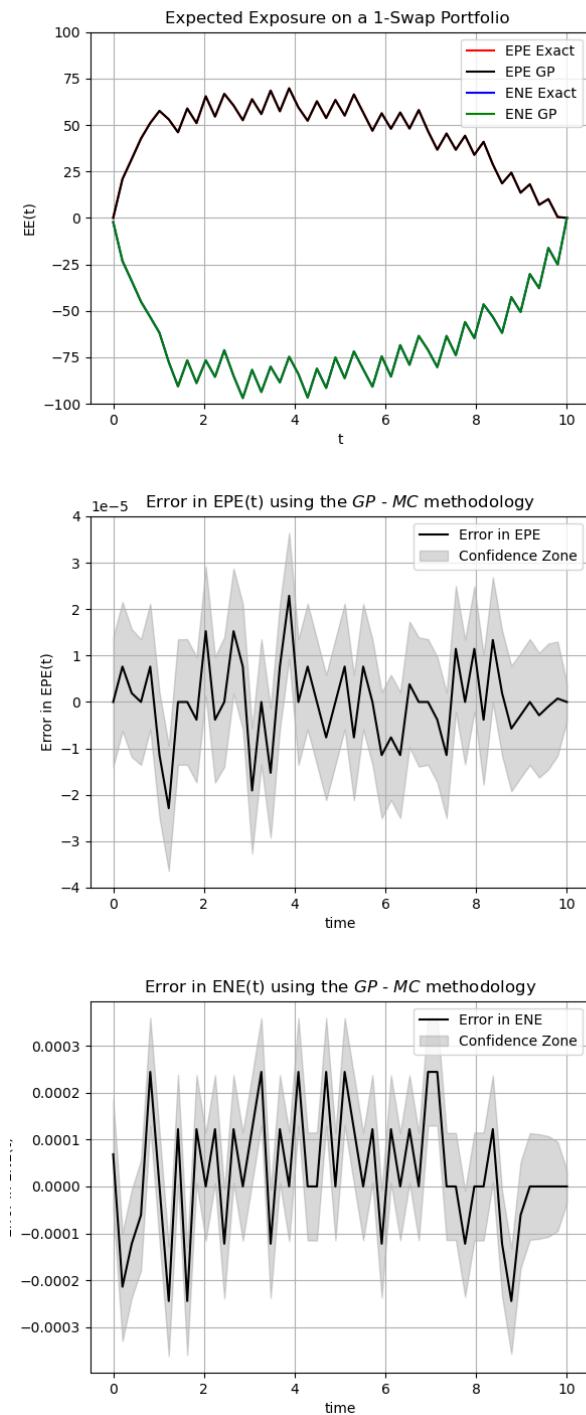
Portfolio 1 :


Figure 6.13: EE profile of a single swap using the $GP - MC$ methodology and associated error in the calculation of the EE profile

We give also in the table below the results of CVA_0 with confidence interval.

Table 6.19: CVA_0 using the $GP - MC$ methodology on the 1st swap portfolio

	True Value	$GP - MC$ estimation	Upper Bound	Lower Bound
CVA_0 on Portfolio 1	2.598366	2.5983661	2.858947	2.33779

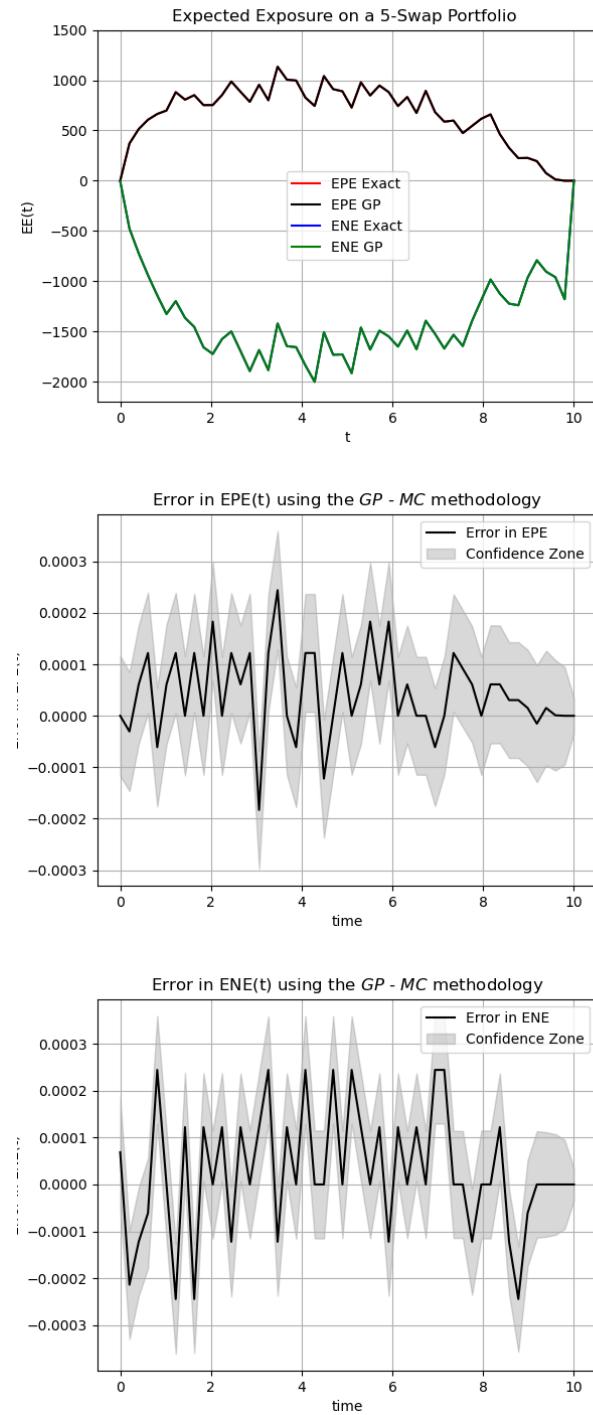
Portfolio 2 :


Figure 6.14: EE profile of a 5-swap Portfolio using the $GP - MC$ methodology and associated error in the calculation of the EE profile

We give also in the table below the results of CVA_0 with confidence interval.

Table 6.20: CVA_0 using the $GP - MC$ methodology on the 2^{nd} swap portfolio

	True Value	$GP - MC$ estimation	Upper Bound	Lower Bound
CVA_0 on Portfolio 2	38.616476	38.616478	42.809036	34.4239205

Remarks on the results on swaps portfolio :

- As we can see from the profile of the exposure, we recover the profile we had in the first chapters of this dissertation. Therefore, we show that the $\mathcal{GP} - MC$ can be extended to the case of swap portfolios with similar accuracy results than in the equity portfolios.
- The computation of CVA_0 shows the remarkable accuracy of the \mathcal{GPR} as the difference between the "true value" and the $\mathcal{GP} - MC$ estimation is almost similar up to 10^{-6} .

Global remarks on the $\mathcal{GPR} - MC$ methodology :

- As we have shown on the plots above, the $\mathcal{GPR} - MC$ has good performance and can therefore be used in order to reduce one level of the nested Monte-Carlo by learning the price surface at every discretization time of the exposure profile.
- As we mentionned in the previous chapter, the learning price surface is very reliable to the quality of the labels we are giving to the \mathcal{GPR} . We see that when feeding the \mathcal{GPR} with highly noisy labels, the performance was of course deteriorated and we had to make sure to be careful when using the methodology that the learning of the price surface is good so we minimize as much as possible the error done by the \mathcal{GPR} in the computation of exposure profile such that the global error comes from the outer MC computation.
- We didn't illustrate the case of \mathcal{GPR} for high-dimensionnal assets but it has been show in [23] by Goudenège, Molent and Zanette that even for bermudan derivatives this method can be well suited for CVA and FVA computations as they both rely on the computation of the expected exposure profile.

6.3 Deep XVA Solver

In this subsection, we will mostly refer to the article [19] from Gnoatto, Picarelli and Reisinger named *Deep xVA solver – A neural network based counterparty credit risk management framework* as they provide an entire framework for the use of the Deep BSDE solver in the *XVA* context. As we mentionned, the *Deep BSDE Solver* is an algorithm which can help to overcome the curse of dimensionality issue and is therefore interesting to study⁷. For this, we will need to write the *XVAs* in a convenient way to make the *Deep BSDE Solver* usable in practice. The core assumption in the *XVA* context is to assume that the default times of the counterparty C and A are exponentially distributed with time-dependent intensity such that we can write :

$$\begin{aligned}\tau^C &= \inf\{t \geq 0 : \int_0^t \lambda_s^C ds \geq \nu^1\}. \\ \tau^A &= \inf\{t \geq 0 : \int_0^t \lambda_s^A ds \geq \nu^2\}.\end{aligned}$$

where λ^C and λ^A are two positive processes adapted to the filtration \mathcal{F}_t and ν^1 and ν^2 are 2 exponential random variables assumed to be independant from \mathbb{G} the enlarged filtration by the potential defaults of A and C .

As we already showed previously in this report, we know that we can write *CVA* and *FVA*⁸ as follows (on the event that no default has been occured before t) :

$$\begin{aligned}CVA_t &= (1 - R^C) \mathbb{E}^Q \left[\int_t^T e^{- \int_t^s (r_u + \lambda_u^C + \lambda_u^A) du} (V_s)^+ \lambda_s^C ds \mid \mathcal{F}_t \right]. \\ FVA_t &= \mathbb{E}^Q \left[\int_t^T e^{- \int_t^s (r_u + \lambda_u^C + \lambda_u^A) du} (s_s^B (V_s)^+ - s_s^L (V_s)^-) ds \mid \mathcal{F}_t \right].\end{aligned}$$

Under markovian models, the information given by \mathcal{F}_t can be simplified by X_t where X denotes the market risk factors associated with the diffusion, we can therefore write by noting ϕ^{CVA} the associated value function of *CVA* :⁹

$$\phi^{CVA}(t, x) = (1 - R^C) \mathbb{E}^Q \left[\int_t^T e^{- \int_t^s (r_u + \lambda_u^C + \lambda_u^A) du} (V_s)^+ \lambda_s^C ds \mid X_t = x \right]. \quad (6.11)$$

Theses forms are really convenient as we can use the Feymann-Kac theorem to write the following *PDE* for ϕ^{CVA} the associated value function of the *CVA* :

$$\partial_t \phi^{CVA}(t, x) + \mathcal{L} \phi^{CVA}(t, x) - (r_t + \lambda_t^C + \lambda_t^A) \phi^{CVA}(t, x) + (1 - R^C) (V_t)^+ \lambda_t^C = 0, \quad \forall (t, x) \in [0, T] \times \Omega. \quad (6.12)$$

$$\phi^{CVA}(T, .) = 0, \quad \forall x \in \Omega. \quad (6.13)$$

where \mathcal{L} is the generator of the diffusion characterized by the diffusion process X .

⁷They provide all their basecode at <https://github.com/AlessandroGnoatto/Deep-xVA-Solver>.

⁸In the article, they used the recursive *FVA* form which we talked in the first chapters but we stick to the *FVA* form we used in this chapter.

⁹A similar PDE can be derived for *FVA*.

Therefore, by setting the following quantities :

- $f(t, X_t, Y_t, Z_t) = (1 - R^C)\lambda_t^C(V_t)^+ - (r_t + \lambda_t^C + \lambda_t^A)Y_t.$
- $g(X_T) = 0.$

We see that we are in the setting of the *Deep BSDE Solver*.

Remark. At this point, we need to be able to compute the quantity $(V_t)^+$ which is the exposure of the portfolio on which the CVA is computed. In the numerical results, we will use a $B - S$ model on simple options and the price will be determined using the Deep BSDE Solver.¹⁰

6.3.1 EE profile computation of some derivatives

In this subsection, we will describe the exposure profile calculation for a portfolio using the *Deep BSDE* approach. For this, we rely to the following Algorithm introduced in [19] where all the notations are also introduced. They start by learning the parameters associated to the financial products on which the *CVA* will be computed. Once the training is done, they simulate the value of the portfolio from the *NN* approximation and they compute the *XVAs* based on this first computation of the exposure profile. This procedure can be summarized in the 2 following algorithms.

Algorithm 6.3 Deep algorithm for exposure simulation (Algorithm from [19])

Set parameters: N, L, B . N time steps, L paths for inner Monte-Carlo loop, B batch size
Fix architecture of ANN

Deep BSDE Solver (N, L, B):

Simulate L paths $(\tilde{S}_n^{(\ell)})_{n=0,\dots,N}$, $\ell = 1, \dots, L$ of the forward dynamics.

Define the neural networks $(\varphi_n^\rho)_{n=1,\dots,N}$.

for $m = 1, \dots, M$ **do**

 Minimize over ξ and ρ

$$\frac{1}{L} \sum_{\ell=1}^L \left(g_m(\tilde{S}_N^{(\ell)}) - \hat{\mathcal{V}}_N^{m,\xi,\rho,(\ell)} \right)^2,$$

 subject to

$$\begin{cases} \hat{\mathcal{V}}_{n+1}^{m,\xi,\rho,(\ell)} = \hat{\mathcal{V}}_n^{m,\xi,\rho,(\ell)} + r_{t_n} \hat{\mathcal{V}}_n^{m,\xi,\rho,(\ell)} \Delta t + (\hat{\mathcal{Z}}_n^{m,\rho,(\ell)})^\top \Delta W_n^{(\ell)}, \\ \hat{\mathcal{V}}_0^{m,\xi,\rho,(\ell)} = \xi, \\ \hat{\mathcal{Z}}_n^{\rho,(\ell)} = \varphi_n^\rho(\tilde{S}_n^{(\ell)}). \end{cases} \quad (6.14)$$

 Save the optimizer (ξ_m^*, ρ_m^*) .

end

end

¹⁰Indeed, the *Deep BSDE Solver* is also well suited for the pricing of derivatives as we already mentionned in the previous chapter.

Algorithm 6.4 Deep XVA Solver for non-recursive valuation adjustments (Algorithm from [19])

Apply Algorithm 6.3

Set parameters: P paths for the outer Monte-Carlo loop

Simulate, for $m = 1 \dots M$, $(\hat{V}_n^{m,\xi_m^*,\rho_m^*(p)})_{n=0 \dots N, p=1 \dots P}$ by means of (6.14) with $(\xi, \rho) = (\xi_m^*, \rho_m^*)$.

Define $\hat{V}_n^{*,(p)} := \sum_{m=1}^M \hat{V}_n^{m,\xi_m^*,\rho_m^*(p)}$ for $n = 0 \dots N, p = 1 \dots P$.

Compute the adjustment as

$$\frac{1}{P} \sum_{i=1}^P \left(\sum_{n=0}^N \eta_n \Phi_{t_n}(\hat{V}_n^{*,(p)}) \right).$$

Numerical Results :

We propose 2 numerical experiments in the $B - S$ setting :

- Calculation of the EE profile for a forward contract following Algorithm 6.3.
- Calculation of the EE profile on a very high dimensional basket call option with $d = 100$ assets following Algorithm 6.3.

EE profile of a forward

We will consider the case of a classic forward contract with the following payoff :

$$g(S_T) = (S_T - K).$$

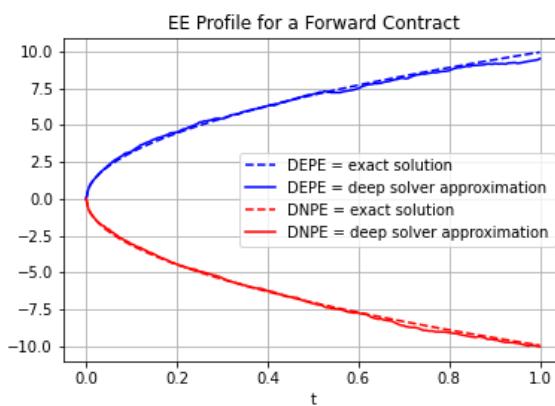


Figure 6.15: EE profile computation of a forward contract under B-S using the *Deep BSDE Solver*

EE profile of a basket option

We consider now the case of a basket call option on $d = 100$ assets with the following payoff :

$$g(S_T^1, \dots, S_T^d) = \left(\sum_{i=1}^d S_T^i - dK \right)^+.$$

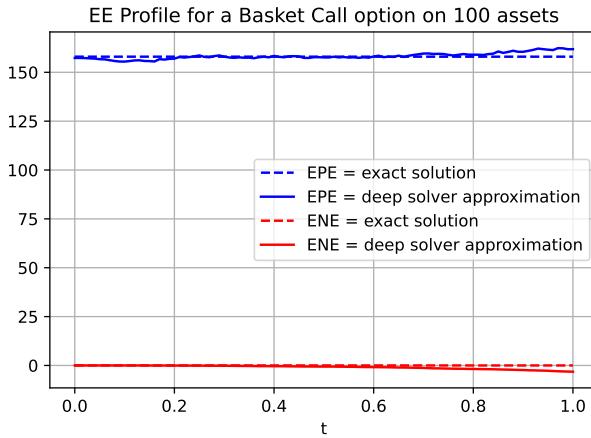


Figure 6.16: *EE* profile computation of a basket option on $d = 100$ assets under B-S using the *Deep BSDE Solver*

Some remarks on the results

- As we can see from the plots above, the *Deep BSDE Solver* performs well even in the high dimensionnal setting which highlights the potential overcome of the curse of dimensionality we are normally facing when dealing with classical Monte-Carlo methods.
- The main issue of the methodology is the need of a semi-linear *PDE* representation of financial derivatives and *XVAs*, otherwise we can't apply the methodology we introduced in this dissertation.
- In the article, they propose even further computations by computing risk measures on *CVA* which is not the main topic of this dissertation that's why we are not focusing on this aspect but the *Deep XVA Solver* framework seems to be really efficient as it can permit to price almost all the *XVAs* in a same framework just by relying to their *PDE* representation.

6.3.2 Direct computation of XVAs using PDE representation

As we wrote in the introduction of this section, CVA and FVA are related to a PDE representation according to Feymann-Kac formula and we show that on the event $\mathbb{1}_{\tau^C > t}$, the process CVA can be reformulated as the solution of the following PDE with ϕ^{CVA} the associated value function.¹¹

$$\begin{aligned} \partial_t \phi^{CVA} + \mathcal{L}\phi^{CVA} - (r_t + \lambda_t^C + \lambda_t^A)\phi^{CVA} + (1 - R^C)(V_t)^+ \lambda_t^C = 0, \quad \forall(t, x) \in [0, T[\times \mathbb{R}_*^+. \\ \phi^{CVA}(T, .) = 0, \quad \forall x \in \mathbb{R}_*^+. \end{aligned}$$

where the generation is given in the $B - S$ framework by :

$$\mathcal{L}u(t, x) = rS\partial_S u(t, x) + \frac{1}{2}\sigma^2 S^2 \partial_{S^2}^2 u(t, x).$$

For the process FVA , using also the Feymann-Kac formula, we have the following PDE with ϕ^{FVA} the associated value function.

$$\begin{aligned} \partial_t \phi^{FVA} + \mathcal{L}\phi^{FVA} - (r_t + \lambda_t^C + \lambda_t^A)\phi^{FVA} + (s_t^b(V_t)^+ - s_t^L(V_t)^-) = 0, \quad \forall(t, x) \in [0, T[\times \mathbb{R}_*^+. \\ \phi^{FVA}(T, .) = 0, \quad \forall x \in \mathbb{R}_*^+. \end{aligned}$$

Remark.

- In the case of ϕ^{CVA} ,¹² the function f is given by $f(t, S_t, Y_t, Z_t) = (1 - R^C)(V_t)^+ \lambda_t^C - (r_t + \lambda_t^C + \lambda_t^A)Y_t$.
- In the case of ϕ^{FVA} , the function f is given by $f(t, S_t, Y_t, Z_t) = (s_t^b(V_t)^+ - s_t^L(V_t)^-) - (r_t + \lambda_t^C + \lambda_t^A)Y_t$.
- In both cases, we see that the terminal value is always 0.

¹¹We also study another Deep Learning algorithm named *Deep Galerkin* introduced in [24] for which we provided some informations in the Annex D and some numerical results on XVA computations.

¹²If we want to incorporate DVA such that we aim to calculate the $BCVA$ instead of the unilateral CVA , the PDE representation is also straightforward with the function f given by $f(t, S_t, Y_t, Z_t) = (1 - R^A)(V_t)^- - (r_t + \lambda_t^C + \lambda_t^A)Y_t$ and null terminal value.

6.3 Deep XVA Solver

We give below the algorithm introduced in [19] where the *Deep BSDE Solver* is used to solve directly the *PDEs* associated with *XVAs*;

Algorithm 6.5 Deep XVA Solver (Algorithm from [19])

Apply Algorithm 6.3.

Set parameters: P . P paths for outer Monte-Carlo loop

Fix architecture of ANN.

intrinsically defines the number of parameters \bar{R} (in general $\bar{R} \neq R$)

Deep XVA-BSDE solver (N, P):

Simulate P paths $(\underline{x}_n^{(p)})_{n=0,\dots,N}$, $p = 1, \dots, P$, of the portfolio value.

Define the neural networks $(\psi_n^\zeta)_{n=1,\dots,N}$.

Minimize over γ and ζ

$$\frac{1}{P} \sum_{p=1}^P \left(\overline{\mathcal{X}}_N^{\gamma, \zeta, (p)} \right)^2,$$

subject to

$$\begin{cases} \overline{\mathcal{X}}_{n+1}^{\gamma, \zeta, (p)} = \overline{\mathcal{X}}_n^{\gamma, \zeta, (p)} - \bar{f}(t_n, \widehat{\mathcal{V}}_n^{(p)}, \overline{\mathcal{X}}_n^{\gamma, \zeta, (p)}) \Delta t + (\overline{\mathcal{Z}}_n^{\zeta, (p)})^\top \Delta W_n^{(p)}, \\ \overline{\mathcal{X}}_0^{\gamma, \zeta, (p)} = \gamma, \\ \overline{\mathcal{Z}}_n^{\zeta, (p)} = \psi_n^\zeta(\widehat{\mathcal{V}}_n^{(p)}). \end{cases} \quad (6.15)$$

end

Numerical results :

We will calculate CVA_0 and FVA_0 in our setting and compare it with the nested Monte-Carlo to check the accuracy of the algorithm 6.5.

Table 6.21: Parameters used in the numerical experiments for the forward contract with the *Deep XVA Solver*

Parameters	S_0	K	r	s_b	s_l	σ	T	λ^C	λ^A	R^C
CVA_0	100	100	0.0	X	X	0.25	1	0.1	0	0.3
FVA_0	100	100	X	0.02	0	0.25	1	0.1	0.01	X

Table 6.22: XVA_0 estimation from *Deep XVA Solver* for a forward contract with *MC* performed on 100000 samples

XVA	Monte-Carlo	Upper Bound ($M - C$)	Lower Bound ($M - C$)	NN Approximation
CVA	0.45229	0.45807	0.44650	0.45520
FVA	0.12757	0.12920	0.12595	0.12698

- As we can see from the numerical experiments on the forward contract, the algorithm 6.5 looks to perform very well as the *NN* approximation falls in the 95% confidence interval of the Monte-Carlo.

Table 6.23: Parameters used in the numerical experiments for the basket call option with $d = 100$ assets with the *Deep XVA Solver*

Parameters	S_0	K	r	s_b	s_l	σ	T	λ^C	λ^A	R^C
CVA_0	100	100	0.01	X	X	0.25	1	0.1	0	0.3
FVA_0	100	100	X	0.02	0	0.25	1	0.1	0.01	X

Table 6.24: XVA_0 estimation from *Deep XVA Solver* for a basket call option on $d = 100$ assets with MC performed on 100000 samples

XVA	Monte-Carlo	Upper Bound ($M - C$)	Lower Bound ($M - C$)	NN Approximation
CVA	10.522	10.582	10.462	10.510
FVA	1.497	1.506	1.489	1.500

- As we can see, the accuracy of the NN from the *Deep XVA Solver* is still very accurate even in a very large dimensionnal setting with the NN approximation falling in the 95% confidence interval which shows the efficiency of the algorithm.

Global remarks on the chapter :

- In the first section of this chapter, we used the *Deep Conditional Learning* algorithm to perform some XVA computations. We saw that it needed improvements to be really useful in practice for computations of CVA_0 or FVA_0 but we show his efficiency in the computation of DIM and MVA_0 with an illustration with swap portfolios. We show that the global error was really low and that the DIM profile was really similar between the brute force nested Monte-Carlo approach and our NN estimation. Therefore, this algorithm can be a tool to overcome the nested Monte-Carlo associated with the MVA_0 computation.
- In the second section, we used \mathcal{GPR} to calculate efficiently exposure profile. We also show how it could be used combined with a classic Monte-Carlo procedure to provide accurate estimation of CVA_0 and we provide numerical examples on equity european derivatives and swap portfolios. We also took advantage of both the techniques which provide error bounds and we could therefore analyze the impact of both the methods in the final computation of CVA_0 showing that the \mathcal{GPR} was particularly efficient in learning the price surface. Therefore, the $\mathcal{GPR} - MC$ algorithm can be a tool to overcome the nested Monte-Carlo associated with the CVA_0 and FVA_0 computation.
- Finally, in the last section , we used the *Deep XVA Solver* to compute efficient exposure profile and associated CVA_0 by solving directly the PDE related to the $XVAs$. In both cases, the *Deep XVA Solver* showed particularly good results once being trained. Moreover, we illustrate it in a really high dimensionnal setting with a basket call on $d = 100$ assets and we showed that the performance was still particularly accurate showing the scalability of the approach.

Chapter 7

Towards methods to hedge CCR

In this chapter, we will introduce hedging strategies for the counterparty credit risk exposure using a credit derivative : *Credit Default Swaps (CDS)*. The outline of the chapter is then as follows. In the 1st section, we introduce *CDS*, the major financial instrument to hedge the *CCR*. In the 2nd section, we talk about a dynamic hedging strategy : the *Mean-Variance Minimizing Strategy* initiated in [25] by Ceci, Colaneri, Frey and Köck. We illustrate their approach with examples arising from the financial market. We also follow their article where they study the case of the counterparty risk of a reinsurer. In the last section, we will discuss about a static hedging method based on the expected utility theory where an insurer looks to solve his optimization problem by finding optimal reinsurance and hedging contracts in the case of the potential default of the reinsurer.

7.1 An introduction to Credit Default Swaps

A *CDS* is a financial product that pays the buyer of protection, at a default time, the difference between the face value and its recovery value, assuming that default occurs before some maturity T . This instrument is particularly well suited for *CCR* since it makes a payment at the default time τ^C . Moreover, the *CDS* market is liquid so these products are well adapted to hedge *CCR*.

Let's describe the cashflows of such a financial product from the point of view of the protection buyer.

- if $\tau < T$, then at τ , the protection buyer receives the *Loss Given Default* which is referred as **Default Leg**.
- The buyer of protection makes periodic coupons payments at time $(T_i)_{i \in \llbracket 1; N \rrbracket}$. If $\tau > T_i$, then an amount C_0 is payed at T_i which results in the following discounted payoff : $C_0 e^{-\int_0^{T_i} r_s ds} \mathbf{1}_{\tau > T_i}$. The quantity $C_0 \sum_{i=1}^N \mathbb{E}^\mathbb{Q}[e^{-\int_0^{T_i} r_s ds} \mathbf{1}_{\tau > T_i}]$ is then referred as the **Premium Leg**

We will suppose in the following of this dissertation ¹ that the coupon payments are made continuously represented by a *running spread premium* $\xi > 0$. This implies that instead of periodic coupons payments at time $(T_i)_{[1; N]}$ the cashflow payment streams is then given by $(\int_0^{\min(t, \tau)} \xi ds)_{t \geq 0}$. We will also considerer that at default time τ^C , a deterministic protection R^{CDS} is given.

We recall that $H_t = \mathbf{1}_{\tau^C \leq t}$. We can then denote the cumulative cashflows C_t in our setting as the following :

¹This is a common assumption in the litterature.

$$C_t = R^{CDS} H_t - \xi \int_0^{\min(t, \tau)} ds = R^{CDS} H_t - \xi \int_0^t (1 - H_u) du. \quad (7.1)$$

From the process C_t , we can now derive the present value of the CDS considering the augmented filtration $\mathbb{G} = (\mathcal{G}_t)_{t \in [0, T]}$ on a probability space $(\Omega, \mathcal{G}, \mathbb{Q})$ where \mathbb{Q} is assumed to be a risk neutral pricing measure. We then have that the present value of the future payments of the CDS are given by :

$$D_t = \mathbb{E}^{\mathbb{Q}}[\int_t^T e^{-\int_s^u r_s ds} dC_u | \mathcal{G}_t] = \mathbb{E}^{\mathbb{Q}}[R^{CDS} \int_t^T e^{-\int_s^u r_s ds} dH_u - \xi \int_t^T e^{-\int_s^u r_s ds} (1 - H_u) du | \mathcal{G}_t].$$

We can now considerer the discounted gain process associated to the CDS which will be denoted as the process $CDS = (CDS_t)_{t \in [0, T]}$ which can be written as follows :

$$CDS_t = e^{-\int_0^t r_s ds} D_t + \int_0^t e^{-\int_0^u r_s ds} dC_u. \quad (7.2)$$

We can note from the definition of D_t and using the fact $C = (C_t)_{t \in [0, T]}$ is a \mathbb{G} -adapted process that $CDS_t = \mathbb{E}^{\mathbb{Q}}[\int_0^T e^{-\int_0^u r_s ds} dC_u | \mathcal{G}_t]$. We can see that the process CDS_t then defined a \mathbb{Q} martingale adapted to the filtration \mathbb{G} .

Let's now considerer a trading strategy $\xi = (\xi^0, \xi^1)$ where ξ_t^0 represents the position in cash at time t and ξ_t^1 the position in the CDS at time t so we can define the strategy associated to position ξ as $V_t^\xi = \xi_t^1 D_t + \xi_t^0$ and his discounted value $\tilde{V}_t^\xi = e^{-\int_0^t r_s ds} V_t^\xi$.

As in the pricing theory, the strategy $(V_t^\xi)_{t \in [0, T]}$ is said to be self-financing if we have the following representation :

$$\tilde{V}_t^\xi = V_0^\xi + \int_0^t \xi_s^1 d(CDS)_s.a \quad (7.3)$$

The representation of the portfolio \tilde{V}_t^ξ defined by the equation (7.3) will help us to determine strategies ξ which minimize an hedging criterion which will be the question of the next section.

7.2 Introduction to the Mean-Variance hedging framework

The *Mean-Variance Hedging Criterion* is a part of a larger class of hedging methods called **Quadratic Hedging**. This class is called like this so because it consists in the minimization of a quadratic problem.²

7.2.1 Mathematical framework

First, we need to introduce what is called an admissible strategy ξ for a self-financing portfolio V^ξ .

Definition 7.2.1. A strategy $\xi = (\xi^0, \xi^1)$ is said to be admissible if ξ^0 is a \mathbb{G} -adapted process and ξ^1 is a \mathbb{G} -predictable process and we have the following integrability assumption : $\mathbb{E}^{\mathbb{Q}}[\int_0^T (\xi_u^1)^2 d\langle CDS \rangle_u] < +\infty$ where $\langle CDS \rangle$ represents the covariation of the process CDS with himself.

The *Mean-Variance Hedging Problem* consists then in finding an admissible strategy ξ^* with initial value $V_0^{\xi^*}$ such as the following quantity is minimized :

²Another quadratic hedging method called *Local Risk Minimization* has been heavily studied in the litterature. See [26] for an application in our context of defaultable claims with random recovery process.

$$\mathbb{E}^{\mathbb{Q}}[(e^{-\int_0^{\tau^C} r_s ds}(1 - R^C)(V_{\tau^C})^+ \mathbf{1}_{\tau^C \leq T} - (V_0^\xi + \int_0^T \xi_t^1 dCDS_t))^2]. \quad (7.4)$$

We will now focus on the method to find the strategy ξ^* . For this sake, we need to introduce the process M^{CL} which represents the discounted gain associated with credit loss and which can be written as by noting the process $CL_t = (1 - R^C)(V_{\tau^C})^+ \mathbf{1}_{\tau^C \leq t}$

$$M_t^{CL} = \mathbb{E}^{\mathbb{Q}}[\int_0^T e^{-\int_0^s r_u du} dCL_s | \mathcal{G}_t] = \int_0^t e^{-\int_0^s r_u du} dCL_s + e^{-\int_0^t r_s ds} CVA_t. \quad (7.5)$$

As we can see by its definition the process M^{CL} defines a (\mathbb{G}, \mathbb{Q}) martingale. As the process CDS defined in (7.2) defines also a (\mathbb{G}, \mathbb{Q}) martingale, we can state the following representation for $\xi^{1,*}$.

Proposition 7.1. *There exists a predictable process $\xi^{1,*}$ satisfying the integrability assumption defined in Definition 7.2.1 and a martingale A with $A_0 = 0$ such as $(CDS_t A_t)_{0 \leq t \leq T}$ is a martingale or equivalently $\langle CDS, A \rangle$ is the null-process such that :*

$$M_t^{CL} = M_0^{CL} + \int_0^t \xi_u^{1,*} d(CDS)_u + A_t \quad \mathbb{Q} - a.s. \quad 0 \leq t \leq T. \quad (7.6)$$

The strategy ξ^* with position in CDS equal to $\xi^{1,*}$ and initial value $V_0(\xi^*) = M_0^{CL} = CVA_0$ is a mean-variance minimizing strategy.

Proof. The proof of this result is based on the *Galtchouk-Kunita Wanatabe* and the *Föllmer-Schweizer* decomposition of M^{CL} on CDS and can be found in [27] from Schweizer. \square

Following the representation of equation 7.6, and using the orthogonality property between the processes A and CDS , it follows that we have :

$$\langle M^{CL}, CDS \rangle_t = \int_0^t \xi_u^{1,*} d\langle CDS \rangle_u \quad 0 \leq t \leq T.$$

We then see that $\xi^{1,*}$ can be computed as $\frac{d\langle M^{CL}, CDS \rangle}{d\langle CDS \rangle}$ and the calculation of this quantity is then required to fully characterize the minimizing strategy.

7.2.2 An application to the CCR in the financial market

In this subsection, we are going to illustrate the *Mean-Variance Methodology* for some financial products in the $B - S$ model under the following dynamics (with $\lambda = (\lambda_t)_{t \geq 0}$ representing the default intensity such that we have $\mathbb{Q}(\tau^C > t | \mathcal{F}_t) = e^{-\int_0^t \lambda_s ds}$).

$$\begin{aligned} dS_t &= S_t(rdt + \sigma dW_t^1), \quad S_0 \in \mathbb{R}_*^+. \\ d\lambda_t &= b(\lambda_t)dt + \sigma(\lambda_t)(\rho dW_t^1 + \sqrt{1 - \rho^2} dW_t^2), \quad \lambda_0 \in \mathbb{R}_*^+. \end{aligned}$$

We will considerer the case of an european derivative of payoff $\phi(S_T)$ where T is the maturity of the contract and K the potential strike associated with the european derivative. As we said in the previous section, finding the optimal strategy ξ^* relies on finding the *Galtchouk-Kunita Wanatabe* of the process M^{CL} . First, we need to derive the dynamics of M^{CL} . For every $0 \leq t \leq T$, we have :

$$M_t^{CL} = (1 - R^C) \int_0^t e^{-rs} (V_s)^+ dH_s + e^{-rt} CVA_t.$$

so that :

$$dM_t^{CL} = e^{-rt} ((1 - R^C)(V_t)^+ dH_t - rCVA_t dt + dCVA_t).$$

We see at this point that we will need to characterize the dynamics of the process CVA . From (2.14), we know that in our setting CVA is given by :

$$CVA_t = (1 - H_t)(1 - R^C) \mathbb{E}^{\mathbb{Q}} \left[\int_t^T e^{-\int_t^u (r + \lambda_s) ds} (V_u)^+ \lambda_u du \mid \mathcal{F}_t \right]. \quad (7.7)$$

Let's note the term $f^{CVA}(t, S_t, \lambda_t) = \mathbb{E}^{\mathbb{Q}}[\int_t^T e^{-\int_t^u (r + \lambda_s) ds} (V_u)^+ \lambda_u du \mid \mathcal{F}_t]$ ³ such that we have :

$$CVA_t = (1 - H_t)(1 - R^C)f^{CVA}(t, S_t, \lambda_t). \quad (7.8)$$

We can therefore write the dynamics of M^{CL} as follows :

$$\begin{aligned} dM_t^{CL} &= (1 - R^C)e^{-rt} ((V_t)^+ - f^{CVA}(t, S_t, \lambda_t)) dH_t, \\ &\quad - r(1 - H_t)f^{CVA}(t, S_t, \lambda_t) dt + (1 - H_t)df^{CVA}(t, S_t, \lambda_t)). \end{aligned} \quad (7.9)$$

Applying Itô formula to the term $f^{CVA}(t, S_t, \lambda_t)$, we have :

$$\begin{aligned} df^{CVA} &= (\sigma S \partial_S f^{CVA} + \rho \sigma(\lambda_t) \partial_{\lambda} f^{CVA}) dW_t^1 + \partial_{\lambda} f^{CVA} \sigma(\lambda_t) \sqrt{1 - \rho^2} dW_t^2 \\ &\quad + (\partial_t f^{CVA} + rS \partial_S f^{CVA} + \frac{1}{2} \sigma^2 S^2 \partial_S^2 f^{CVA} \\ &\quad + \partial_{\lambda} f^{CVA} b(\lambda_t) + \frac{1}{2} \sigma^2(\lambda_t) \partial_{\lambda}^2 f^{CVA} + \rho \sigma \sigma(\lambda_t) \partial_{S\lambda}^2 f^{CVA}) dt. \end{aligned}$$

Moreover, we already show that f^{CVA} is solution of the following PDE according to Feymann-Kac formula.

$$\begin{aligned} \partial_t f^{CVA} + \mathcal{L}^{S,\lambda} f^{CVA} + \lambda(V(t, s))^+ &= (\lambda + r) f^{CVA}, \quad \forall (t, s, \lambda) \in [0, T] \times \mathbb{R}_*^+ \times \mathbb{R}_*^+. \\ f^{CVA}(T, s, \lambda) &= 0, \quad \forall (s, \lambda) \in \mathbb{R}_*^+ \times \mathbb{R}_*^+. \end{aligned}$$

Therefore, we can write the dynamics of dM_t^{CL} as follows :

$$\begin{aligned} dM_t^{CL} &= (1 - R^C)e^{-rt} ((V_t)^+ dH_t + f^{CVA}(\lambda_t dt - H_t) \\ &\quad + (\sigma S \partial_S f^{CVA} + \rho \sigma(\lambda_t) \partial_{\lambda} f^{CVA}) dW_t^1 + \partial_{\lambda} f^{CVA} \sigma(\lambda_t) \sqrt{1 - \rho^2} dW_t^2). \end{aligned}$$

Deriving the dynamics of CDS :

Let's focus now on the dynamics of CDS . From equation (7.2), we can write :

$$dCDS_t = e^{-rt} (-rdD_t + dC_t).$$

³The form is justified by the markovian property of the processes S and λ

Proposition 7.2. *The process D is given by :*

$$D_t = (1 - H_t)g(t, Y_t). \quad (7.10)$$

where $g : [0, T] \times \mathbb{R} \rightarrow \mathbb{R}$ can be written as follows :

$$g(t, Y_t) = \mathbb{E}^Q \left[\int_t^T e^{-\int_t^u r + \lambda_s ds} (R^{CDS} \lambda_u - \xi) du \mid \mathcal{F}_t \right]. \quad (7.11)$$

Proof. The proof of this result is based on similar arguments as used in [25] \square

By characterizing the term , we arrive at the following dynamics of CDS :

$$d(CDS)_t = e^{-rt} ((R^{CDS} - g(t, \lambda_t)) (dH_t - \lambda_t dt) + (1 - H_t) \sigma(\lambda_t) \partial_\lambda g(t, \lambda_t) (\rho dW_t^1 + \sqrt{1 - \rho^2} dW_t^2)). \quad (7.12)$$

Therefore, we are able to calculate the optimal hedging strategy $\xi^{1,*}$. We have therefore by noting $g = g(t, \lambda_t)$ and $f^{CVA} = f^{CVA}(t, S_t, \lambda_t)$:

$$\begin{aligned} d\langle M^{CL}, CDS \rangle_t &= (1 - R^C) e^{-2rt} (R^{CDS} - g) (V(t, S_t)^+ - f^{CVA}) dH_t \\ &\quad + (1 - R^C) e^{-2rt} (1 - H_t) \sigma^2(\lambda_t) \partial_\lambda f^{CVA} \partial_\lambda g dt \\ &\quad + (1 - R^C) e^{-2rt} (1 - H_t) \rho \sigma S \sigma(\lambda_t) \partial_S f^{CVA} \partial_\lambda g dt. \\ d\langle CDS \rangle_t &= e^{-2rt} (R^{CDS} - g)^2 dH_t + e^{-2rt} (1 - H_t) \sigma(\lambda_t)^2 (\partial_\lambda g)^2 dt. \end{aligned}$$

A special case for numerical results :

We will set for sake of simplicity for the numerical results :

- $\sigma(\lambda_t) = 0$ and $b(\lambda_t) = 0$ such that $\lambda_t = \lambda > 0 \quad \forall t \in [0, T]$.

Under this assumption, we can write the optimal strategy as follows :

$$\xi_t^{1,*} = (1 - H_t) \frac{(1 - R^C)(V(t, S_t)^+ - f^{CVA}(t, S_t, \lambda))}{(R^{CDS} - g(t, \lambda))}. \quad (7.13)$$

In the case of an european option with payoff $\phi(S_T)$, we already show that :

$$\begin{aligned} f^{CVA}(t, S_t, \lambda) &= V(t, S_t) (1 - e^{-\lambda(T-t)}). \\ V(t, S_t) &= \mathbb{E}^Q [e^{-r(T-t)} \phi(S_T) \mid \mathcal{F}_t]. \\ \xi_t^{1,*} &= (1 - H_t) \frac{(1 - R^C)V(t, S_t)e^{-\lambda(T-t)}}{R^{CDS} - g(t, \lambda)}. \end{aligned} \quad (7.14)$$

In our setting, we can calculate the function $g(t, \lambda)$ as follows :

$$g(t, \lambda) = \mathbb{E}^Q \left[\int_t^T e^{-\int_t^u (r + \lambda) ds} (R^{CDS} \lambda - \xi) du \mid \mathcal{F}_t \right] = R^{CDS} (1 - e^{-\lambda(T-t)}) + \frac{\xi}{\lambda} (e^{-\lambda(T-t)} - 1).$$

For the computation of $(d(CDS)_t)_{t \geq 0}$, we need to distinct 3 cases.

- Before the jump characterized by $t < \tau^C$, we have $d(CDS)_t = (g'(t, \lambda) - \xi) dt$.
- At the jump τ^C , we have $dCDS_{\tau^C} = R^{CDS} - g(\tau^C, \lambda)$.⁴
- After the jump characterized by $t > \tau^C$, we have $d(CDS)_t = 0$.

⁴This case is particularly important as the CDS strategy provides at the default time τ^C of size $(1 - R^C)(V(\tau^C, S_{\tau^C})^+ - f^{CVA}(\tau^C, S_{\tau^C}, \lambda))$ which provides a perfect hedge against the counterparty risk loss at τ^C .

Let's now illustrate the algorithm we just described with some numerical results. For this, we will compute the tracking error of the strategy defined by the process :

$$e_t = \int_0^t (1 - R)(V_s)^+ dH_s - (CVA_0 + \int_0^t \xi_s^1 dCDS_s) \quad \forall t \in [0, T]. \quad (7.15)$$

- A positive value of e_T will correspond to a loss for the associated contract
- We will assume a rebalancing portfolio twice a week for the numerical result.

We will consider 2 cases of hedging ξ^1 which are :

- $\xi^1 = 0$ which means no hedging of CCR .
- $\xi_t^1 = \xi_t^{1*}$ which is the optimal hedging strategy derived previously.

To check the accuracy of the optimal hedging strategy, we will calculate on the event $\{\tau^C < T\}$ the mean of the error defined as $\mathbb{E}^Q[e_T^2]$ as it will be the indicator of the hedging. We expect of course that this value will be the lowest for the dynamic mean-variance minimizing strategy in comparison with the others as it should be concentrated around zero with a small mass in the tails of the distribution.

We give in the table below the following parameters we used for the numerical experiments where N stands for the number of timesteps and M the number of Monte-Carlo samples.

Table 7.1: Parameters used in the numerical experiments for the *Mean Variance Minimizing* strategy in the $B - S$ model

Parameters	S_0	K	r	T	λ	ξ	R^C	R^{CDS}	N	M
Value	100	100	0.0	1	0.5	0.5	0	1	52	2000

We consider also 2 cases for the value of σ :

- Case 1 when σ is supposed to be 0.2 considering a stressed market.
- Case 2 when σ is supposed to be 0.1 in an unstressed market.

Application to an european call option:

We compute the optimal strategy according to (7.14) with $\phi(S_T) = (S_T - K)^+$:

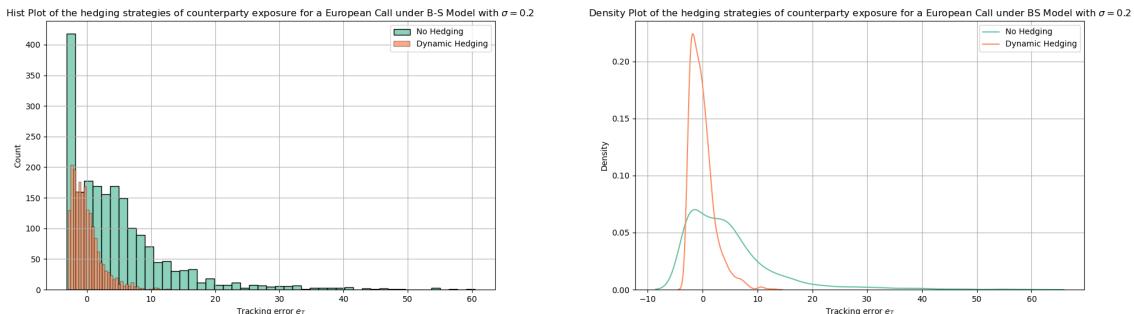


Figure 7.1: Comparison of 2 hedging strategies in order to hedge the CCR on a call option with the *Mean Variance Minimizing* framework in Case 1

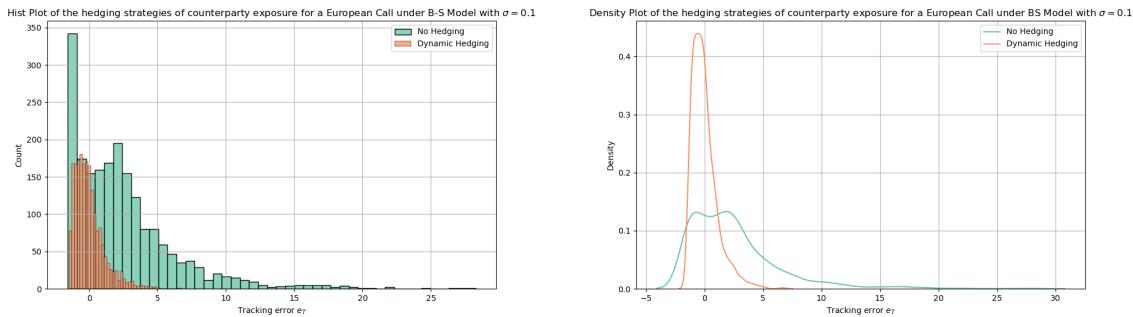


Figure 7.2: Comparison of 2 hedging strategies in order to hedge the CCR on a call option with the *Mean Variance Minimizing* framework in Case 2

We also provide the study of the tracking error e_T ⁵ :

Table 7.2: Norm 2 of e_T in case of an european call option in the $B - S$ model

	No Hedging	Dynamic Hedging
$\mathbb{E}[(e_T)^2]$ in Case 1	88.13	5.04
$\mathbb{E}[(e_T)^2]$ in Case 2	20.18	1.19

Application to a european put option :

We compute the optimal strategy according to (7.14) with $\phi(S_T) = (K - S_T)^+$

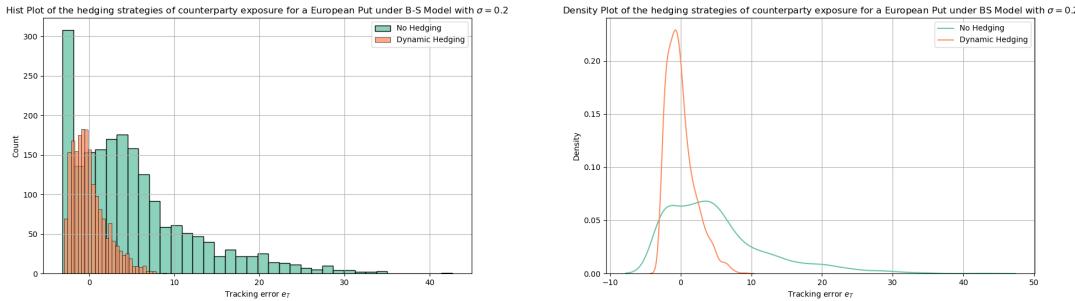


Figure 7.3: Comparison of 2 hedging strategies in order to hedge the CCR on a put option with the *Mean Variance Minimizing* framework in Case 1

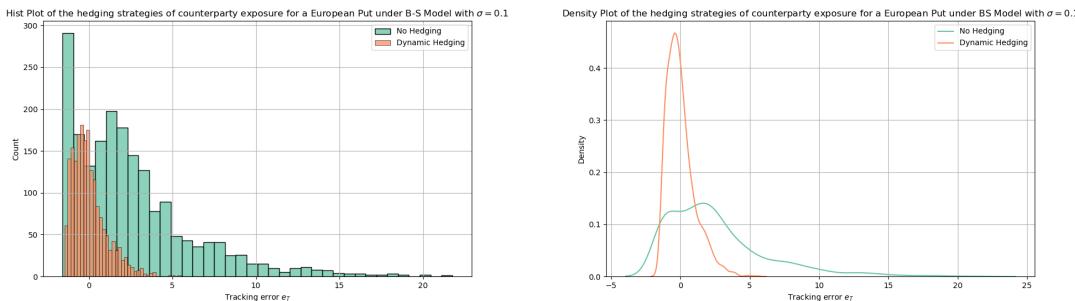


Figure 7.4: Comparison of 2 hedging strategies in order to hedge the CCR on a put option with the *Mean Variance Minimizing* framework in Case 2

⁵In the Annex E, similar results are given in the case of a Heston model for the underlying

Table 7.3: Norm 2 of e_T in case of an european put option in the $B - S$ model

	No Hedging	Dynamic Hedging
$\mathbb{E}[(e_T)^2]$ in Case 1	73.25	4.07
$\mathbb{E}[(e_T)^2]$ in Case 2	19.42	1.03

Application to a forward contract :

In this case where the payoff is given by $\phi(S_T) = (S_T - K)$, we need to calculate the f^{CVA} all paths long as we don't have a simplification like for the european case. We give below the numerical results obtained while computing CVA of the forward following equation (2.9).

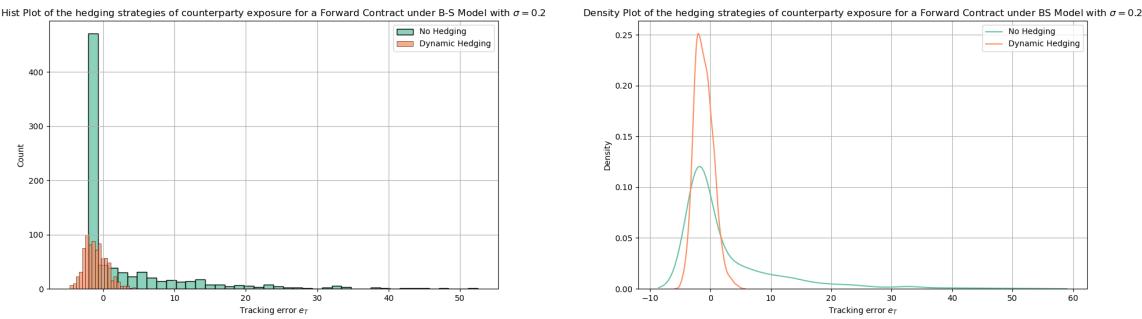


Figure 7.5: Comparison of 2 hedging strategies in order to hedge the CCR on a forward contract with the *Mean Variance Minimizing* framework in Case 1

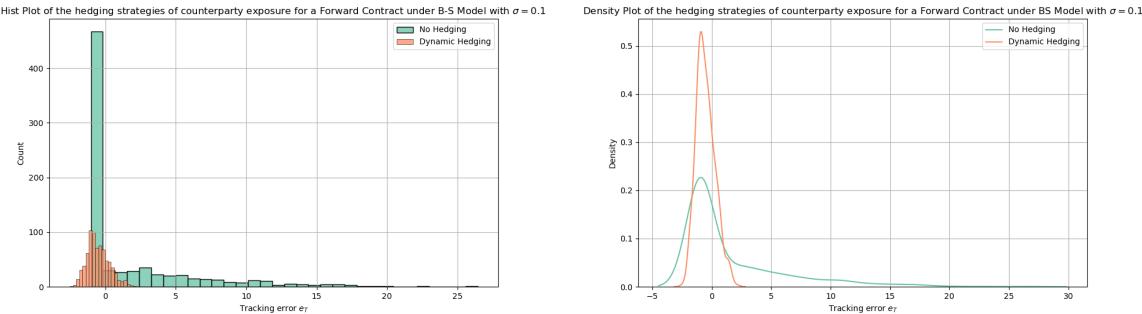


Figure 7.6: Comparison of 2 hedging strategies in order to hedge the CCR on a forward contract with the *Mean Variance Minimizing* framework in Case 2

Table 7.4: Norm 2 of e_T in case of a forward contract in the $B - S$ model

	No Hedging	Dynamic Hedging
$\mathbb{E}[(e_T)^2]$ in Case 1	72.45	3.48
$\mathbb{E}[(e_T)^2]$ in Case 2	20.75	0.89

- We show the overall performance of the dynamic hedging strategy against the no hedging strategy on the 3 different examples we provided. Losses in the extreme tails are particularly important in the case of the no hedging strategy whereas they are very measured in the case of dynamic hedging. Moreover, as expected when the market is more stressed , the tracking error e_T is more volatile but the dynamic hedging still provides good results.

7.2.3 An application to the CCR in the reinsurance market

We will now considerer the case of the hedging of the CCR in the reinsurance market. First, we need to introduce the model we will considerer. For this, we will follow the approach from [25] by referring to their notations.

Let's considerer an insurance company, labelled by I and an reinsurer labelled by R who enters in a reinsurance contract with a maturity denoted by T . We will consider that the losses in the insurance portfolio which will be the underlying of the reinsurance contract are of the form :

$$L_t = \sum_{i=1}^{N_t} Z_i \quad t \geq 0.$$

with :

- A counting process N such as $N_t = \sum_{n=1}^{+\infty} \mathbf{1}_{T_n \leq t}$ with $(T_n)_{n \in \mathbb{N}}$ \mathbb{G} -stopping times represent-ing the arrival time of claims.
- $(Z_n)_{n \in \mathbb{N}}$ strictly positive random variables representing the claim sizes and are \mathcal{G}_{T_n} measurable.

We will assume that the payment is of the form $\phi(L_T)$ for a specified ϕ assumed to be an increasing, lipschitz and continuous function. For the form of ϕ , it can cover various reinsurance contracts such as :

- A stop loss SL contract with priority m and upper limit M such as the payoff becomes $\phi(L_T^{SL}) = \min\{M, (L_T^{SL} - m)^+\}$.
- An excess-of-loss named EL with retention level m and upper limit M such as if we note $L_T^{EL} = \sum_{T_n \leq t, Z_n > M} Z_n - M$, the payoff becomes $\phi(L_T^{EL}) = \min\{M, L_T^{EL}\}$.

The model construction

We fix a filtered probability space $(\Omega, \mathcal{G}, \mathbb{Q})$. We now consider $W = (W_t)_{t \geq 0}$ a two-dimensional brownian motion $(W_t^1, W_t^2)_{t \geq 0}$, $\eta = (\eta_t)_{t \geq 0}$ a poisson process with intensity 1 independant of W and Z_n i.i.d random variables with distribution given by ν ⁶, also independant from W and η . We now consider the following process $M^L = (M_t^L)_{t \geq 0}$ defined as $M_t^L = \sum_{i=1}^{\eta_t} Z_i$ which represents a compound Poisson process. We will assume that the process N_t which represents the claim arrival is given by an intensity cadlag loss process \mathbb{G} -adapted $\lambda^L = (\lambda_t^L)_{t \geq 0}$. We will finally assume that the process H admits a default intensity $\lambda^R = (\lambda_t^R)_{t \geq 0}$.

We now define the process $Y = (Y_t)_{t \geq 0}$ as the unique solution of the following SDE ^{7 8}:

$$dY_t = b^Y(Y_t)dt + \sigma^Y(dt)(\rho dW_t^1 + \sqrt{1 - \rho^2}dW_t^2), \quad Y_0 = y_0 \in \mathbb{R}_*^+.$$

We assume as usually that there is a random variable $\phi \sim \mathcal{E}(1)$ \mathcal{G} -measurable independant with W and M^L and we define τ as :

$$\tau = \inf\{t \geq 0 : \int_0^t \lambda_s^R ds \geq \phi\}. \quad (7.16)$$

By defining τ like this, we already saw that τ is doubly stochastic with hazard rate the process $(\lambda^R(Y_t))_{t \geq 0}$

⁶supposed to be absolutely continuous with respect to Lebesgue measure.

⁷Assuming that b^Y and σ^Y verify classical hypothesis ensuring existence and unicity.

⁸The parameter ρ will represent the impact of the *Wrong Way Risk*.

We can now define the process $(H_t = \mathbb{1}_{\tau \geq t})_{t \geq 0}$ and we introduce the process $X = (X_t)_{t \geq 0}$ as the unique solution of the following SDE⁹ :

$$dX_t = \gamma^X(X_{t-})dH_t + b^X(X_t)dt + \sigma^X(X_t)dW_t^1, \quad X_0 = x_0 \in \mathbb{R}. \quad (7.17)$$

We will now assume that λ^L and λ^R are of the forms $\lambda_t^L = \lambda^L(X_t)$ and $\lambda_t^R = \lambda^R(Y_t)$. To define the loss process, we now introduce the process $\theta = (\theta_t)_{t \geq 0}$ defined as $\theta_t = \int_0^t \lambda^L(X_{s-})ds$ such that we have :

$$\begin{aligned} \tilde{N}_t &= \eta_{\theta_t}, \quad t \geq 0. \\ L_t &= M_{\theta_t}^L = \sum_{i=1}^{N_t} Z_i, \quad t \geq 0. \end{aligned}$$

From now, we can define the filtration $\mathbb{G} = (\mathcal{G}_t)_{t \leq 0}$ defined as :

$$\mathcal{G}_t = \mathcal{F}_t^W \vee \mathcal{F}_t^L \vee \mathcal{H}_t, \quad t \geq 0. \quad (7.18)$$

where F_t^W is the filtration generated by W , F_t^L by L and H_t by H . Under this setting, we have Z_n which are \mathcal{G}_{T_n} -measurable random variables and τ is a random time with respect to \mathbb{G} . In [25], they introduce the so called **Contagion Free Market** which will be used in the computation of the optimal hedging strategy. For this, let's define the process $\tilde{X} = (\tilde{X}_t)_{t \geq 0}$ the unique solution to the following SDE :

$$d\tilde{X}_t = b^X(\tilde{X}_t)dt + \sigma^X(\tilde{X}_t)dW_t^1, \quad \tilde{X}_0 = x_0 \in \mathbb{R}.$$

From the definition of the process, we see that \tilde{X} has the same dynamics with X except from the jump at τ . As previously, we define the analogue process $\tilde{\theta}$ to θ such that $\tilde{\theta}_t = \int_0^t \lambda^L(\tilde{X}_s)ds$ such that we can define the Cox Process $\tilde{N} = (\tilde{N}_t)_{t \geq 0}$ and the loss process $\tilde{L} = (\tilde{L}_t)_{t \geq 0}$ and as follows :

$$\begin{aligned} \tilde{N}_t &= \eta_{\tilde{\theta}_t} \quad t \geq 0. \\ \tilde{L}_t &= \sum_{i=1}^{\tilde{N}_t} Z_i \quad t \geq 0. \end{aligned}$$

After having defined the model, we can now focus on deriving the equations associated with the quantities of interest. We will assume a constant risk-free interest rate and we can define therefore the price of the reinsurance contract and the associated *CVA* as follows under the \mathbb{Q} risk neutral pricing measure.

$$\begin{aligned} V_t &= \mathbb{E}^{\mathbb{Q}}[e^{-r(T-t)}\phi(L_T)|\mathcal{G}_t]. \\ CVA_t &= (1 - R^C)\mathbb{E}^{\mathbb{Q}}[\mathbb{1}_{t < \tau \leq T} e^{-r(\tau-t)}(V_{\tau})^+|\mathcal{G}_t]. \end{aligned} \quad (7.19)$$

For the following, we will resume the main results from [25] where proofs can be found in their article.¹⁰

⁹The function $\gamma^X(X_t^-)$ represents the pricing contagion which aims to represent the fact that if R defaults than the supply for the reinsurance market will reduce by making the price of the reinsurance contract going up

¹⁰The proof of each result can be found in their article if the reader wants more details. We just give here the main quantities which are necessary to derive the optimal hedging strategy.

Proposition 7.3. *There exists a unique bounded classical solution v (i.e. continuous, \mathcal{C}^1 in t and \mathcal{C}^2 in x) of the following backward PIDE¹¹*

$$\frac{\partial v}{\partial t}(t, l, x) + \mathcal{L}^{(\tilde{L}, \tilde{X})}v(t, l, x) = rv(t, l, x), \quad (t, l, x) \in [0, T] \times \mathbb{R}^+ \times \mathbb{R}, \quad (7.20)$$

with the generator of the diffusion given by :

$$\mathcal{L}^{(\tilde{L}, \tilde{X})}f(t, l, x) = \frac{\partial f}{\partial x}(t, l, x)b^X(x) + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(t, l, x)(\sigma^X(x))^2 \quad (7.21)$$

$$+ \int_{\mathbb{R}^+} (f(t, l+z, x) - f(t, l, x))\lambda^L(x)\nu(dz). \quad (7.22)$$

with terminal condition $v(T, l, x) = \phi(l)$. Moreover, it holds for $\tau \leq T$ that

$$V_\tau = v(\tau, \tilde{L}_\tau, \tilde{X}_\tau + \gamma^X(\tilde{X}_\tau)).$$

Proposition 7.4. *The value of the CVA is given by :*

$$CVA_t = (1 - R^C)(1 - H_t)f^{CVA}(t, L_t, X_t, Y_t). \quad (7.23)$$

where $f^{CVA} : [0, T] \times \mathbb{R}^+ \times \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}^+$ can be written as follows :

$$f^{CVA}(t, L_t, X_t, Y_t) = \mathbb{E}^\mathbb{Q}[\int_t^T v(s, \tilde{L}_s, (\tilde{X}_s + \gamma^X(\tilde{X}_s))\lambda^Y(Y_s)e^{-\int_t^s(r+\lambda^Y(Y_u))du}ds | \mathcal{F}_t] \quad (7.24)$$

Similarly to the previous section, we will assume the same form of hedging strategy so we have the following :

Proposition 7.5. *The process D is given by :*

$$D_t = (1 - H_t)g(t, Y_t). \quad (7.25)$$

where $g : [0, T] \times \mathbb{R} \rightarrow \mathbb{R}$ can be written as follows :

$$g(t, Y_t) = \mathbb{E}^\mathbb{Q}[\int_t^T e^{-\int_t^u r + \lambda^Y(Y_s)ds} (R^{CDS}\lambda^Y(Y_u) - \xi)du | \mathcal{F}_t]. \quad (7.26)$$

Theorem 7.1. *The Q -mean-variance minimizing strategy is characterized by the initial value $V_0(\xi^*) = CVA_0$ and by the CDS position $\xi_t^{1,*} = \frac{d\langle M^{CL}, CDS \rangle}{d\langle S \rangle}$ for every $0 \leq t \leq T$, where*

$$\begin{aligned} \frac{d\langle M, S \rangle_t}{dt} &= (1 - R)e^{-2rt}(1 - H_{t-}) \left\{ \rho\sigma^X(X_{t-})\sigma^Y(Y_t) \frac{\partial f^{CVA}}{\partial x}(t, L_{t-}, X_{t-}, Y_t) \frac{\partial g}{\partial y}(t, Y_t) \right. \\ &\quad + (\sigma^Y(Y_t))^2 \frac{\partial f^{CVA}}{\partial y}(t, L_{t-}, X_{t-}, Y_t) \frac{\partial g}{\partial y}(t, Y_t) \\ &\quad \left. + \lambda^R(Y_t)(R^{CDS} - g(t, Y_t))(v^\phi(t, L_{t-}, X_{t-} + \gamma^X(X_{t-})) - f^{CVA}(t, L_{t-}, X_{t-}, Y_t)) \right\}. \end{aligned} \quad (7.27)$$

and

$$\frac{d\langle CDS \rangle_t}{dt} = e^{-2rt}(1 - H_{t-}) \left\{ \lambda^R(Y_t)(R^{CDS} - g(t, Y_t))^2 + (\sigma^Y(Y_t))^2 \left(\frac{\partial g}{\partial y}(t, Y_t) \right)^2 \right\}. \quad (7.28)$$

Proof. The proof is very similar to the ones we did in the previous section leading to similar computations except that as the process X can jump, we have to take care of this. \square

¹¹PIDE refers to *Partial Integro Differential Equation* which is a *PDE* where it can involve also the integral of a function.

Some numerical results :

We first for our numerical results the case of a stop loss contract with the following payoff $\phi(L_T) = (L_T - K)^+$ with $K = 90$ and $T = 1$.

For the calculation of CVA_0 , we will consider the following model.

$$\begin{aligned} dY_t &= (0.05 - Y_t)dt + 0.1\sqrt{Y_t}(\rho dW_t^1 + \sqrt{1-\rho^2}dW_t^2), \quad Y_0 = 0.05. \\ dX_t &= \gamma X_{t-}dH_t + \kappa(100 - X_t)dt + \sigma X_t dW_t^1, \quad X_0 = 100. \end{aligned}$$

As we mentioned, we have 2 possible representations of the CVA with equation (7.19) and (7.23) but as we know how to simulate τ here, we decided to use the equation (7.19).

- Claim sizes are $\Gamma(1, 1)$ distributed. ¹²
- $\kappa = 0.5$ and $\sigma = 0.2$.

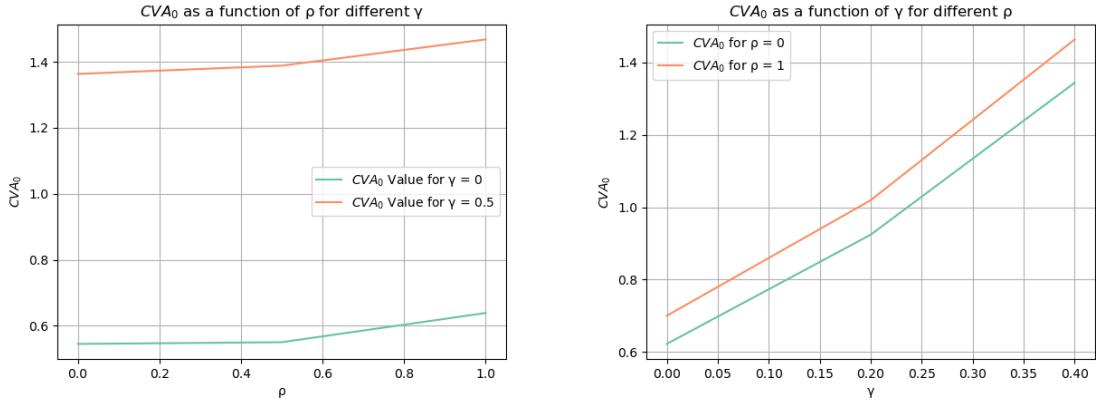


Figure 7.7: Evolution of CVA_0 as a function of ρ for different values of γ (left) and as a function of γ for different values of ρ (right)

Some remarks on the results :

- The parameter ρ which captures the wrong way risk in our model shows that indeed where ρ increases, CVA_0 also increases from figure 7.7 but the impact of the price contagion looks way more pronounced in CVA_0 than the wrong way risk and it looks to be an important feature in the modelling of the reinsurance counterparty credit risk.

¹²The density function of $X \sim \Gamma(\alpha, \beta)$ is given by $f_X(x) = x^{\alpha-1} \frac{\beta^\alpha e^{-\beta x}}{\Gamma(\alpha)} \mathbf{1}_{x \geq 0}$ where Γ is the Gamma function of Euler.

Comparison of hedging strategies :

For the hedging strategies, we will like in the previous section suppose that the loss intensity is such that we have $Y_t = Y_0 \quad \forall t \in [0, T]$ as it doesn't require the computations of derivatives of f^{CVA} or g .

We therefore have a simplified expression for f^{CVA} and g which are given by :

$$f^{CVA}(t, L_t, X_t, \lambda) = \mathbb{E}^{\mathbb{Q}} \left[\int_t^T v(s, L_s, (1 + \gamma)\tilde{X}_s) \lambda e^{-\int_t^s (r + \lambda) du} ds \mid \mathcal{F}_t \right] = v(t, L_t, (1 + \gamma)\tilde{X}_t)(1 - e^{-\lambda(T-t)}).$$

$$g(t, \lambda) = \mathbb{E}^{\mathbb{Q}} \left[\int_t^T e^{-\int_t^u (r + \lambda) ds} (R^{CDS} \lambda - \xi) du \mid \mathcal{F}_t \right] = R^{CDS}(1 - e^{-\lambda(T-t)}) + \frac{\xi}{\lambda}(e^{-\lambda(T-t)} - 1).$$

From this simplified formula, it follows that the optimal strategy $\xi_t^{1,*}$ is given for $0 \leq t \leq T$ by :

$$\xi_t^{1,*} = (1 - H_t) \frac{(1 - R^C)(v(t, L_t, (1 + \gamma)\tilde{X}_t)e^{-\lambda(T-t)})}{R^{CDS} - g(t, \lambda)}. \quad (7.29)$$

We will now give some numerical results by calculating like in the previous subsection the tracking error e_T for one case with no hedging of the CVA and the other with the formula (7.29).

We give in the following table the parameters we used in the different cases. ¹³

Table 7.5: Parameter used in the numerical experiments in the *Mean Variance Minimizing strategy* for the stop loss contract

Parameters	X_0	L_0	λ	ρ	ξ	α	β	γ	R^C	R^{CDS}	N	M
Case 1	100	0	0.2	0	0.2	1	1	0	0	1	52	2000
Case 2	10	0	0.2	0	0.2	10	1	0	0	1	52	2000

- The first case where we choose $X_0 = 100$ and $\alpha = 1$ will correspond to a case where we suppose that the insurer suffer frequent small losses whereas the second case $X_0 = 10$ and $\alpha = 10$ correspond to a case where the insurer suffers infrequent but large losses. Therefore, we expect that we will have a more volatile tracking error e_T in the second case with a dynamic hedging strategy performing better in the first case.

¹³We took the following parameters similarly to [25].

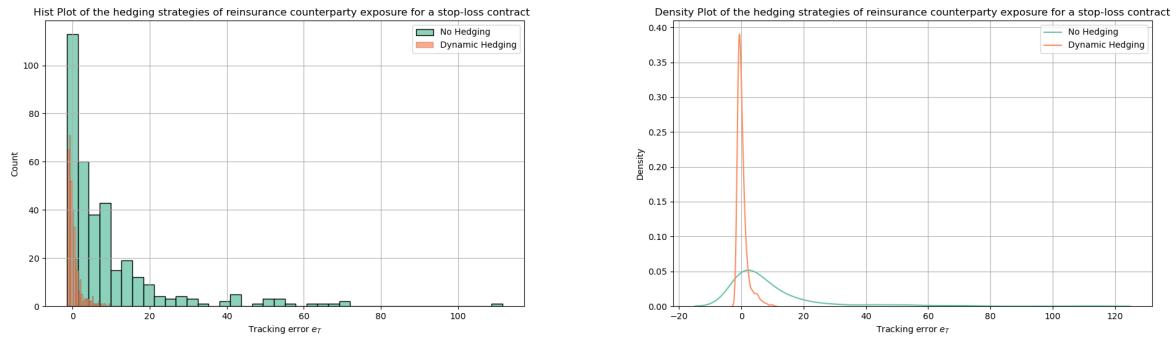


Figure 7.8: Comparison of 2 hedging strategies in order to hedge the CCR on a stop loss contract with the *Mean Variance Minimizing* framework in Case 1

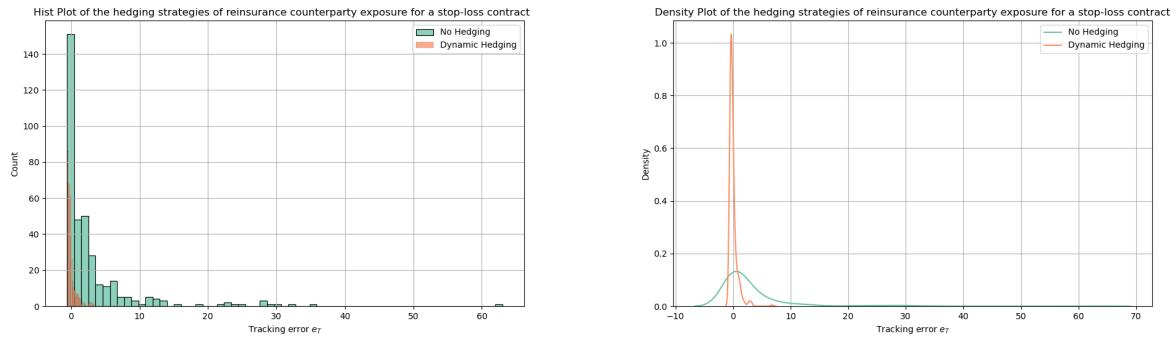


Figure 7.9: Comparison of 2 hedging strategies in order to hedge the CCR on a stop loss contract with the *Mean Variance Minimizing* framework in Case 2

Table 7.6: Norm 2 of e_T in case of a stop-loss contract

	No Hedging	Dynamic Hedging
$\mathbb{E}[(e_T)^2]$ in Case 1	283.65	2.90
$\mathbb{E}[(e_T)^2]$ in Case 2	50.84	0.52

Some remarks on the results :

- We observe that the dynamic hedging strategy performs well better than the no hedging strategy with the norm 2 of the tracking error being way less important in the dynamic hedging strategy.
- Moreover, we observe that in the case of infrequent but large losses which correspond to the case 1, the losses that the insurer can suffer are way more higher in the case of frequent but small losses. Moreover, we see that the dynamic hedging performs worse than in the case 1 which is also intuitive as the potential losses are largest. Anyways in this case, we clearly see that not performing an hedging strategy can lead to disastrous losses whereas with performing the dynamic strategy, we can clearly have a better protection against the potential default of the reinsurer.

7.3 Introduction to a static hedging approach for the CCR in the reinsurance market

In this subsection, we will quickly introduce the static approach introduced in [28] by Chi, Hu and Huang in their article *Optimal risk management with reinsurance and its counterparty risk hedging* where they propose an optimal risk management point of view for an insurer who has to deal with the potential counterparty risk of a reinsurer and who can manage this with an hedging instrument assumed to be default-free like we did in the previous section. The purpose of the insurer is to solve his optimization problem by maximizing his expected utility wealth under the potential default of the reinsurer.

7.3.1 Mathematical framework

Let's consider an insurer who has an initial wealth w and who faces a random variable risk X in a fixed time period assumed to be bounded on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

Let's consider that the insurer can reduce his risk exposure by ceding an amount $r(X)$ which is assumed to be such that $0 \leq r(X) \leq X$ a.s¹⁴. Moreover, they add the following condition : $r(X)$ and $X - r(X)$ are both increasing random variables.

- The condition $0 \leq r(X) \leq X$ a.s is called *Indemnity Principle*.
- The condition $r(X)$ and $X - r(X)$ both increasing functions is called *Non Sabotage condition* to reduce the ex-post moral hazard on the insurer who can misreport the true value of the sinister.

These 2 conditions allow us to define $\mathcal{R} = \{r : r(0) = 0 \text{ and } 0 \leq r'(X) \leq 1 \text{ a.s}\}$.

We now consider the potential default of the reinsurance denote by D such as $\mathbb{1}_D$ is the indicator variable representing the default. We also assume that in case of a default the *LGD* is given by $1 - R^C = \tau$.

The reinsurance premium is then given by the expected value principle for fair pricing :

$$\pi_R(r) = (1 + \rho_R) \mathbb{E}[(1 - \tau \mathbb{1}_D)r(X)]. \quad (7.30)$$

where ρ_R is the classic safety chargement.

If we make the assumption that the default D is independant from the risk X , and assuming that the default happens with a probability p , we therefore have :

$$\pi_R(r) = (1 + \rho_R)(1 - \tau p) \mathbb{E}[r(X)]. \quad (7.31)$$

We consider that to hedge the potential default of the reinsurer, we can use a financial hedging instrument $h(X)$ with payoff $h(X)\mathbb{1}_D$ with h assumed to be a positive function.

The fair value of this contract is therefore given by :

$$\pi_H(h) = 1 + \rho_H \mathbb{E}[h(X)\mathbb{1}_D] = (1 + \rho_H)p \mathbb{E}[h(X)]. \quad (7.32)$$

where ρ_H is a safety chargement.

We can therefore write the final wealth W of the insurer at the end of the period :

$$W_{r,h}(X, D) = w - X + (1 - \tau \mathbb{1}_D)r(X) + h(X)\mathbb{1}_D - \pi_R(r) - \pi_H(h). \quad (7.33)$$

¹⁴as means *almost surely* meaning that $\mathbb{P}(0 \leq r(X) \leq X) = 1$.

Under the expected utility theory, we assume that the insurer is risk averse and is characterized by an utility function u who verifies the classic hypothesis.¹⁵
Therefore the expected utility of the insurer's final wealth is given by :

$$\mathcal{L}(r, h) = \mathbb{E}[u(W_{r,h}(X, D))] = p\mathbb{E}[u(W_{r,h}^d(X))] + (1-p)\mathbb{E}[u(W_{r,h}^f(X))]. \quad (7.34)$$

with :

$$\begin{aligned} W_{r,h}^d(X) &= w - X + (1-\tau)r(X) + h(X) - \pi_R(r) - \pi_H(h). \\ W_{r,h}^f(X) &= w - X + r(X) - \pi_R(r) - \pi_H(h). \end{aligned}$$

The objective for the insurer is to seek an optimal policy (r, h) which will maximize his expected utility. Therefore, we seek to solve the following problem :

$$\max_{r \in \mathcal{R}, h \geq 0} \mathcal{L}(r, h). \quad (7.35)$$

Optimal Forms of Reinsurance and Hedging :

We will now give optimal forms of r and h that solve the problem (7.35). For this, we will mainly refer to the article [28].

Proposition 7.6. For $\tau \in]0, 1[$, an optimal solution (r^*, h^*) to (7.35) is given by :

$$\begin{aligned} r^*(x) &= (x - l)^+ - (x - m)^+ + (x - t)^+. \\ h^*(x) &= (x - c)^+ - (1 - \tau)(x - t)^+. \end{aligned}$$

where l, m, c, t are parameters such as $0 \leq l \leq m \leq c \leq t \leq M$ with M the supremum of X . If $\tau = 1$, an optimal solution is given by :

$$\begin{aligned} r^*(x) &= (x - a)^+. \\ h^*(x) &= (x - b)^+. \end{aligned}$$

where $a, b \in [0, M]$.

Proof. The proof can be found in [28] and they show the existence and unicity of the optimal solution by simplifying the infinite dimensional problem (7.35) into a problem with finite decision variables. \square

They also shows some interesting forms of reinsurance and hedging in multiple cases which we will try to summarize here :

- The optimal reinsurance/hedging form in the case of $\rho_R = \rho_H = 0$ is given by :

$$(r^*(x), h^*(x)) = (x, \tau x)$$

- In the case of $\frac{u'(w-M)}{\mathbb{E}[u'(w-X)]} \leq 1 + \rho_R \wedge \rho_H$, the optimal form is given by :

$$(r^*(x), h^*(x)) = (0, 0) \quad (7.36)$$

¹⁵u is assumed to be concave with $\lim_{x \rightarrow +\infty} u'(x) = 0$

These 2 cases are already interesting as they help to show some limit cases. Indeed, the first case says that if both reinsurance and hedging contracts are fair priced, then it's optimal to cede all the risk to the insurance and to hedge all the defaultable risk. This result is quite intuitive as it will lead to a fair value of the wealth of the insurer to be equal to $w - \pi_R - \pi_H$ in both cases no matter default happened or not. The second case is also informative in the sense that it says if both reinsurance contract and hedging contract are too costly for the insurer, then it's preferable to not buy a reinsurance contract and so on an hedging contract.

They consider also non-trivial cases where we suppose the following assumptions :

$$\begin{aligned} 1 &< (1 + \rho_R)(1 + \rho_H). \\ (1 + \rho_R \wedge \rho_H) &< \frac{u'(w - M)}{\mathbb{E}[u'(w - X)]}. \\ S(X) &= [0, M]. \end{aligned} \tag{7.37}$$

By considering this, we don't take in consideration the both previous forms obtained in simple cases. and considering that $S(X) = [0, M]$ allows for a potential jump at $X = 0$.

Proposition 7.7. *Under Assumptions 7.37, we have :*

(i) *If $\rho_R > \rho_H$, then :*

$$\begin{cases} r^*(x) = (x - t)^+. \\ h^*(x) = (x - c)^+ - (1 - \tau)(x - t)^+. \end{cases} \tag{7.38}$$

where $0 < c < t \leq M$

(ii) *If $\rho_R < \rho_H$, then :*

$$\begin{cases} r^*(x) = (x - l)^+. \\ h^*(x) = \tau(x - t)^+. \end{cases} \tag{7.39}$$

where $0 \leq l < t \leq M$

(iii) *If $\rho_R = \rho_H$, then :*

$$\begin{cases} r^*(x) = (x - d^*)^+. \\ h^*(x) = \tau(x - d^*)^+. \end{cases} \tag{7.40}$$

where $d^* = \sup\{d \in [0, M] : \frac{\mathbb{E}[u'(w - X \wedge d - (1 + \rho_R)\mathbb{E}[(X - d)^+])]}{u'(w - d - (1 + \rho_R)\mathbb{E}[(X - d)^+])} \geq \frac{1}{1 + \rho^R}\}$.

Some remarks on the model

- We assume that X is a bounded random variable but of course, it could not be the case and we could potentially have to deal with a potential ruin of the insurer which can also be taking account in the modelisation.
- An interesting point is the form of optimal contracts. We can see that optimal reinsurance and hedging contracts are function of stop-loss contracts which shows that these type of contracts are particularly well suited in the design of an optimal risk management policy.

7.3.2 Some numerical results

To illustrate the previous algorithm, we will consider the following set of characteristics :

- The risk function X assumed to have a density function $f_X(x) = \frac{\lambda e^{-\lambda x}}{1-e^{-\lambda M}} \mathbb{1}_{0 \leq x \leq M}$.
- Utility Function : $u_\alpha(w) = -e^{-\alpha w}$.
- $\rho_R = \rho_H = 0.1$, $\lambda = 0.6$, $p = 0.2$, $\tau = 0.6$, $M = 20$ and $w = M + 10 = 30$.

According to Proposition 7.7, we need to calculate :

$$d^* = \sup\{d \in [0, M] : \frac{\mathbb{E}[u'(w - X \wedge d - (1 + \rho_R)\mathbb{E}[(X - d)^+])]}{u'(w - d - (1 + \rho_R)\mathbb{E}[(X - d)^+])} \geq \frac{1}{1 + \rho_R}\}. \quad (7.41)$$

Lemma 7.1. Assuming X has a density function $f_X(x) = \frac{\lambda e^{-\lambda x}}{1-e^{-\lambda M}} \mathbb{1}_{0 \leq x \leq M}$, Then, $\forall 0 < d < M$, we have :

$$\begin{aligned} \mathbb{E}[e^{\alpha(X \wedge d)}] &= \int_0^d e^{\alpha x} f_X(x) dx + \int_d^M e^{\alpha d} f_X(x) dx, \\ &= \frac{\lambda}{1 - e^{-\lambda M}} \int_0^d e^{(\alpha - \lambda)x} dx + \int_d^M e^{\alpha d} e^{-\lambda x} dx, \\ &= \frac{\lambda}{1 - e^{-\lambda M}} \left(\frac{1}{\lambda + \alpha} (1 - e^{(\alpha - \lambda)d}) + \frac{e^{\alpha d}}{\lambda} (e^{-\lambda d} - e^{-\lambda M}) \right), \\ &= \frac{\lambda}{(\lambda + \alpha)(1 - e^{-\lambda M})} (1 - e^{(\alpha - \lambda)d}) + \frac{e^{\alpha d}}{1 - e^{-\lambda M}} (e^{-\lambda d} - e^{-\lambda M}). \end{aligned} \quad (7.42)$$

As $u'_\alpha(w) = \alpha e^{-\alpha w}$, the left term in (7.43) reduces to the following computation :

$$\frac{\mathbb{E}[u'_\alpha(X \wedge d)]}{u'_\alpha(d)} = e^{-\alpha d} \mathbb{E}[e^{\alpha(X \wedge d)}].$$

According to the Lemma 7.1, we therefore have :

$$\begin{aligned} \frac{\mathbb{E}[u'_\alpha(X \wedge d)]}{u'_\alpha(d)} &= \frac{1}{1 - e^{-\lambda M}} \left\{ \frac{\lambda}{\lambda + \alpha} (e^{-\alpha d} - e^{-\lambda d}) + e^{-\lambda d} - e^{-\lambda M} \right\}, \\ &= \frac{1}{1 - e^{-\lambda M}} \left\{ \left(\frac{\lambda}{\lambda + \alpha} e^{-\alpha d} + \frac{\alpha}{\lambda + \alpha} e^{-\lambda d} \right) - e^{-\lambda M} \right\}. \end{aligned}$$

Finally we see that d^* is defined as :

$$\sup\{d \in [0, M] : \frac{\lambda}{\lambda + \alpha} e^{-\alpha d} + \frac{\alpha}{\lambda + \alpha} e^{-\lambda d} \geq \frac{1 + e^{-\lambda M} \rho_R}{1 + \rho_R}\}. \quad (7.43)$$

From equation (7.43), we can now derive the optimal level d^* for the optimal reinsurance and hedging contract.¹⁶

¹⁶ Finding d^* has been performed using the classic Newton-Raphson algorithm to find the roots of a smooth function f (See <https://personal.math.ubc.ca/~anstee/math104/newtonmethod.pdf> for an introduction to the algorithm).

Some numerical results

We will provide the evolution of d^* with respect to both α and ρ^R to see how they affect the optimal reinsurance and hedging contracts.¹⁷

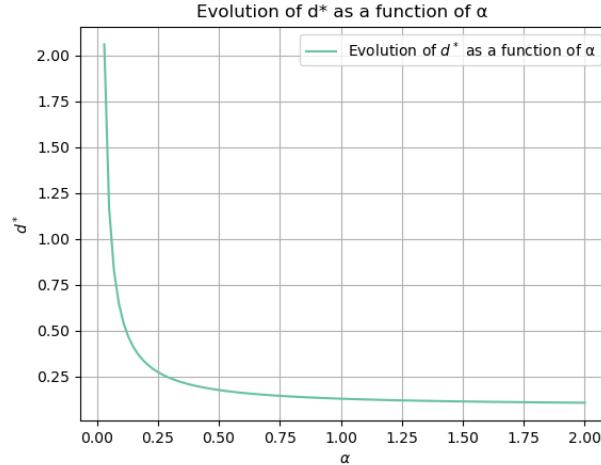


Figure 7.10: Evolution of d^* as a function of the aversion parameter α with $\rho_R = \rho_H = 0.1$

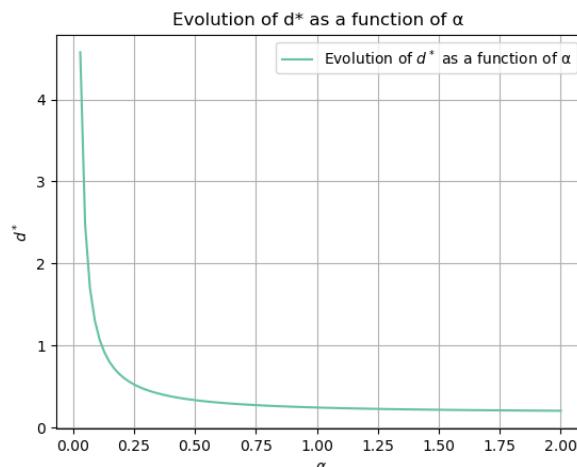


Figure 7.11: Evolution of d^* as a function of the aversion parameter α with $\rho_R = \rho_H = 0.2$

¹⁷Other numerical results are presented in the article in less trivial cases but involve more numerical computations to find the optimal levels of reinsurance and hedging.

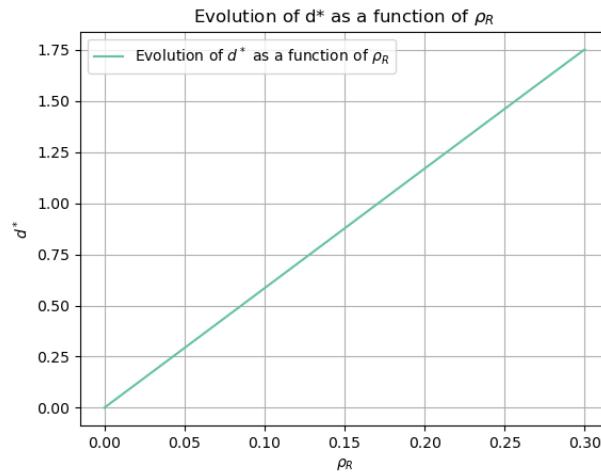


Figure 7.12: Evolution of d^* as a function of the aversion parameter ρ_R with $\alpha = 0.1$

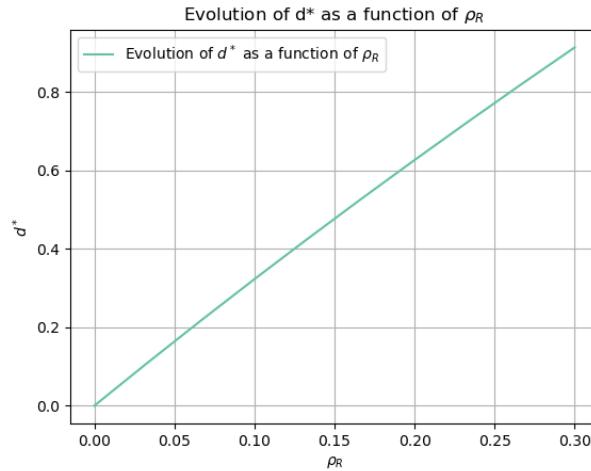


Figure 7.13: Evolution of d^* as a function of the aversion parameter ρ_R with $\alpha = 0.2$

Some remarks on the results :

- The optimal type of reinsurance contracts according to the algorithm are stop loss contracts which is similar to Arrow's result from 1963 with in this case the hedging instrument where the insurer covers the tailed risk for any level of τ .
- We see that d^* is decreasing with the aversion parameter α which is intuitive in the sense that if an insurer is really risk averse, he wants to transfer the most risk as possible to the reinsurer which will lead to a lower level of d^* .
- We can see that when both the reinsurance and the hedging contracts become expensive with an higher ρ^R , d^* increases which can be understood as trying to reduce the actuarial prime by setting the level of d^* higher for both contracts.
- In our special case of utility function u_α , we see that d^* is not a function of the initial wealth which is a nontrivial term to analyse and if we choose another type of utility function such as $u(w) = \sqrt{w}$, we would see an other dependency of d^* with respect to w .

Globals remarks on the chapter :

- Through this chapter, we talked about efficient hedging methods of the counterparty exposure. We first introduce a dynamic hedging strategy with a possible investment in a *CDS* called *Mean-Variance Hedging strategy*. We illustrate it through numerical examples on the financial market by deriving analytic formulas for the self-financing portfolio. We showed how it globally reduces the error and why it was important to cover the counterparty exposure by showing the loss we could face in case of no hedging. After talking about financial market, we talked about the reinsurance market by dealing with the potential default of a reinsurance involved in a stop loss contract. For this, we used the model introduced by Ceci, Colanery, Frey and Köck in [25] where they propose a stochastic model of the loss of an insurer and where they also derive the CVA_0 on the stop loss contract. They also propose the same investment on a *CDS* written on the potential default of the reinsurer and how a dynamic hedging could be performed in order to reduce drastically the counterparty exposure. The numerical results we obtained are similar to them showing that it could indeed be used by the insurance market to better cover their exposition to reinsurance *CCR*.
- Finally, in this chapter, we also analyze a static hedging strategy based on the article by Chi, Hu and Huang [28] where they analyze optimal risk management with reinsurance in case of counterparty risk. To perform this, they consider their problem as an optimization problem where the insurer tries to maximise his utility function through the choice of the reinsurance and hedging contracts. They show that the solutions will depend of some model parameters but they rely on stop loss contracts for the reinsurance contracts and similarly for the hedging contracts. We illustrate numerically one optimal reinsurance and hedging contracts where the loss of the insurer is given by a truncated exponential random variable and with a *CARA* utility function. In this sample case, we could derive optimal reinsurance and hedging contracts and perform some analysis on the impact of each parameter of the model on the optimal solutions.
- We note that the hedging strategies we derived in this chapter are highly dependent of the choice of the model we are dealing with and could clearly be improved by potentially adding transaction costs and rebalancing costs in the dynamic hedging setup for example but they provide a mathematical overview of the current analysis of the counterparty exposure which exists in the litterature. We also observe that other strategies in order to mitigate the *CCR* also exist like *Collateralization* or *Netting* but the litterature is really scarce on these concepts.

Conclusion

This dissertation provided an overview of numerical methods, transitioning from classical approaches to recent supervised learning methods, addressing the challenges associated with the computation of *XVAs*.

Firstly, a specific mathematical framework, inspired by academic literature, was proposed for the calculation of various *XVAs* and formed the foundation of this dissertation. From this framework, numerous illustrations were provided for the calculation of counterparty risk across different types of financial products, including european equity options and interest rate swaps under the *G2++* and Hull & White models, as well as more exotic products such as bermudan options, evaluated via a put and a swaption using the *Least Square Monte-Carlo* method. This dissertation also emphasized the significant impact of *Wrong Way Risk* in the valuation of *XVAs* by presenting two modelling approaches based on research articles. One of these approaches incorporates a change of measure, allowing the integration of *WWR* into the pricing of *CVA* under the *Wrong Way Measure*.

Secondly, this dissertation demonstrated the relevance of supervised learning algorithms for the calculation of *XVAs*, particularly in overcoming the classical issues of the nested Monte-Carlo approach. A detailed study was conducted on the contribution of Gaussian Process Regression, highlighting both the strengths and limitations of the algorithm. It was shown how combining *Gaussian Process Regression* with classical Monte-Carlo methods could efficiently learn the expected exposure profile of European derivatives or interest rate swap portfolios, thereby avoiding the need for nested Monte-Carlo procedures. An important part of this dissertation was also devoted to the study of fully connected deep neural networks and their usefulness in the calculation of *XVAs*. Two algorithms were studied : one which was based on the *PDE* representation of *XVAs* called *Deep XVA Solver* where we computed the expected profile of a very high dimensional european derivative which showed that it can overcome the curse of dimensionality and the other based on their probabilistic representation called *Deep Conditional Expectation Solver* where an efficient calculation of MVA_0 was computed avoiding also the nested Monte-Carlo procedure.

Finally, this dissertation focused on the hedging aspects of counterparty risk. A dynamic hedging approach was studied using quadratic hedging methods, with an investment in a benchmark hedging instrument for counterparty risk : Credit Default Swaps (*CDS*). An application to reinsurance counterparty risk in a stop-loss contract was presented, where we derived an analytical formula for the optimal investment strategy in the *CDS* to minimize the tracking error. A static approach, based on expected utility theory, was also proposed. In this approach, an insurer seeks to maximize the expected utility of its wealth by subscribing to a reinsurance contract and a hedging instrument to guard against the reinsurer's potential default. The insurer then determines the optimal contracts to solve his optimization problem.

In conclusion, this dissertation explored various quantitative aspects related to the management of *XVAs*, which have become a critical topic for banks and insurers since the 2008

financial crisis. The work highlighted the computational challenges of pricing *XVAs* and proposed new efficient numerical methods to address these challenges. Finally, it introduced key concepts in counterparty risk hedging, using both dynamic and static approaches.

Potential further research

- The financial industry also seeks to calculate in addition to the average exposure profile EE the exposure profile at a given percentile $\alpha \in [0.1]$ given by:

$$PFE_t^\alpha = \inf\{y : P((V_t)^+ \leq y) \leq \alpha\}$$

This complementary measure echoes the definition of *Value-at-Risk* and recently supervised learning methods have emerged for the calculation of these risk measures based on a dual representation of the *Value-at-Risk* and *Expected Shortfall* as minimization problems as introduced in the article [29] *Learning Value-at-Risk and Expected Shortfall* from Barrera, Crépey, Gobet, Nguyen and Saadeddine.

- Deep neural networks are currently studied for the valuation of life insurance options indexed to stocks, in particular as introduced in the article [30] *Pricing equity-linked life insurance contracts with multiple risk factors by neural networks* from Barigou and Delong.
- Supervised learning algorithms are also studied for the valuation of high-dimensional Bermudan options as introduced in the article [31] *Deep Optimal Stopping* from Becker, Cheridito and Jentzen where the optimal exercise time is learned on a sample of data as well as the calculation of exposure profiles associated with these options as introduced in the article [32] *A deep learning approach for computations of exposure profiles for high-dimensional Bermudan options* from Andersson and Oosterlee.
- Finally, the academic literature is also focusing on neural networks for hedging objectives as introduced in the article [33] *Deep Quadratic Hedging* from Gnoatto, Picarelli and Lavagnini where quadratic hedging strategies are learned based on the *Deep BSDE Solver*. Recently, methods based on reinforcement learning have emerged as introduced in the article [34] *Reinforcement Learning for CVA hedging* from Alonso and Zhdankin.

Bibliography

- [1] J.Gregory, 2020, *The xVA Challenge : Counterparty Risk, Funding, Collateral, Capital and Initial Margin*, Third Edition, Wiley Finance Series.
- [2] A.Green, C.Kenyon, C.Dennis, 2014, *KVA: Capital Valuation Adjustment*, arXiv : 1405.0515.
- [3] T. R. Bielecki, M. Rutkowski, 2002, *Credit Risk: modelling, Valuation and Hedging, Series: Springer Finance*, Springer.
- [4] A.Nikeghbali, 2006, *An essay on the general theory of stochastic processes*, arXiv : 0506581.
- [5] T. Van de Zwaard, L.A. Grzelak, C.W. Oosterlee, 2023, *Relevance of Wrong-Way Risk in Funding Valuation Adjustments*, Finance Research Letters, Volume 49,103091.
- [6] C.Albanese, S.Caenazzo, S.Crépey, 2017, *Funding, Margin and Capital Valuation Adjustments for Bilateral Trade Portfolios*, American Institute of Mathematical Sciences.
- [7] D.Brigo, F.Mercurio, 2001, *Interest Rate Models – Theory and Practice*, Springer.
- [8] J.N.Tsitsiklis, B.Van Roy, 2001, *Regression methods for pricing complex American-style options*, IEEE Transactions on Neural Networks, Volume 12.
- [9] F.Longstaff, E.Schwartz, 2001, *Valuing American options by simulation: a simple least-squares approach*, Review of Financial Studies, 113-147.
- [10] P. Glasserman, 2003, *Monte-Carlo methods in financial engineering*, Springer.
- [11] D.Brigo, F.Vrins, 2016, *Disentangling wrong-way risk: pricing CVA via change of measures and drift adjustment*, European Journal of Operational Research, Volume 269, 1154-1164.
- [12] T.R. Bielecki, M.Jeanblanc, M.Rutkowski, 2011, *Credit risk modelling. Technical report, Center for the Study of Finance and Insurance*
- [13] B.Huge and A.Savine, 2020, *Differential Machine Learning*, arXiv : 2005.02347.
- [14] C.E.Rasmussen, C.K. I. Williams, 2006, *Gaussian Processes for Machine Learning. The MIT Press*, Adaptive Computation and Machine Learning series.
- [15] E.Weinan, J.Han, A.Jentzen, 2017, *Deep learning-based numerical methods for high-dimensional parabolic partial differential equations and backward stochastic differential equations*, Springer, Commun. Math. Stat. 5, 349–380.
- [16] K.P.Murphy, 2012, *Machine Learning: A Probabilistic Perspective*
- [17] H.Pham, 2009, *Continuous-time Stochastic Control and Optimization with Financial Applications*, Springer.

- [18] A.Bachouch, C.Huré, N.Langrené, H.Pham, 2017, *Deep neural networks algorithms for stochastic control problems on finite horizon: numerical applications*, Springer, Methodology and Computing in Applied Probability, Volume 24, 143-178.
- [19] A.Gnoatto, A.Picarelli, C.Reisinger, 2022, *Deep XVA Solver – A Neural Network Based CounterParty Credit Risk Management Framework*, SIAM, Volume 14.
- [20] ISDA SIMM Methodology version 2.6 available on the ISDA website.
- [21] J.P. Villarino, A.Leitao, 2024, *On Deep Learning for computing the Dynamic Initial Margin and Margin Value Adjustment*, arXiv : 2407.16435.
- [22] S.Crépey, M.F.Dixon, 2019, *Gaussian Process Regression for Derivative Portfolio modelling and Application to CVA Computations*, arXiv : 1901.11081.
- [23] L.Goudenège, A.Molent, A.Zanette, 2023, *Computing XVA for American basket derivatives by Machine Learning techniques*, arXiv : 2209.06485.
- [24] J.Sirignano, K.Spiliopoulos, 2018, *DGM: A deep learning algorithm for solving partial differential equations*, Journal of Computational Physics, Volume 375, 1339-1364.
- [25] C.Ceci, K.Colanery, R.Frey, V.Köck, 2019, *Value Adjustments And Dynamic Hedging of Reinsurance Counterparty Risk*, SIAM, Volume 11.
- [26] F.Biagini, A.Cretarola, 2012, *Local Risk-Minimization for Defaultable Claims with Recovery Process*, Springer, Applied Mathematics & Optimization, Volume 65, 293-314.
- [27] M.Schweizer, 2001, *A guided tour through quadratic hedging approaches*
- [28] Y.Chi, T.Hu, Y.Huang, 2023, *Optimal risk management with reinsurance and its counterparty risk hedging*, Insurance: Mathematics and Economics, Volume 113, 274-292.
- [29] J.D.B Cano, S.Crépey, E.Gobet, H-D.Nguyen, B.Saadeddine, 2022, *Learning Value-at-Risk and Expected Shortfall*, arXiv : 2209.06476.
- [30] K.Barigou, L.Delong, 2021, *Pricing equity-linked life insurance contracts with multiple risk factors by neural networks*, Journal of Computational and Applied Mathematics, Volume 404, 113922.
- [31] S.Becker, P.Cheridito, A.Jentzen, 2020, *Deep Optimal Stopping*, Journal of Machine Learning Research.
- [32] K.Andersson, C.W.Oosterlee, 2020, *A deep learning approach for computations of exposure profiles for high-dimensional Bermudan options* Applied Mathematics and Computation, Volume 408, 126332.
- [33] A.Gnoatto, S.Lavagnini, A.Picarelli, 2022, *Deep Quadratic Hedging*, arXiv : 2212.12725.
- [34] M.N.Alonso, I.Zhdankin, 2022, *Reinforcement Learning for CVA hedging*, Social Science Research Network.
- [35] I. Karatzas, S.E. Shreve, 1991, *Brownian Motion and Stochastic Calculus*, Springer.
- [36] A.Alfonsi, 2005, *On the discretization schemes for the CIR (and Bessel squared) processes*, Monte Carlo Methods and Applications.

- [37] K.Hornik, 1988, *Multilayer feedforward networks are universal approximators*, Neural Networks, Volume 2 Issue 5, 359-366.
- [38] M.F.Dixon, I.Halperin, P.Bilokon, 2020, *Machine Learning in Finance: From Theory to Practice*, Springer.
- [39] E. Ben Slama, 2022, Mémoire d'Actuariat, *Deep pricing and deep calibration: applications in finance and insurance*
- [40] A. Mohamed Karim, 2022, Master's Thesis, *Numerical methods for Expected Exposure computation*
- [41] A.Colin, 2017, Mémoire d'Actuariat, *Marge initiale : impact de la méthodologie sur la valorisation des produits OTC non compensés*

Glossary

XVA	X Valuation Adjustment
CVA	Credit Valuation Adjustment
DVA	Debt Valuation Adjustment
BCVA	Bilateral Credit Valuation Adjustment
FVA	Funding Valuation Adjustment
FCA	Funding Cost Adjustment
FBA	Funding Benefit Adjustment
KVA	Capital Valuation Adjustment
MVA	Margin Valuation Adjustment
IM	Initial Margin
DIM	Dynamic Initial Margin
ISDA	International Swaps and Derivatives Association
SIMM	Standard Initial Margin Model
EC	Economic Capital
MPOR	Margin Period of Risk
LGD	Loss Given Default
EAD	Exposure at Default
PD	Probability of Default
CCR	Counterparty Credit Risk
EE	Expected Exposure
PE	Positive Exposure
NE	Negative Exposure
EPE	Expected Positive Exposure
ENE	Expected Negative Exposure
PFE	Potential Future Exposure

ML	Machine Learning
DL	Deep Learning
FFNN	Feed Forward Neural Networks
ANN	Artificial Neural Network
GPR	Gaussian Processes Regression
PDE	Partial Differential Equation
PIDE	Partial Integro-Differential Equation
SDE	Stochastic Differential Equation
BSDE	Backward Stochastic Differential Equation
FBSDE	Forward Backward Stochastic Differential Equation
CARA	Constant Absolute Risk Aversion
CDS	Credit Default Swap
WWR	Wrong Way Risk
IRS	Interest Rate Swap
MC	Monte-Carlo
LSMC	Least Square Monte-Carlo
GMMB	Guarantee Maximum Maturity Benefit
B-S	Black-Scholes
GBM	Geometric Brownian Motion
CDF	Cumulative Distribution Function
MSE	Mean Squared Error
MAE	Mean Absolute Error
VaR	Value at Risk
ES	Expected Shortfall

Annexes

A Key propositions used in this dissertation

Proposition A.1. Unidimensionnal Feymann-Kac Formula

Let's consider a probability space $(\Omega, \mathcal{F}, \mathbb{Q})$ supported by a one-dimensional brownian motion $W^{\mathbb{Q}} = (W_t^{\mathbb{Q}})_{t \geq 0}$ and the filtration $\mathbb{F} = (\mathcal{F}_t)_{t \geq 0}$ generated by the brownian motion. We consider $X = (X_t)_{t \in [0, T]}$ with $T > 0$ the unique¹⁸ process valued in \mathbb{R} solution of the following SDE :

$$dX_t = \mu(t, X_t)dt + \sigma(t, X_t)dW_t^{\mathbb{Q}}.$$

Let's consider now the following partial differential equation of which u is solution :

$$\begin{aligned} \partial_t u(t, x) + \mu(t, x)\partial_x u(t, x) + \frac{1}{2}\sigma^2(t, x)\partial_{x^2} u(t, x) - r(t, x)u(t, x) + f(t, x) &= 0, \quad \forall (t, x) \in [0, T] \times \mathbb{R}. \\ u(T, x) &= g(x), \quad \forall x \in \mathbb{R}. \end{aligned}$$

Then, $u(t, x)$ can be rewritten as the following :

$$u(t, x) = \mathbb{E}^{\mathbb{Q}}[e^{-\int_t^T r(s, X_s)ds} g(X_T) + \int_t^T e^{-\int_t^u r(s, X_s)ds} f(u, X_u)du | X_t = x]. \quad (\text{A.1})$$

Proof. We will just give the proof of the form of the solution u assuming existence has been done. Let's consider the following process $Y = (Y_s)_{s \in [t, T]}$ defined as :

$$Y_s = e^{-\int_t^s r(u, X_u)du} u(s, X_s) + \int_t^s e^{-\int_t^u r(l, X_l)dl} f(u, X_u)du.$$

By differentiating dY_s , we therefore have :

$$dY_s = e^{-\int_t^s r(u, X_u)du} (-r(s, X_s)u(s, X_s)ds + du(s, X_s) + f(s, X_s)ds). \quad (\text{A.2})$$

By applying Itô formula to u as it is assumed to be $\mathcal{C}^{1,2}$, we can therefore write :

$$du(s, X_s) = (\partial_s u(s, X_s) + \mu(s, X_s)\partial_x u(s, X_s) + \frac{1}{2}\sigma^2(s, X_s)\partial_{x^2} u(s, X_s))ds + \sigma(s, X_s)\partial_x u(s, X_s)dW_s^{\mathbb{Q}}.$$

By now using the PDE of which u is solution, we can therefore write (A.2) as :

$$dY_s = e^{-\int_t^s r(u, X_u)du} \sigma(s, X_s)\partial_x u(s, X_s)dW_s^{\mathbb{Q}}. \quad (\text{A.3})$$

By integrating and using the fact according to (A.3), the process Y defines a local martingale (which will be assumed to be a true martingale), we can write :

$$\mathbb{E}^{\mathbb{Q}}[Y_T | X_t = x] = \mathbb{E}^{\mathbb{Q}}[Y_t | X_t = x] = u(t, x).$$

Going back to the definition of Y_T , we recover the form of $u(t, x)$ of the theorem. \square

¹⁸Assuming that μ and σ verify the classical assumptions ensuring existence and unicity.

Proposition A.2. Bayes Lemma

Let's consider $(\Gamma_t)_{t \in [0, T]}$ a strictly positive martingale on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in [0, T]})$ with a probability measure \mathbb{P} on this probability space such that $\mathbb{E}^{\mathbb{P}}[\cdot]$ denotes the expectation of \cdot under \mathbb{P} and we assume that $\mathbb{E}^{\mathbb{P}}[\Gamma_t] = 1$. Let's now define the probability measure \mathbb{Q} on (Ω, \mathcal{F}) such that we have :

$$\mathbb{Q}(A) = \mathbb{E}^{\mathbb{P}}[\Gamma_T \mathbf{1}_A] \quad \forall A \in \mathcal{F}. \quad (\text{A.4})$$

Therefore, we have for X random variable such that $\mathbb{E}^{\mathbb{Q}}[|X|] < +\infty$:

$$\mathbb{E}^{\mathbb{Q}}[X | \mathcal{F}_t] = \frac{\mathbb{E}^{\mathbb{P}}[X \Gamma_T | \mathcal{F}_t]}{\mathbb{E}^{\mathbb{P}}[\Gamma_T | \mathcal{F}_t]} = \frac{\mathbb{E}^{\mathbb{P}}[X \Gamma_T | \mathcal{F}_t]}{\Gamma_t} \quad \forall t \in [0, T]. \quad (\text{A.5})$$

Moreover, we can define the Radon-Nikodym process $(\Gamma_t)_{t \in [0, T]}$ ¹⁹ such that we have for any $A \in \mathcal{F}_t$:

$$\mathbb{Q}(A) = \mathbb{E}^{\mathbb{P}}[\Gamma_T \mathbf{1}_A] = \mathbb{E}^{\mathbb{P}}[\Gamma_t \mathbf{1}_A]. \quad (\text{A.6})$$

Assuming now a process $X = (X_t)_{t \in [0, T]}$ which is an $(\mathcal{F}_t)_{t \in [0, T]}$ adapted process, we can now write:

$$\mathbb{E}^{\mathbb{Q}}[X_t | \mathcal{F}_s] = \frac{\mathbb{E}^{\mathbb{P}}[X_t \Gamma_t | \mathcal{F}_s]}{\Gamma_s} \quad \forall s < t < T. \quad (\text{A.7})$$

Proof. First, it's clear that \mathbb{Q} defines a probability measure as \mathbb{P} is assumed to be a probability measure. Now let's consider a set $A \in \mathcal{F}_t$ and X random variable such that $\mathbb{E}^{\mathbb{Q}}[|X|] < +\infty$. Let's show that :

$$\mathbb{E}^{\mathbb{Q}}[\mathbf{1}_A X] = E^{\mathbb{Q}}[\mathbf{1}_A \frac{\mathbb{E}^{\mathbb{P}}[X \Gamma_T | \mathcal{F}_t]}{\Gamma_t}].$$

For this, let's start from the term $\mathbb{E}^{\mathbb{Q}}[\mathbf{1}_A \frac{\mathbb{E}^{\mathbb{P}}[X \Gamma_T | \mathcal{F}_t]}{\Gamma_t}]$.

$$\begin{aligned} E^{\mathbb{Q}}[\mathbf{1}_A \frac{\mathbb{E}^{\mathbb{P}}[X \Gamma_T | \mathcal{F}_t]}{\Gamma_t}] &= E^{\mathbb{P}}[\mathbf{1}_A \Gamma_T \frac{\mathbb{E}^{\mathbb{P}}[X \Gamma_T | \mathcal{F}_t]}{\Gamma_t}] \\ &= E^{\mathbb{P}}[\mathbb{E}^{\mathbb{P}}[\mathbf{1}_A \Gamma_T \frac{\mathbb{E}^{\mathbb{P}}[X \Gamma_T | \mathcal{F}_t]}{\Gamma_t} | \mathcal{F}_t]] \\ &= E^{\mathbb{P}}[[\mathbf{1}_A \frac{\mathbb{E}^{\mathbb{P}}[X \Gamma_T | \mathcal{F}_t]}{\Gamma_t} \mathbb{E}^{\mathbb{P}}[\Gamma_T | \mathcal{F}_t]]] \\ &= E^{\mathbb{P}}[\mathbf{1}_A \mathbb{E}^{\mathbb{P}}[X \Gamma_T | \mathcal{F}_t]] \\ &= E^{\mathbb{P}}[\mathbb{E}^{\mathbb{P}}[\mathbf{1}_A X \Gamma_T | \mathcal{F}_t]] \\ &= \mathbb{E}^{\mathbb{P}}[\mathbf{1}_A X \Gamma_T] \\ &= E^{\mathbb{Q}}[\mathbf{1}_A X]. \end{aligned}$$

The first and last lines come from the Radon Nikodym density between \mathbb{Q} and \mathbb{P} . The second and sixth lines come from the *tower rule property*. The third and fifth lines come from the \mathcal{F}_t -measurability of $\mathbf{1}_A$ and $\mathbb{E}^{\mathbb{P}}[X \Gamma_T | \mathcal{F}_t]$. The fourth line come from the martingale property of the process Γ which ends the proof of equation (A.5). The equation (A.6) comes from conditioning on \mathcal{F}_t and the proof of the equation (A.7) is the same as the one of equation (A.5).

□

¹⁹In the literature, we can see the notation $\frac{d\mathbb{Q}}{d\mathbb{P}}|_{\mathcal{F}_t} = \Gamma_t$

Proposition A.3. Girsanov Theorem

Let's consider a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in [0, T]}, \mathbb{P})$. Let's also consider a process $L = (L_t)_{t \in [0, T]}$ with $T > 0$ a local martingale with $L_0 = 0$ such that the process $\mathcal{E}(L)_t = e^{L_t - \frac{1}{2}\langle L, L \rangle_t}$ is a true martingale (where $\langle ., . \rangle$ denotes the quadratic variation of the process). Let's now define the probability measure \mathbb{Q} such that $\forall A \in \mathcal{F}_t$, we have :

$$\mathbb{Q}(A) = \mathbb{E}^{\mathbb{P}}[\mathbf{1}_A \mathcal{E}(L)_t]. \quad (\text{A.8})$$

Now, let's consider $(W_t)_{t \in [0, T]}$ a brownian motion under \mathbb{P} , then the process $\tilde{W} = (\tilde{W}_t)_{t \in [0, T]}$ defined by :

$$\tilde{W}_t = W_t - \langle W, L \rangle_t. \quad (\text{A.9})$$

is a brownian motion under \mathbb{Q} . More generally, if we consider a local martingale $M = (M_t)_{t \in [0, T]}$ under \mathbb{P} , then the process $\tilde{M} = (\tilde{M}_t)_{t \in [0, T]}$ defined by :

$$\tilde{M}_t = M_t - \langle M, L \rangle_t. \quad (\text{A.10})$$

is a local martingale under \mathbb{Q} with the same quadratic variation as M .

Proof. For the case of a brownian motion, several proofs of this result exist in the litterature. we will do a proof by using the following characterization of a brownian motion.

Lemma A.1. *A continuous stochastic process $W = (W_t)_{t \geq 0}$ such that $W_0 = 0$ on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$ is a one dimensional Brownian motion if and only if :*

- W is a martingale with respect to the filtration $(\mathcal{F}_t)_{t \geq 0}$.
- $(W_t^2 - t)_{t \geq 0}$ is a martingale with respect to the filtration $(\mathcal{F}_t)_{t \geq 0}$.

Proof. The proof of this lemma can be found in [35]. □

Now, let's consider W_t a brownian motion under \mathbb{P} and let's define the process $\tilde{W}_t = W_t - \langle W, L \rangle_t$. Let's show that $\tilde{W} = (\tilde{W}_t)_{t \geq 0}$ is a brownian motion under \mathbb{Q} which means that we have to check according to the lemma (A.1) that \tilde{W} and $(\tilde{W}_t^2 - t)_{t \geq 0}$ are martingales under \mathbb{Q} with the filtration generated by W . For this, we need to show by using the Bayes Lemma and noting the process $M_t = \mathcal{E}(L)_t$ that $\forall 0 \leq s < t$

$$\mathbb{E}^{\mathbb{Q}}[\tilde{W}_t | \mathcal{F}_s] = \frac{\mathbb{E}^{\mathbb{P}}[\tilde{W}_t M_t | \mathcal{F}_s]}{M_s} = \tilde{W}_s. \quad (\text{A.11})$$

We see that we just need to show that the process $\tilde{W}M$ is a martingale under \mathbb{P} . Or, we have that $dM_t = M_t dL_t$, $d\tilde{W}_t = dW_t - d\langle W, L \rangle_t$, $d\langle \tilde{W}, M \rangle_t = M_t d\langle W, L \rangle_t$. Therefore, by applying Itô formula for a product, we have that :

$$d(\tilde{W}M)_t = \tilde{W}_t M_t dL_t + M_t d\tilde{W}_t + d\langle \tilde{W}, M \rangle_t = \tilde{W}_t M_t dL_t + M_t dW_t.$$

Therefore the process $\tilde{W}M$ is a true martingale as the sum of 2 martingales under \mathbb{P} (assuming L is a true martingale under \mathbb{P}) which proves the equation (A.11). The proof of the martingale property of $(\tilde{W}_t^2 - t)_{t \geq 0}$ under \mathbb{Q} is the same by considering the process $((\tilde{W}_t^2 - t)M_t)_{t \geq 0}$ and showing this is a martingale under \mathbb{P} by the same arguments. Therefore, the process \tilde{W} is a continuous process starting from 0 verifying the properties of the previous lemma so is a one dimensional brownian motion under \mathbb{Q} . □

Proposition A.4. Conditional Expectation Representation

Consider 2 random variables Y and X in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such as $\mathbb{E}[Y|X]$ is in $L^2(X)$. Then, $\mathbb{E}[Y|X]$ is the unique solution to the following optimization problem :

$$\operatorname{argmin}_{f \in L^2(X)} \mathbb{E}[(Y - f(X))^2].$$

Proof. To make a proof, we will need the following lemma :

Lemma A.2. Let H be a closed vector sub-space of L^2 and $X \in L^2$.

Then, there exists an unique Z such that $\mathbb{E}[(X - Z)^2] = \inf\{\mathbb{E}[(X - Y)^2] : Y \in H\}$ and $\forall Y \in H$, we have $\mathbb{E}[XY] = \mathbb{E}[ZY]$.

Proof. Let's note by $a = \inf\{\mathbb{E}[(X - Y)^2] : Y \in H\}$. Let's denote by Y' and $Z' \in L^2$. Therefore, we have the median formula :

$$\mathbb{E}[(Z' - Y')^2] + \mathbb{E}[(Z' + Y')^2] = 2(\mathbb{E}[Z'^2] + \mathbb{E}[Y'^2]).$$

Let's define $(X_n)_{n \in \mathbb{N}}$ a sequence of H such that $\lim_{n \rightarrow +\infty} \mathbb{E}[(X - X_n)^2] = a$. Defining now $Z' = X_n - X$ and $Y' = X_m - X$, we then have by defining $I = \frac{X_n + X_m}{2} \in H$ the following :

$$\mathbb{E}[(X_n - X_m)^2] = 2\mathbb{E}[(X - X_n)^2] + 2\mathbb{E}[(X - X_m)^2] - 4\mathbb{E}[(X - I)^2]$$

By the definition of a , we see that $(X_n)_{n \in \mathbb{N}}$ defines a Cauchy sequence in L^2 which is a complete space, then it means that $(X_n)_{n \in \mathbb{N}}$ converges in L^2 , let's say to X_H . But as H is closed, we have that $X_H \in H$. Let's consider $Y \in H$ and the function $t \rightarrow \mathbb{E}[(X - X_H + tY)^2]$ which is a polynomial function minimal for $t = 0$ with value a . Or, the derivative at $t = 0$ is also 0 and equals $\mathbb{E}[(X - X_H)Y] = 0$ which gives $\mathbb{E}[XY] = \mathbb{E}[XHY]$.

Let's proof the unicity now by considering another variable Z such that we have $\mathbb{E}[(X - Z)^2] = a$. According to the equality, we have $\forall Y \in H$, $\mathbb{E}[(X - Z)Y] = 0$. Therefore, we can write :

$$\begin{aligned} \mathbb{E}[(X_H - Z)Y] &= \mathbb{E}[(X_H - X + X - Z)Y] \\ &= \mathbb{E}[(X_H - X)Y] + \mathbb{E}[(X - Z)Y] \\ &= 0. \end{aligned}$$

By taking now $Y = X_H - Z \in H$, we have $\mathbb{E}[(X_H - Z)^2] = 0$ meaning that $X_H = Z$ a.s and the unicity is proved. □

Going back now to the definition of $\mathbb{E}[Y|X]$, we know that it means that for every $A \in \sigma(X)$, we have that :

$$\mathbb{E}[Y \mathbf{1}_A] = \mathbb{E}[E[Y|X] \mathbf{1}_A]. \quad (\text{A.12})$$

Moreover, the set $L^2(X)$ correspond to the space L^2 on the probability space $(\Omega, \sigma(X), \mathbb{P})$ which is a closed space. Therefore, $L^2(X)$ is a closed vector subspace of L^2 and following the property of the Lemma A.2 which verifies $\mathbb{E}[Y|X]$ by his definition A.12 and the fact that $\mathbb{E}[Y|X] \in L^2(X)$ and $\mathbf{1}_A \in L^2(X)$, we must have that $\mathbb{E}[Y|X]$ is the orthogonal projection of $Y \in L^2$ on $\sigma(X)$ which ends the proof. □

Proposition A.5. Martingale Representation Theorem

Let's consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and W a n -dimensional Brownian motion on this space such that $W = (W^1, W^2, \dots, W^n) \in \mathbb{R}^n$ with for each $i \in \llbracket 1; n \rrbracket$, W^i is a standard 1-dimensional Brownian motion and for each $i \neq j$, W^i and W^j are independant. We also suppose that the space is filtered by the canonical filtration generated by W meaning that $\mathcal{F}_t = \sigma((W_s)_{s \leq t}) \quad \forall t \geq 0$.

Let $T > 0$ and $M = (M_t)_{t \in [0, T]}$ a squared adapted integrable martingale valued in \mathbb{R} with $M_0 \in \mathbb{R}$ defined on the filtered probability space. Therefore, there exists a unique constant c and a unique progressif process $\phi = (\phi_t)_{t \in [0, T]}$ ²⁰ such that :

$$M_t = c + \int_0^t \phi_s^T dW_s, \quad \mathbb{P} - a.s. \quad (\text{A.13})$$

Moreover, we have that $c = M_0$ and $\mathbb{E}[\int_0^T \|\phi_t\|^2 dt] < +\infty$.

The proof of this really important result is based on the following lemma which is a more general case of the martingale decomposition on a brownian filtration.

Lemma A.3. We consider the same setup as in the proposition. Let $T > 0$ and $\xi \in L^2(\mathcal{F}_T)$. Therefore, there exists a unique constant $c = \mathbb{E}[\xi]$ and a unique progressif process $\phi = (\phi_t)_{t \in [0, T]}$ valued in \mathbb{R}^n with $\mathbb{E}[\int_0^T \|\phi_t\|^2 dt] < +\infty$ such that

$$\xi = c + \int_0^T \phi_t^T dW_t, \quad \mathbb{P} - a.s. \quad (\text{A.14})$$

Proof. We start by the unicity by supposing that the existence has been proved. By taking $t = 0$ in equation A.13, it's clear that c is unique and equals to M_0 . Let's now consider another progressif process $\psi = (\psi_t)_{t \in [0, T]}$ such that we have :

$$M_t = c + \int_0^t \psi_s^T dW_s, \quad \mathbb{P} - a.s.$$

The result therefore follows from the unicity in the decomposition of an Itô Process from the previous Lemma leading to the equality of the processes $\phi_t = \psi_t \quad 0 \leq t \leq T, \quad ds \otimes d\mathbb{P}$. For the existence, we will refer to the previous lemma applied to the random variable M_T which is by assumption \mathcal{F}_T -adapted and squared integrable. Therefore, there exists $(\phi = (\phi_t)_{t \in [0, T]} \in L^2([0, T])$ such that we have :

$$M_T = \mathbb{E}[M_T] + \int_0^T \phi_s^T dW_s, \quad \mathbb{P} - a.s. \quad (\text{A.15})$$

As the process ϕ is $\in H^2([0, T])$, the process $\int_0^t \phi_s^T dW_s$ defines a true martingale and by taking the conditional expectation with respect to \mathcal{F}_t in A.15, we therefore have that :

$$\begin{aligned} \mathbb{E}^Q[M_T | \mathcal{F}_t] &= M_0 + \mathbb{E}\left[\int_0^T \phi_s^T dW_s | \mathcal{F}_t\right], \quad \mathbb{P} - a.s \\ M_t &= M_0 + \int_0^t \phi_s^T dW_s, \quad \mathbb{P} - a.s. \end{aligned} \quad (\text{A.16})$$

Therefore, the existence is proven with $c = M_0$ and wih $\phi \in L^2$.

□

²⁰A process ϕ is said to be progressively measurable if $\forall t \geq 0$ the application X defined on $[0, t] \times \Omega$ by $X(s, w) = \phi_s(w)$ is $\mathcal{B}([0, t]) \times \mathcal{F}_t$ -measurable

B A quick introduction to discretization schemes

In this annex, we will discuss a bit around the discretisation schemes as we make an heavy use of them in this dissertation. For this, we will consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and W a one-dimensionial brownian motion. We now consider a process $X = (X_t)_{t \in [0, T]}$ valued in \mathbb{R} given by the following *SDE* :

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t, \quad X_0 \in \mathbb{R}.$$

Let's now consider a timegrid π over $[0, T]$ defined by :

$$\pi = \{0 = t_0 < t_1 < \dots < t_N = T\}. \quad (\text{B.1})$$

Remark. In this dissertation, we used an equidistant timestep $h = \frac{T}{N}$ such that $t_n = nh$.

Definition B.1. An Euler approximation of the equation B.1 associated with the timegrid π is a discrete time process $X^\pi = \{X_t^\pi \mid t \in \pi\}$ defined by the following iterative scheme $\forall n \in \llbracket 0; N-1 \rrbracket$:

$$X_{t_{n+1}}^\pi = X_{t_n}^\pi + b(t_n, X_{t_n}^\pi)(t_{n+1} - t_n) + \sigma(t_n, X_{t_n}^\pi)(W_{t_{n+1}} - W_{t_n}), \quad X_0^\pi = X_0 \quad (\text{B.2})$$

Remark. $W_{t_{n+1}} - W_{t_n}$ is a Gaussian random variable with zero mean and variance given by $t_{n+1} - t_n = h$. To generate the increments $W_{t_{n+1}} - W_{t_n}$, we can therefore use a sequence of independant variables $(G_i)_{i \geq 1} \sim \mathcal{N}(0, h)$ as increments of a brownian motion are independant.

The main result about the Euler discretization scheme is given by the following proposition which gives the rate of convergence of X^π towards the true solution X .

Proposition B.1. Assume b and σ continous and lipschitz functions in t and x , then we have the following error control :

$$\mathbb{E}[\sup_{t \in [0, T]} |X_t - X_t^\pi|^2] \leq C|\pi|. \quad (\text{B.3})$$

where $|\pi| = \max_{n \in \llbracket 0; N-1 \rrbracket} t_{n+1} - t_n$ and $C > 0$ a constant which doesn't depend in $|\pi|$.

Remark. In this dissertation, we used the Euler scheme when we had to discretize diffusion processes but other choices could have been done to improve the rate of convergence.²¹ For the case of the CIR process where the SDE is given with the following choices of b and σ :

- $b(t, x) = \kappa(\theta - x)$ with $(\kappa, \theta) \in \mathbb{R}^+ \times \mathbb{R}$.
- $\sigma(t, x) = \sigma\sqrt{x}$ with $\sigma \in \mathbb{R}^+$.

There could be an issue in the discretization scheme over a timegrid π due to the brownian increment term becoming negative. Therefore, we used in this dissertation a well known scheme for CIR process in the litterature called the Symmetrized Euler²² scheme given by :

$$X_{t_{n+1}}^\pi = |X_{t_n}^\pi + b(t_n, X_{t_n}^\pi)(t_{n+1} - t_n) + \sigma(t_n, X_{t_n}^\pi)(W_{t_{n+1}} - W_{t_n})|, \quad X_0^\pi = X_0. \quad (\text{B.4})$$

²¹For instance, the Milstein scheme is another discretization scheme which can improve the rate of convergence of X^π towards X .

²²There exists a lot of other scheme discussed in the litterature. For instance see [36].

C An introduction to neural networks

In this annex, we will make an overview of the most famous Deep Learning algorithm : *Neural Networks* asas we used them in some of the supervised learning algorithms we present in this dissertation. We will focus on the most common architecture of neural networks : the so called Feed Forward Neural Networks (FFNN) or Fully Connected Networks (FCN). There exists a lot of other architectures in the litterature such as :

- Convolutional Neural Networks (CNN) which are well suited for image processing
- Recurrent Neural Networks (RNN) which are used for sequentiel data like time series.

Presentation of FFNN

Architecture of a Neural Network

A feedforward neural network is a parametric function $f(\cdot, \theta) : \mathbb{R}^{N_0} \rightarrow \mathbb{R}^{N_M}$ defined as the composition of functions called layers and can be defined as follows :

$$f(\cdot, \theta) = L^{(M)} \circ \dots \circ L^{(1)}.$$

Each $(L^{(m)})_{m \in [1, M]}$ constitutes a layer and can be defined as follows :

$$\begin{aligned} L^{(m)} &: \mathbb{R}^{N_{m-1}} \rightarrow \mathbb{R}^{N_m} \\ x &\mapsto \phi_m(W^{(m)}x + b^{(m)}). \end{aligned}$$

with :

- $\phi_m : \mathbb{R} \rightarrow \mathbb{R}$ is called activation function of the layer m . Here in the definition, it's an ease of notation to define for a vector $x = (x_1, \dots, x_{N_m}) \in \mathbb{R}^{N_m}$, $\phi_m(x) = (\phi_m(x_1), \dots, \phi_m(x_{N_m}))$.
- $W^{(m)} \in \mathcal{M}_{N_m \times N_{m-1}}(\mathbb{R})$ is the weight matrix of the layer m .
- The vector $b^{(m)} \in \mathbb{R}^{N_m}$ is called the bias of the layer m .
- The dimension N_m defines the number of neurons of the layer m .
- The parameter θ describes all the parameters of the neural network , biais and weights such that $\theta = (W^{(1)}, b^{(1)}, \dots, W^{(M)}, b^{(M)})$ is a vector of dimension $\sum_{m=1}^M N_m(N_{m-1} + 1)$.

In this definition of a FFNN, there is one input layer $L^{(1)}$ which will be fed with the inputs of the user and there is one output layer $L^{(M)}$. Both dimensions are fixed before defining a FFNN. The other layers defined in the architecture are usually called hidden layers and help to connect the data between them. For example, a neural network with no hidden layer is called a *Perceptron* and is the easier architecture you can build for a neural network. In the case of a neural network with multiple layers, it's common to call it a *Multi Layer Perceptron*.

Remark. *The output dimension N_M can either be unidimensional or multidimensional depending on the problem. For instance, in this dissertation we used neural networks in both unidimensional and multidimensional settings.*

Let's focus a bit now on the activation functions $(\phi_m)_{m \in [1, M-1]}$ as they are an important part of the architecture of a neural network. We will cite some of the most used activations functions in the litterature.

- The identity function such as $\phi(x) = x$ which doesn't bring a nonlinearity.
- The sigmoid function such as $\phi(x) = \frac{1}{1+e^{-x}}$ used in classification problems as it returns a value between 0 and 1.
- The Hyperbolic tangent function such as $\phi(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$
- The ReLu function such as $\phi(x) = \max(x, 0) = x^+$ which is one of the most used activation functions in the literature but whose gradient is 0 for negative values.
- The LeakyRelu function of parameter α such as $\phi_\alpha(x) = x\mathbb{1}_{x>0} + \alpha x\mathbb{1}_{x\leq 0}$ which allows a gradient value non zero for negative values.

The use of neural networks has been increased tenfold in recent years because of power of computers as well as efficient ecosystems such as *Tensorflow* or *PyTorch* who can help us to create Neural Networks from scratch easily.

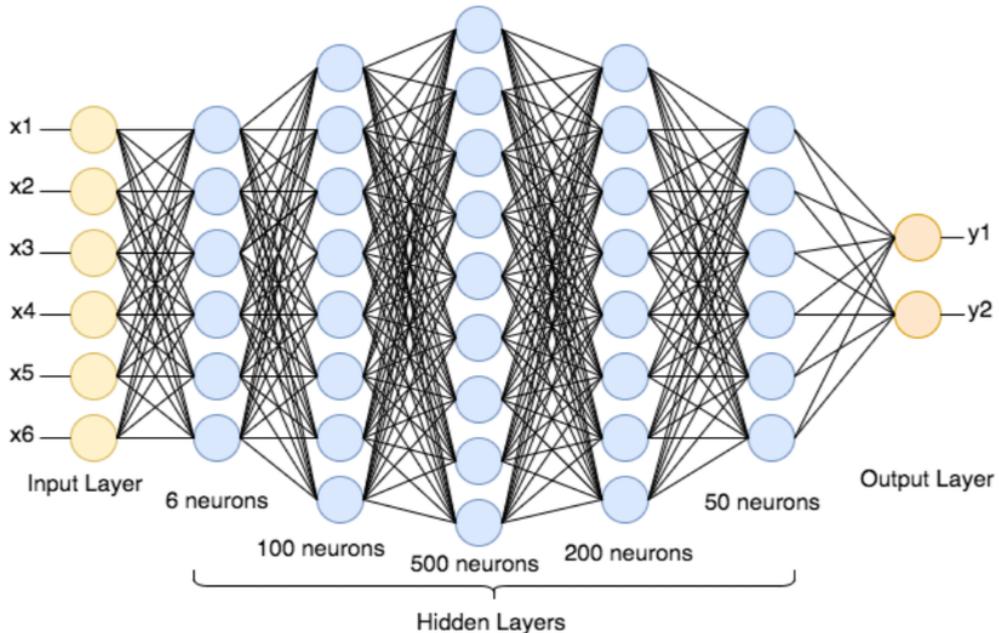


Figure C.1: Neural network architecture with an input vector $\in \mathbb{R}^6$ with 5 hidden layers and with an output layer $\in \mathbb{R}^2$.

The Universal Approximation Theorem

The main result which explains the popularity of neural networks is the universal approximation theorem. We will give a version of the theorem for the case of a bounded and smooth function $f : \mathbb{R}^d \rightarrow \mathbb{R}$ but the theorem can be extended to the multidimensional case.

Theorem C.1.²³ Consider ϕ a bounded function, continuous and non increasing function. Consider K a compact in the space \mathbb{R}^d and $\mathcal{C}(K)$ the space of the continuous functions on K . Let $f \in \mathcal{C}(K)$. Then $\forall \epsilon > 0$, there exists $N \in \mathbb{N}$ real numbers a_i , b_i and \mathbb{R}^d -vectors w_i such that by defining the function F as follows :

$$F(x) = \sum_{i=1}^N a_i \phi(\langle w_i, x \rangle + b_i).$$

²³The proof of this theorem can be found in [37] *Multilayer feedforward networks are universal approximators* by Kur' Hornik published in 1988

we have :

$$\sup_{x \in K} \|F(x) - f(x)\| < \epsilon.$$

Some remarks on the theorem :

From a theoretical point of view, this result is particularly interesting because it says that any continuous function f can be approximated by a neural network with one hidden layer with a finite number of neurons with the same activation function ϕ .

From a practical point of view, this result lacks of interest because it doesn't provide any idea about the number of neurons we have to consider to approximate f with a certain level of confidence ϵ . In practice, as the number N of neurons can be very large, it's preferable to consider multi-layer neural networks.

The Training process of a NN

The training process of a neural network is the most crucial step if we aim to find the optimal parameter θ to achieve a good accuracy. As always in a supervised machine learning algorithm, we will assume that we are able to sample $(X_i, Y_i)_{i \in \mathbb{N}}$ independant copies of an unknown distribution μ who generates (X, Y) . The aim will be to minimize a criterion or a loss function called l over the parameter θ which leads to the following minimization problem.

$$\min_{\theta} \mathbb{E}_{(X,Y) \sim \mu}[l(f(X, \theta), Y)]. \quad (\text{C.1})$$

As, μ is unknown, we will approximate the quantity $\mathbb{E}_{(X,Y) \sim \mu}[l(f(X, \theta), Y)]$ by his empirical counterparty assuming being able to have N samples $(X_i, Y_i)_{i \in [1, N]}$ following the law of large numbers.

$$\mathbb{E}_{(X,Y) \sim \mu}[l(f(X, \theta), Y)] \approx \frac{1}{N} \sum_{i=1}^N l(f(X_i, \theta), Y_i).$$

As for every supervised learning algorithm, the entire dataset can be divided either in 2 ways:

- Training, validation and test data if you need to perform hyperparameters tuning and/or model selection.
- Training and test data otherwise.

As we won't focus on hyperparameter tuning in this dissertation, we won't use a validation test and we will refer equivalently to validation test as the test data in all this dissertation.

The training dataset which represents usually 80% of the entire dataset where we will try to learn the optimal θ which minimizes our problem and the remaining 20% for the test data on which we will test the accuracy of the model.

The loss l defines the type of metric we have to choose in order to check the accuracy of the model. We give below the main metrics we used in this dissertation as the most common loss we deal with is the *Mean Squared Error (MSE)*.

- *MSE* or *RMSE* defined as $MSE(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$ and $RMSE(y, \hat{y}) = \sqrt{MSE(y, \hat{y})}$.
- *MAE* defined $MAE(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$.

where if we note $\hat{\theta}$ as the actual solution of our minimization problem, we then have $\hat{y}_i = f(x_i, \hat{\theta})$ where f is the neural network we defined from the previous section and y_i is the actual value.

Initialization of parameters of the NN

During the training of the neural networks, the vector θ needs to be initialized and then updated at each step of the training process. It has been seen in the litterature that this initialization is particularly important and can lead to problem during the training phase as the gradient can either vanish or explode. Here are the most common techniques to initialize the parameters of the NN :

- Uniform or Normal Initialization.
- Xavier/Goriot Initialization.

In the Xavier/Goriot Initialization, each weight $(W^{(m)})_{m \in [1, M]}$ from each layer m is assumed to follow the following distribution $W^{(m)} \sim \mathcal{N}(0, \frac{2}{N_{m-1} + N_m})$ and $b^{(m)}$ is initialized at 0. We used mostly the Xavier/Goriot initialization in the next chapters of this report as it is the most common practice allowing to avoid vanishing or exploding gradients in the training process.

The learning algorithm : Gradient Descent

As we said we aim to find θ which solves our minimization problem, we will refer to the classical gradient descent algorithm which is the way most popular algorithm to find minima of functions.

Here is below the algorithm for the classical gradient descent algorithm.

Algorithm C.1 Gradient descent algorithm

Input Parameters : θ_0 defined from the previous section, ϵ tolerance threshold, sequence $(\rho_n)_{n \in \mathbb{N}}$ of learning rates, $k = 0$

```

while  $\|\nabla_\theta l(x, y, \theta_k)\|^2 \geq \epsilon$  do
     $\theta_{k+1} = \theta_k - \rho_k \nabla_\theta l(x, y, \theta_k)$ 
     $k = k + 1$ 
end

```

Output Parameters : Parameter θ optimized

As we can see from the algorithm, it is required to calculate the gradient of the loss with respect to θ but also to define an appropriate learning rate sequence $(\rho_n)_{n \in \mathbb{N}}$. Indeed, the learning rate is a particularly important hyperparameter in the Gradient Descent algorithm as it can either not converge or even diverge. Let's illustrate the concept on the figure below.

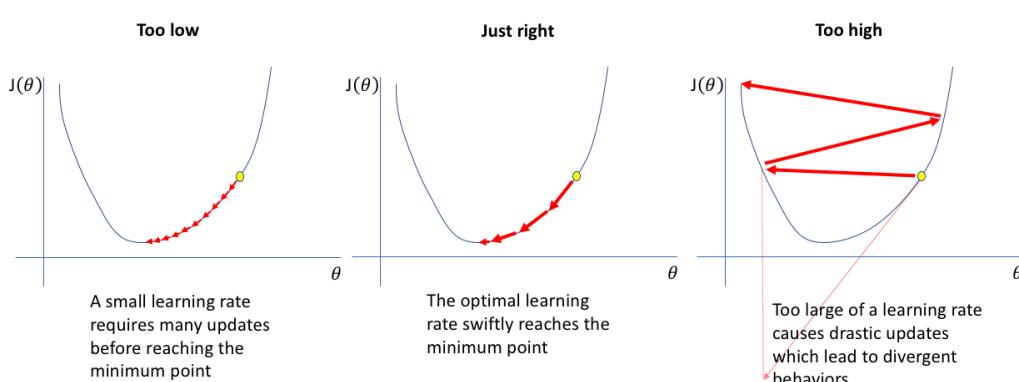


Figure C.2: Learning rate impact in the gradient descent algorithm

As we can see from Figure C.2, choosing an appropriate learning rate sequence is particularly important. If it's "*too high*", we see that it will fail to converge to the actual global minimum to the function. But if it's "*too low*", the algorithm looks unefficient as it will take too much steps to actually converge and it can also end stuck in a local minima.

As we said, we also need to calculate the gradient of the loss with respect to θ at each timestep of the algorithm. However, in practice, the Gradient Descent Algorithm is slightly modified in the way they compute the actual gradient. We give the most popular methods below :

- **Batch Gradient Descent :** In this version of the Gradient Descent algorithm the gradient $\nabla_{\theta} l(x, y, \theta_k)$ is calculated over all the training samples as follows where N_{Train} defines the number of samples we have in the training dataset.

$$\nabla_{\theta} l(x, y, \theta_k) = \frac{1}{N_{Train}} \nabla_{\theta} \sum_{i=1}^{N_{Train}} l(x_i, y_i, \theta_k).$$

As we can see, this method is computationally intense as at each step, we need to calculate the gradient over all the training dataset and this is not the choice which is done in practice when N_{Train} becomes too big.

- **Stochastic Gradient Descent :** In this version of the Gradient Descent algorithm, the gradient $\nabla_{\theta} l(x, y, \theta_k)$ is calculated over the gradient on one unique sample from the training dataset choosen randomly from the data at the timestep k . Let's note by (x_k, y_k) the choosen sample, then the gradient is calculated as follows :

$$\nabla_{\theta} l(x, y, \theta_k) = \nabla_{\theta} l(x_k, y_k, \theta_k).$$

The convergence of the Stochastic Gradient Descent can look surprisingly as it requires only one sample at each iteration of the Gradient Descent but it has been show that under convex minimization that it converges almost surely to a local minimum under a suitable choice of the learning sequence rate $(\rho_n)_{n \in \mathbb{N}}$.

The main advantage of this algorithm is his speedness and his low memory required but his main disadvantage is the accuracy of the actual minimizor as the gradient is calculated over one random sample from the entire dataset at each step of the algorithm.

- **Mini Batch Gradient Descent :** The Mini Batch Gradient Descent is a mix between Batch Gradient Descent and Stochastic Gradient Descent. Indeed, we define a *batch-size* $1 < M < N_{Train}$ which defines the number of samples on which the gradient will be performed at each step of the algorithm. The gradient is therefore calculated as follows :

$$\nabla_{\theta} l(x, y, \theta_k) = \frac{1}{M} \nabla_{\theta} \sum_{i=1}^M l(x_i, y_i, \theta_k).$$

This method is the most used in practice as it is a good comprise between a good accuracy given by the Batch Gradient Descent method and the computationally efficiency given by the Stochastic Gradient Descent.

However, in practice there is not an optimal batch-size M . It will highly depend from the dataset structure and the neural network architecture.

In the implementation, we used the Adam optimizer which is currently the most used optimizer and who is seen has an improved version of the classic SGD. Indeed, the Adam Optimizer allows to maintain a learning rate for each network parameter who is separately updated during the learning process. This learning rate is estimated following an estimation of the 2 first moments of the gradient. The parameters are updated as follows :

$$\begin{aligned}
g_k &= \nabla_{\theta} l(x, y, \theta_k) \\
m_{k+1} &= \beta_1 m_k + (1 - \beta_1) g_k \\
v_{k+1} &= \beta_2 v_k + (1 - \beta_2) g_k^2 \\
\theta_{k+1} &= \theta_k - \alpha \frac{m_k}{1 - \beta_1^{k+1}} \frac{1}{\sqrt{\frac{v_k}{1 - \beta_2^{k+1}} + \epsilon}}.
\end{aligned}$$

with $(\beta_1, \beta_2) \in [0, 1]^2$, ϵ and α which are default parameters and initialization $m_0 = v_0 = 0$. Classical values for $\beta_1, \beta_2, \epsilon$ and α are respectively 0.9, 0.999, 10^{-8} and 0.001.

The backpropagation principle

We will now focus on the actual way of how gradients are calculated to be applied in the Gradient Descent Algorithm. As we said, we need to be able to calculate the gradient of the loss function over the parameter $\theta = (W^{(1)}, b^{(1)}, \dots, W^{(M)}, b^{(M)})$.

As we can see, it can be really challenging to calculate the gradient of the loss function over each parameter of the NN but the idea behind neural networks is to use the chain rule to efficiently calculate the gradients. This approach allows in only one pass from the output layer to the input layer to compute the gradients of the loss function of every parameter and that's why neural networks can be efficiently used. If gradients of each parameter were calculated separately, it would make the optimization problem impossible in practice.²⁴

²⁴See <https://web.stanford.edu/class/cs224n/readings/gradient-notes.pdf> which explains for a simple neural network architecture how gradients are calculated efficiently using backpropagation principle.

D Another deep PDE solver : The Deep Galerkin algorithm

The Deep Galerkin method has been introduced in [24] in 2018 by Sirignano and Spiliopoulos in their publication *DGM: A deep learning algorithm for solving partial differential equations* in order to solve partial differential equations in a general setting. Let's consider u an unknown function defined on the region $[0, T] \times \Omega$ where $\Omega \subset \mathbb{R}^d$ with $d \in \mathbb{N}^*$. We assume that u is solution of the following PDE with \mathcal{L} a generator :

$$\begin{aligned} (\partial_t + \mathcal{L})u(t, x) + f(t, x) &= 0 \quad \forall (t, x) \in [0, T] \times \Omega. \\ u(T, x) &= g(x) \quad \forall x \in \Omega. \end{aligned}$$

The idea of the Deep Galerkin is to approximate the function u by a neural network parametrized by θ namely u^θ such that we have $u^\theta \approx u$ by trying u^θ by "forcing" u to verify the PDE of which he is solution. For this, we will define the following :

- 2 random variables τ and \mathcal{X} which are respectively defined on $[0, T]$ and Ω .
- We will have one error term L_θ^1 in the loss which characterizes the fact that u^θ has to verified the PDE on $[0, T] \times \Omega$
- We will have a second error term L_θ^2 in the loss which characterizes the terminal condition for u^θ

Therefore, we can define L_θ^1 and L_θ^2 as follows :

$$\begin{aligned} L^1(\theta) &= \mathbb{E}[(\partial_t + \mathcal{L})u^\theta(\tau, \mathcal{X}) + f(\tau, \mathcal{X})^2]. \\ L^2(\theta) &= \mathbb{E}[(u^\theta(T, \mathcal{X}))^2]. \end{aligned}$$

Therefore, we will look for the following optimisation problem where $L^\theta = L^1(\theta) + L^2(\theta)$

$$\theta^* = \operatorname{argmin}_\theta L^\theta. \tag{D.1}$$

Some remarks on the Deep Galerkin algorithm :

- For the Deep Galerkin, the choice of the sampling of τ and \mathcal{X} is particularly crucial. Indeed, as the learning process is based of the samples of τ and \mathcal{X} , if we want the neural network to approximate as much as possible the solution of the *PDE*, we therefore need to choice adapted distributions for τ and \mathcal{X} .
- The Deep Galerkin is well suited for high dimensional problems as the learning process goes from random sampling of (τ, \mathcal{X}) . It can therefore overcome the classic *curse of dimensionnality* issue when we are dealing with finite difference methods when the *meshgrid* grows exponentially with the dimension of the risk factors.
- In the following, we will just focus on computing *CVA* and *FVA* for financial contracts on one underlying risk factor and illustrate how this method could possibly be used for *XVA* computations. We will also show how the method can be used for an efficient derivatives pricing setup.

Numerical Results :

We now illustrate the use of the Deep Galerkin Algorithm in the pricing of the *CVA* on an european call and *FVA* profile for a forward contract under $B - S$ and we will rely on the same *PDEs* as we derived in (6.12) for *CVA* and *FVA*.

For the sampling choice of (τ, \mathcal{X}) , we decided to sample the distribution on an uniform meshgrid for both variables as we are in a low dimensionnal setting with 100 points for τ on $[0, T]$ and 100 points for \mathcal{X} on $[20, 200]$. For convenience, we rewrite below the associated PDEs.

$$\begin{aligned} \partial_t \phi^{CVA}(t, x) + \mathcal{L} \phi^{CVA}(t, x) - (r_t + \lambda_t^C) \phi^{CVA}(t, x) + (1 - R_t^C)(V_t)^+ \lambda_t^C = 0, \quad \forall (t, x) \in [0, T] \times \mathbb{R}_*^+ \\ \phi^{CVA}(T, .) = 0, \quad \forall x \in \mathbb{R}_*^+ \end{aligned}$$

$$\begin{aligned} \partial_t \phi^{FVA}(t, x) + \mathcal{L} \phi^{FVA}(t, x) - (r_t + \lambda_t^C + \lambda_t^A) \phi^{FVA}(t, x) + (s_t^b(V_t)^+ - s_t^L(V_t)^-) = 0, \quad \forall (t, x) \in [0, T] \times \mathbb{R}_*^+ \\ \phi^{FVA}(T, .) = 0, \quad \forall x \in \mathbb{R}_*^+ \end{aligned}$$

We will assume a $B - S$ model and we give in the table below the parameters used in the numerical experiments:

Table D.1: Parameters used in the *Deep Galerkin Method* illustrations

Parameters	r	σ	K
Value	0.02	0.2	100

Learning the CVA profile of an european call with $\lambda^C = 0.4$ and $R^C = 0$:

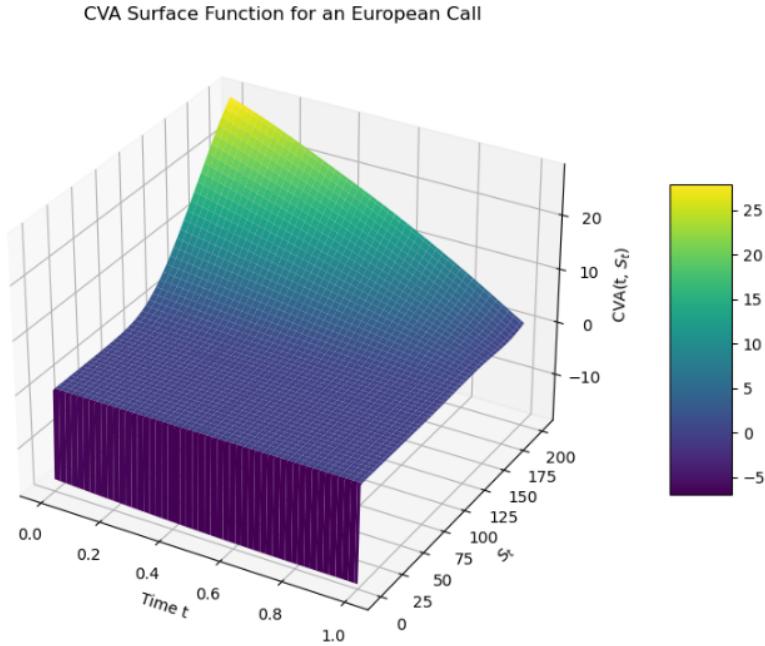


Figure D.1: $CVA(t, S_t)$ surface for an european call with the following parameters : ($\lambda^C = 0.4$ and $R^C = 0$)

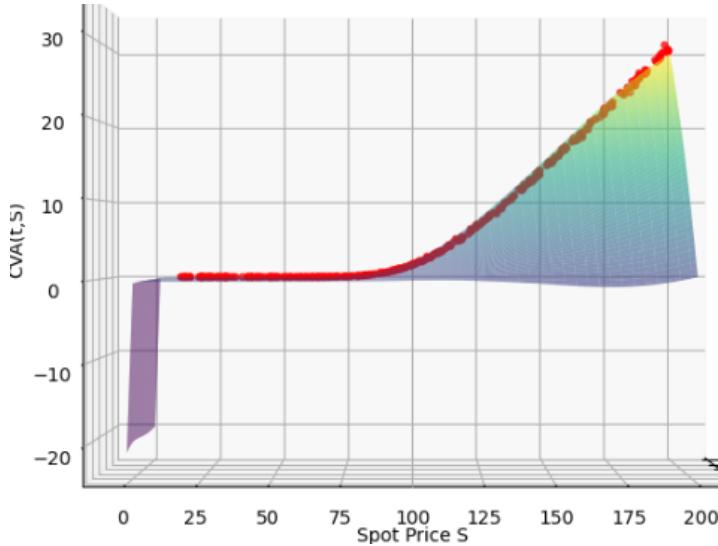


Figure D.2: Projection of the CVA surface of an european call on $t = 0$ with in red line the $M - C$ estimation of $CVA(0, S_0)$ with the following parameters : ($\lambda^C = 0.4$ and $R^C = 0$)

Some remarks on the result :

- As we can see from the red line which represents the $M - C$ estimation , the accuracy of the Deep Galerkin is quite good as it replicates the value on $t = 0$.
- As we can observe from the CVA surface profile, there is quite an error on the terminal value of CVA which is supposed to be equal at 0 but overall the *Deep Galerkin* seems to perform well.

Learning the CVA profile of an european call with $\lambda^C = 0.1$ and $R^C = 0$:

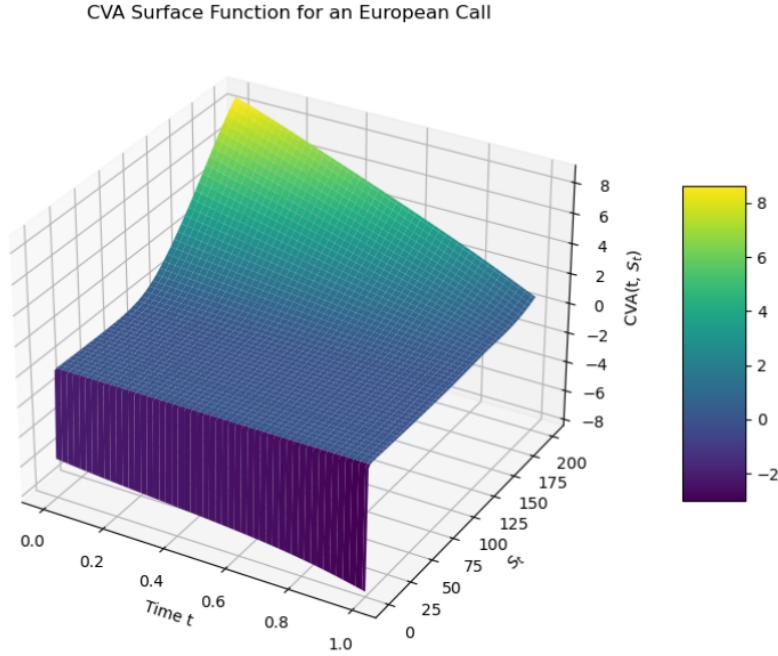


Figure D.3: $CVA(t, S_t)$ surface for an european call with the following parameters : ($\lambda^C = 0.1$ and $R^C = 0$)

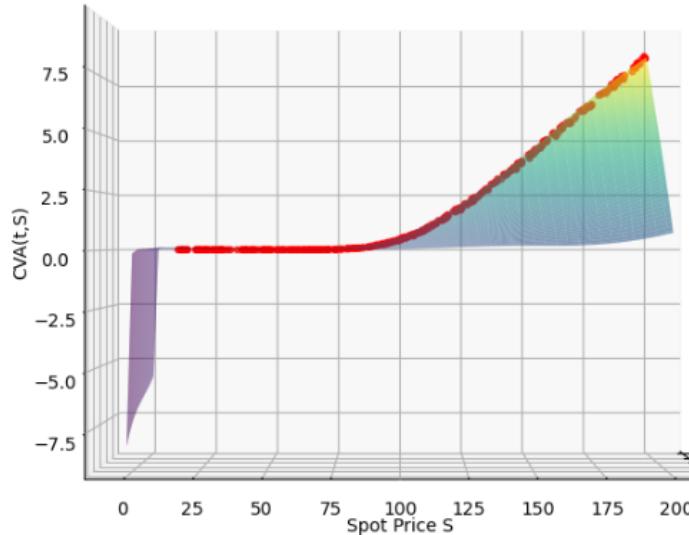


Figure D.4: Projection of the CVA surface of an european call on $t = 0$ with in red line the $M - C$ estimation of $CVA(0, S_0)$ with the following parameters : ($\lambda^C = 0.1$ and $R^C = 0$)

Some remarks on the result :

- We have globally the same conclusions than the previous example except that for the learning of $CVA(T, .)$ which seems to be better as the function seems to be really close to 0 in this case.
- We can also observe that when $\lambda^C = 0.1$, the overall CVA is reduced as the intensity default of the counterparty is lower than in the previous example.

Learning the FVA profile of a forward contract with $s^B = 0.02$ and $s^L = 0$:

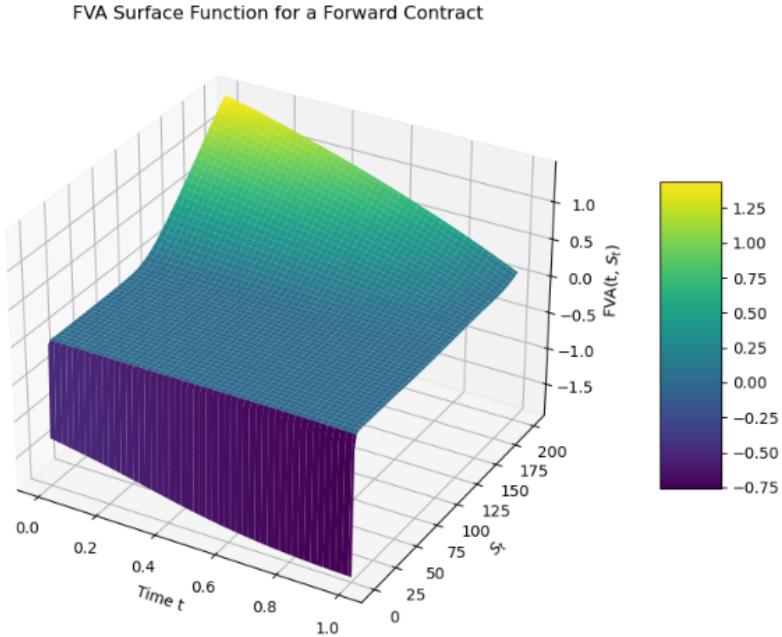


Figure D.5: $FVA(t, S_t)$ surface for a forward contract with the following parameters : ($s^B = 0.02$, $s^L = 0$, $\lambda^C = 0.4$ and $\lambda^A = 0.1$)

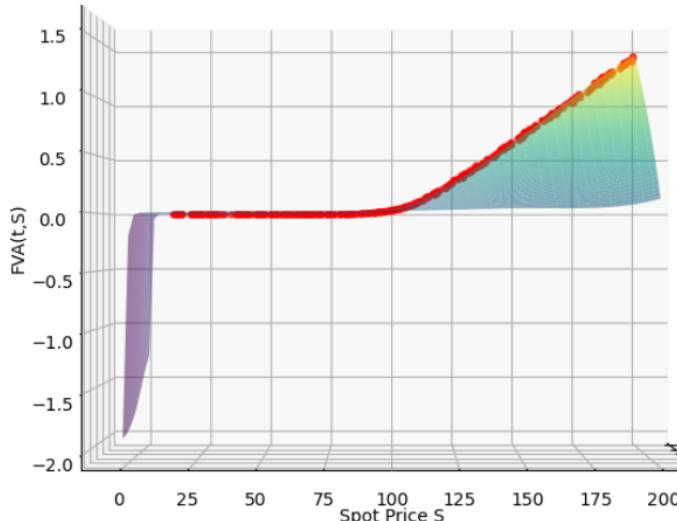


Figure D.6: Projection of the FVA surface of a forward contract on $t = 0$ with in red line the $M - C$ estimation of $FVA(0, S_0)$ with the following parameters : ($s^B = 0.02$, $s^L = 0$, $\lambda^C = 0.4$ and $\lambda^A = 0.1$)

Some remarks on the result :

- In the case of the FVA surface learning, we have similar results when we set $s^L = 0$ which means that $FBA = 0$ so we expect a similar profile than to the CVA which is something that we end up with. The red line Monte-Carlo allows to assess the learning of the FVA surface for the forward contract.

Learning the FVA profile of a forward contract with $s^B = 0.02$ and $s^L = 0.02$:

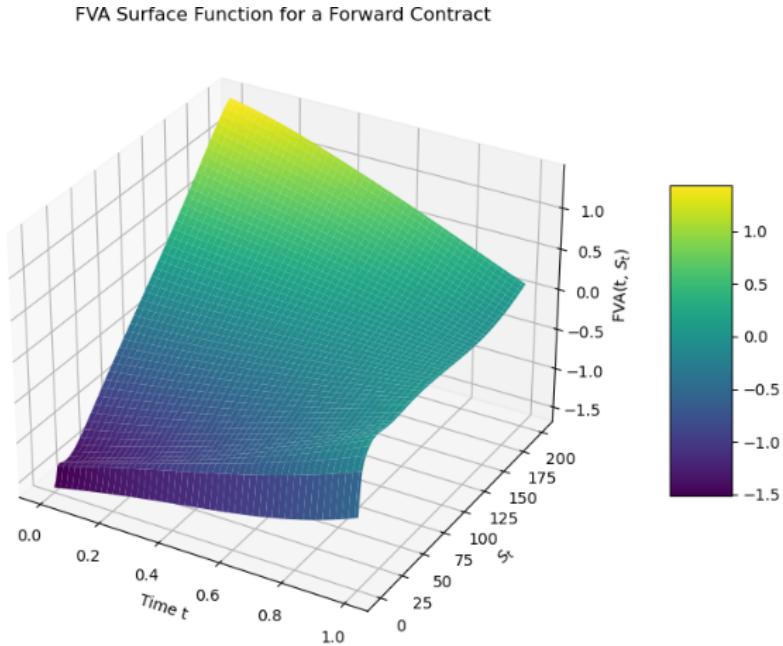


Figure D.7: $FVA(t, S_t)$ Surface for a forward contract with the following parameters : ($s^B = 0.02$, $s^L = 0.02$, $\lambda^C = 0.4$ and $\lambda^A = 0.1$)

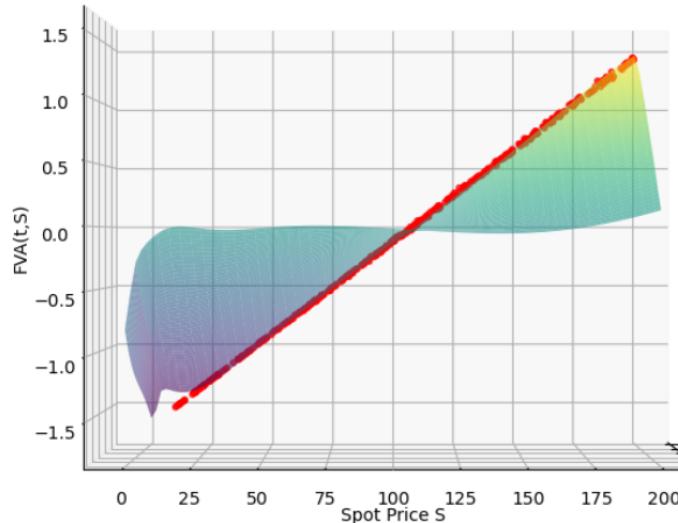


Figure D.8: Projection of the FVA surface of a forward contract on $t = 0$ with in red line the MC estimation of $FVA(0, S_0)$ with the following parameters : ($s^B = 0.02$, $s^L = 0.02$, $\lambda^C = 0.4$ and $\lambda^A = 0.1$)

Some remarks on the result and the method :

- In this setting, we set $s^L = 0.02$ such that $FVA \neq 0$ and therefore we can have negative FVA . We see in this case that the learning of the FVA is not that good with some error for low values of S as we can see from the red line $M - C$ estimation. Moreover, the value for terminal T is close to 0 but it can be improved.
- In a PDE approach, the *Deep Galerkin Method* can be a good alternative to the *Deep XVA Solver* as it is a more straightforward approach and easier but the accuracy of the neural network solution is still a challenge that needs to be improved.

Learning the surface profile of an european call under $B - S$:

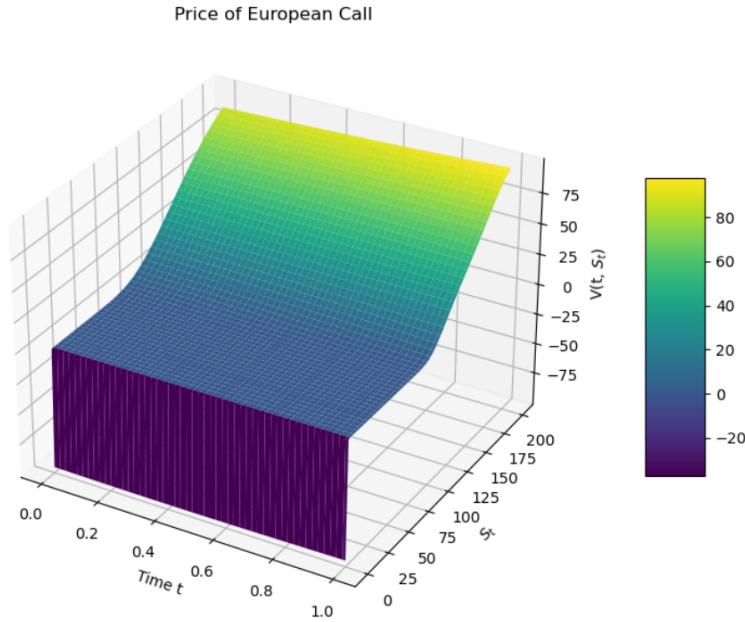


Figure D.9: $V(t, S_t)$ price surface for an european call

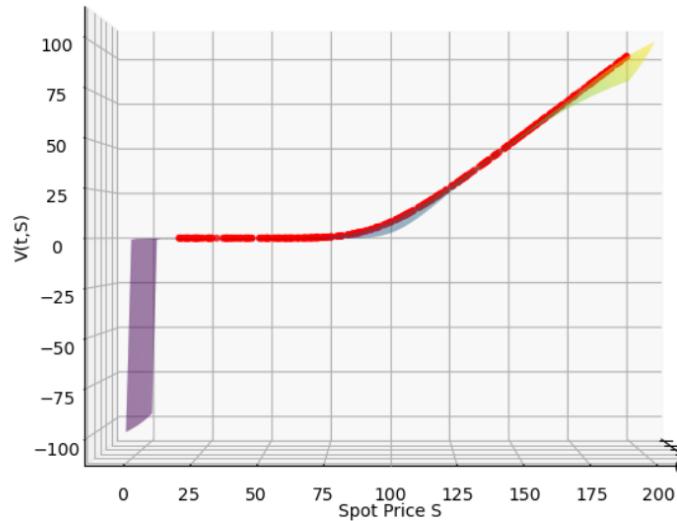


Figure D.10: Projection of the price surface of an european call on $t = 0$ with in red line the MC estimation of $V(0, S_0)$

Some remarks on the result :

- In this setting, we see that the Deep Galerkin performs really well in reproducing the price surface of the european call $V(t, S_t)$ with the red line representing the $M - C$ price being really close to the estimate of the Deep Galerkin.
- The Deep Galerkin looks to be a good candidate for the first loop in the XVA nested $M - C$ computation as we can use the price of the derivative learned by the *Deep Galerkin* on the whole surface $(t, s) \in [0, T] \times D$ where D denotes here the interval $[20, 200]$.

Learning the surface profile of a forward contract under $B - S$:

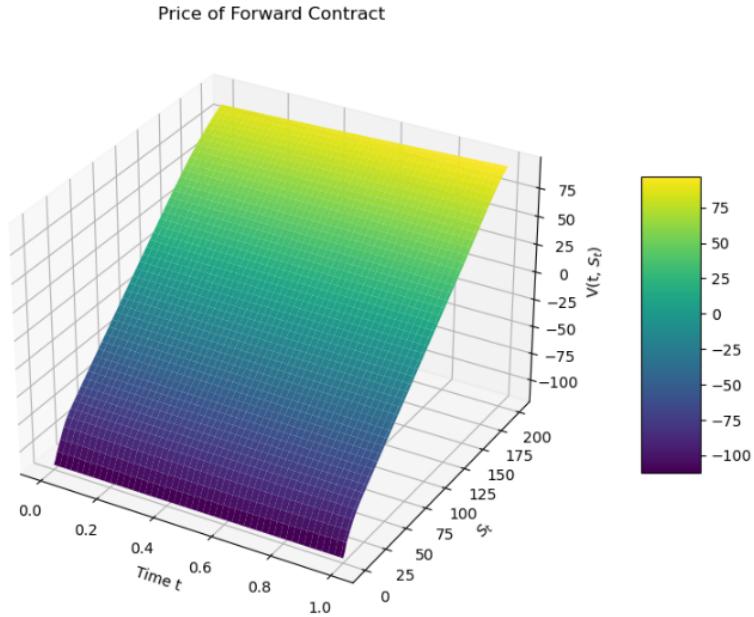


Figure D.11: $V(t, S_t)$ price surface for a forward contract

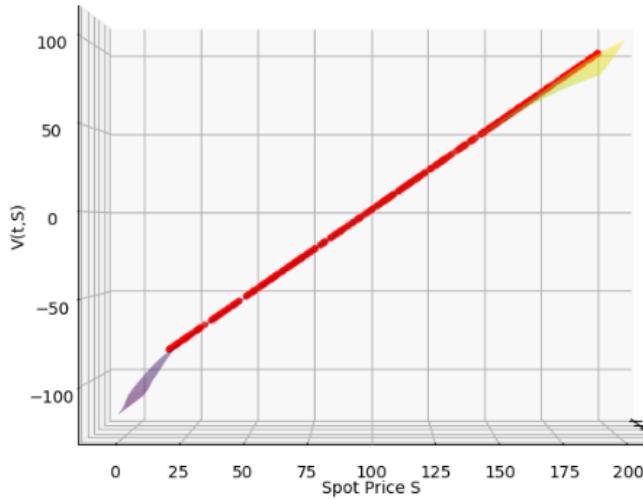


Figure D.12: Projection of the price surface of a forward contract on $t = 0$ with in red line the MC estimation of $V(0, S_0)$

Some remarks on the result :

- In this setting, we see that the Deep Galerkin performs really well in reproducing the price surface of the forward contract $V(t, S_t)$ with the red line representing the $M - C$ price being really close to the estimate of the Deep Galerkin. Moreover, we see from the price surface of the forward clearly the *Delta One* structure of the forward.
- The Deep Galerkin looks to be a good candidate for the first loop in the *XVA* nested $M - C$ computation as we can use the price of the derivative learned by the *Deep Galerkin* on the whole surface $(t, s) \in [0, T] \times D$ where D denotes here the interval $[20, 200]$.

E Additional plots

LSMC Plots for Bermudan Options

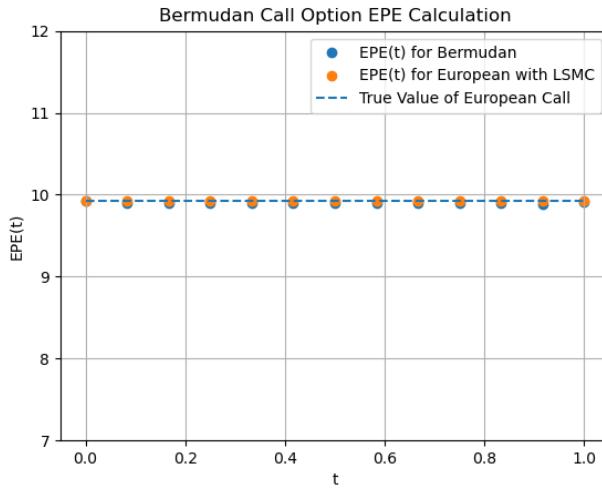


Figure E.1: Computation of the expected exposure profile of a bermudan call under $B - S$ model with the following parameters : ($S_0 = 100$, $K = 100$, $r = 0.04$, $q = 0$, $\sigma = 0.2$, $T = 1$ and $N = 13$)

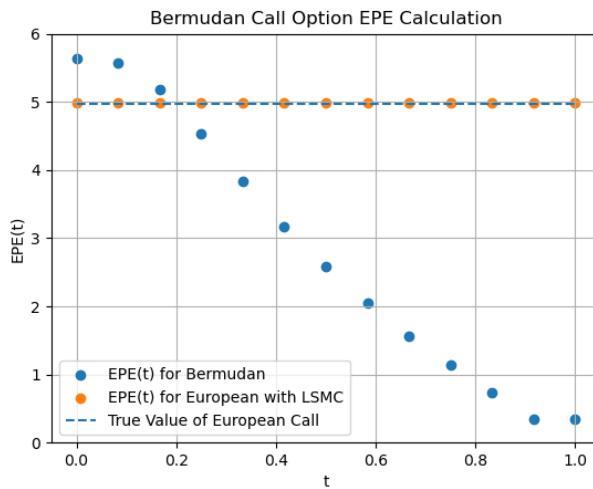


Figure E.2: Computation of the expected exposure profile of a bermudan call under $B - S$ model with dividends and with the following parameters : ($S_0 = 100$, $K = 100$, $r = 0.04$, $q = 0.10$, $\sigma = 0.2$, $T = 1$ and $N = 13$)

Some remarks on the result :

- We can see the impact of the dividend q on the EPE profile of a bermudan. As it is well known under the $B - S$ model, it's never optimal to exercise an american / bermudan call option when $r - q > 0$ as we can see from the figure above.
- Moreover, as we mentioned when $q = 0.10$, we see that the bermudan call option price is different from his european version and being superior at $t = 0$ which is expected.

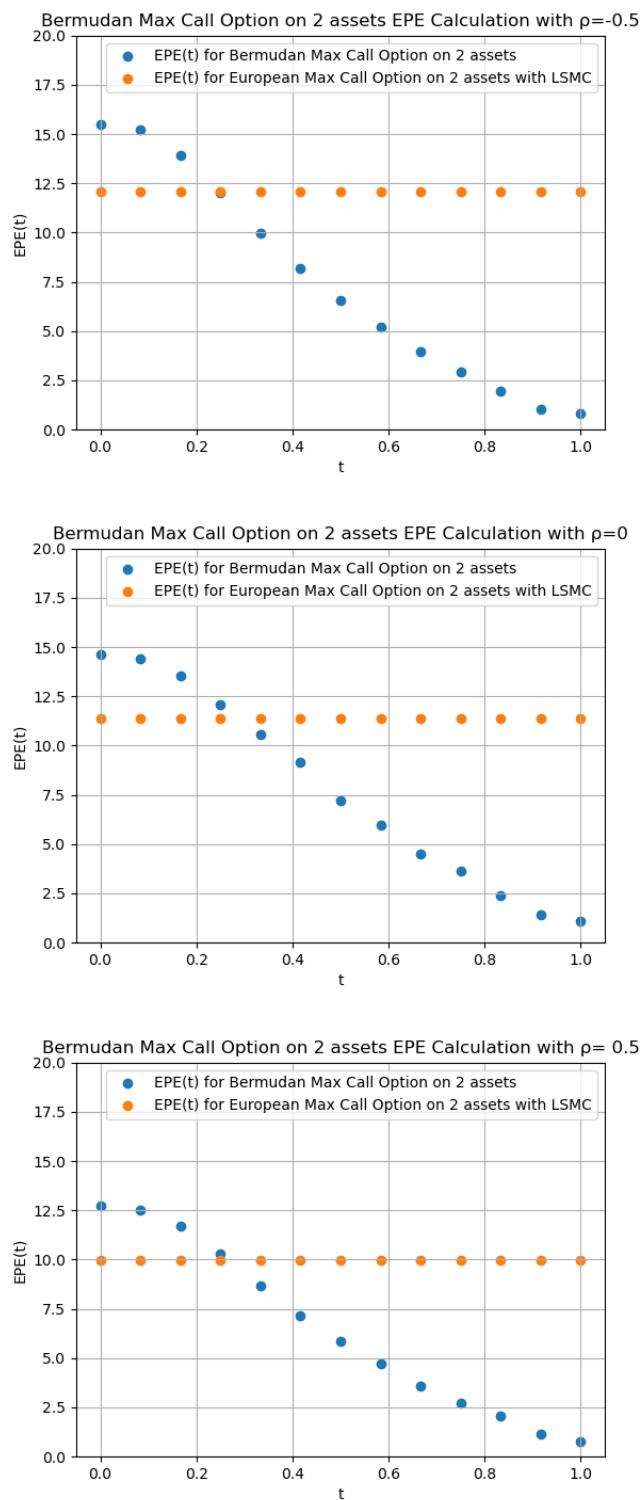


Figure E.3: Computation of the expected exposure profile of a bermudan max call on 2 assets under $B - S$ model for different ρ with dividends and with the following parameters : ($S_0^1 = S_0^2 = 100$, $K = 100$, $r = 0.04$, $q^1 = q^2 = 0.10$, $\sigma^1 = \sigma^2 = 0.2$, $T = 1$ and $N = 13$)

Some remarks on the result :

- As we can see from the plot above, the price of the max call option is a decreasing function of ρ which is expected from the intuition of the payoff structure $(\max_{i \in \{1,2\}} S_T^i - K)^+$
- We also recover that the price of the bermudan option is higher at $t = 0$.

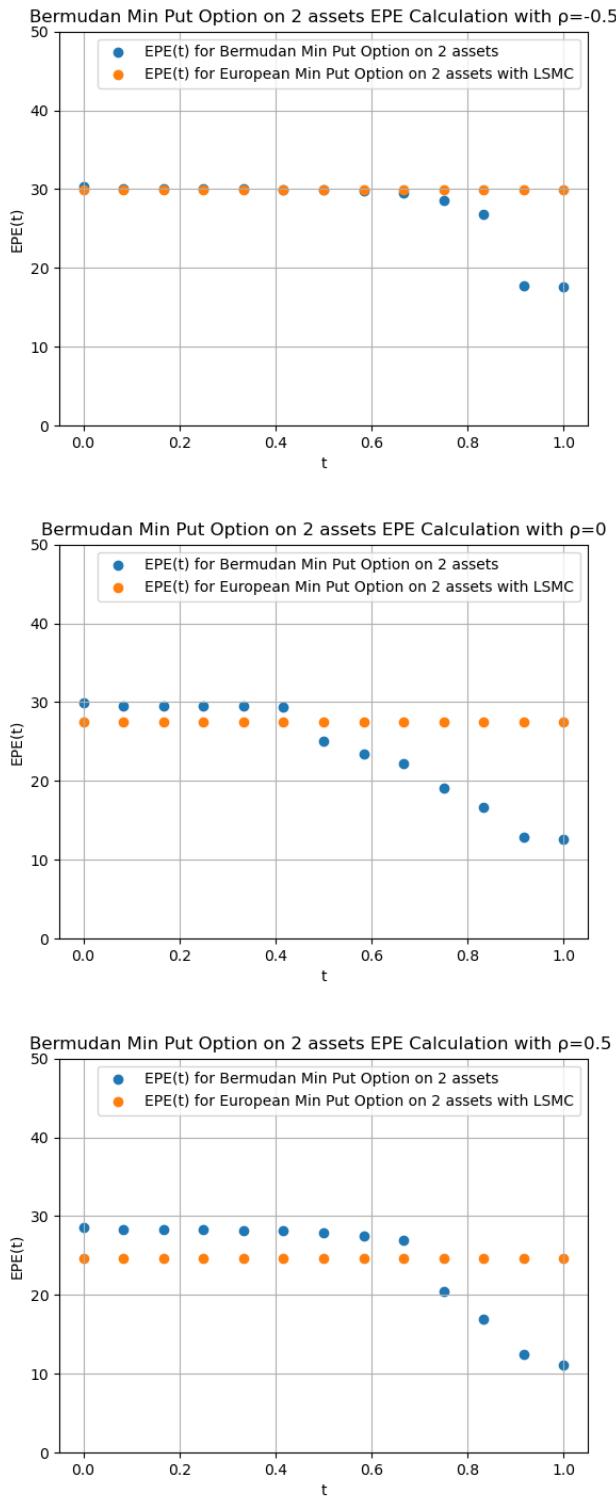


Figure E.4: Computation of the expected exposure profile of a bermudan min put option on 2 assets under $B - S$ model for different ρ with dividends and with the following parameters : $(S_0^1 = S_0^2 = 100, K = 100, r = 0.04, q^1 = q^2 = 0.10, \sigma^1 = \sigma^2 = 0.2, T = 1$ and $N = 13)$

Some remarks on the result :

- As we can see from the plot above, the price of the Min Put option is a decreasing function of ρ which is expected from the intuition of the payoff structure $(K - \min_{i \in \{1,2\}} S_T^i)^+$
- We also recover that the price of the bermudan option is higher at $t = 0$.

Neural networks learning process

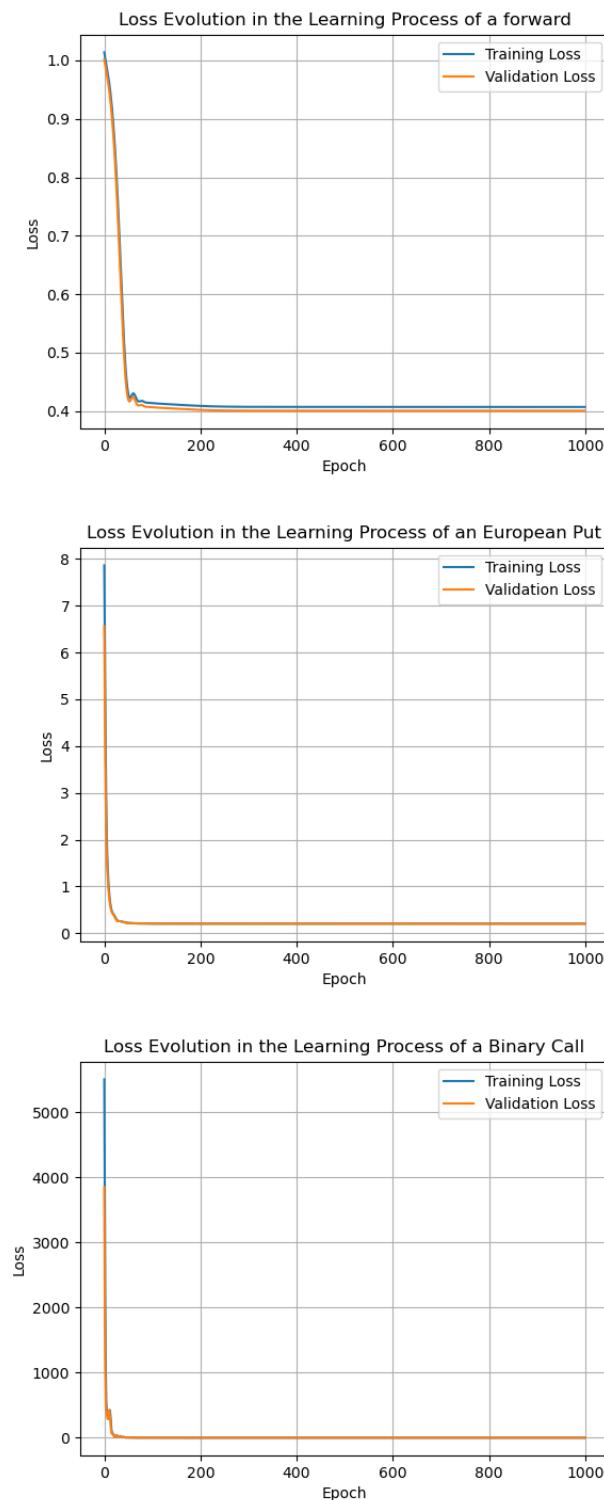


Figure E.5: Learning processes of neural networks for various european options from plots of chapter 5

Hedging strategies in an Heston model

We consider the case of a more sophisticated model to show that the hedging strategy still can be computed.²⁵ We consider the case of european call and put which we will assume to price using $2000 = MC$ Paths with $N = 100$ discretization time steps. We recall that the Heston model is given by the following dynamics :

$$\begin{aligned} dS_t &= S_t(rdt + \sqrt{v_t}dW_t^1), & S_0 \in \mathbb{R}_*^+ \\ dv_t &= \kappa(\theta - v_t)dt + \sigma\sqrt{v_t}dW_t^2, & v_0 \in \mathbb{R}_*^+ \\ d\langle W^1, W^2 \rangle_t &= \rho dt. \end{aligned}$$

We give in the table below the parameters used in the numerical experiments for both cases.

Table E.1: Parameters used in the numerical experiments in the *Mean Variance Minimizing strategy* in the Heston model

Parameters	ξ	λ	r	κ	θ	σ	S_0	v_0	ρ	K
Case 1	0.5	0.2	0	0.5	0.15	0.02	100	0.2	-0.9	100
Case 2	0.5	0.2	0	0.5	0.15	0.02	100	0.1	-0.9	100

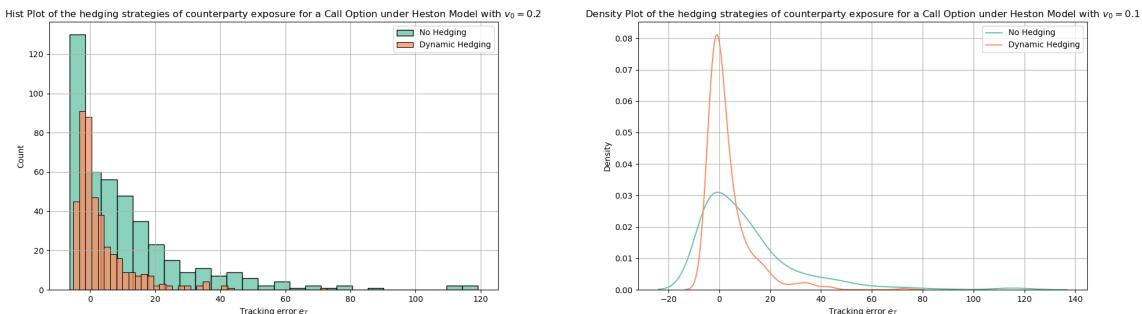


Figure E.6: Comparison of 2 hedging strategies in order to hedge the CCR on a call option under Heston model in Case 1

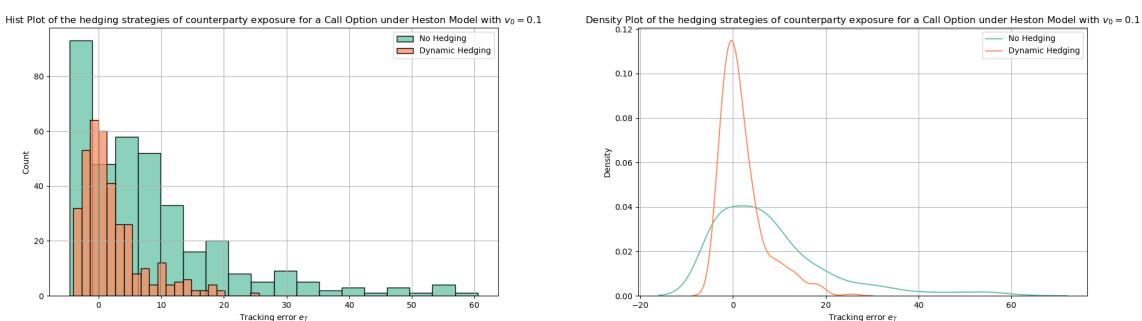


Figure E.7: Comparison of 2 hedging strategies in order to hedge the CCR on a call option under Heston model in Case 2

²⁵Doing the computation shows that the optimal hedging strategy as the same form in the Heston Model than in the $B - S$ version. Therefore, we only need to price the financial product under Heston to compute the optimal strategy

Table E.2: Norm 2 of e_T in case of an european call in the Heston model

	No Hedging	Dynamic Hedging
$\mathbb{E}[(e_T)^2]$ in Case 1	501.64	89.19
$\mathbb{E}[(e_T)^2]$ in Case 2	212.98	28.26

We also provide some numerical results in the case of an european put under the Heston model

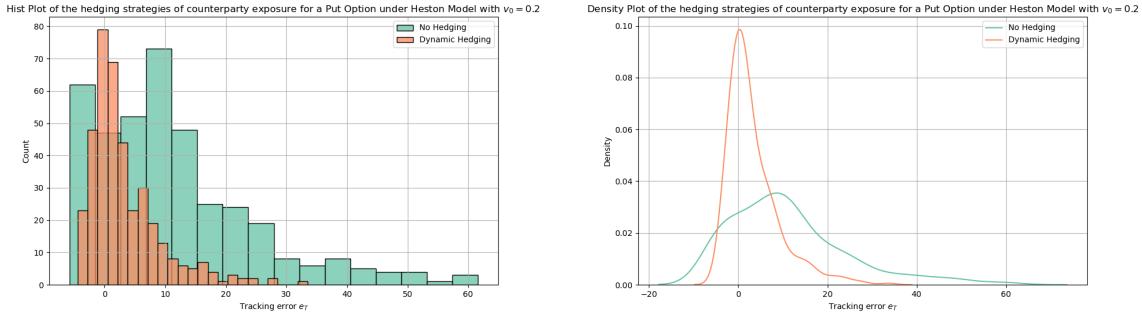


Figure E.8: Comparison of 2 hedging strategies in order to hedge the CCR on a put option under Heston model in Case 1

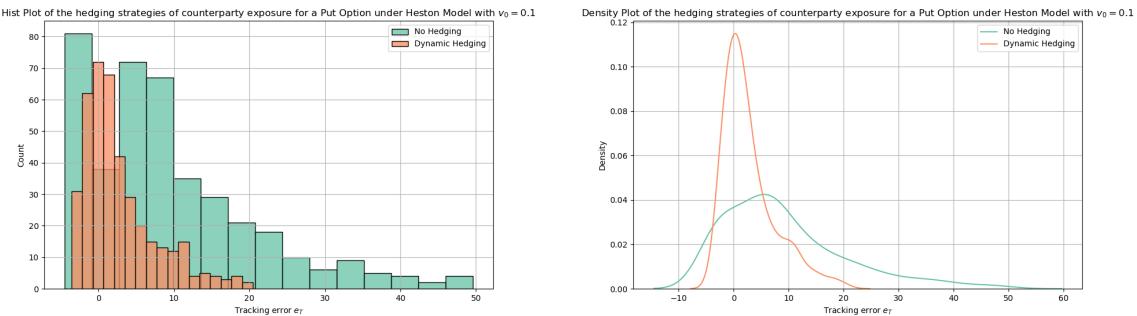


Figure E.9: Comparison of 2 hedging strategies in order to hedge the CCR on a put option under Heston model in Case 2

Table E.3: Norm 2 of e_T in case of an european put in the Heston model

	No Hedging	Dynamic Hedging
$\mathbb{E}[(e_T)^2]$ in Case 1	299.00	45.94
$\mathbb{E}[(e_T)^2]$ in Case 2	206.58	30.32

Some remarks on the results :

- As we see also when the market is more stressed corresponding to the first case with $v_0 = 0.2$, the tracking error is overall superior.
- We see that the tracking error e_T is globally higher in the Heston model than in the $B - S$ model which can be explained due to the fact that we priced the options using MC paths and not analytical formulas and also due to the relatively low number of MC paths used. Moreover, we choose a value of θ which is the mean reversion equal to 0.15 which is relatively high so we expected more error in the Heston model.