

Metamodeling de l'Initial Margin/Margin Value Adjustment

Adrien Misko, Amine Zarbi

Nexialog Consulting

Paris, France

16 mars 2021



Résumé

L'Initial Margin pour les dérivés OTC non compensés a pris de plus en plus d'importance suite aux exigences de la BCBS-IOIS. Cet appel de marge a pour but de réduire l'exposition au risque de marché pendant la période nécessaire pour la liquidation/remplacement du portefeuille négocié suivant le défaut d'une contrepartie. Pour intégrer l'IM dans les calculs des xVAs, il est nécessaire de simuler une **IM Forward Dynamique** tout au long de la durée du portefeuille. Cependant, l'approche directe pour cette estimation est onéreuse en matière de temps de calcul, surtout pour des portefeuilles complexes et assez larges. On s'intéressera dans cette étude à une estimation en temps raisonnable de l'IM dynamique (Discounted Dynamic Initial Margin) via metamodeling. Ainsi, nous proposerons une adaptation du Least Square Monte Carlo et de la régression par processus gaussiens dans le cadre de l'estimation de quantiles. La performance de ces approches sera évaluée par rapport au calcul direct par Monte Carlo.

Keywords— Initial Margin, Margin Valuation Adjustment, Metamodeling, Least Square Monte Carlo, Gaussian Processes

Table des matières

| | |
|---|-----------|
| Introduction | 4 |
| 1 Préliminaires | 5 |
| 2 Calcul de l'IM | 5 |
| 2.1 Modèle interne | 5 |
| 2.2 Méthode standard : SIMM | 6 |
| 3 Calcul de la MVA | 8 |
| 3.1 Présentation | 8 |
| 3.2 Probabilité historique VS Mesure annuité : Coefficient de scaling | 10 |
| 3.3 Calcul de la DDIM | 10 |
| 3.4 Complexité Algorithmique du calcul direct | 12 |
| 4 Metamodeling de l'IM/MVA | 12 |
| 4.1 Least Square Monte Carlo | 12 |
| 4.1.1 Adaptation de la LSMC au calcul de la DDIM | 12 |
| 4.1.2 Choix des régresseurs | 13 |
| 4.1.3 Algorithme LSMC pour la DDIM | 13 |
| 4.1.4 Complexité algorithmique | 14 |
| 4.1.5 Limitations | 14 |
| 4.2 Régression par GPs | 14 |
| 4.2.1 Présentation | 14 |
| 4.2.2 Application des GPs au calcul de la MVA | 16 |
| 4.2.3 MC-GP : Complexité Algorithmique | 19 |
| 4.2.4 Limitations | 19 |
| 5 Implémentation et résultats numériques | 19 |
| 5.1 Calcul de l'IM | 19 |
| 5.2 Calcul de la MVA | 20 |
| 5.2.1 Calcul de la MVA par méthode directe (MC²) | 20 |
| 5.2.2 Calcul de la MVA par Least Square Monte Carlo | 21 |
| 5.2.3 Calcul de la MVA par MC-GP | 22 |
| 5.2.4 Comparaison et interprétation des résultats | 23 |
| 5.2.5 Estimation des moments par Processus Gaussien Stochastique | 25 |
| 6 Conclusion et améliorations possibles | 26 |
| 7 Annexes | 29 |
| 7.1 Annexe A : Généralités sur le marché des taux, mesures de probabilités et formules de Pricing | 29 |
| 7.1.1 Généralités | 29 |
| 7.1.2 Changement de numéraire et mesures de probabilité | 31 |
| 7.1.3 Formule de Pricing Swap et Swaption Vanille | 32 |
| 7.2 Annexe B : Modélisation stochastique des taux swap | 33 |
| 7.2.1 Modèle de Bachelier | 33 |
| 7.2.2 Modèle de Black | 33 |
| 7.2.3 Modèle SABR | 34 |
| 7.3 Annexe B : Grecques sous le modèle SABR λ -shifté | 36 |

| | | |
|-------|--|-----------|
| 7.3.1 | Grecques Bachelier | 36 |
| 7.3.2 | Grecques SABR | 36 |
| 7.3.3 | Grecques ajustés sous SABR shifté | 37 |
| 7.4 | Annexe C : Changement de numéraire | 38 |
| 7.4.1 | Changement de numéraire : de la mesure annuité à la mesure risque neutre | 38 |
| 7.4.2 | Calcul des drifts sous la mesure annuité commune - Preuve | 38 |
| | Contacts | 40 |

Introduction

Les appels de marge jouent un rôle clé dans la limitation du risque de défaut de la contrepartie (CCR : Counterparty Credit Risk). Dans le marché des produits dérivés OTC, les deux principaux types de collatéral postés sont la Variation Margin (VM) qui garantit les engagements de la contrepartie et l'Initial Margin (IM) qui couvre la variation de prix entre le défaut et la contraction d'un nouveau contrat similaire.

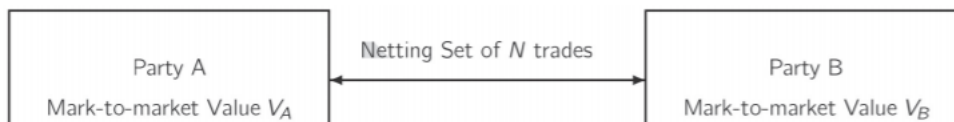


Figure 1 Netting Set entre deux contreparties A et B

Pour un ensemble de transactions entre deux contreparties ayant un accord, appelé "Netting Set", les appels de marge réguliers sont effectués sur la base de l'évaluation quotidienne de la valeur de marché (Mark-to-Market : MtM). La VM à l'instant t sera postée unilatéralement par la contrepartie dont la MtM est négative, et sera égale à la valeur absolue de cette dernière. La VM peut être reçue sous forme de note et peut être réutilisé par le détenteur.

Bien que la VM puisse réduire considérablement les expositions aux risques de contrepartie, des expositions résiduelles subsistent dues aux variations de la MtM sur un ensemble de périodes appelé Margin Period of Risk (MPoR) regroupant l'intervalle de temps durant :

- l'émission de la VM et la réception du collatéral $\simeq 1J$
- la période nécessaire pour confirmer un évènement de défaut $\simeq 2J$
- le temps nécessaire pour remplacer les positions suite au défaut $\simeq 7J$

La réglementation BCBS-IOSCO de 2015 a apporté un changement important sur le marché des dérivés en rendant obligatoire la modélisation de ce risque au travers de l'IM. De plus, la nouvelle réglementation définit cette dernière comme une Value-at-Risk à 99% sur un horizon de 10 jours ouvrables représentant la MPoR. Enfin, contrairement à la VM, l'IM ne peut être réutilisée et est constituée uniquement de cash ou d'instruments ultra-liquides. Ce dernier point comporte la contrainte la plus forte pour une entité financière en terme de liquidité.

La réglementation sur les appels de marges (VM-IM) autorise aux différentes contreparties qualifiées l'utilisation de modèles internes pour le calcul de l'IM. Cependant, dans un but de réconcilier les calculs de l'IM et éviter les litiges, une méthode standard a été proposée par l'ISDA (2014-2019) (International Swaps and Derivatives Association) : la SIMM (Standard Initial Margin Model). Il s'agit d'un modèle de Value at Risk paramétrique qui prend en entrée les sensibilités du Netting Set aux différents facteurs de risque auxquels il est exposé, puis les agrège selon des corrélations calibrées historiquement.

Ainsi, le calcul de l'IM a un fort impact sur la liquidité des institutions financières. il est donc important de pouvoir modéliser son coût total pendant la durée de vie du produit, appelé MVA (Margin Valuation Adjustment). Le calcul de cette métrique se base essentiellement sur une projection d'une IM forward : DIM (Dynamic Initial Margin). Cette dernière est particulièrement délicate à mettre en oeuvre et est souvent impossible par un calcul direct. Des alternatives ont fait l'objet de plusieurs recherches au cours de ces dernières années. Plusieurs papiers, dont Anfuso et al. (2000) et Chan et al. (2017) ont proposé l'utilisation de différentes méthodes de régression (linéaire, polynomiale, kernel smoothing), d'autres proposent d'estimer les sensibilités futures par utilisation de méthodes d'interpolation plus élaborées, à savoir les tenseurs de Chebychev étudié par Zeron and Ruiz (2020) et les réseaux de neurones proposés par Ma et al. (2019). Dans cette étude, on se propose d'adapter les algorithmes du Least Square Monte Carlo et de régression par processus gaussiens. On testera la performance des deux méthodes en terme de précision et de temps de calcul.

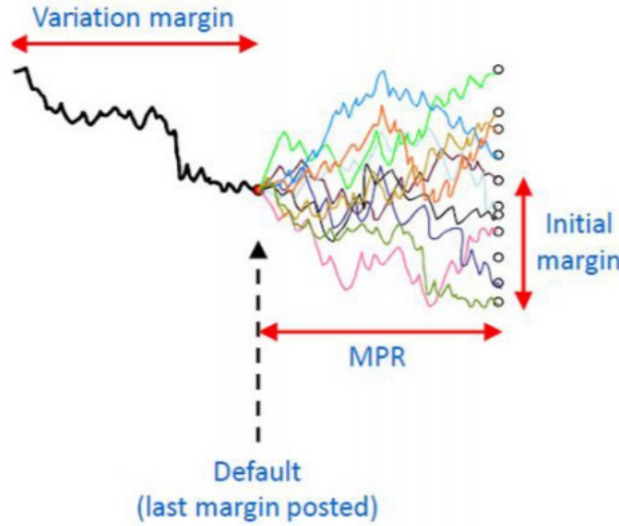


Figure 2 Différence entre VM et IM

1 Préliminaires

L'ensemble de l'étude sera basé sur un portefeuille de swaptions européennes. Il est rappelé en [Annexe A](#) quelques notions sur le marché de taux ainsi que les produits dérivés nous concernant. De plus, on y présente les différentes mesures de probabilité dans lesquelles se place cette étude. Enfin, on présente en [Annexe B](#) le modèle SABR-shifté utilisé pour cette étude.

2 Calcul de l'IM

2.1 Modèle interne

Le calcul de l'IM par modèle interne est opté par quelques grandes institutions financières. Cette méthode se base sur le calcul d'une **Value-at-Risk Monte Carlo**. Pour définir cette quantité, on pose le cadre de notations suivant :

On se place dans un espace de probabilité $(\Omega, \mathbb{P}, \mathcal{F})$ où \mathbb{P} désigne la probabilité historique. On considère un horizon fini $T > 0$ et on munit notre espace d'une filtration $(\mathcal{F}_t)_{0 \leq t \leq T}$. On désigne par $(X_t)_t$ le processus de \mathbb{R}^d qui représente l'évolution des facteurs de risque auxquels notre portefeuille P est exposé (prix d'actions, taux d'intérêts, taux de change, volatilité stochastique, etc..), finalement, on note par $(V_t)_t$ le processus $(\mathcal{F}_t)_t$ -adapté qui représente le prix du portefeuille négocié P .

On définit l'IM à l'instant t , avec une MPoR égale à $\delta < T - t$ comme suit :

$$IM_{t,\delta} = \mathcal{Q}_{99\%}(\Delta V_t \mid \mathcal{F}_t) \quad (1)$$

Avec $\Delta V_t = V_{t+\delta} - V_t$ la fluctuation du prix du portefeuille pendant la période de risque et $\mathcal{Q}_\alpha(\cdot \mid \mathcal{F}_t)$ le quantile α de la distribution conditionnelle, sous la mesure historique \mathbb{P} .

Dans notre cas, on souhaite calculer l'IM d'un portefeuille P de swaptions européennes de différentes maturités, tenor et strike, modélisées sous un modèle SABR shifté. Les facteurs de risque sont alors les couples (taux swap forward, volatilité stochastique) pour chaque couple maturité-ténor (m, n) .

L'IM du portefeuille en tant que quantile est **non linéaire**. La modélisation stochastique des différents facteurs de risque (taux swap, vol) pour chaque couple (m, n) doit donc prendre compte de la structure de corrélation entre ces facteurs de risque.

Modèle SABR-shifté multivarié : On définit le modèle SABR-shifté multidimensionnel sous une mesure commune \mathbb{Q} et pour N différents couples de maturité-tenor $i = (m, n)$, pour $i \in \{1, \dots, N\}$, comme suit :

$$\begin{aligned} dS_t^i &= \mu^i(t)dt + \alpha_t^i(S_t^i + \lambda)^{\beta_i} dW_t^i \\ d\alpha_t^i &= \nu_i \alpha_t^i dZ_t^i \\ S_0^i &= s_0^i - \lambda \\ \langle dW_i(t), dW_j(t) \rangle &= \rho_{ij} dt \\ \langle dW_i(t), dZ_j(t) \rangle &= \gamma_{ij} dt \\ \langle dZ_i(t), dZ_j(t) \rangle &= \xi_{ij} dt \end{aligned}$$

Avec (S, α) le couple de taux swap et de volatilité stochastique pour un couple maturité-tenor donné et W, Z deux mouvements browniens standards sous la mesure commune \mathbb{Q} .

Les $\rho_{i,j}$ représentent les corrélations entre les browniens des taux swap, les $\gamma_{i,j}$ les corrélations croisées entre les browniens des taux swap et les volatilités stochastiques et les $\xi_{i,j}$ les corrélations entre les browniens des volatilités. Pour chaque couple $i = (m, n)$, $\mu_i(t)$ représente le drift du taux swap $S^{m,n}$ sous la mesure commune \mathbb{Q} , traduisant le changement de mesure en AOA. λ est un paramètre permettant de traiter le cas des taux négatifs et ν_i désigne la volol du modèle SABR.

Dans notre cas, on choisit une mesure commune $\mathbb{Q}^{1,2}$ (la mesure swap), associée au numéraire annuité $A^{1,2}(t)$. Sous ce numéraire, seul le taux swap $S^{1,2}(t)$ est une martingale avec un drift nul.

Pour les autres taux swaps, la formule du drift $\mu^{m,n}(t)$ est définie par :

$$\begin{aligned} \mu^{m,n}(t)dt &= -\frac{A_t^{1,2}}{A_t^{m,n}} \times d < S_t^{m,n}, \frac{A_t^{m,n}}{A_t^{1,2}} > \\ &= \frac{A_t^{1,2}}{A_t^{m,n}} \sum_{i=(m+1) \wedge 3}^n \sum_{j=3}^i \frac{P_i(t)}{P_2(t)} \frac{\alpha_t^{m,n}(S_t^{m,n} + \lambda)^{\beta_{m,n}} \times \alpha_t^{j-1,j}(S_t^{j-1,j} + \lambda)^{\beta_{j-1,j}} \times \rho_{(m,n),(j-1,j)}}{1 + S_t^{j-1,j}} dt \end{aligned} \quad (2)$$

La preuve est détaillée en [Annexe C](#).

2.2 Méthode standard : SIMM

Nous présentons ici succinctement la formule standard. Le lecteur intéressé pourra se référer à [ISDA \(2014-2019\)](#) pour plus de détails. Pour permettre une comparaison des valeurs d'IM entre chaque contrepartie, l'ISDA a développé un modèle standard présentant les avantages suivants :

- **Facilité de réplication des calculs :** Pour éviter les litiges entre les contreparties.
- **Transparence :** Les calculs doivent pouvoir être vérifiés.
- **Coût opérationnel modéré :** La mise en œuvre, la maintenance et l'utilisation ne doivent pas être onéreuses.
- **Rapidité de calcul :** Possibilité d'une publication quotidienne.
- **Extensibilité :** Le modèle doit permettre la prise en compte d'un nouveau produit/facteur de risque.

En pratique, ce modèle est basé sur une formule de Value-at-Risk paramétrique, avec un jeu de paramètres calibrés historiquement sur 5 ans, incluant une période de stress des marchés financiers. Seul les sensibilités du portefeuille sont nécessaires pour les contreparties. Le calcul de l'IM par la méthode SIMM peut être décomposé en trois étapes :

1/ Classification des portefeuilles d'une contrepartie parmi les catégories suivantes :

- Taux et FX
- Crédit
- Action
- Matières premières

Le calcul de l'IM totale du portefeuille se fait en sommant l'IM des différentes classes de produit :

$$IMP = SIMM_{Taux\&FX} + SIMM_{Crédit} + SIMM_{Actions} + SIMM_{Matières\ premières} \quad (3)$$

On agrège les IM calculées pour chaque classe de risque¹ auquel est exposé le produit selon la formule :

$$SIMM_{product} = \sqrt{\sum_r IM_r^2 + \sum_r \sum_{s \neq r} \psi_{rs} IM_r IM_s} \quad (4)$$

1. devises, courbes de taux, ...

Pour chaque classe de risque, l'IM est calculée comme une somme de quatre quantités : la Delta Margin, la Vega Margin, la Curvature Margin, et la BaseCorr Margin. Si on note X la classe de risque considérée, on a :

$$IM_X = \text{DeltaMargin}_X + \text{VegaMargin}_X + \text{CurvatureMargin}_X + \text{BaseCorrMargin}_X \quad (5)$$

Dans notre cas, notre portefeuille P est uniquement composé de Swaptions d'une même devise, avec ψ_{rs} la corrélation des facteurs de risque, la SIMM du portefeuille s'écrit alors ² :

$$IM_P = \text{DeltaMargin}_{\text{Taux}} + \text{VegaMargin}_{\text{Taux}} + \text{CurvatureMargin}_{\text{Taux}} \quad (6)$$

2/ Détermination des facteurs de risque pour la classe taux :

Les facteurs de risque sont les 12 taux de rendement pour les maturités 2 semaines, 1 mois, 3 mois, 6 mois, 1 an, 2 ans, 3 ans, 5 ans, 10 ans, 15 ans, 20 ans, 30 ans, et ce pour chaque devise "b". Les courbes de taux considérées par la norme sont le "OIS", les Libors "1 mois", "3 mois", "6 mois", "12 mois", puis les courbes "Prime" et "Municipal" pour l'USD.

Ici, on considère que notre portefeuille est dans le marché Euro et qu'on est en mono-curve, (i.e on travaille avec une seule courbe de taux). Les facteurs de risque sont donc uniquement les 12 taux de rendement actuariels cités ci-dessus.

3/ Calcul des Margin pour chaque classe de risque

a) Calcul de la Delta Margin

On calcule la sensibilité de notre portefeuille s_k à chaque rendement R_k pour $k = [2W, 1M, 6M, 1Y, 2Y, 3Y, 5Y, 10Y, 15Y, 20Y, 30Y]$:

$$s_k = \sum_{\text{swaption} \in \text{portfolio}} s_{\text{swaption},k} \quad (7)$$

Pour chaque swaption, la sensibilité par rapport au taux de rendement R_k est définie comme :

$$s_{\text{swaption},k} = V^{\text{Swpt}}(R_k + 1bp) - V^{\text{Swpt}}(R_k) \quad (8)$$

Avec $R_k = R(t=0, k) = -\frac{1}{k} \ln B(t=0, k)$ et $B(t, k)$ le prix à t du ZC de maturité k .

Puis on calcule les sensibilités pondérées WS :

$$WS_{k,i} = RW_k s_{k,i} CR_b \quad (9)$$

Avec ici $i = 1$ en mono-curve. Les RW_k correspondent aux coefficients de pondération pour chaque facteur de risque et le CR_b correspond au facteur de concentration reflétant la MPOR. Ces différentes quantités sont calibrées annuellement par l'ISDA.

On notera $WS_{1,k}$: WS_k pour alléger les notations.

A partir de l'ensemble des sensibilités pondérées, on peut calculer la DeltaMargin :

$$\text{DeltaMargin}_{\text{Taux}} = \sqrt{\sum_k WS_k^2 + \sum_k \sum_{l \neq k} \rho_{k,l} WS_k WS_l} \quad (10)$$

Avec ρ la matrice de corrélation entre les 12 différentes maturités, calibrée par l'ISDA.

b) Calcul de la Vega Margin

Tout d'abord, on agrège les swaptions du portefeuille selon chaque maturité k et on calcule le risque de Vega correspondant :

$$VR_k = VRW \left(\sum_i VR_{ik} \right) VCR \quad (11)$$

Avec les VR_{ik} les différents risques de vega des swaptions de maturité égale à k et VRW , VT_b respectivement la pondération en vega et le risque de concentration calibrés par l'ISDA.

$VR_{i,k}$ correspond au risque de vega de la swaption i par rapport à la maturité k défini comme le produit de la volatilité à la monnaie (SABR) multipliée par le Vega SABR :

$$VR_{i,k} = \sigma_{kj} \frac{\partial V_i}{\partial \sigma} \quad (12)$$

$$\frac{\partial V_i}{\partial \sigma} = V_i(\sigma + 1) - V_i(\sigma)$$

Avec σ_{kj} la volatilité ATM de la swaption i , de maturité k et de tenor j

2. On ne prend pas en compte les buckets et la corrélation entre les devises

Une fois les risques de vega pour chaque maturité du portefeuille calculés, on peut en déduire la VegaMargin comme suit :

$$\text{VegaMargin}_{\text{Taux}} = \sqrt{\sum_k V R_k^2 + \sum_k \sum_{l \neq k} \rho_{kl} V R_k V R_l} \quad (13)$$

Avec ρ la matrice de corrélations définie précédemment.

c) Calcul de la CurvatureMargin

On regroupe les swaptions selon chaque maturité :

$$CVR_k = \sum_j SF(t_{kj}) \sigma_{kj} \frac{dV_i}{d\sigma} \quad (14)$$

$SF(t_{kj})$: **scaling function** donnée par l'ISDA, reliant le Vega et le Gamma pour les options vanilles
Puis, on agrège le résultat des différentes Curvatures selon la formule :

$$CVR = \sqrt{\sum_k CVR_k^2 + \sum_k \sum_{(l) \neq (k)} \rho_{kl}^2 CVR_k CVR_l} \quad (15)$$

On calcule finalement la CurvatureMargin à l'aide de la formule :

$$\text{CurvatureMargin} = HVR * \max \left\{ \sum_k CVR_k + \lambda \times CVR, 0 \right\} \quad (16)$$

Avec le scalaire λ et HVR fourni par l'ISDA.

Un manque de réactivité du modèle SIMM La crise du Covid-19 a fortement impacté les marchés financiers, et notamment celui des dérivés OTC où de forts mouvements liés à des niveaux de volatilité extrêmes ont été constatés. Ces conditions de volatilité ont soulevé quelques inquiétudes concernant le modèle SIMM. Contrairement aux modèles de calcul de l'IM au sein des chambres de compensation, calibrés en permanence et qui ont pu refléter les conditions d'un marché tendu, les exigences pour les produits non compensés calculées par le SIMM n'ont pratiquement pas augmenté. Ce qui indique une faible réactivité du modèle aux conditions réelles du marché. En effet, cela s'explique par le fait que les paramètres du SIMM sont calibrés sur une base annuelle avec une latence d'un an, ce qui veut dire que la volatilité extrême du marché survenue en Mars 2020 ne sera reflétée que sur le calcul de l'IM de 2021. Cette faible réactivité a soulevé quelques préoccupations chez plusieurs acteurs du marché quant à la cohérence de cette fréquence de calibration.

Un modèle contra-cyclique Cependant, la calibration annuelle des paramètres du modèle SIMM vise à s'affranchir d'effets procycliques non désirés. Dans les modèles IM au sein des CCPs se basant sur une VaR historique, les montants d'IM ont tendance à diminuer en période de prospérité, puis à augmenter drastiquement en période de stress, ce qui aggrave les problèmes de liquidité des membres (+25% Q1-2020). Les montants de marge calculés par le SIMM sont assez stables en comparaison et permettent alors d'éviter ces problématiques de liquidité.

3 Calcul de la MVA

3.1 Présentation

Le coût de l'Initial Margin peut être compris dans le prix d'un produit dérivé OTC, tout comme les autres xVA (Credit, Funding et Capital). Les premières recherches sur la quantification des risques de contrepartie ont été développées par [Burgard and Kjaer \(2011\)](#). En découle un premier framework CVA-FVA-KVA développé par [Green \(2015\)](#).

Après la mise en vigueur de l'IM sur les marchés OTC, [Green and Kenyon \(2015\)](#) ont inclus la MVA dans leur framework xVA. Cette MVA est calculée en introduisant un portefeuille \hat{V} répliquant les actifs avec risque de contrepartie. Si on pose $\hat{V} = V + U$ avec V la MtM du portefeuille et U l'ensemble des xVA, on aboutit aux différentes formules des xVAs par application du théorème de Feynman-Kac.

Le but de la MVA est de quantifier le coût total associé au posting de l'IM tout au long de la durée de vie du portefeuille négocié. Ce calcul se base sur des projections futures de l'IM (Forward ou Dynamic IM). Pour la modélisation, on se place dans un espace de probabilité filtré $(\Omega, \mathbb{P}, F, (F_t)_{0 \leq t \leq T})$ où \mathbb{P} désigne la probabilité historique et $T > 0$ un horizon fixé. La **Margin Valuation Adjustment** se calcule comme suit :

$$MVA_t = \int_t^T ((1 - R)\lambda_B(s) - S_I(s)) e^{-\int_t^s (\lambda_B + \lambda_c)(u) du} \tilde{\mathbb{E}}[D(t, s) IM(s) | F_t] ds \quad (17)$$

Avec :

- R : le taux de recouvrement de la banque B .
- $S_I(s)$: le spread de la rémunération de l'IM.
- T : la maturité du portefeuille négocié.
- $\lambda_B(s)$ et $\lambda_C(s)$ les intensités de défaut de la banque B et de sa contrepartie C respectivement.
- $D(t, s)$: le facteur d'actualisation stochastique entre les dates t et s .
- $IM(s)$: le montant de l'IM Forward à poster en s .
- $\tilde{\mathbb{E}}[\cdot | F_t]$ l'espérance conditionnelle sous la mesure risque neutre.

En $t = 0$:

$$\begin{aligned} MVA &= \int_0^T ((1-R)\lambda_B(s) - S_I(s)) e^{-\int_0^s (\lambda_B + \lambda_C)(u) du} \tilde{\mathbb{E}}[D(0, s) IM(s) | F_0] ds \\ &= \int_0^T ((1-R)\lambda_B(s) - S_I(s)) e^{-\int_0^s (\lambda_B + \lambda_C)(u) du} \tilde{\mathbb{E}}\left[\frac{IM(s)}{B(s)}\right] ds \end{aligned}$$

En se plaçant sous la mesure annuité $\mathbb{Q}^{1,2}$ associée au numéraire $A^{1,2}(t)$:

$$\tilde{\mathbb{E}}\left[\frac{IM(s)}{B(s)}\right] = A^{1,2}(0) \cdot \mathbb{E}^{\mathbb{Q}^{1,2}}\left[\frac{IM(s)}{A^{1,2}(s)}\right] \quad (18)$$

Pour alléger les notations, on notera $A^{1,2}(t) = An(t)$ et sa mesure associée $\mathbb{Q}^{1,2} = \mathbb{Q}$. La MVA en $t = 0$ s'exprime alors comme :

$$MVA = \int_0^T ((1-R)\lambda_B(s) - S_I(s)) e^{-\int_0^s (\lambda_B + \lambda_C)(u) du} An(0) \cdot \mathbb{E}^{\mathbb{Q}}\left[\frac{IM(s)}{An(s)}\right] ds \quad (19)$$

3.2 Probabilité historique VS Mesure annuité : Coefficient de scaling

Le calcul de la MVA se base sur l'espérance de l'IM sous une mesure annuité en chaque instant futur $t > 0$. Ainsi, pour calculer cette espérance en un instant t futur, on diffuse les facteurs de risque entre 0 et t selon la mesure annuité commune, puis on calcule l'IM pour chaque scénario généré, en diffusant les facteurs de risque entre t et $t + \delta$ selon la probabilité **historique**.

Le problème qu'induit ce genre de configuration est que l'utilisation de metamodeling (notamment l'algorithme **Least Square Monte Carlo**) n'est pas autorisée puisqu'on ne peut pas inférer le résultat d'une simulation **en monde réel** à partir de différentes simulations générées en **risque neutre**.

Une solution proposée dans la littérature par [Anfuso et al. \(2000\)](#) est d'estimer l'IM non pas sous la probabilité historique mais en risque neutre. En effet, une fois l'IM estimée en t sous la mesure annuité, on lui applique une fonction de scaling dynamique $\alpha(t)$ afin qu'elle soit conforme aux données du marché réel. Ce scaling se base sur le montant de l'IM en $t_0 = 0$ retournée par le modèle SIMM proposé par l'ISDA, ainsi que d'un jeu d'hyper-paramètres à calibrer :

$$\alpha_t = (1 - h(t)) [\alpha_\infty + (\alpha_{t_0} - \alpha_\infty) \exp^{-\beta(t)}] \quad (20)$$

Avec α_{t_0} le coefficient de scaling en $t = 0$ calculé comme le ratio entre le montant de l'IM en $t_0 = 0$ retournée par le modèle SIMM proposé par l'ISDA et celui calculé en risque neutre :

$$\alpha_{t_0} = \frac{SIMM(t_0)}{IM(t_0)} \quad (21)$$

Le terme α_∞ correspond au niveau de scaling sur le long terme quand $t \rightarrow \infty$, le terme β définit la transition entre les deux niveaux de scaling, tandis que la fonction h peut être interprétée comme étant un "haircut". Dans notre cadre d'étude, et par défaut de données historiques nécessaires pour la calibration des hyper-paramètres, on se contentera d'appliquer le coefficient de scaling α_{t_0} afin de jongler entre les deux mesures.

En appliquant ce coefficient, l'IM devient :

$$\widehat{IM}(t) = \alpha_{t_0} \times IM(t) \quad (22)$$

On calculera alors l'IM sous notre mesure annuité \mathbb{Q} et on appliquera le coefficient d'ajustement α_0 .

La MVA en $t = 0$ s'écrit alors :

$$\begin{aligned} MVA_0 &= \int_0^T ((1 - R)\lambda_B(s) - S_I(s)) e^{-\int_0^s (\lambda_B + \lambda_c)(u) du} An(0) \cdot \mathbb{E}^{\mathbb{Q}} \left[\frac{\widehat{IM}(s)}{An(s)} \right] ds \\ &= \int_0^T ((1 - R)\lambda_B(s) - S_I(s)) e^{-\int_0^s (\lambda_B + \lambda_c)(u) du} \alpha_0 An(0) \mathbb{E}^{\mathbb{Q}} \left[\frac{IM(s)}{An(s)} \right] ds \\ &= \int_0^T ((1 - R)\lambda_B(s) - S_I(s)) e^{-\int_0^s (\lambda_B + \lambda_c)(u) du} \alpha_0 An(0) \mathbf{DDIM}_{0,s,\delta} ds \end{aligned}$$

Avec $\mathbf{DDIM}_{0,s,\delta}$ (pour *Discounted Dynamic Initial Margin*) représente l'estimation en 0 de la future IM (actualisée) à poster en s pour une MPoR égale à δ :

$$\begin{aligned} \mathbf{DDIM}_{0,s,\delta} &= \mathbb{E}^{\mathbb{Q}} \left[\frac{IM(s)}{An(s)} \right] \\ &= \mathbb{E}^{\mathbb{Q}} \left[\frac{1}{An(s)} \times \mathcal{Q}_{99\%}(\Delta V_s \mid F_s) \right] \end{aligned}$$

La complexité du calcul de la MVA provient directement du calcul de la $\mathbf{DDIM}_{0,t,\delta}$ en chaque instant t , qui repose sur plusieurs simulations Monte Carlo imbriquées.

3.3 Calcul de la DDIM

Le calcul de la $\mathbf{DDIM}_{0,s,\delta}$ se fait via la formule ci-dessus. Au temps s futur, le quantile est calculé conditionnellement à la sous-tribu \mathcal{F}_s qui représente la donnée du marché jusqu'en s . Si on note X_t le vecteur des facteurs de risque (taux swap + volatilité stochastique pour chaque couple (m, n) de maturité \times tenor) en t , on remarque que la valeur du portefeuille négocié n'est sensible qu'à la donnée de ce vecteur, on peut alors supposer que $(\mathcal{F}_t)_{T \geq t \geq 0}$

correspond à la filtration naturelle du processus $(X_t)_{t \geq 0}$, i.e $\mathcal{F}_t = \sigma(X_u, u \leq t)$ pour tout t entre 0 et T . La $\mathbf{DDIM}_{0,s,\delta}$ s'écrit alors :

$$\mathbf{DDIM}_{0,s,\delta} = \mathbb{E}^{\mathbb{Q}} \left[\frac{1}{An(s)} \times \mathcal{Q}_{99\%}(\Delta V_s \mid \sigma(X_u, u \leq s)) \right] \quad (23)$$

En outre, le processus $(X_t)_t$ des facteurs de risque est solution de l'EDS SABR multivarié, il est donc un processus de Markov. On peut donc écrire :

$$\mathcal{Q}_{99\%}(\Delta V_s \mid \sigma(X_u, u \leq s)) = \mathcal{Q}_{99\%}(\Delta V_s \mid X_s) \quad (24)$$

La \mathbf{DDIM} s'écrit donc sous la forme :

$$\mathbf{DDIM}_{0,s,\delta} = \mathbb{E}^{\mathbb{Q}} \left[\frac{1}{An(s)} \times \mathcal{Q}_{99\%}(\Delta V_s \mid X_s) \right] \quad (25)$$

Le calcul de la \mathbf{DDIM} se base alors sur une simulation Monte-Carlo de trajectoires futures (simulations externes) $\{\omega, \omega \in \Omega\}$ des facteurs de risque pendant la durée de vie du portefeuille négocié. Pour chaque instant t futur, et pour chaque trajectoire ω , l'IM se calcule comme un quantile à 99% entre t et $t + \delta$ (simulations imbriquées). Il s'agit une simulation MC^2 .

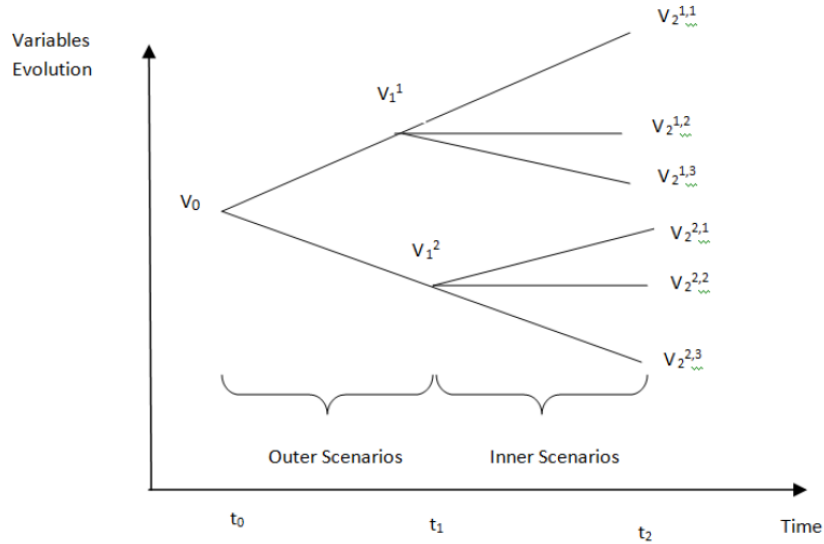


Figure 3 Simulations MC^2

On note N_{p1} le nombre de simulations MC externes utilisées et N_{t1} le nombre de points entre t_0 et t_1 pour une simulation. On note N_{p2} le nombre de simulations MC imbriquées et N_{t2} le nombre de points entre t_1 et t_2 pour une simulation. On considère un schéma de discrétisation entre $t = 0$ et T à subdivision régulière : $\Delta_{N_t} = \{0 = t_0 \leq t_1 \leq \dots \leq t_{N_t} = T\} = \{t_k = \frac{kT}{N_t}, k \in 0, \dots, N_t\}$.

$$\mathbf{DDIM}_{0,t_k,\delta} \approx \frac{1}{N_{p1}} \sum_{i=1}^{N_{p1}} \frac{1}{An(t_k, \omega_i)} \mathcal{Q}_{99\%}(\Delta V_{t_k}(\omega_i) \mid X_{t_k}(\omega_i) = x_{t_k}(\omega_i))$$

$\mathcal{Q}_{99\%}(\Delta V_{t_k}(\omega_i) \mid X_{t_k}(\omega_i) = x_{t_k}(\omega_i))$ correspond à l'IM du portefeuille négocié, à poster en date t_k si le marché emprunte le scénario ω_i . Ce quantile à 99% se calcule par une simulation Monte-Carlo imbriquée des facteurs de risque entre t_k et $t_k + \delta$. Il faut s'assurer que le nombre de simulations imbriquées (N_{p2}) soit suffisamment grand pour capturer la structure de la queue de distribution et ainsi obtenir un estimateur du quantile de faible variance.

3.4 Complexité Algorithmique du calcul direct

En notant d le nombre de couples (maturité, tenor), on peut décomposer la complexité algorithmique du calcul de la $\mathbf{DDIM}_{0,t_k,\delta}$ en deux parties principales :

- 1- Simulation des scénarios externes des facteurs de risque entre 0 et t_k en $C_{simulation} = O(d^3 + N_{p_1} \times d)$. La partie en $O(d^3)$ correspondant au calcul de la factorisation de Cholesky de la matrice de corrélation des Browniens.
- 2- Calcul de l'IM en t_k pour tous les scénarios $(\omega_i)_i$ en $C_{nested-IM} = O(N_{p_1} \times N_{t_2} \times N_{p_2} \times d)$. Finalement on obtient :

$$\begin{aligned} C_{MC^2}^{DDIM} &= C_{simulation} + C_{nested-IM} \\ &\approx O(d^3 + N_{p_1} \times N_{t_2} \times N_{p_2} \times d) \end{aligned}$$

Le calcul de la MVA se basant sur le calcul de la $\mathbf{DDIM}_{0,t_k,\delta}$ en chaque instant t_k entre 0 et T , on a :

$$\begin{aligned} C_{MC^2}^{MVA} &\approx O(d^3 + N_{t_1} \times N_{p_1} \times N_{t_2} \times N_{p_2} \times d) \\ &\approx O(N_{t_1} \times N_{p_1} \times N_{t_2} \times N_{p_2} \times d) \end{aligned}$$

4 Metamodeling de l'IM/MVA

En pratique, il n'est souvent pas possible de calculer la MVA avec une approche directe. En réponse à ce défi, plusieurs méthodes d'approximation peuvent être adoptées afin de réduire le coût total du calcul. On se propose de comparer deux méthodes pour l'approximation de la \mathbf{DDIM} et donc de la \mathbf{MVA} , à savoir l'estimation des quantiles par une régression LSMC et la régression par processus gaussiens. Les deux méthodes proposées assureront un niveau de complexité C_{LSMC} et C_{MC-GP} bien inférieur à C_{MC^2} . On comparera les résultats en terme de précision en utilisant la méthode MC^2 comme benchmark.

4.1 Least Square Monte Carlo

La régression LSMC (Least Square Monte Carlo) est la méthode la plus utilisée en finance pour la réduction de complexité des problèmes de simulations Monte Carlo imbriquées. La première application de cette régression, présentée dans l'article originel de [Longstaff and Schwartz \(2001\)](#) était dans le cadre de la valorisation des options américaines. Dans ce papier, les deux auteurs proposent de réduire le coût de calcul d'une MC^2 en simulant uniquement les scénarios externes et en utilisant des régressions par fonctions polynomiales pour inférer le résultat des simulations imbriquées. L'utilisation des régressions est ici mathématiquement justifiée dans la mesure où il s'agit d'une approximation d'espérance conditionnelle. Cependant, l'IM étant un quantile et non une espérance, il sera nécessaire d'adapter cet algorithme.

4.1.1 Adaptation de la LSMC au calcul de la DDIM

On rappelle que la $\mathbf{DDIM}_{0,s,\delta}$ s'écrit de la forme :

$$\mathbf{DDIM}_{0,s,\delta} = \mathbb{E}^Q \left[\frac{1}{An(s)} \times \mathcal{Q}_{99\%}(\Delta V_s \mid F_s) \right] \quad (26)$$

Notre but ici est d'estimer la quantité $\mathcal{Q}_{99\%}(\Delta V_s(\omega) \mid F_s)$ sans avoir à simuler N_{p_2} trajectoires MC internes entre s et $s + \delta$. Pour se ramener dans le cadre de l'estimation de l'espérance conditionnelle, l'idée de base sera de supposer une distribution de probabilité paramétrique pour $\Delta V_s(\omega) \mid F_s$ pour tout scénario externe ω et tout s entre 0 et T . Il suffira ensuite de formuler le quantile de cette distribution en fonction des moments qui s'expriment sous forme d'espérances conditionnelles.

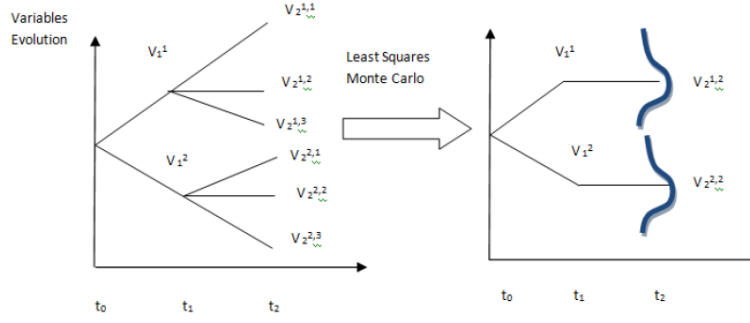


Figure 4 Simulations sous LSMC

On fait ici l'hypothèse que pour tout s entre 0 et T , la distribution conditionnelle $\Delta V_s(\omega) \mid F_s$ est localement gaussienne (c'est à dire à scénario ω fixé), cette approximation est pratique puisque la loi gaussienne est entièrement déterminée par deux paramètres : moyenne et variance . Cette hypothèse sera vérifiée dans la section résultat. Sous cette hypothèse, on peut montrer que pour tout scénario de marché ω et tout $s \in [0, T]$:

$$\mathcal{Q}_{99\%}(\Delta V_s(\omega) \mid F_s) = \mathbb{E}^{\mathbb{Q}}[\Delta V_s(\omega) \mid F_s] + \Phi^{-1}(99\%) \sqrt{\mathbb{E}^{\mathbb{Q}}[\Delta V_s(\omega)^2 \mid F_s] - \mathbb{E}^{\mathbb{Q}}[\Delta V_s(\omega) \mid F_s]^2} \quad (27)$$

où $\Phi^{-1}(99\%)$ le quantile à 99% de la loi normale centrée réduite.

Notre problème se ramène alors à l'estimation des deux premiers moments de la distribution conditionnelle. On notera alors $M_1(\omega, s) = \mathbb{E}^{\mathbb{Q}}[\Delta V_s(\omega) \mid F_s]$ et $M_2(\omega, s) = \mathbb{E}^{\mathbb{Q}}[\Delta V_s(\omega)^2 \mid F_s]$. $M_1(\omega, s)$ et $M_2(\omega, s)$ peuvent ainsi être estimés par une régression LSMC sur un ensemble de fonctions de base évaluées en \mathbf{X}_s .

4.1.2 Choix des régresseurs

Lorsque la taille du vecteur d'état X_s est grande, le nombre de coefficients β_i de la régression peut vite devenir important et il est souvent nécessaire de réduire la dimension (ACP,...). Une autre idée est de projeter sur les valeurs du portefeuille considéré $(V_t)_{t \geq 0}$ à la place de X_s . i.e $\mathcal{F}_s^V = \sigma(V_u, u \leq t)$:

$$\begin{aligned} M_i(\omega, s) &= \mathbb{E}^{\mathbb{Q}}[(\Delta V_s(\omega))^i \mid F_s] \\ &\approx \mathbb{E}^{\mathbb{Q}}[(\Delta V_s(\omega))^i \mid \mathcal{F}_s^V] \\ &\approx \mathbb{E}^{\mathbb{Q}}[(\Delta V_s(\omega))^i \mid V_s] \quad \forall i \in 1, 2 \\ &\approx \sum_{r=1}^R \hat{\beta}_r \phi_r(V_s) \end{aligned}$$

Avec $\hat{\beta} = [\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_R] \in \mathbb{R}^{R+1}$ les coefficient de régression et $\{\phi_r, 0 \leq r \leq R\}$ une famille de polynômes orthogonaux (Legendre, Laguerre ou Chebychev).

Si on choisit le processus multidimensionnel $(X_t)_{t \geq 0}$ des facteurs de risque comme régresseur, on a :

$$\begin{aligned} M_i(\omega, s) &= \mathbb{E}^{\mathbb{Q}}[(\Delta V_s(\omega))^i \mid X_s] \quad \forall i \in 1, 2 \\ &\approx \sum_{\substack{(i_1, \dots, i_d) \in \{0, \dots, R\}^d \\ i_1 + \dots + i_d \leq R}} \hat{\beta}_{i_1, \dots, i_d} \prod_{j=1}^d (X_s^j)^{i_j} \end{aligned}$$

4.1.3 Algorithme LSMC pour la DDIM

On présente ici l'algorithme LSMC adapté pour le calcul de la $\mathbf{DDIM}_{0, t_k, \delta}$ en 0 pour un instant de discrétisation futur t_k :

1. Générer N_{p_1} scénarios externes $(\omega_1, \dots, \omega_{N_{p_1}})$ entre 0 et $t + \delta$. On obtient alors 3 vecteurs :
 $X_{t_k} = [X_{t_k}(\omega_1), X_{t_k}(\omega_2), \dots, X_{t_k}(\omega_{N_{p_1}})]^{\top}$ le vecteur des facteurs risque simulés jusqu'en date t_k
 $V_{t_k} = [V_{t_k}(\omega_1), V_{t_k}(\omega_2), \dots, V_{t_k}(\omega_{N_{p_1}})]^{\top}$ le vecteur des MtM en date t_k suivant chaque simulation.

$\Delta V_{t_k} | \mathcal{F}_{t_k} = [\Delta V_{t_k}(\omega_1), \Delta V_{t_k}(\omega_2), \dots, \Delta V_{t_k}(\omega_{N_{p_1}})]^\top$ les fluctuations de la MtM entre t_k et $t_k + \delta$ suivant chaque simulation

2. Pour $i \in \{1, 2\}$:

2.1. Noter $\mathbf{Y}^{(i)}$ le vecteur $(\Delta V_{t_k})^i | \mathcal{F}_{t_k}$.

2.2. Estimer le vecteur β des coefficients par moindres carrés.

2.3. Pour chaque scénario ω_j , utiliser les coefficients estimés pour calculer :

$$M_i(\omega_j, t_k) = \mathbb{E}^Q[(\Delta V_s(\omega_j))^i | V_{t_k}(\omega_j) = v_{t_k}]$$

ou

$$M_i(\omega_j, t_k) = \mathbb{E}^Q[(\Delta V_s(\omega_j))^i | X_{t_k}(\omega_j) = X_{t_k}]$$

3. Pour chaque scénario externe ω_j , utiliser $M_1(\omega_j, t_k)$ et $M_2(\omega_j, t_k)$ pour estimer le quantile $\mathcal{Q}_{99\%}(\Delta V_{t_k}(\omega_j) | F_{t_k})$:

$$\mathcal{Q}_{99\%}(\Delta V_{t_k}(\omega_j) | F_{t_k}) = M_1(\omega_j, t_k) + \Phi^{-1}(99\%) \times \sqrt{M_2(\omega_j, t_k) - M_1(\omega_j, t_k)^2} \quad (28)$$

4.1.4 Complexité algorithmique

La complexité algorithmique C_{LSMC}^{DDIM} du calcul de la $\mathbf{DDIM}_{0,t_k,\delta}$ en t_k peut aussi se décomposer en deux parties principales :

1- Simulation des scénarios externes des facteurs de risque entre 0 et $t + \delta$ en $C_{simulation} = O(d^3 + N_{t_2} \times N_{p_1} \times d)$. La partie en $O(d^3)$ correspondant au calcul de la factorisation de Cholesky de la matrice de corrélation des Browniens.

2- Calcul des quantiles par une méthode de régression à moindres carrés ordinaires (MCO) en $C_{regression} = O(D^3 + N_{p_1} \times D)$ où $D = D(d, R)$ si on utilise le vecteur des facteurs de risque comme régresseur et $D = R$ si la valeur du portefeuille est utilisée comme régresseur. On rappelle que la complexité d'une régression MCO de dimension p et de nombre d'échantillons n est en $O(p^3 + np)$. On écrit alors

$$\begin{aligned} C_{LSMC}^{DDIM} &= C_{simulation} + C_{regression} \\ &\approx O(d^3 + N_{t_2} \times N_{p_1} \times d + D^3 + N_{p_1} \times D) \end{aligned}$$

La complexité algorithmique pour le calcul de la MVA par LSMC est donc :

$$C_{LSMC}^{MVA} = O(d^3 + N_{t_1} \times N_{t_2} \times N_{p_1} \times d + N_{t_1} \times (D^3 + N_{p_1} \times D)) = \begin{cases} O(d^3 + N_{t_1} \times N_{t_2} \times N_{p_1} \times d) & \text{si } D = R \\ O(d^3 + N_{t_1} \times (D^3 + N_{p_1} \times D)) & \text{si } D = D(d, R) \end{cases}$$

4.1.5 Limitations

L'idée du calcul de la \mathbf{DDIM} en adaptant la régression LSMC est assez simple à mettre en oeuvre mais présente quelques limitations :

- L'hypothèse paramétrique (normale) sur la distribution conditionnelle de $\Delta V_s | \mathcal{F}_s$.
- Le résultat de la régression peut fortement dépendre du degré R des polynômes.
- Pour garantir une meilleure estimation, la taille de l'échantillon N_{p_1} doit être supérieur au nombre de simulations externes pour un calcul direct de la \mathbf{DDIM} .
- Si on choisit le vecteur X_s des facteurs de risque comme régresseur, le nombre de coefficients β est quasi-exponentiel en la dimension de X . Cette méthode n'est donc pas **scalable**.

4.2 Régression par GPs

La régression par processus gaussiens est une technique d'inférence statistique bayésienne.

4.2.1 Présentation

L'inférence statistique repose sur l'apprentissage d'une fonction $Y = f(X)$ à partir des données, $(X, Y) := \{(\mathbf{x}_i, y_i) | i = 1, \dots, n\}$. L'idée derrière la régression par GP est de considérer un a priori sur l'espace des fonctions explicatives \mathcal{H} , sans avoir à les paramétrer. On désignera par hyper-paramètre, toute inconnue autre que X .

Notations et définitions

On dit que Y est un processus gaussien (GP) dont la fonction moyenne est $\tilde{\mu}$ et la fonction covariance, appelée "kernel" est k , si pour tout vecteur $X = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)^\top$ à composantes dans \mathbb{R}^d , le vecteur des images

$(f(\mathbf{x}_1), f(\mathbf{x}_2), \dots, f(\mathbf{x}_n))^T$ est un vecteur gaussien :

$$[f(\mathbf{x}_1), f(\mathbf{x}_2), \dots, f(\mathbf{x}_n)]^T \sim \mathcal{N}(\mu, K_{X,X}) \quad (29)$$

Avec μ le vecteur des moyennes tel que $\mu_i = \tilde{\mu}(\mathbf{x}_i)$ et $K_{X,X}$ la matrice de covariance vérifiant $(K_{X,X})_{i,j} = k(\mathbf{x}_i, \mathbf{x}_j)$. Le kernel peut être toute fonction symétrique semi-définie positive $k : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$:

$$\sum_{i,j=1}^n k(\mathbf{x}_i, \mathbf{x}_j) \xi_i \xi_j \geq 0 \quad \forall n \in \mathbb{N}, (\mathbf{x}_k)_k \in (\mathbb{R}^d)^{\mathbb{N}}, (\xi_k)_k \in (\mathbb{R})^{\mathbb{N}} \quad (30)$$

Il existe quelques types de Kernels qui sont communément utilisés, notamment les Radial basis functions (RBF) qui dépendent uniquement de la distance entre deux inputs $(\mathbf{x} - \mathbf{x}')$. Un exemple de Kernel RBF est le Kernel exponentiel (SE) défini par :

$$k(\mathbf{x}, \mathbf{x}') = \exp\left\{-\frac{1}{2l^2}(\mathbf{x} - \mathbf{x}')^2\right\} \quad (31)$$

Une autre famille connue de Kernel sont les Kernels Matérn (généralisation des kernels SE), définis par :

$$k(\mathbf{x}, \mathbf{x}') = \frac{1}{\Gamma(\nu)2^{\nu-1}} \left(\frac{\sqrt{2\nu}}{l}\mathbf{x} - \mathbf{x}'\right)^\nu K_\nu\left(\frac{\sqrt{2\nu}}{l}\mathbf{x} - \mathbf{x}'\right) \quad (32)$$

Avec un paramètre supplémentaire ν qui contrôle la régularité de la fonction, K_ν la fonction de Bessel modifiée évaluée en ν , et Γ la fonction Gamma.

On peut montrer que quand $\nu \rightarrow +\infty$, la fonction Kernel Matérn(ν, l) converge vers la fonction Kernel exponentiel SE(l). Les Kernels Matérn $\frac{3}{2}$ et $\frac{5}{2}$ sont les plus communs :

Pour $\nu = 3/2$:

$$k(\mathbf{x}, \mathbf{x}') = \left(1 + \frac{\sqrt{3}}{l}\mathbf{x} - \mathbf{x}'\right) \exp\left(-\frac{\sqrt{3}}{l}\mathbf{x} - \mathbf{x}'\right)$$

Pour $\nu = 5/2$

$$k(\mathbf{x}, \mathbf{x}') = \left(1 + \frac{\sqrt{5}}{l}\mathbf{x} - \mathbf{x}' + \frac{5}{3l}\mathbf{x} - \mathbf{x}'^2\right) \exp\left(-\frac{\sqrt{5}}{l}\mathbf{x} - \mathbf{x}'\right)$$

On rappelle que le but est d'inférer, en utilisant un meta-modèle, une fonction f qui existe entre l'input et l'output (apprentissage supervisé).

Pour cela, on dispose d'une base d'apprentissage S , de cardinal $|S| = n : S = (\mathbf{X}, Y) = \{(\mathbf{x}_i, y_i) \mid i = 1, \dots, n\}$. Dans notre cadre, on suppose que les données (observations de la vraie fonction f) sont bruitées, le modèle GPR s'écrit sous la forme :

$$Y = f(\mathbf{X}) + \epsilon \quad (33)$$

où f la fonction à inférer que l'on suppose *a priori* dans la classe \mathcal{H} des processus gaussiens, i.e $f \sim \mathcal{GP}(\tilde{\mu}, k)$ et $\epsilon = (\epsilon_1, \dots, \epsilon_n)^T$ est un vecteur de variables aléatoires i.i.d qui représentent le bruit.

ϵ est communément appelé le **bruit intrinsèque**. Dans notre cadre, on fait l'hypothèse que pour tout i dans $\{1, \dots, n\}$, ϵ_i suit une loi normale centrée de variance σ^2 .

Sous ces hypothèses, on peut montrer facilement que :

$$Y \mid \mathbf{X} \sim \mathcal{N}(\mu, \mathbf{K}_Y = \mathbf{K}_{X,X} + \sigma^2 I) \quad (34)$$

Avec I la matrice identité de dimension n et $K_{X,X}$ la matrice du kernel évalué en X , définie précédemment. Pour simplifier les notations, on suppose que le vecteur μ de l'évaluation du GP aux points d'apprentissage, est nul ($\mu = 0$).

Distribution prédictive à posteriori

On garde la même notation $S = (\mathbf{X}, Y)$ pour la base d'apprentissage. Le but ici est de trouver la distribution prédictive *à posteriori* du GP bruité, évalué à un ensemble de points de test \mathbf{X}_* , i.e la $\mathbf{p}(f_* \mid \mathbf{X}, Y, \mathbf{X}_*)$.

Sous ces notations, on peut écrire la loi jointe de Y et de f_* comme suit :

$$\begin{pmatrix} Y \\ f_* \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ \mu_{X_*} \end{pmatrix}, \begin{pmatrix} \mathbf{K}_Y & \mathbf{K}_{X,X_*} \\ \mathbf{K}_{X,X_*}^\top & \mathbf{K}_{X_*,X_*} \end{pmatrix}\right) \quad (35)$$

Avec $\mathbf{K}_{X,X_*}, \mathbf{K}_{X_*,X_*}$ les matrices d'évaluation du Kernel aux points \mathbf{X} et \mathbf{X}_* , et μ_{X_*} la fonction moyenne du GP évaluée aux points de test X_* . On déduit alors, en appliquant le conditionnement des Gaussiens, que la distribution *à posteriori* $\mathbf{p}(f_* \mid \mathbf{X}, Y, \mathbf{X}_*)$ s'écrit comme :

$$\mathbf{p}(f_* \mid \mathbf{X}, Y, \mathbf{X}_*) = \mathcal{N}(\mathbb{E}[f_* \mid X, Y, X_*], \text{Var}[f_* \mid X, Y, X_*]) \quad (36)$$

Avec les moments *à posteriori* :

$$\begin{aligned}\mathbb{E}[f_* | X, Y, X_*] &= \mu_{X_*} + \mathbf{K}_{\mathbf{X}_*, \mathbf{X}} [\mathbf{K}_{\mathbf{X}, \mathbf{X}} + \sigma^2 I]^{-1} Y \\ \text{Var}[f_* | X, Y, X_*] &= \mathbf{K}_{\mathbf{X}_*, \mathbf{X}_*} - \mathbf{K}_{\mathbf{X}_*, \mathbf{X}} [\mathbf{K}_{\mathbf{X}, \mathbf{X}} + \sigma^2 I]^{-1} \mathbf{K}_{\mathbf{X}, \mathbf{X}_*}\end{aligned}\quad (37)$$

L'espérance calculée correspond à l'estimation de la fonction f évaluée sur les points de test \mathbf{X}_* , la variance correspond à l'erreur de l'interpolation. On voit bien que cette erreur est nulle si et seulement si les points X_* sont inclus dans la base d'apprentissage et que les observations sont sans bruit ($\sigma = 0$).

Le modèle de régression par GP permet alors une inférence par formule fermée, ainsi qu'une quantification de l'erreur par formule fermée, contrairement à la plupart des autres algorithmes de Machine Learning.

Estimation des hyper-paramètres

L'estimation des hyper-paramètres du meta-modèle se fait en maximisant la vraisemblance marginale. Si on note θ le vecteur qui représente l'ensemble des hyper-paramètres, l'expression de cette vraisemblance s'écrit sous la forme :

$$\mathbf{p}(Y | \mathbf{X}, \theta) = \int \mathbf{p}(Y | f, \mathbf{X}, \theta) \mathbf{p}(f | \mathbf{X}, \theta) df \quad (38)$$

Le log de la vraisemblance s'écrit sous la forme suivante :

$$\log \mathbf{p}(Y | \mathbf{X}, \theta) = - \left[Y^\top (\mathbf{K}_{\mathbf{X}, \mathbf{X}} + \sigma^2 I)^{-1} Y + \log \det (\mathbf{K}_{\mathbf{X}, \mathbf{X}} + \sigma^2 I) \right] - \frac{n}{2} \log 2\pi \quad (39)$$

La minimisation de l'opposé du log-vraisemblance se fait par Multistart SGD (Stochastic Gradient Descent) vu que la fonction n'est pas strictement convexe. Le gradient en un point θ s'écrit :

$$\partial_\theta \log \mathbf{p}(Y | \mathbf{X}, \theta) = \text{tr} \left(\alpha \alpha^\top - (\mathbf{K}_{\mathbf{X}, \mathbf{X}} + \sigma^2 I)^{-1} \right) \partial_\theta (\mathbf{K}_{\mathbf{X}, \mathbf{X}} + \sigma^2 I)^{-1} \quad (40)$$

Avec $\alpha := (\mathbf{K}_{\mathbf{X}, \mathbf{X}} + \sigma^2 I)^{-1} Y = (\mathbf{K}_{\mathbf{Y}})^{-1} Y$

Complexité et considérations numériques

La première étape est l'estimation des hyper-paramètres. Le calcul du gradient de la log-vraisemblance se base sur le calcul de l'inverse de la matrice carrée $\mathbf{K}_{\mathbf{Y}} = \mathbf{K}_{\mathbf{X}, \mathbf{X}} + \sigma^2 I$ de dimension M . Pour des raisons de stabilité numérique, on préfère ne pas inverser la matrice $\mathbf{K}_{\mathbf{Y}}$ directement, une approche plus robuste est de calculer la factorisation de Cholesky de cette matrice $\mathbf{K}_{\mathbf{Y}} = L^\top L$, ceci se fait avec un coût en $O(M^3)$. Une fois ce calcul fait, l'inverse $\mathbf{K}_{\mathbf{Y}}^{-1} = (L^\top)^{-1} L^{-1}$ se calcule en $O(M^2)$ et on calcule la quantité $\alpha = \mathbf{K}_{\mathbf{Y}}^{-1} Y$ en $O(M^2)$. Ce calcul d'inverse et de α étant réitéré à chaque étape de l'algorithme d'optimisation, l'estimation des hyper-paramètres se fait alors avec un coût en $O(M^3)$. Le dernier résultat de l'inversion de la matrice est conservé à l'issue de l'estimation des hyper-paramètres.

Une fois les hyper-paramètres estimés, le calcul de la prédiction, ie espérance et variance, de l'output sur l'ensemble de test X^* se fait en $O(M)$ et $O(M^2)$ respectivement.

On déduit alors que globalement, le coût de la régression par processus gaussiens est en $C_{GP} = O(N_{optim} M^3)$ où M est la taille de l'échantillon d'apprentissage et N_{optim} la complexité de descente du gradient.

4.2.2 Application des GPs au calcul de la MVA

L'optimisation du calcul de la MVA passe par l'optimisation du calcul de la $\mathbf{DDIM}_{0,s,\delta}$ en tout instant s de posting d'IM entre 0 et T . On rappelle la formule de la $\mathbf{DDIM}_{0,s,\delta}$ avec les notations du problème :

$$\mathbf{DDIM}_{0,s,\delta} = \mathbb{E}^{\mathbb{Q}} \left[\frac{1}{An(s)} \times \mathcal{Q}_{99\%}(\Delta V_s | X_s) \right] \quad (41)$$

Où X_s le vecteur aléatoire (\mathcal{F}_s -mesurable) des facteurs de risque en s , et l'espérance calculée sur N_{p_1} scénarios externes $\{\omega_1, \dots, \omega_{N_{p_1}}\}$ entre 0 et s , générés sous la mesure annuité \mathbb{Q} .

L'annuité $An(s)$ étant \mathcal{F}_s -mesurable, notre fonction f qu'on essaiera d'inférer est :

$$X(\omega) \mapsto \mathcal{Q}_{99\%}(\Delta V_s(\omega) | X_s = X(\omega))$$

Où $\omega \in \Omega$ un scénario externe du marché et $2d$ la dimension du vecteur aléatoire des facteurs de risque ; $d = \text{Card}(\{\text{maturities}\} \times \{\text{tenors}\})$.

L'idée de base ici est de ne pas calculer le quantile conditionnel (l'IM) pour tout scénario ω_i afin de réduire le coût du calcul en $O(N_{p_1} \times N_{p_2})$. En effet, on note \mathcal{E}_s un compact de \mathbb{R}^{2d} dans lequel les valeurs $(X_s(\omega_1), \dots, X_s(\omega_{N_{p_1}}))$ sont comprises. On propose alors de construire une base d'apprentissage en calculant le quantile conditionnel sur

un échantillon qui soit à la fois réduit et assez représentatif de \mathcal{E}_s , inférer la fonction f à partir de cette base d'apprentissage et puis calculer la sortie du méta modèle sur l'ensemble des simulations $(X_s(\omega_1), \dots, X_s(\omega_{N_{p1}}))$.

On note $X_s^{\text{Tr}} = [X_s^{(1)}, X_s^{(2)}, \dots, X_s^{(M)}]$ notre échantillon d'apprentissage de taille $M \ll N_{p1}$. Cet échantillon doit être choisi de telle sorte qu'il soit représentatif de l'espace \mathcal{E}_s . On détaillera ce choix dans la section suivante. Pour tout élément de cet échantillon $X_s^{(i)}$, on calcule par Monte Carlo **une estimation** $\hat{Q}_{99\%N_{p2}}(\Delta V_s | X_s = X_s^{(i)})$ du quantile conditionnel $Q_{99\%}(\Delta V_s | X_s = X_s^{(i)})$. On obtient alors notre base d'apprentissage $S := (X_s^{\text{Tr}}, Y_s^{\text{Tr}})$ de taille $M \ll N_{p1}$ avec :

$$\begin{aligned} X_s^{\text{Tr}} &= [X_s^{(1)}, X_s^{(2)}, \dots, X_s^{(M)}] \\ Y_s^{\text{Tr}} &= [Y_s^{(1)}, Y_s^{(2)}, \dots, Y_s^{(M)}] \\ \forall i \in \{1, \dots, M\} : Y_s^{(i)} &= \hat{Q}_{99\%N_{p2}}(\Delta V_s | X_s = X_s^{(i)}) \\ &= Q_{99\%}(\Delta V_s | X_s = X_s^{(i)}) + \epsilon_i \\ &= f(X_s^{(i)}) + \epsilon_i \end{aligned}$$

et ϵ_i représente l'erreur de l'estimateur Monte-Carlo du quantile conditionnel pour chaque i entre 1 et M . En exprimant cela de façon plus compacte, On a :

$$Y_s^{\text{Tr}} = f(X_s^{\text{Tr}}) + \epsilon \quad (42)$$

Avec $\epsilon = [\epsilon_1, \epsilon_2, \dots, \epsilon_M]$

On se place alors dans ce cadre en notant $f \sim \mathcal{GP}(\tilde{\mu}, k)$ avec une fonction kernel choisie. On entraîne ce modèle (i.e on estime l'ensemble des hyper-paramètres) en utilisant la base d'apprentissage $S := (X_s^{\text{Tr}}, Y_s^{\text{Tr}})$.

Une fois le modèle entraîné, on l'utilise pour estimer les quantiles conditionnels pour l'ensemble des simulations $X^* = (X_s(\omega_1), \dots, X_s(\omega_{N_{p1}}))$.

Si on note $Y^* = (y^*(\omega_1), y^*(\omega_2), \dots, y^*(\omega_{N_{p1}}))^T \in \mathbb{R}^{N_{p1}}$ l'output du modèle, on calcule la **DDIM** $_{0,s,\delta}$:

$$\text{DDIM}_{0,s,\delta} \approx \frac{1}{N_{p1}} \sum_{i=1}^{N_{p1}} \frac{1}{An(s, \omega_i)} \times y^*(\omega_i) \quad (43)$$

Choix de l'échantillon d'apprentissage

Le choix de l'échantillon d'apprentissage est crucial puisqu'on doit garantir que cet échantillon couvre bien l'espace \mathcal{E}_s des valeurs prises par l'ensemble des simulations $(X_s(\omega_1), \dots, X_s(\omega_{N_{p1}}))$, et ce pour tout instant s entre 0 et T .

Pour ce faire, on introduit les notations suivantes :

- On note $\mathcal{C} = \{(m, n); m \in \{\text{Maturities}\}, n \in \{\text{Tenors}\}\}$ l'ensemble des couples maturité \times tenor, de cardinal $|\mathcal{C}| = d$.
- On a alors pour tout $i \in \{1, \dots, N_{p1}\}$, le vecteur des facteurs de risque simulés selon le scénario ω_i $X_s(\omega_i) = [(S_s^{m,n}(\omega_i), \sigma_s^{m,n}(\omega_i))_{(m,n) \in \mathcal{C}}] \in \mathbb{R}^{2d}$. Avec $S_s^{m,n}(\omega_i)$ la valeur du taux swap (m, n) et $\sigma_s^{m,n}(\omega_i)$ la valeur de la volatilité stochastique de ce taux swap, en s et suivant le scénario ω_i .

Une première idée serait de prendre des quantiles espacés de la distribution de chaque facteur de risque sur les N_{p1} scénarios du marché. Pour bien couvrir l'espace \mathcal{E}_s , on doit aussi prendre en compte les quantiles extrêmes de chaque distribution. Le total des quantiles pris étant égal à M

En effet, pour un couple $(m, n) \in \mathcal{C}$ donné, la distribution du taux swap (m, n) et de sa volatilité en s s'écrivent :

$$\begin{aligned} \mathcal{D}(S_s^{m,n}) &= [S_s^{m,n}(\omega_1), \dots, S_s^{m,n}(\omega_{N_{p1}})] \\ \mathcal{D}(\sigma_s^{m,n}) &= [\sigma_s^{m,n}(\omega_1), \dots, \sigma_s^{m,n}(\omega_{N_{p1}})] \end{aligned}$$

A partir de ces deux distributions, on fixe deux seuils extrêmes q_1 (=5% par exemple) et q_2 (=95% par exemple). On prend alors $M' \ll M$ quantiles espacés des deux côtés de q_1 et q_2 . Puis $M - 2M'$ quantiles à α entre q_1 et q_2 .

On réitère cette procédure pour chaque couple (m, n) afin d'obtenir un échantillon qui couvre bien l'espace \mathcal{E}_s . L'inconvénient majeur de cette méthode est que le choix de l'échantillon peut fortement dépendre des quantités q_1 et q_2 fixés, ainsi que le nombre de quantiles extrêmes M' . De plus, pour assurer une bonne couverture de \mathcal{E}_s , ces quantités même doivent dépendre des distributions $\mathcal{D}(S_s^{m,n})$ et $\mathcal{D}(\sigma_s^{m,n})$ pour les différents couples (m, n) et les instants s .

Une idée alternative serait d'essayer de couvrir l'espace \mathcal{E}_s en utilisant des suites (quasi-aléatoires) à discrétion faible. On rappelle tout d'abord la notion de discrétion.

Définition 1. Discrépance La discrépance d'un ensemble $P = \{x_1, \dots, x_N\}$ de N points par rapport à un hypercube J de dimension s est définie par :

$$D_N(P) = \sup_{B \in J} \left| \frac{A(B; P)}{N} - \lambda_s(B) \right| \quad (44)$$

avec :

· λ_s est la mesure de Lebesgue de dimension s ,

· $A(B; P)$ est le nombre de points de P appartenant à B ,

· J est l'ensemble des pavés de dimension s , de la forme

$$\prod_{i=1}^s [a_i, b_i[= \{\mathbf{x} \in \mathbb{R}^s : a_i \leq x_i < b_i\}$$

avec : $0 \leq a_i < b_i \leq 1$

On travaillera avec la suite de Sobol sur l'hypercube $H = ([0, 1] \times [0, 1])^d$, on admettra ici que cette suite est à discrépance faible. Pour la preuve, on se référera à l'article originel de [Sobol \(1967\)](#).

L'idée est de tirer un ensemble fini $P = \{p_1, \dots, p_M\}$ de M points d'une suite de Sobol sur $H = ([0, 1] \times [0, 1])^d$. Puis effectuer une transformation pour se ramener à l'espace \mathcal{E}_s qu'on souhaite représenter. Une approximation de \mathcal{E}_s peut s'écrire sous la forme d'un hypercube $\hat{\mathcal{E}}_s$ de dimension $2d$ tel que :

$$\hat{\mathcal{E}}_s = \prod_{(m,n) \in \mathcal{C}} ([a_{m,n}^{(1)}, b_{m,n}^{(1)}] \times [a_{m,n}^{(2)}, b_{m,n}^{(2)}]) \quad (45)$$

Avec :

$$a_{m,n}^{(1)} = \min_{i \leq N_{p1}} S_s^{m,n}(\omega_i)$$

$$b_{m,n}^{(1)} = \max_{i \leq N_{p1}} S_s^{m,n}(\omega_i)$$

$$a_{m,n}^{(2)} = \min_{i \leq N_{p1}} \sigma_s^{m,n}(\omega_i)$$

$$b_{m,n}^{(2)} = \max_{i \leq N_{p1}} \sigma_s^{m,n}(\omega_i)$$

La transformation ξ qui nous permet alors de ramener notre échantillon P à l'espace $\hat{\mathcal{E}}_s$ est :

$$\hat{p}_i = \xi(p_i) = \left(p_i^{(1)}(b_{m,n}^{(1)} - a_{m,n}^{(1)}) + a_{m,n}^{(1)}, p_i^{(2)}(b_{m,n}^{(2)} - a_{m,n}^{(2)}) + a_{m,n}^{(2)} \right) \forall i \in \{1, \dots, M\} \quad (46)$$

L'ensemble des points $\hat{P} = (\hat{p}_i)_{1 \leq i \leq M}$ peut alors être utilisé comme échantillon d'apprentissage.

Cet échantillon d'apprentissage peut aussi être obtenu en appliquant la même transformation sur un ensemble de points \tilde{P} obtenu par "Mini-max Latin Hypercube Sampling" sur l'hypercube H . Cette méthode garantit aussi une bonne couverture du domaine $\hat{\mathcal{E}}_s$. Pour plus de détails concernant cette méthode, on se référera à l'article originel de [McKay et al. \(1979\)](#).

Avec les notations posées dans les sections précédentes, on présente ici l'algorithme MC-GP pour le calcul de la $\mathbf{DDIM}_{0, \mathbf{t}_k, \delta}$ en 0 pour un instant de discrétisation futur t_k :

1. Générer N_{p1} scénarios externes du marché $(\omega_1, \dots, \omega_{N_{p1}})$ entre 0 et t_k . On obtient : $X_{t_k}^* = (X_{t_k}(\omega_1), \dots, X_{t_k}(\omega_{N_{p1}}))$ le vecteur des facteurs de risque simulés en t_k .
2. Choisir l'échantillon d'apprentissage $X_{t_k}^{\text{Tr}} = [X_{t_k}^{(1)}, X_{t_k}^{(2)}, \dots, X_{t_k}^{(M)}]$ qui soit assez représentatif de l'espace $\hat{\mathcal{E}}_s$ (Choix des quantiles, Sobol, Mini-max LHS).
3. Pour tout élément de l'échantillon d'apprentissage, calculer par Monte Carlo une estimation de l'IM. On obtient alors la base d'apprentissage $S := (X_s^{\text{Tr}}, Y_s^{\text{Tr}})$ de taille M .
4. Entraîner le modèle de régression par GP en utilisant la base S .
5. Appliquer le modèle pour l'estimation de l'IM sur notre input $X_{t_k}^*$.
6. Si on note $Y_{t_k}^* = [y_{t_k}^*(\omega_1), y_{t_k}^*(\omega_2), \dots, y_{t_k}^*(\omega_{N_{p1}})]^\top \in \mathbb{R}^{N_{p1}}$ l'output du modèle. Calculer la $\mathbf{DDIM}_{0, \mathbf{t}_k, \delta}$:

$$\mathbf{DDIM}_{0, \mathbf{t}_k, \delta} \approx \frac{1}{N_{p1}} \sum_{i=1}^{N_{p1}} \frac{1}{An(t_k, \omega_i)} \times y_{t_k}^*(\omega_i) \quad (47)$$

4.2.3 MC-GP : Complexité Algorithmique

Le calcul peut se décomposer en trois parties principales :

- 1- Simulation des scénarios externes des facteurs de risque entre 0 et t_k en $C_{simulation} = O(d^3 + N_{p_1} \times d)$.
- 2- Calcul de l'IM sur l'échantillon d'apprentissage de taille $M \ll N_{p_1}$ en $C_{nested-IM} = O(M \times N_{t_2} \times N_{p_2} \times d)$. N_{t_2} étant le nombre d'instant de discrétisation entre t_k et $t_k + \delta$.
- 3- Estimation des hyper-paramètres et calcul prédictif de l'IM sur l'ensemble des simulations en $C_{GP} = O(N_{optim}M^3)$.

Finalement, le coût du calcul de la $\mathbf{DDIM}_{0,t_k,\delta}$ en un instant t_k s'écrit :

$$\begin{aligned} C_{MC-GP}^{DDIM} &= C_{simulation} + C_{nested-IM} + C_{GP} \\ &\approx O(d^3 + N_{p_1} \times d) + O(M \times N_{t_2} \times N_{p_2} \times d) + O(N_{optim}M^3) \end{aligned}$$

La complexité algorithmique pour le calcul de la MVA via l'algorithme MC-GP est donc :

$$\begin{aligned} C_{MC-GP}^{MVA} &= O(d^3 + N_{t_1} \times N_{p_1} \times d + N_{t_1} \times (M \times N_{t_2} \times N_{p_2} \times d + N_{optim}M^3)) \\ &= O(d^3 + N_{t_1} \times (M \times N_{t_2} \times N_{p_2} \times d + N_{optim}M^3)), N_{p_1} \ll M \times N_{t_2} \times N_{p_2} \end{aligned}$$

4.2.4 Limitations

L'inconvénient le plus important de la régression par processus gaussiens est le temps de calcul. En effet, la complexité temporelle de l'estimation des hyper-paramètres est cubique en fonction de la taille des observations M . Cela signifie qu'en pratique, il est difficile d'utiliser les GPs dans le cas où la dimension du problème est importante.

5 Implémentation et résultats numériques

5.1 Calcul de l'IM

Dans cette première partie, on s'intéresse au calcul de l'IM par les deux méthodes, à savoir le calcul par modèle interne (VaR Monte-Carlo) et celui par le modèle SIMM (VaR paramétrique)

Notre portefeuille P contient un ensemble de 356 swaptions européennes (payer, receiver et ATM) de différents schémas de (maturité, tenor) et de différents strikes. La jambe variable du swap sous-jacent est l'**Euribor 6 mois**.

On scinde le portefeuille P en deux sous-portefeuilles : un portefeuille P_1 de swaptions payeuses et un deuxième portefeuille P_2 de swaptions receivers. Puis on calcule l'IM sur chaque sous-portefeuille par les deux méthodes : le **SIMM** en utilisant les grecques SABR de Barlett, et le **modèle interne** en utilisant 10000 scénarios Monte Carlo des facteurs de risque diffusés sous le modèle SABR shifté multivarié. On obtient les résultats présentés dans la figure ci-dessous :

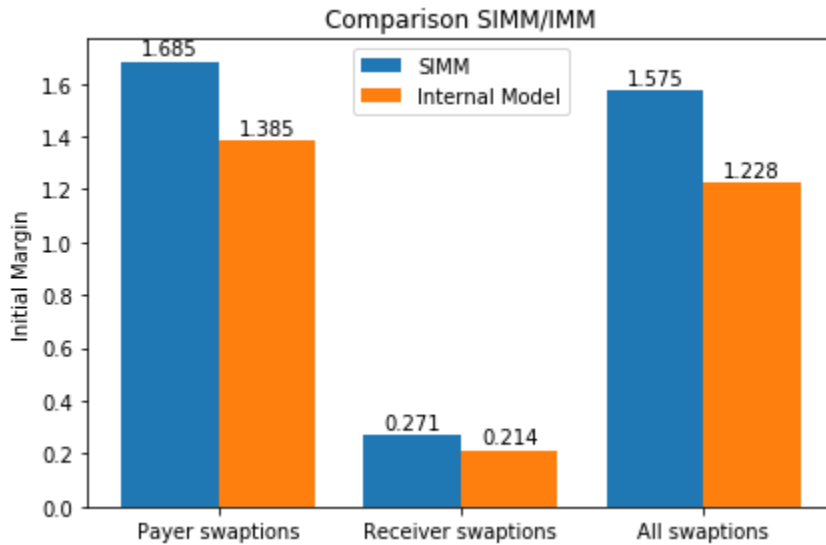


Figure 5 Comparaison : SIMM vs IMM

On remarque que notre modèle interne a tendance à sous-estimer le montant de l'IM à poster en $t_0 = 0$ par rapport à celui calculé par la norme SIMM. Cette différence est principalement due au fait que la corrélation du modèle SIMM est fournie par l'ISDA tandis que notre corrélation modèle a été estimée sur une base historique de taux swap. De plus, on remarque qu'il y a un effet de "Netting" entre les deux portefeuilles.

Le **coefficient de scaling** qui permet de rendre l'IM calculée par le modèle interne plus conforme aux données du marché est donc :

$$\alpha_{t_0}(P) = \frac{SIMM_{t_0}(P)}{IM_{t_0}(P)} = 1.2826 \quad (48)$$

5.2 Calcul de la MVA

On propose de calculer la MVA sur deux portefeuilles P_1 et P_2 réduits.

Le portefeuille P_1 contient un total de 10 swaptions européennes : trois swaptions de maturité égale à 3 mois et de tenor égal à un an, et sept swaptions de maturité 1 an et de tenor 1 an.

Le portefeuille P_2 comporte un total de 18 swaptions européennes : 9 swaptions de maturité 20 ans et de tenor 20 ans et 9 swaptions de maturité 20 ans et de tenor 30 ans. Ce portefeuille nous permettra d'avoir une idée sur la performance des différentes approches de Metamodeling sur le long terme.

Les swaptions de même schéma (maturité, tenor) sont de différents types (Payer, Receiver, ATM) et différents strikes

Les paramètres qui interviennent dans la **formule de la MVA** sont choisis comme suit : Le taux de recouvrement $R = 0.4$, les intensités $\lambda_B = \lambda_C = 0.015$ et le spread de la rémunération de l'IM $S_I = 0$. A noter que le Metamodeling est effectué sur la DDIM et que par conséquent, ce jeu de paramètre n'influence pas l'étude

5.2.1 Calcul de la MVA par méthode directe (MC²)

Pour calculer la MVA via la méthode **MC²**, on prend un nombre $N_t = 20$ d'instantes de discrétisation entre l'instant $t_0 = 0$ et T la maturité du portefeuille négocié ; Le portefeuille P_1 a une durée de vie $T = 1$ an et P_2 est de maturité $T = 20$ ans. Par ailleurs, on choisit $N_{p_1} = 5000$ le nombre de scénarios MC externes et $N_{p_2} = 1000$ le nombre de scénarios MC imbriqués.

On calcule d'abord le **coefficient de scaling** à l'instant t_0 pour les deux portefeuilles. On agrège les résultats dans le tableau suivant :

| IM à t_0 | P_1 | P_2 |
|----------------|-------|--------|
| MC | 16.8 | 1450.6 |
| SIMM | 35.9 | 2929.2 |
| α_{t_0} | 2.13 | 2.01 |

En utilisant le coefficient de scaling, on trace la fonction ϕ représentant l'intégrande de la MVA. On obtient les profils suivants :

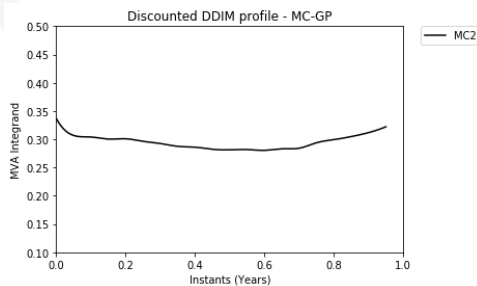


Figure 6 Profil de la DDIM-MC2 pour P_1

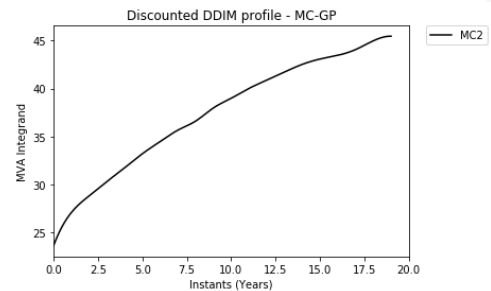


Figure 7 Profil de la DDIM-MC2 pour P_2

En intégrant ϕ par rapport au temps via la méthode de trapèzes, on obtient les montants de MVA pour les deux portefeuilles :

$$MVA_{t_0}(P_1) = 0.2803$$

$$MVA_{t_0}(P_2) = 710.11$$

5.2.2 Calcul de la MVA par Least Square Monte Carlo

On se propose de valider empiriquement l'hypothèse de la normalité locale de la distribution conditionnelle $\Delta V_s(\omega) \mid F_s$ en $s = 0$. Pour cela, on génère par Monte-Carlo un échantillon de la distribution $\Delta V_0 \mid F_0$.

On trace le graphe Q-Q plot ci-dessous :

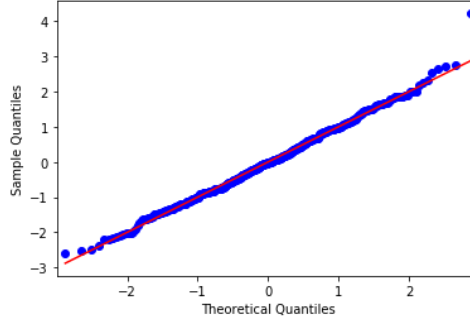


Figure 8 Q-Q plot

La relation entre les quantiles empiriques de l'échantillon normalisé et ceux théorique de la loi gaussienne centrée réduite est quasi-linéaire. Entre autres, on a $skew = 0.204$ et $kurtosis = 3.21$. L'hypothèse semble ici visuellement vérifiée.

Pour aller plus loin, on peut effectuer un test de Kolmogorov-Smirnov avec \mathcal{H}_0 : l'échantillon suit une loi normale. On obtient $p_{valeur} = 0.86 \gg 1\%$ et l'hypothèse \mathcal{H}_0 ne peut être rejetée.//

En regardant cette hypothèse au dernier temps T pour un $kurtosis$, $skew$ et p_{valeur} moyen sur l'ensemble des chemins, nous obtenons respectivement $\{0.20; 3.22; 0.66\}$. Nous pouvons en conclure que l'hypothèse de normalité est assez plausible dans la réalité mais de moins en moins vérifiée sur le long terme.

On calcule alors la MVA pour les deux portefeuilles P_1 et P_2 en utilisant l'approche LSMC adaptée au calcul des quantiles. On conserve le même jeu de paramètres pour les simulations. Par ailleurs, on tient à préciser que chacun des deux portefeuilles est exposé à **quatre facteurs de risque**.

En effet, le portefeuille P_1 est exposé aux fluctuations du taux swap et volatilité pour les couples de maturité X tenor $\{(3 \text{ mois}, 1 \text{ an}), (1 \text{ an}, 1 \text{ an})\}$, tandis que le portefeuille P_2 est exposé aux mêmes facteurs pour les couples $\{(20 \text{ ans}, 20 \text{ ans}), (20 \text{ ans}, 30 \text{ ans})\}$. **La dimension du problème de Metamodeling pour chacun des deux portefeuilles est $d = 4$.**

On se propose de comparer le résultat du calcul de la MVA par LSMC en utilisant en premier lieu le vecteur des facteurs de risque puis ensuite la valeur du portefeuille comme **régresseur**.

Dans le cas où le vecteur des facteurs de risque est choisi comme régresseur, on utilise une **régression polynomiale multidimensionnelle de degré $R_{multi-d} = 3$** , le nombre de coefficients β à estimer est de $D = D(3, d) = 35$. Sinon, dans le cas où la valeur du portefeuille est choisie comme régresseur, on utilise une **régression par polynômes orthogonaux (Laplace, Laguerre, Chebychev) de degré $R = 8$** . le nombre de coefficients β à estimer est $R + 1 = 9$.

En utilisant les mêmes coefficients de scaling, on obtient les profils de la MVA suivants :

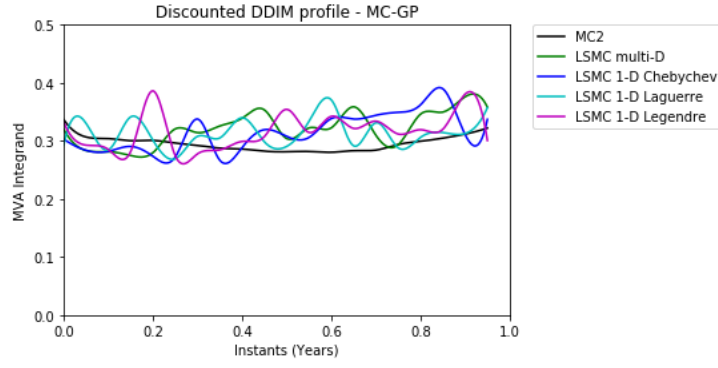


Figure 9 Profil de la MVA-LSMC pour P_1

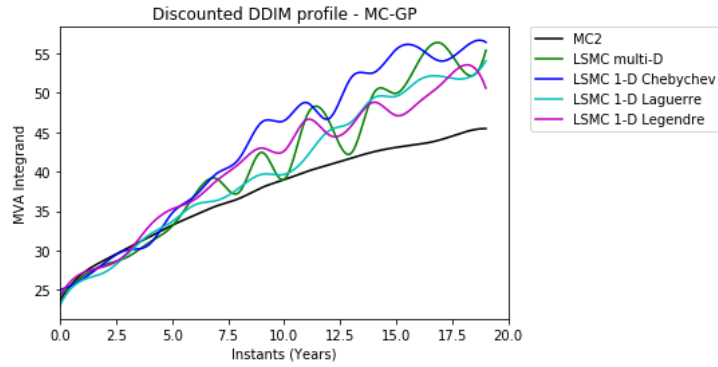


Figure 10 Profil de la MVA-LSMC pour P_2

On constate une divergence à partir de 10 ans due au fait que la calibration devient de moins en moins bonne, détériorant ainsi l'estimation des moments. Pour réduire cet écart, on peut augmenter le nombre de chemins et imposer quelques contraintes.

De plus, on remarque que les performances obtenues sont supérieures en prenant le vecteur de facteurs de risque comme régresseur. En effet, en prenant le processus $(V_t)_t$ représentant la valeur du portefeuille, on suppose que l'information du marché peut se résumer en une unique variable dans \mathbb{R} et que ce processus est Markovien. Ces deux hypothèses sont fausses et induisent une erreur supplémentaire.

5.2.3 Calcul de la MVA par MC-GP

On calcule la MVA sur les portefeuilles P_1 et P_2 en utilisant l'algorithme **MC-GP**. On fixe la taille de la base d'apprentissage à $M = 100 \ll N_{p_1}$ et on compare les résultats générés par les 3 méthodes de **choix de la base d'apprentissage**, à savoir par génération des suites quasi-aléatoires (Sobol, LHS) ou bien par sélection de quantiles. Pour le modèle de régression par GP, on choisit de travailler avec le Kernel exponentiel (SE). En utilisant les mêmes coefficients de scaling, on obtient les profils de la MVA suivants :

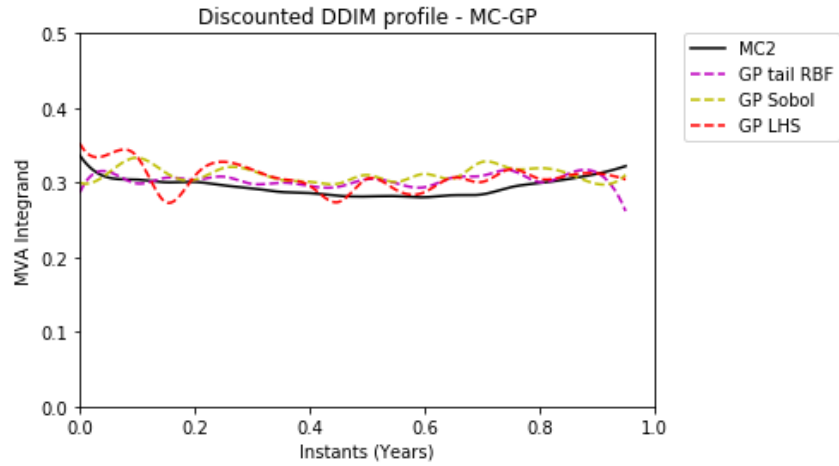


Figure 11 Profil de la MVA-GP pour P_1

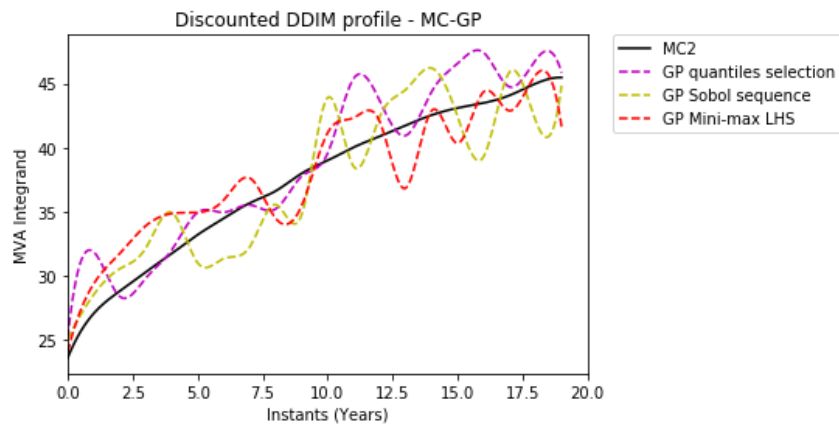


Figure 12 Profil de la MVA-GP pour P_2

On remarque que les différentes méthodes ondulent autour de la solution et ne semblent pas diverger sur le long terme.

5.2.4 Comparaison et interprétation des résultats

On compare les profils de la MVA retournés par les différents algorithmes (\mathbf{MC}^2 , LSMC, MC-GP), et ce pour les deux portefeuilles P_1 et P_2 :

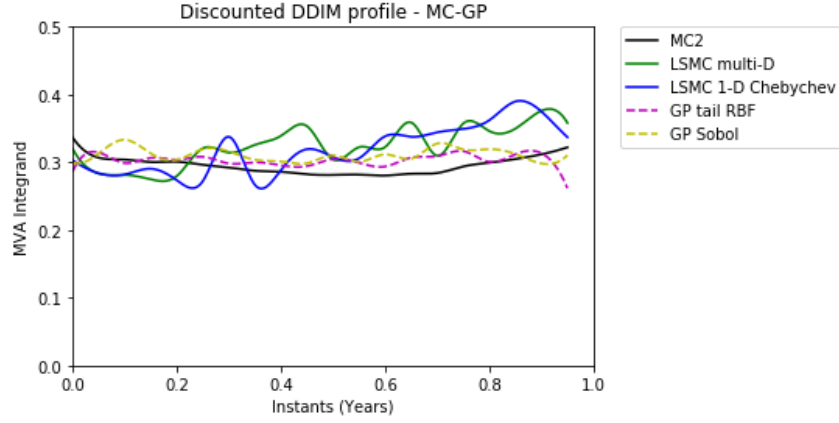


Figure 13 Profils selon algo de la MVA pour P_1

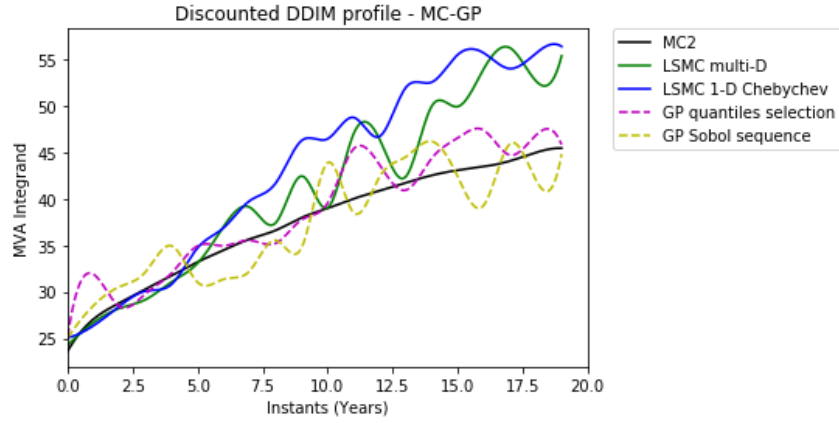


Figure 14 Profils selon algo de la MVA pour P_2

On remarque que pour le portefeuille P_1 de maturité courte, les deux approches (LSMC 1-D et multi-D, MC-GP) approche bien le montant de la DDIM calculé par MC^2 , avec une fluctuation moins importante pour l'approche MC-GP.

Cependant, pour le portefeuille P_2 , l'estimation de la DDIM estimée par LSMC a tendance à diverger contrairement à une regression GP qui converge. Cette sur-estimation non souhaitable peut s'expliquer par l'erreur de calibration pour des horizons lointains. En effet, la complexité du Metamodel doit augmenter avec la maturité et la forme retenue des courbes pour le LSMC ne permet plus d'interpoler correctement tous les points.

On résume les résultats dans le tableau ci-dessous :

| Algorithm | Relative Error P_1 | Relative Error P_2 | Computing time | Complexity |
|----------------------|----------------------|----------------------|----------------|--|
| MC^2 | \ | \ | 16h47min | $O(N_{t_1} \times N_{p_1} \times N_{t_2} \times N_{p_2} \times d)$ |
| LSMC Multi-D | 7.83% | 8.97% | 11.5min | $O(N_{t_1} \times N_{t_2} \times N_{p_1} \times D)$ |
| LSMC 1-D Chebychev | 8.81% | 9.90% | 7.7min | $O(N_{t_1} \times N_{t_2} \times N_{p_1} \times d)$ |
| LSMC 1-D Laguerre | 8.99% | 9.64% | 7.7min | $O(N_{t_1} \times N_{t_2} \times N_{p_1} \times d)$ |
| LSMC 1-D Legendre | 9.09% | 9.68% | 7.7min | $O(N_{t_1} \times N_{t_2} \times N_{p_1} \times d)$ |
| MC-GP quantiles | 4.29% | 3.54% | 21min | $O(N_{t_1} \times (M \times N_{t_2} \times N_{p_2} \times d + N_{optim} M^3))$ |
| MC-GP Sobol sequence | 3.44% | 3.14% | 21min | $O(N_{t_1} \times (M \times N_{t_2} \times N_{p_2} \times d + N_{optim} M^3))$ |
| MC-GP Mini-max LHS | 4.62% | 4.32% | 21min | $O(N_{t_1} \times (M \times N_{t_2} \times N_{p_2} \times d + N_{optim} M^3))$ |

Avec $D = D(d, R) = |\{(i_1, i_2, \dots, i_d), i_j \leq R, \sum_{j=1}^d i_j \leq R\}|$ le nombre de coefficients de la régression LSMC multidimensionnelle.

En comparant les performances obtenues par LSMC à celles obtenues par l'algorithme MC-GP, on s'aperçoit que ce dernier fournit de meilleurs résultats puisque aucune hypothèse sur la distribution n'est faite contrairement à l'algorithme LSMC.

Finalement, On remarque un compromis entre le temps de calcul et la précision de l'approche utilisée. D'après nos résultats, l'algorithme MC-GP est celui qui assure le meilleur compromis pour les deux portefeuilles P_1 et P_2 , avec une dimension du problème de $d = 4$. En analysant les complexités algorithmiques, on a $M \times N_{t_2} \times N_{p_2} \times d + N_{optim}M^3 \ll N_{p_1} \times N_{t_2} \times N_{p_2} \times d$ et donc $C_{MC-GP}^{MVA} \ll C_{MC^2}^{MVA}$.

Cependant, si la dimension du problème est grande, on aura besoin d'une base d'apprentissage de taille M conséquente et le terme en M^3 ne pourra alors être négligé, rendant le GP inutilisable. Une réduction de dimension (ACP, Impact en sensibilité,...) sera alors nécessaire.

5.2.5 Estimation des moments par Processus Gaussien Stochastique

Pour finir, on peut se demander ce qu'il se passerait si on estimait directement les moments de la méthode LSMC par processus gaussien. Nous ne rentrerons pas dans les détails mais le lecteur intéressé pourra se référer à [Bruce Ankenman \(2008\)](#) pour la théorie.

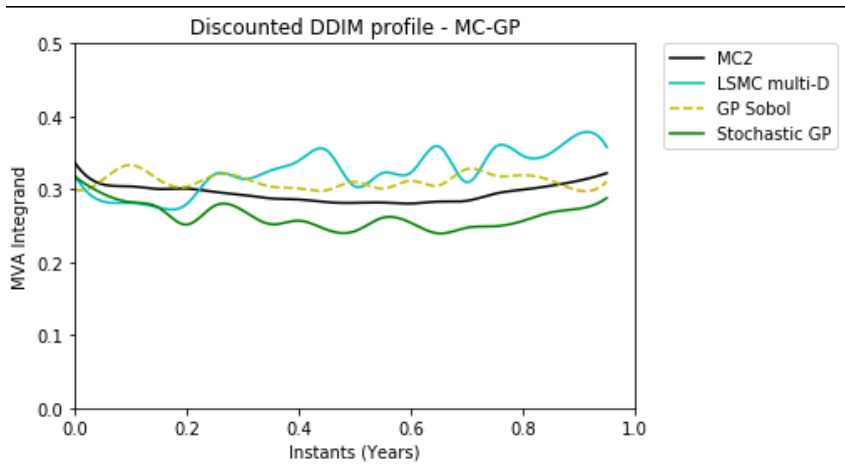


Figure 15 Profils en GP sto de la MVA pour P_1 et $N_{P1} = 200$

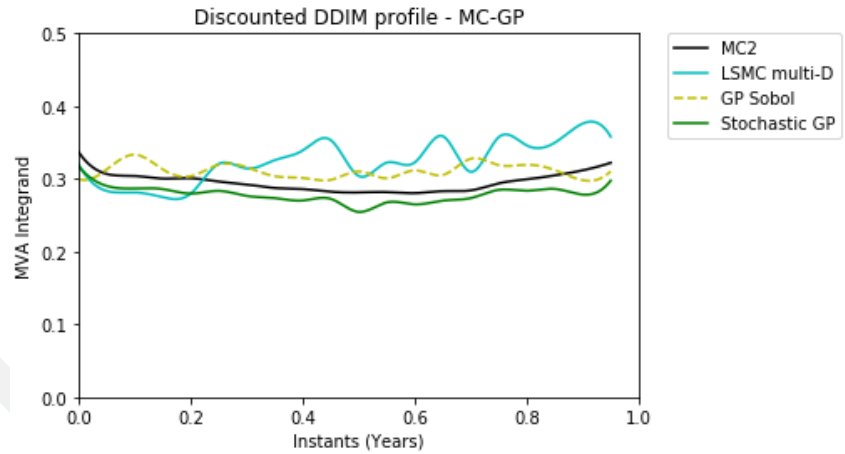


Figure 16 Profils en GP sto de la MVA pour P_1 et $N_{P1} = 500$

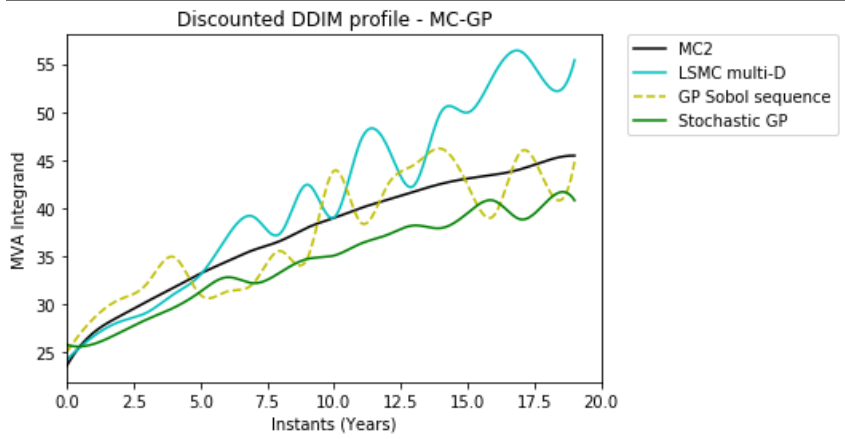


Figure 17 Profils en GP sto de la MVA pour P_2 et $N_{P1} = 200$

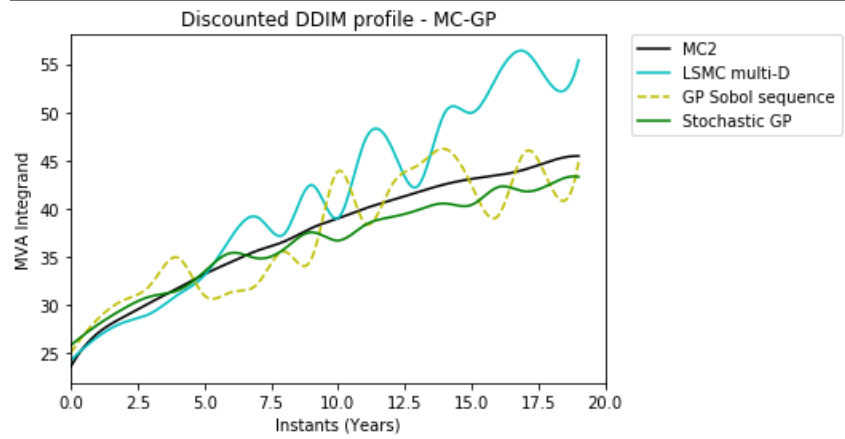


Figure 18 Profils en GP sto de la MVA pour P_2 et $N_{P1} = 500$

| Algorithm | Relative Error P_1 | Relative Error P_2 | Computing time |
|---------------------------------|----------------------|----------------------|----------------|
| MC ² | \ | \ | 16h47min |
| LSMC Multi-D | 7.83% | 8.97% | 11.5min |
| MC-GP Sobol sequence | 3.44% | 3.14% | 21min |
| MC-GP sto quantile N_{P1} 200 | 10.85% | 7.96% | 51min |
| MC-GP sto quantile N_{P1} 500 | 5.63% | 2.36% | 2h18 |

On remarque que l'estimation est très bonne mais est cependant très coûteuse en temps de calcul du fait de la calibration du processus gaussien en N_{P1}^3 . En effet, la regression par processus gaussien permet d'améliorer l'estimation des moments par rapport au LSMC. Cependant, cette dernière méthode impose une dimension faible du problème du fait que le nombre de chemins générés doit rester petit. Enfin, on remarque une bonne stabilité de l'estimation ; contrairement aux autres méthodes, les oscillations sont considérablement réduites.

6 Conclusion et améliorations possibles

Cette étude propose l'utilisation du Metamodeling pour le calcul de la MVA par projections futures des montants de l'IM en risque neutre. L'avantage étant de pouvoir utiliser par la suite des méthodes de Machine Learning afin de diminuer la complexité du problème. Pour ce faire, un facteur de conversion "monde réel/risque neutre" est appliqué en accord avec les pratiques de place.

Les méthodes de régression qui reposent sur l'algorithme du Least Square Monte Carlo sont souvent proposées dans la littérature comme alternative au calcul direct vu la simplicité de les mettre en place. De plus, elles garantissent un très bon compromis entre temps de calcul et précision. Cependant, cette méthode implique de supposer localement la loi de probabilité sous-jacente. En effet, la méthode LSMC repose sur l'estimation de moments et une estimation du quantile à l'aide de ces derniers est donc nécessaire. Nous avons vu que cette hypothèse ne

semble pas être problématique pour des maturités courtes mais est cependant de moins en moins vérifiée avec la durée de vie du portefeuille. De plus, la calibration des polynômes de régression semble elle aussi se détériorer avec la maturité.

Afin de s'affranchir de cette hypothèse et améliorer l'estimation du quantile, nous proposons d'estimer directement le quantile via un algorithme de processus gaussien. Si les performances sont meilleures, ce dernier impose néanmoins un temps de calcul beaucoup plus conséquent. Pour diminuer cet effet, une méthode d'optimisation de points de calibration a été proposée. Enfin, nous avons regardé succinctement ce que donnerait l'estimation des moments par processus gaussien stochastique. En effet, ce dernier semble pouvoir mieux estimer les moments que les polynômes orthogonaux. Néanmoins, même si les performances sont meilleures, le temps de calcul est cependant trop important en pratique. Une amélioration possible consisterait à traiter ce problème de la dimension plus spécifiquement en allouant plus de temps de calcul dans la recherche des points optimaux. A titre d'exemple, [Liu \(2010\)](#) propose de rajouter au fur et à mesure les points qui améliorent le plus l'estimation sur l'ensemble des simulations.

Références

- F. Anfuso, D. Aziz, P. Giltinan, and K. Loukopoulou. Monte carlo statistical method. *Technometrics*, 2000.
- B. Barlett. Hedging under sabr model. *Wilmott magazine*, 2006.
- J. S. Bruce Ankenman, Barry L. Nelson. Stochastic kriging for simulation metamodeling. *operations research*, 2008.
- C. Burgard and M. Kjaer. Pde representations of options with bilateral counterparty risk and funding costs. *Journal of Credit Risk*, 2011).
- J. Chan, S. Zhu, and B. Tourtzevitch. Practical approximation approaches to forecasting and backtesting initial margin requirements. 2017.
- H. Geman, N. E. Karoui, and J. Rochet. Changes of numeraire, changes of probability measure and option pricing. *Journal of Applied Probability*, 1995.
- A. Green. Xva :credit, funding and capital valuation adjustments. 2015.
- A. Green and C. Kenyon. Mva : Initial margin valuation adjustment by replication and regression adjustments. 2015.
- P. Hagan, D. Kumar, A. Lesniewski, and D. Woodward. Managing smile risk. *Wilmott 84-108*, 2002.
- ISDA. Isda standard initial margin model for non-cleared derivatives, international swaps and derivatives association. 2014-2019.
- M. Liu. Stochastic kriging for efficient nested simulation of expected shortfall. *National Science Foundation*, 2010.
- F. A. Longstaff and E. S. Schwartz. Valuing american options by simulation : A simple least-squares approach. *The Review of Financial Studies*, 2001.
- X. Ma, S. Spinner, A. Venditti, Z. Li, and S. Tang. Initial margin simulation with deep learning. 2019.
- M. McKay, R. Beckman, and W. Conover. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 1979.
- I. Sobol. Distribution of points in a cube and approximate evaluation of integrals. *U.S.S.R Comput. Maths. Math. Phys*, 1967.
- M. Zeron and I. Ruiz. Dynamic initial margin via chebychev tensors. 2020.

7 Annexes

7.1 Annexe A : Généralités sur le marché des taux, mesures de probabilités et formules de Pricing

7.1.1 Généralités

Définition 2. Compte du marché monétaire *Un compte de marché monétaire, plus communément appelé 'Bank account' est un investissement sans risque où les gains sont cumulés de manière continue au taux sans risque du marché. On note sa valeur $B(t)$, et on admet qu'elle suit l'équation différentielle suivante :*

$$dB(t) = r_t B(t) dt, \quad B(0) = 1 \quad (49)$$

Où r_t est une fonction positive du temps qui représente le taux sans risque instantané. On obtient alors par lemme d'Itô :

$$B(t) = \exp \left(\int_0^t r_s ds \right) \quad (50)$$

Définition 3. Facteur d'actualisation stochastique *Le facteur d'actualisation stochastique, plus communément appelé 'Discount Factor' $D(t, T)$ entre deux instants t et T , est le montant en date t qui est équivalent à une unité monétaire payable à la date T , il est donné par :*

$$D(t, T) = \frac{B(t)}{B(T)} = \exp \left(- \int_t^T r_s ds \right) \quad (51)$$

On remarque d'après la formule ci-dessus que si le processus du taux d'intérêt instantané (court) r_t est stochastique, le DF est à son tour stochastique.

Définition 4. Obligation Zéro-Coupon *Une obligation zéro-coupon de maturité T est un actif qui garantit à son détenteur le paiement d'une unité de monnaie à l'instant T sans aucun paiement intermédiaire. La valeur de ce contrat à un instant $t < T$ est notée $P(t, T)$. En découle alors que $P(T, T) = 1$ pour tout T .*

Si on se place alors dans un instant t , une obligation zéro-coupon de maturité T est un contrat qui établit la valeur présente d'une unité de monnaie payable en T , sa maturité. On voit bien que si le taux sans risque r est déterministe, le DF est aussi déterministe et on a $D(t, T) = P(t, T)$ pour tout couple (t, T) . Cependant, si le taux est stochastique, $D(t, T)$ à l'instant t est une variable aléatoire qui dépend de l'évolution future du taux sans risque entre t et T , par contre, le prix $P(t, T)$ de l'obligation zéro-coupon à l'instant t est le prix d'un contrat, il doit alors être déterministe, non aléatoire.

En effet, dans un univers risque-neutre, on montre que les prix actualisés au taux sans risque forment des martingales sous \mathbb{Q} .

Ainsi, si on suppose que H est le prix d'un actif qui suit un processus (\mathcal{F}_t) -adapté, intégrable de payoff H_T à maturité T , sa valeur en $t < T$ est donnée par :

$$H_t = \mathbb{E}^{\mathbb{Q}} [D(t, T) H_T \mid \mathcal{F}_t] \quad (52)$$

En l'occurrence, si on considère la valeur à $t < T$ du contrat Obligation Zéro-Coupon $P(t, T)$, On a $P_T = P(T, T) = 1$ et donc :

$$P(t, T) = \mathbb{E}^{\mathbb{Q}} [D(t, T) \mid \mathcal{F}_t] = \mathbb{E}^{\mathbb{Q}} \left[e^{-\int_t^T r_s ds} \mid \mathcal{F}_t \right] \quad (53)$$

Le prix de l'obligation zéro-coupon est donc l'espérance sous la mesure risque-neutre du DF.

Définition 5. Taux de rendement continu *Le taux de rendement continu à un instant $t < T$ est le taux constant auquel un investissement de $P(t, T)$ unité de monnaie à t croît continûment pour donner une unité de monnaie à T . On le note $R(t, T)$ et il est donné par la formule suivant qui traduit la définition :*

$$P(t, T) e^{R(t, T) \tau(T, t)} = 1 \quad (54)$$

Ou sous une forme équivalente :

$$R(t, T) := - \frac{\ln P(t, T)}{\tau(t, T)} \quad (55)$$

Avec $\tau(t, T)$ le temps entre t et T , exprimée dans une convention de base annuelle, par exemple $\tau(t, T) = \frac{30}{360}$

on peut montrer que $r(t) = \lim_{T \rightarrow t+} R(t, T)$

Définition 6. Courbe des taux La courbe des taux, communément appelée 'term structure' ou 'yield curve' à l'instant t est définie par le graphe de la fonction :

$$T \mapsto R(t, T), \quad T > t \quad (56)$$

On peut déduire la courbe Zéro-Coupon à partir de la Yield Curve grâce à la relation (7) qui lie les deux.

Définition 7. Tenor structure Une tenor structure est définie par la suite finie croissante $(T_i)_{i \in \{0, \dots, N\}}$ d'instants tels que :

$$0 \leq T_0 < T_1 < \dots < T_N < \infty \quad (57)$$

On définit alors le i^{eme} pas de temps :

$$\delta_i := T_{i+1} - T_i, \quad i \in \{0, \dots, N-1\} \quad (58)$$

Ici, on suppose implicitement que les instants (T_i) sont équidistants dans un but d'alléger les notations.

Taux swap

On utilise dans cette section la notion de "tenor structure" définie dans la section précédente. On définira le taux swap à partir du contrat Swap.

Définition 8. Contrat Swap Un Swap, de maturité T_m et de tenor $T_n - T_m$, défini sur une structure $(T_i)_{i \in \{m, \dots, n\}}$ est un contrat entre deux contreparties qui échangent sur chaque période le taux forward F^i , $i \in \{m, \dots, n-1\}$ contre un taux fixe K .

Dans le cadre d'un Swap **payeur**, l'acheteur du contrat paie le taux fixe K et reçoit les taux variables F_t^i . Dans un swap **receiver**, l'acheteur reçoit le taux fixe K et paie les taux variables F_t^i .

Si on se place en $t \leq T_m$, on note $\text{Swap}_t^{m,n}$ la valeur du contrat :

$$\begin{aligned} \text{Swap}_{t,\phi}^{m,n} &:= \sum_{i=m}^{n-1} \delta_i P(t, T_{i+1}) \phi (F_t^i - K) \\ &= \sum_{i=m}^{n-1} P(t, T_i) - P(t, T_{i+1}) - \delta_i P(t, T_{i+1}) K \\ &= \phi (P(t, T_m) - P(t, T_n) - K \sum_{i=m}^{n-1} \delta_i P(t, T_{i+1})) \end{aligned} \quad (59)$$

Avec $\phi = 1$ pour un swap payeur et $\phi = -1$ pour un swap receiver.

On remarque que, par AOA, la valeur du taux fixe K est choisie de telle sorte que la valeur du contrat swap à t soit nulle :

$$\begin{aligned} 0 &= P(t, T_m) - P(t, T_n) - K_{eq} \sum_{i=m}^{n-1} \delta_i P(t, T_{i+1}) \\ \Leftrightarrow K_{eq} &= \frac{P(t, T_m) - P(t, T_n)}{\sum_{i=m}^{n-1} \delta_i P(t, T_{i+1})} \end{aligned} \quad (60)$$

Définition 9. Taux Swap Forward Le taux Swap Forward défini en $t \leq T_m$ pour la période $[T_m, T_n]$ est donné par :

$$S_t^{m,n} := \frac{P(t, T_m) - P(t, T_n)}{A_t^{m,n}} \quad (61)$$

Où $A_t^{m,n}$ est le numéraire swap, souvent appelé l'annuité, est donné par :

$$A_t^{m,n} := \sum_{i=m}^{n-1} \delta_i P(t, T_{i+1}) \quad (62)$$

Remarque : On peut écrire le taux swap forward comme un panier de taux forward :

$$\begin{aligned} S_t^{m,n} &= \frac{P(t, T_m) - P(t, T_n)}{A_t^{m,n}} \\ &= \frac{\sum_{i=m}^{n-1} P(t, T_i) - P(t, T_{i+1})}{A_t^{m,n}} \\ &= \sum_{i=m}^{n-1} \frac{\delta_i P(t, T_{i+1})}{A_t^{m,n}} F_t^i \\ &=: \sum_{i=m}^{n-1} \omega_i^{m,n}(t) F_t^i \end{aligned} \quad (63)$$

Avec les coefficients de pondération donnés par :

$$\omega_i^{m,n}(t) := \frac{\delta_i P(t, T_{i+1})}{A_t^{m,n}} \quad (64)$$

Définition 10. Swaption Une swaption est une option d'entrer dans un contrat swap à une date future T_E . Une swaption **payeuse** donne le droit à son détenteur de mettre en place un swap payeur à la date T_m , de durée $T_n - T_m$ au taux fixe K contre un taux forward variable. Une swaption **receiver** donne le droit à son détenteur de mettre en place un swap receiver de mêmes caractéristiques. On a en général $T_E = T_m$ la date d'exercice (maturité) de l'option qui coïncide avec la première date du tenor structure. A maturité (T_E), le payoff d'une swaption s'écrit :

$$V^{Swpt}(T_E) = \left[\phi \left(\sum_{i=m}^{n-1} \delta_i P(T_E, T_{i+1}) (F_{T_E}^i - K) \right) \right]^+ \quad (65)$$

Avec $\phi = 1$ pour une swaption payeuse et $\phi = -1$ pour une swaption receiver.

Remarque : On peut réécrire le payoff de la swaption en fonction du taux swap observé à la date T_E , on écrit :

$$\begin{aligned} V^{Swpt}(T_E) &= \left[\phi \left(\sum_{i=m}^{n-1} \delta_i P(T_E, T_{i+1}) (F_{T_E}^i - K) \right) \right]^+ \\ &= \left[\phi \left(\sum_{i=m}^{n-1} \delta_i P(T_E, T_{i+1}) F_{T_E}^i - K \sum_{i=m}^{n-1} \delta_i P(T_E, T_{i+1}) \right) \right]^+ \\ &= \left[\phi \left(\sum_{i=m}^{n-1} \delta_i P(T_E, T_{i+1}) F_{T_E}^i - A_{T_E}^{m,n} K \right) \right]^+ \\ &= A_{T_E}^{m,n} \left[\phi \left(\sum_{i=m}^{n-1} \frac{\delta_i P(T_E, T_{i+1})}{A_{T_E}^{m,n}} F_{T_E}^i - K \right) \right]^+ \\ &= A_{T_E}^{m,n} \left[\phi \left(\sum_{i=m}^{n-1} \omega_i^{m,n}(T_E) F_{T_E}^i - K \right) \right]^+ \\ &= A_{T_E}^{m,n} [\phi (S_{T_E}^{m,n} - K)]^+ \end{aligned} \quad (66)$$

On écrit donc le payoff de la swaption :

$$V^{Swpt}(T_E) = A_{T_E}^{m,n} [\phi (S_{T_E}^{m,n} - K)]^+$$

7.1.2 Changement de numéraire et mesures de probabilité

Définition 11. Numéraire On appelle **Numéraire** tout actif financier strictement positif et ne payant pas de dividendes.

Remarque : Généralement, le numéraire est un actif de référence qu'on choisit pour normaliser les prix des différents actifs financiers. Si on choisit par exemple un numéraire Z , on ne considère plus les prix $S_k, k \in 1, \dots, K$ des actifs eux mêmes mais plutôt leurs prix relatifs dans le numéraire Z : $\frac{S_k}{Z}, k \in 1, \dots, K$

Théorème 1. Changement de Numéraire Soit $(\Omega, \mathcal{F}, (\mathcal{F}_t)_t, \mathbb{P})$ un espace probabiliste muni d'une filtration $(\mathcal{F}_t)_t$, défini par la donnée Ω et la probabilité historique \mathbb{P} .

On suppose qu'il existe un numéraire N et une probabilité associée \mathbb{Q}^N équivalente à la probabilité historique \mathbb{P} , tels que le prix de tout actif financier X dans le numéraire N soit une martingale sous \mathbb{Q}^N ; ie pour tout horizon $T \geq 0$ et tout $0 \leq t \leq T$:

$$\frac{X_t}{N_t} = \mathbb{E}^{\mathbb{Q}^N} \left[\frac{X_T}{N_T} \mid \mathcal{F}_t \right] \quad (67)$$

Alors : Pour tout numéraire U , il existe une mesure de probabilité \mathbb{Q}^U , équivalente à \mathbb{P} telle que le prix de tout actif financier Y dans le numéraire U soit une martingale sous cette mesure; ie pour tout horizon $T \geq 0$ et tout $0 \leq t \leq T$:

$$\frac{Y_t}{U_t} = \mathbb{E}^{\mathbb{Q}^U} \left[\frac{Y_T}{U_T} \mid \mathcal{F}_t \right] \quad (68)$$

De plus, la dérivée de Radon-Nikodym qui définit la mesure \mathbb{Q}^U s'écrit :

$$\frac{d\mathbb{Q}^U}{d\mathbb{Q}^N} = \frac{U_T N_0}{U_0 N_T} \quad (69)$$

Preuve : Geman et al. (1995)

Corollaire 1. Changement de Numéraire *Le théorème 1 assure le fait que le prix de tout actif financier X dans un numéraire N soit une martingale sous la mesure de probabilité \mathbb{Q}^N associée à ce numéraire.*

Mesure Forward-Neutre

On se place tout d'abord dans l'espace de probabilité historique et filtré $(\Omega, \mathcal{F}, (\mathcal{F}_t)_t, \mathbb{P})$.

On considère le taux forward F avec les trois dates t, T, S définis précédemment et on considère l'actif Obligation Zéro-Coupon $P(., S)$. Cet actif est strictement positif et ne verse pas de dividende, il peut alors servir de numéraire. D'après le Corollaire du Théorème 1, il existe une mesure de probabilité \mathbb{Q}^S équivalente à la probabilité historique \mathbb{P} sous laquelle le prix de n'importe quel actif financier X dans ce numéraire est une martingale.

On pose $X_t = P(t, T) - P(t, S)$ et on écrit alors :

$$\mathbb{E}^{\mathbb{Q}^S} \left[\frac{P(t, T) - P(t, S)}{P(t, S)} \mid \mathcal{F}_u \right] = \frac{P(u, T) - P(u, S)}{P(u, S)}, \text{ avec } 0 \leq u \leq t \leq T \leq S \quad (70)$$

On a donc d'après la définition du taux forward :

$$\mathbb{E}^{\mathbb{Q}^S} [F(t, T, S) \mid \mathcal{F}_u] = F(u, T, S) \quad (71)$$

Le taux forward est donc une martingale sous cette mesure de probabilité. D'où l'appellation **Forward-Neutre**.

Mesure Swap

On se place encore dans l'espace de probabilité historique et filtré $(\Omega, \mathcal{F}, (\mathcal{F}_t)_t, \mathbb{P})$ et on considère le taux swap $S_t^{m,n}$, on rappelle l'écriture de sa définition :

$$S_t^{m,n} := \frac{P(t, T_m) - P(t, T_n)}{A_t^{m,n}} \quad (72)$$

Avec l'annuité $A_t^{m,n} := \sum_{i=m}^{n-1} \delta_i P(t, T_{i+1})$.

De même que précédemment, l'annuité est un actif strictement positif ne versant pas de dividendes, il est donc un numéraire.

D'après le corollaire du théorème 1, il existe une mesure de probabilité $\mathbb{Q}^{m,n}$ équivalente à la probabilité historique \mathbb{P} , sous laquelle le prix dans le numéraire annuité de tout actif financier X est une martingale.

Si on prend alors $X_t = P(t, T_m) - P(t, T_n)$ pour $t \leq T_m$, alors le **taux swap est une martingale sous la mesure $\mathbb{Q}^{m,n}$** . On écrit alors :

$$\mathbb{E}^{\mathbb{Q}^{m,n}} [S_s^{m,n} \mid \mathcal{F}_t] = S_t^{m,n} \text{ avec } 0 \leq t \leq s \leq T_m \quad (73)$$

La mesure $\mathbb{Q}^{m,n}$ est appelée la **Mesure Swap** ou bien la **Mesure Annuité**.

7.1.3 Formule de Pricing Swap et Swaption Vanille

Swap de taux

La définition présentée dans la section précédente nous donne la forme du prix d'un swap de taux à une date $0 \leq t \leq T_m$:

$$Swap_{t,\phi}^{m,n} = \sum_{i=m}^{n-1} \delta_i P(t, T_{i+1}) \phi(F_t^i - K) \quad (74)$$

Cette écriture est équivalente à :

$$Swap_{t,\phi}^{m,n} = A_t^{m,n} \phi(S_t^{m,n} - K) \quad (75)$$

Swaption Vanille

D'après les résultats présentés dans la section précédente. Le taux swap dans le numéraire Annuité est une martingale sous la Mesure $\mathbb{Q}^{m,n}$.

Le payoff d'une swaption Vanille s'écrit de la forme :

$$V^{Swpt}(T_E) = A_{T_E}^{m,n} [\phi(S_{T_E}^{m,n} - K)]^+ \quad (76)$$

Le prix d'une swaption dans le numéraire Annuité est une martingale sous la mesure de probabilité $\mathbb{Q}^{m,n}$. On écrit pour tout $0 \leq t \leq T_E$:

$$\begin{aligned} \frac{V^{Swpt}(t)}{A_t^{m,n}} &= \mathbb{E}^{\mathbb{Q}^{m,n}} \left[\frac{V^{Swpt}(T_E)}{A_{T_E}^{m,n}} \mid \mathcal{F}_t \right] \\ &= \mathbb{E}^{\mathbb{Q}^{m,n}} \left[[\phi(S_{T_E}^{m,n} - K)]^+ \mid \mathcal{F}_t \right] \end{aligned} \quad (77)$$

On aboutit alors à la formule de pricing d'une Swaption Vanille pour tout $0 \leq t \leq T_E$:

$$V^{Swpt}(t) = A_t^{m,n} \mathbb{E}^{\mathbb{Q}^{m,n}} \left[[\phi(S_{T_E}^{m,n} - K)]^+ \mid \mathcal{F}_t \right] \quad (78)$$

On observe que le pricing d'une Swaption nécessite la modélisation de la distribution terminale de $S_{T_E}^{m,n}$.

Remarque : Relation de parité

De même que pour les options Vanille sur Equity, on remarque que les prix d'une swaption payeuse et d'une swaption receiver sont liés par une relation de parité comme suit :

$$V_{payer}^{swpt}(t) - V_{receiver}^{swpt}(t) = A_t^{m,n}(S_t^{m,n} - K) \quad (79)$$

7.2 Annexe B : Modélisation stochastique des taux swap

On se place dans le nouvel espace de probabilités défini par la mesure Swap $\mathbb{Q}^{m,n} : (\Omega, \mathcal{F}, \mathbb{Q}^{m,n})$. On considère aussi un Mouvement Brownien Standard W sur cet espace et on note $(\mathcal{F}_t)_t$ sa filtration naturelle associée.

On sait que sous cette mesure, le taux swap $S_t^{m,n}$ est une \mathcal{F}_t -martingale. Donc d'après le **théorème de représentation des martingales** : Il existe un processus $\sigma_{m,n}(t, \omega)$ \mathcal{F}_t -adapté tel que :

$$dS_t^{m,n} = \sigma_{m,n}(t, \omega) dW_t \quad (80)$$

Ce théorème fondamental nous permet alors de poser un cadre général pour tous les modèles de diffusion des taux swap qui en découlent. Chaque modèle est basé sur une hypothèse sur le processus σ .

7.2.1 Modèle de Bachelier

Dans le cadre du modèle de Bachelier, on suppose que le processus $(\sigma_t)_t$ est presque sûrement constant et égal à un scalaire $\sigma_{m,n} > 0$. Le modèle s'écrit donc :

$$dS_t^{m,n} = \sigma_{m,n} dW_t \quad (81)$$

Si on considère alors deux instants $t \geq 0$ et $T \geq t$ alors le taux swap à T conditionnellement à sa donnée en t suit une loi normale. On écrit :

$$S^{m,n}(T) | S^{m,n}(t) = s \hookrightarrow \mathcal{N}(s, \sigma_{m,n}^2(T - t)) \quad (82)$$

Théorème 2. Formule de Bachelier On se place dans le modèle de Bachelier où le taux swap vérifie l'EDS :

$$dS_t^{m,n} = \sigma_{m,n} dW_t \quad (83)$$

Alors le prix de la swaption défini par :

$$\frac{V^{Swpt}(t)}{A_t^{m,n}} = \mathbb{E}^{\mathbb{Q}^{m,n}} \left[[\phi(S_{T_E}^{m,n} - K)]^+ | \mathcal{F}_t \right] \quad (84)$$

Est donné par la formule de Bachelier :

$$\frac{V^{Swpt}(t)}{A_t^{m,n}} = \text{Bachelier}(S_t^{m,n}, K, \sigma_{m,n} \sqrt{T_E - t}, \phi) \quad (85)$$

Avec :

$$\text{Bachelier}(F, K, \nu, \phi) = \nu \cdot \left[\Phi \left(\frac{\phi[F - K]}{\nu} \right) \cdot \frac{\phi[F - K]}{\nu} + \Phi' \left(\frac{\phi[F - K]}{\nu} \right) \right] \quad (86)$$

Et $\Phi(\cdot)$ la fonction de répartition de la loi normale centrée réduite.

Remarque : Dans ce modèle, on remarque que les taux swap peuvent être négatifs avec une probabilité non nulle.

7.2.2 Modèle de Black

Dans le modèle de Black, on suppose que le processus σ_t est une fonction linéaire du taux swap. On écrit :

$$dS_t^{m,n} = \sigma_{m,n} S_t^{m,n} dW_t \quad (87)$$

Si on considère alors deux instants $t \geq 0$ et $T \geq t$, alors le taux swap à T conditionnellement à sa donnée en t suit une loi log-normale, ie :

$$\ln(S_T^{m,n}) | \ln(S_t^{m,n}) = s \hookrightarrow \mathcal{N} \left(s - \frac{1}{2} \sigma_{m,n}^2(T - t), \sigma_{m,n}^2(T - t) \right) \quad (88)$$

Théorème 3. Formule de Black On se place dans le modèle de Black où le taux swap vérifie l'EDS :

$$dS_t^{m,n} = \sigma_{m,n} S_t^{m,n} dW_t \quad (89)$$

Alors le prix de la swaption défini par :

$$\frac{V^{Swpt}(t)}{A_t^{m,n}} = \mathbb{E}^{Q^{m,n}} \left[[\phi(S_{T_E}^{m,n} - K)]^+ \mid \mathcal{F}_t \right] \quad (90)$$

Est donné par la formule de Black :

$$\frac{V^{Swpt}(t)}{A_t^{m,n}} = \text{Black}(S_t^{m,n}, K, \sigma_{m,n} \sqrt{T_E - t}, \phi) \quad (91)$$

Avec :

$$\begin{aligned} \text{Black}(F, K, \nu, \phi) &= \phi \cdot [F \cdot \Phi(\phi \cdot d_1) - K \cdot \Phi(\phi \cdot d_2)] \\ d_{1,2} &= \frac{\ln(F/K)}{\nu} \pm \frac{\nu}{2} \end{aligned} \quad (92)$$

Et $\Phi(\cdot)$ la fonction de répartition de la loi normale centrée réduite.

Remarque : Contrairement au modèle de Bachelier, le modèle de Black ne permet pas la modélisation des taux swaps négatifs.

7.2.3 Modèle SABR

Le modèle SABR (Stochastic Alpha Beta Rho) est un modèle de volatilité stochastique qui vient pallier à la problématique du smile/skew observée sur les marchés financiers. En effet, les deux modèles présentés ci-dessus supposent que la volatilité des différentes swaptions est constante en fonction du strike K et de la maturité T_E . Le modèle SABR propose de modéliser les taux swaps forward avec une volatilité stochastique qui permet de fitter le smile observé sur les marchés. Le plus gros avantage du modèle SABR est le fait qu'on puisse obtenir une approximation de la volatilité implicite par formule fermée (travaux de [Hagan et al. \(2002\)](#)) en fonction du strike, de la maturité et des différents paramètres du modèle.

Pour un couple de maturité \times tenor donné (m, n) , le taux swap forward $S_t^{m,n}$ est modélisé sous la mesure Swap $\mathbb{Q}^{m,n}$ comme suit :

$$\begin{aligned} dS_t^{m,n} &= \alpha_t^{m,n} \cdot (S_t^{m,n})^{\beta_{m,n}} \cdot dW_t \\ S_0^{m,n} &= s_{m,n} \\ d\alpha_t^{m,n} &= \nu_{m,n} \cdot \alpha_t^{m,n} \cdot dZ_t \\ \alpha_0^{m,n} &= \alpha_{m,n} \\ d\langle W, Z \rangle_t &= \rho_{m,n} \cdot dt \end{aligned} \quad (93)$$

Avec :

- $s_{m,n}$ le taux swap initial.
- $\beta_{m,n} \in (0, 1)$ le paramètre d'élasticité
- $\alpha_t^{m,n}$ la volatilité stochastique, sa volatilité est notée $\nu_{m,n} > 0$ (la vol-vol) et $\alpha_{m,n}$ sa valeur initiale
- W et Z sont deux mouvements browniens sous la mesure swap, et corrélés par un coefficient $\rho_{m,n} \in (-1, 1)$

Remarques :

- Les paramètres du modèle SABR sont β , α , ρ , et ν .
- En pratique, le choix du β se fait *a priori* en amont. Souvent, on choisit $\beta = 0$ pour le modèle normal, $\beta = 1$ pour le modèle log-normal et $\beta = \frac{1}{2}$ pour le modèle CIR. Le choix du β n'a pas un grand effet sur la structure du smile produite, mais à un effet non négligeable sur les grecques SABR.

Modèle SABR shifté

On se propose d'étendre le modèle SABR aux taux swap négatifs en introduisant un shift $\lambda > 0$ bien choisit. On pose $\tilde{S}_t^{m,n} = S_t^{m,n} - \lambda$ où $S_t^{m,n}$ suit un modèle SABR standard sous la mesure swap. $\tilde{S}_t^{m,n}$ vérifie alors l'EDS :

$$\begin{aligned} d\tilde{S}_t^{m,n} &= \alpha_t^{m,n} \cdot (\tilde{S}_t^{m,n} + \lambda)^{\beta_{m,n}} \cdot dW_t \\ \tilde{S}_0^{m,n} &= s_{m,n} - \lambda \\ d\alpha_t^{m,n} &= \nu_{m,n} \cdot \alpha_t^{m,n} \cdot dZ_t \\ \alpha_0^{m,n} &= \alpha_{m,n} \\ d\langle W, Z \rangle_t &= \rho_{m,n} \cdot dt \end{aligned} \quad (94)$$

Remarque : Le paramètre du shift λ permet d'étendre le domaine des taux swap à $[-\lambda, +\infty[$

Volatilité Implicite SABR et Prix des Options

Hagan et al. (2002) ont présenté une approximation de la volatilité implicite normale dans le cadre du modèle SABR par une formule fermée en fonction du taux swap spot $S_t^{m,n}$, du strike K , de la maturité T et des différents paramètres du modèle SABR. Le prix d'une swaption est obtenu en injectant cette formule de volatilité implicite, qu'on adaptera dans le cadre de notre modèle SABR λ -shifté, dans la formule de Bachelier.

On note $\sigma_B(K, S)$ l'approximation asymptotique de la volatilité implicite normale sous SABR shifté, avec K le strike de l'option et S le taux swap spot :

$$\begin{aligned}\sigma_B(K, S) &= \frac{\alpha \left\{ 1 + \left[\frac{(1-\beta)^2}{24} \frac{\alpha^2}{(S'K')^{1-\beta}} + \frac{1}{4} \frac{\rho\beta v\alpha}{(S'K')^{(1-\beta)/2}} + \frac{2-3\rho^2}{24} v^2 \right] T \right\}}{(S'K')^{(1-\beta)/2} \left[1 + \frac{(1-\beta)^2}{24} \ln^2 \frac{S'}{K'} + \frac{(1-\beta)^4}{1920} \ln^4 \frac{S'}{K'} \right]} \times \frac{z}{\chi(z)} \\ z &= \frac{v}{\alpha} (S'K')^{(1-\beta)/2} \ln \frac{S'}{K'} \\ \chi(z) &= \ln \left[\frac{\sqrt{1-2\rho z + z^2} + z - \rho}{1 - \rho} \right] \\ S' &= S + \lambda \\ K' &= K + \lambda\end{aligned}\tag{95}$$

Avec β et λ fixés à l'avance, et α, ρ, ν les estimations des paramètres du modèle pour un couple (maturité, tenor) fixé.

Remarque : On peut montrer que la volatilité implicite à la monnaie $\sigma_{ATM} = \sigma_B(K, K)$ peut être approchée par la formule :

$$\ln \sigma_{ATM} \approx \ln \alpha - (1 - \beta) \ln(S + \lambda)\tag{96}$$

Une fois la volatilité implicite calculée, on l'injecte dans la formule de Bachelier pour obtenir le prix de la swaption :

$$V^{Swpt}(t) = A_t^{m,n} \times \text{Bachelier} \left(S_t^{m,n}, K, \sigma_B(K, S_t^{m,n}) \times \sqrt{T_E - t}, \phi \right)\tag{97}$$

Calibration du modèle SABR λ -shifté

Comme les deux paramètres β et λ sont fixés a priori, les seuls paramètres à estimer sont $\alpha_{m,n}, \rho_{m,n}, \nu_{m,n}$, et ce pour chaque couple (maturité, tenor) = (m, n) donné. Pour alléger les notations, et en fixant le couple (m, n) , on écrira désormais α, ρ et ν à la place de $\alpha_{m,n}, \rho_{m,n}, \nu_{m,n}$.

Deux approches d'estimation sont possibles :

- Estimer α, ρ et ν directement.
- Estimer ρ et ν directement, puis inférer la valeur de α à partir de ρ, ν et la volatilité à la monnaie σ_{ATM}

$$\sigma_{ATM} = \sigma_B(S, S) = \frac{\alpha \left\{ 1 + \left[\frac{(1-\beta)^2}{24} \frac{\alpha^2}{(S')^{2-2\beta}} + \frac{1}{4} \frac{\rho\beta v\alpha}{(S')^{1-\beta}} + \frac{2-3\rho^2}{24} v^2 \right] T \right\}}{(S')^{1-\beta}}\tag{98}$$

a) Première approche : Estimation directe de α, ρ et ν Pour un couple (m, n) donné, l'estimation des paramètres α, ρ et ν peut se faire en minimisant les erreurs entre les volatilités implicites du modèle et les volatilités du marché $\{\sigma_i^{mkt}\}$:

$$(\hat{\alpha}, \hat{\rho}, \hat{\nu}) = \arg \min_{\alpha, \rho, \nu} \sum_i \left\{ \sigma_i^{mkt} - \sigma_B(S, K_i; \alpha, \rho, \nu) \right\}^2\tag{99}$$

b) Deuxième approche : Estimation de ρ et ν

On peut réduire le nombre de paramètres à estimer en utilisant σ_{ATM} pour obtenir $\hat{\alpha}$ via la relation (98) au lieu de l'estimer directement. On estimera donc uniquement ρ et ν , l'estimateur de α proviendra de la recherche de la racine de l'équation cubique :

$$\left[\frac{(1-\beta)^2 T}{24(S+\lambda)^{2-2\beta}} \right] \alpha^3 + \left[\frac{\rho\beta v T}{4(S+\lambda)^{1-\beta}} \right] \alpha^2 + \left[1 + \frac{2-3\rho^2}{24} v^2 T \right] \alpha - \sigma_{ATM}(S+\lambda)^{1-\beta} = 0\tag{100}$$

$$(\hat{\alpha}, \hat{\rho}, \hat{\nu}) = \arg \min_{\alpha, \rho, \nu} \sum_i \left\{ \sigma_i^{mkt} - \sigma_B(S, K_i; \alpha(\rho, \nu), \rho, \nu) \right\}^2\tag{101}$$

Algorithme d'optimisation : Pour les deux approches, on utilise un algorithme de recherche de minimum global *multistart* avec plusieurs points de départ tirés aléatoirement selon la loi uniforme multidimensionnelle sur un compact bien choisi de \mathbb{R}^3 pour la première approche et de \mathbb{R}^2 pour la deuxième.

7.3 Annexe B : Grecques sous le modèle SABR λ -shifté

7.3.1 Grecques Bachelier

On montre dans cette section le Delta/Gamma/Vega d'un contrat swaption dans le cadre du modèle de Bachelier :

- $S(t)$ le taux swap observé à la date t .
- K le strike du contrat
- T la maturité du contrat
- σ la volatilité du taux, supposée constante sous Bachelier.
- $\phi \in \{-1, +1\}$ selon le type du contrat (payeur, receveur) respectivement.
- $P_{Bach}(t)$ le prix Bachelier du contrat en t
- $An(t)$ l'annuité en t .

Le prix du contrat est donné par la formule :

$$P_{Bach}(t) = An(t)\sigma\sqrt{T-t}[\mathcal{N}(\frac{\phi(S(t)-K)}{\sigma\sqrt{T-t}}) * \frac{\phi(S(t)-K)}{\sigma\sqrt{T-t}} + \mathcal{N}'(\frac{\phi(S(t)-K)}{\sigma\sqrt{T-t}})] \quad (102)$$

Où \mathcal{N} représente la fonction de répartition de la loi normale centrée réduite. On suppose ici que $t = 0$ et pour alléger les notations : $An(t) = 1$.

Le delta Bachelier est donné par la formule :

$$\Delta_{Bach} = \frac{\partial P_{Bach}(t)}{\partial S(t)} = \phi\mathcal{N}(\frac{\phi(S(t)-K)}{\sigma\sqrt{T}}) \quad (103)$$

Le vega Bachelier est donné par la formule :

$$\nu_{Bach} = \frac{\partial P_{Bach}(t)}{\partial \sigma} = \sqrt{T}\mathcal{N}'(\frac{S(t)-K}{\sigma\sqrt{T}}) \quad (104)$$

Le gamma Bachelier est donné par la formule :

$$\Gamma = \frac{\partial^2 P_{Bach}(t)}{\partial S(t)^2} = \frac{1}{\sigma\sqrt{T}}\mathcal{N}''(\frac{S(t)-K}{\sigma\sqrt{T}}) \quad (105)$$

7.3.2 Grecques SABR

Dans le cadre d'un modèle SABR, on approche le prix du contrat swaption par la formule fermée de Bachelier appliquée aux mêmes paramètres définis en amont, sauf la volatilité qui est désormais une fonction locale du taux swap, du strike, et d'autres paramètres du modèle. Sa formule approchée autour de la monnaie est donnée par Hagan *et al.*, notée sous la forme $\sigma_B(S, K, \alpha)$ avec α la valeur de volatilité stochastique SABR à l'instant ($t=0$). Le prix du contrat swaption s'écrit donc sous la forme :

$$P_{SABR}(t) = P_{Bach}(S(t), K, \sigma_B(S(t), K, \alpha)) \quad (106)$$

Le delta SABR est donné en utilisant la règle de dérivation en chaîne :

$$\begin{aligned} \Delta_{SABR} &= \frac{\partial P_{SABR}(t)}{\partial S(t)} \\ &= \frac{\partial P_{Bach}}{\partial S} + \frac{\partial P_{Bach}}{\partial \sigma} \frac{\partial \sigma_B}{\partial S} \\ &= \Delta_{Bach} + \nu_{Bach} \times \frac{\partial \sigma_B}{\partial S} \end{aligned}$$

Par le même raisonnement qu'avant, le Vega sous SABR est donné par la formule :

$$\begin{aligned}\nu_{SABR} &= \frac{\partial P_{SABR}(t)}{\partial \alpha} \\ &= \frac{\partial P_{Bach}}{\partial \sigma} \frac{\partial \sigma_B}{\partial \alpha} \\ &= \nu_{Bach} \times \frac{\partial \sigma_B}{\partial \alpha}\end{aligned}$$

Le Gamma SABR est déduit du Delta SABR :

$$\begin{aligned}\Gamma_{SABR} &= \frac{\partial \Delta_{SABR}}{\partial S(t)} \\ &= \Gamma_{Bach} + \nu_{Bach} \times \frac{\partial^2 \sigma_B}{\partial S^2}\end{aligned}$$

7.3.3 Grecques ajustés sous SABR shifté

Le taux swap et la volatilité SABR étant corrélés par un coefficient ρ , un bump du taux est vraisemblablement accompagné par un bump de volatilité et vis-versa. Pour cela, [Barlett \(2006\)](#) propose un ajustement dans les formules des grecques SABR :

$$\begin{aligned}\Delta_{Barlett} &= \frac{\partial P_{Bach}}{\partial S} + \frac{\partial P_{Bach}}{\partial \sigma} \left(\frac{\partial \sigma_B}{\partial S} + \frac{\partial \sigma_B}{\partial \alpha} \frac{\delta_S \alpha}{\delta S} \right) \\ &= \Delta_{SABR} + \nu_{Bach} \frac{\partial \sigma_B}{\partial \alpha} \frac{\delta_S \alpha}{\delta S}\end{aligned}$$

Avec $\frac{\partial \sigma_B}{\partial \alpha}$ calculé par différences finies et $\frac{\delta_S \alpha}{\delta S}$ qu'on peut facilement déduire de la dynamique de $(S, \alpha(t))$ sous SABR λ -shifté : On rappelle la dynamique SABR λ -shifté :

$$\begin{aligned}dS_t &= \alpha_t (S_t + \lambda)^\beta dW_t \\ d\alpha_t &= v \alpha_t \left(\rho dW_t + \sqrt{1 - \rho^2} dZ_t \right) \\ E \left[dW_t \left(\rho dW_t + \sqrt{1 - \rho^2} dZ_t \right) \right] &= \rho dt\end{aligned}$$

On déduit de cette dynamique que :

$$d\alpha_t = \frac{\rho v}{(S_t + \lambda)^\beta} dS_t + v \alpha_t \sqrt{1 - \rho^2} dZ_t \quad (107)$$

et donc :

$$\frac{\delta_S \alpha}{\delta S} = \frac{\rho v}{(S + \lambda)^\beta} \quad (108)$$

Par le même raisonnement, on écrit :

$$\begin{aligned}\nu_{Barlett} &= \frac{\partial P_{Bach}}{\partial \sigma} \left(\frac{\partial \sigma_B}{\partial \alpha} + \frac{\partial \sigma_B}{\partial S} \frac{\delta_\alpha S}{\delta \alpha} \right) \\ &= \nu_{SABR} + \nu_{Bach} \frac{\partial \sigma_B}{\partial S} \frac{\delta_\alpha S}{\delta \alpha}\end{aligned}$$

Avec :

$$\frac{\delta_\alpha S}{\delta \alpha} = \frac{\rho(S + \lambda)^\beta}{v} \quad (109)$$

$$dS_t = \frac{\rho(S_t + \lambda)^\beta}{v} d\alpha_t + S_t^\beta \alpha_t \sqrt{1 - \rho^2} dZ_t \quad (110)$$

$$\begin{aligned}
\Gamma_{Barlett} &= \frac{\partial \Delta_{Barlett}}{\partial S} \\
&= \frac{\partial}{\partial S} \left[\frac{\partial P_{Bach}}{\partial S} + \frac{\partial P_{Bach}}{\partial \sigma} \left(\frac{\partial \sigma_B}{\partial S} + \frac{\partial \sigma_B}{\partial \alpha} \frac{\delta_S \alpha}{\delta S} \right) \right] \\
&= \frac{\partial}{\partial S} \left[\Delta_{Bach} + \nu_{Bach} \left(\frac{\partial \sigma_B}{\partial S} + \frac{\partial \sigma_B}{\partial \alpha} \frac{\delta_S \alpha}{\delta S} \right) \right] \\
&= \frac{\partial}{\partial S} \left[\Delta_{SABR} + \nu_{Bach} \frac{\partial \sigma_B}{\partial \alpha} \frac{\delta_S \alpha}{\delta S} \right] \\
&= \Gamma_{SABR} + \nu_{Bach} \frac{\partial}{\partial S} \left[\frac{\partial \sigma_B}{\partial \alpha} \frac{\delta_S \alpha}{\delta S} \right]
\end{aligned}$$

$$\begin{aligned}
\frac{\partial}{\partial S} \left[\frac{\partial \sigma_B}{\partial \alpha} \frac{\delta_S \alpha}{\delta S} \right] &= \frac{\partial}{\partial S} \left[\frac{\partial \sigma_B}{\partial \alpha} \frac{\rho \nu}{(S + \lambda)^\beta} \right] \\
&= \frac{\partial^2 \sigma_B}{\partial S \partial \alpha} \frac{\rho \nu}{(S + \lambda)^\beta} - \beta \frac{\rho \nu}{(S + \lambda)^{\beta+1}} \frac{\partial \sigma_B}{\partial \alpha} \\
&= \frac{\rho \nu}{(S + \lambda)^\beta} \left(\frac{\partial^2 \sigma_B}{\partial S \partial \alpha} - \frac{\partial \sigma_B}{\partial \alpha} \frac{\beta}{S + \lambda} \right)
\end{aligned}$$

On a donc :

$$\Gamma_{Barlett} = \Gamma_{SABR} + \nu_{Bach} \frac{\rho \nu}{(S + \lambda)^\beta} \left(\frac{\partial^2 \sigma_B}{\partial S \partial \alpha} - \frac{\partial \sigma_B}{\partial \alpha} \frac{\beta}{S + \lambda} \right) \quad (111)$$

7.4 Annexe C : Changement de numéraire

7.4.1 Changement de numéraire : de la mesure annuité à la mesure risque neutre

Si on se place l'instant $t=0$ on a besoin de calculer $\tilde{\mathbb{E}}(\frac{IM_t}{B(t)} \mid F_0)$, avec un changement de numéraire entre $B(t)$ et une annuité $An(t)$ qui définit une probabilité \mathbb{Q} on a :

$$\tilde{\mathbb{E}}(\frac{IM_t}{B(t)} \mid F_0) = \mathbb{E}^{\mathbb{Q}} \left(\frac{IM_t}{B(t)} \cdot \frac{d\tilde{\mathbb{P}}}{d\mathbb{Q}} \Big|_{\mathcal{F}_t} \mid F_0 \right) \quad (112)$$

Avec la dérivée de Radon-Nikodym donnée par le théorème de changement de numéraire :

$$\frac{d\tilde{\mathbb{P}}}{d\mathbb{Q}} \Big|_{\mathcal{F}_t} = \frac{B(t)}{An(t)} \cdot \frac{An(0)}{B(0)} \quad (113)$$

On obtient alors :

$$\tilde{\mathbb{E}}(\frac{IM_t}{B(t)} \mid F_0) = An(0) \cdot \mathbb{E}^{\mathbb{Q}}(\frac{IM_t}{An(t)} \mid F_0) \quad (114)$$

car $An(0)$ est F_0 -mesurable et $B(0) = 1$

7.4.2 Calcul des drifts sous la mesure annuité commune - Preuve

Le but ici est de calculer les drifts associés aux taux swap $(S^{m,n})_{(m,n)}$ dans le modèle SABR-shifté multivarié défini sous la mesure annuité commune $\mathbb{Q}^{1,2}$, associée au numéraire annuité $(A^{1,2}(t))_t$. Sous ce numéraire, seul le taux swap $S^{1,2}$ est une martingale, son terme de drift est alors nul.

Pour $m \geq 1$ et $n \geq 3$, le drift $\mu^{m,n}(t)$ s'écrit :

$$\begin{aligned}
\mu^{m,n}(t)dt &= -\frac{A_t^{1,2}}{A_t^{m,n}} d \langle S_t^{m,n}, \frac{A_t^{m,n}}{A_t^{1,2}} \rangle \\
&= -\frac{A_t^{1,2}}{A_t^{m,n}} \sum_{i=\max(m+1,3)}^n X_i
\end{aligned} \quad (115)$$

Avec X_i défini pour $3 \leq i \leq n$ comme la différentielle totale de la covariation quadratique $\langle S_t^{m,n}, \frac{P_i(t)}{P_2(t)} \rangle$ qu'on obtient avec la relation de récurrence :

$$\begin{aligned} X_3 &= -\frac{P_3(t)}{P_2(t)(1 + S_t^{2,3})} d \langle S_t^{m,n}, S_t^{2,3} \rangle \\ X_i &= \frac{P_i(t)}{P_{i-1}(t)} X_{i-1} - \frac{P_i(t)}{P_2(t)(1 + S_t^{i-1,i})} d \langle S_t^{m,n}, S_t^{i-1,i} \rangle \text{ pour } i > 3 \end{aligned} \quad (116)$$

Avec $P_i(t) = P(t, T_i)$ le prix en t du ZC de maturité T_i

En utilisant cette formule de récurrence, on obtient une formule générique pour $(X_i)_{i \in \{3, \dots, n\}}$ comme suit :

$$\begin{aligned} X_i &= \sum_{j=3}^i \frac{P_i(t)}{P_2(t)} \frac{d \langle S_t^{m,n}, S_t^{i-1,i} \rangle}{1 + S_t^{j-1,j}} \\ &= \sum_{j=3}^i \frac{P_i(t)}{P_2(t)} \frac{\alpha_t^{m,n} (S_t^{m,n} + \lambda)^{\beta_{m,n}} \times \alpha_t^{j-1,j} (S_t^{j-1,j} + \lambda)^{\beta_{j-1,j}} \times \rho_{(m,n),(j-1,j)}}{1 + S_t^{j-1,j}} dt \end{aligned}$$

Pour prouver cette proposition, on s'appuie sur le lemme suivant :

Lemme 1. Soient X, Y, N trois actifs. Si on note $R = \frac{X}{Y}$, le terme de drift de R dans le numéraire N est donné par :

$$\mu_{R,N} dt = -\frac{N}{Y} d \langle R, \frac{Y}{N} \rangle_t \quad (117)$$

Si de plus on considère un nouveau numéraire M , alors le drift $\mu_{R,M}$ de R dans le numéraire M est relié à celui défini dans le numéraire N par l'équation :

$$\mu_{R,M} dt = \mu_{R,N} dt - d \langle R, \frac{N}{M} \rangle_t \frac{M}{N} \quad (118)$$

On applique cela dans notre cas d'étude, pour un couple (m, n) donné, on a :

$$S_t^{m,n} = \frac{P(t, T_m) - P(t, T_n)}{A_t^{m,n}} \quad (119)$$

On pose alors $X_t = P(t, T_m) - P(t, T_n)$, $Y_t = N_t = A_t^{m,n}$, X , et $M_t = A_t^{1,2}$. On applique la formule du drift à $R = \frac{X}{Y} = S^{m,n}$, on obtient :

$$\mu_{S^{m,n}, A^{1,2}} dt = \mu_{S^{m,n}, A^{n,m}} dt - d \langle S^{m,n}, \frac{A^{n,m}}{A^{1,2}} \rangle_t \frac{A^{1,2}}{A^{m,n}} \quad (120)$$

$\mu_{S^{m,n}, A^{1,2}}$ représente le drift de $S^{m,n}$ dans la mesure annuité commune associée au numéraire $A^{1,2}$, on a alors $\mu_{S^{m,n}, A^{1,2}}(t) = \mu^{m,n}(t)$. Par ailleurs on a :

$$\begin{aligned} \mu_{S^{m,n}, A^{n,m}} dt &= -\frac{A^{n,m}}{A^{n,m}} d \langle S^{m,n}, \frac{A^{n,m}}{A^{n,m}} \rangle_t \\ &= -d \langle S^{m,n}, 1 \rangle_t \\ &= 0 \end{aligned}$$

Ce qui est cohérent avec le fait que $S^{m,n}$ est une martingale sous sa mesure annuité associée au numéraire $A^{m,n}$, son drift est donc nul. Finalement on obtient :

$$\mu^{m,n}(t) dt = -\frac{A_t^{1,2}}{A_t^{m,n}} d \langle S_t^{m,n}, \frac{A_t^{m,n}}{A_t^{1,2}} \rangle \quad (121)$$

Pour $m \geq 1$ et $n \geq 3$, on peut écrire ce drift sous la forme :

$$\begin{aligned} \mu^{m,n}(t) dt &= -\frac{A_t^{1,2}}{A_t^{m,n}} \sum_{i=\max(m+1,3)}^n X_i \\ \text{Avec : } X_i &= d \langle S_t^{m,n}, \frac{P_i(t)}{P_2(t)} \rangle \\ &= \sum_{j=3}^i \frac{P_i(t)}{P_2(t)} \frac{\alpha_t^{m,n} (S_t^{m,n} + \lambda)^{\beta_{m,n}} \times \alpha_t^{j-1,j} (S_t^{j-1,j} + \lambda)^{\beta_{j-1,j}} \times \rho_{(m,n),(j-1,j)}}{1 + S_t^{j-1,j}} dt \end{aligned}$$

Nexialog Consulting est un cabinet de conseil spécialisé en Banque et en Assurance. Organisés autour de 3 domaines d'activité - Risques Bancaires, Financiers & Assurantiels - nous intervenons au sein des équipes métiers afin de les accompagner depuis le cadrage jusqu'à la mise en œuvre de leurs projets. Associant innovation et expertise, le savoir-faire de notre cabinet a permis de consolider notre positionnement sur ce segment et de bénéficier d'une croissance forte et régulière.

Les besoins de nos clients étant en constante évolution, nous nous adaptons continuellement pour proposer le meilleur accompagnement. Le département R&D de Nexialog Consulting se donne pour objectif de proposer des solutions innovantes à des problématiques métier ou d'actualité. Pour cela, nous nous appuyons sur des bibliothèques internes et sur le travail de nos consultants. Le pôle R&D Nexialog a également pour mission de former les collaborateurs sur l'évolution des techniques et la réglementation en lien avec leur activité.

Site web du cabinet : <https://www.nexialog.com>

Publications : <https://www.nexialog.com/publications-nexialog/>

Contacts

Ali BEHBAHANI
Associé, Fondateur
Tél : + 33 (0) 1 44 73 86 78
Email : abebahani@nexialog.com

Christelle BONDOUX
Associée, Directrice commerciale
Tél : + 33 (0) 1 44 73 75 67
Email : cbondoux@nexialog.com

Adrien MISKO
Manager R&D
Email : amisko@nexialog.com