



HAL
open science

Méthodes numériques efficaces pour la valorisation des GMWB

Tarik Ben Zineb

► **To cite this version:**

Tarik Ben Zineb. Méthodes numériques efficaces pour la valorisation des GMWB. Probabilités [math.PR]. Ecole Polytechnique X, 2012. Français. NNT: . pastel-00800761

HAL Id: pastel-00800761

<https://pastel.hal.science/pastel-00800761v1>

Submitted on 14 Mar 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THESE

présentée pour obtenir le titre de

Docteur de l'ECOLE POLYTECHNIQUE

Spécialité : Mathématiques appliquées

Tarik BEN ZINEB

Méthodes numériques efficaces pour la valorisation des GMWB

préparée au CMAP-Ecole Polytechnique et à
AXA Hedging Services

Directeur de thèse : Emmanuel Gobet

Co-directeur : Renaud Caillet

Jury

Bruno Bouchard	Rapporteur
Renaud Caillet	Co-directeur
Emmanuel Gobet	Directeur
Anis Matoussi	Examineur
Nizar Touzi	Président
Alexandre Tsybakov	Examineur
Stephane Villeneuve	Rapporteur

Dédiée à

Mes parents, ma femme et ma famille.

Remerciements

J'aimerais tout d'abord remercier Emmanuel Gobet qui a été un directeur de thèse exceptionnel et visionnaire. Ses qualités scientifiques aussi bien qu'humaines qui se sont manifesté par ses idées originales , ses remarques très pertinentes, sa compréhension et sa patience ont été un facteur primordial pour la réussite de ce travail. Il était toujours présent au bon moment pour relancer ma motivation pendant les moment difficiles. Cette thèse lui doit beaucoup. Je lui exprime aussi ma grande reconnaissance pour sa relecture minutieuse de mes notes qui a débusqué plusieurs erreurs.

Je remercie également Renaud Caillet, mon responsable chez Axa Hedging Services, qui a été très collaboratif et dont les remarques ont beaucoup contribué à l'amélioration de ma thèse. Je lui suis reconnaissant aussi par l'effort qu'il a consacré à lecture de ce manuscrit pendant un temps très limité.

Je voudrais adresser mes chaleureux remerciements à Bruno Bouchard et Stephane Villeneuve pour avoir consacré une grande partie de leur temps précieux afin de rapporter cette thèse. Je suis très touché que Nizar Touzi ait accepté de présider le jury. Je remercie également Anis Matoussi et Alexandre Tsybakov d'avoir voulu faire partie de mon jury.

Merci aux membres du CMAP où j'ai effectué ma thèse et particulièrement Nassera, Alexandra et Sylvain qui ont été toujours disponibles pour m'assister sur un plan administratif. Je remercie aussi mes anciens collègues chez Axa Hedging Services et spécialement Abdelhadi, Ahmed, Hassan, Jonathan, Pierre-Henri et Zakaria avec qui j'ai beaucoup échangé. Un grand merci va aussi à mon frère, mes soeurs ainsi que mes amis proches qui se reconnaîtront certainement.

Je ne saurai oublier ma femme qui m'a soutenu pendant ma thèse et dont les encouragements aux moment difficiles ont joué un rôle primordial pour l'aboutissement de ce travail.

Pour clore, j'aimerais témoigner à mes parents l'expression de ma plus grande gratitude, qui, si grande qu'elle puisse être, ne sera jamais être à la hauteur de leur dévouement.

Résumé

Cette thèse traite du problème de valorisation par des méthodes numériques efficaces de contrats GMWB dans une optique de calcul par formules fermées ou par méthode de Monte Carlo sous contrainte de faible nombre de simulations. Les produits GMWB sont des produits très complexes qui ont connu ces dernières années un grand succès de par la garantie dont bénéficie l'assuré sur les retraits futurs avec un effet upside dépendant de la performance du fond sous-jacent au contrat. En outre, le souscripteur dispose de nombreuses options attrayantes qu'il peut exercer à tout moment dont l'option de racheter partiellement ou totalement son contrat, la possibilité de modifier l'allocation de la prime payée (fund switching) pendant la durée du contrat et enfin l'option d'avancer ou de reporter la date de début des paiements. Cependant, de telles options cumulées avec la complexité du produit et les risques de marché et de mortalité exposent l'assureur qui doit gérer des dizaines de milliers de contrats sous plusieurs contraintes opérationnelles (temps de calcul, faible nombre de simulation, etc.) à une difficulté majeure en terme de valorisation et de couverture.

Une grande partie de cette thèse (chapitres de 2, 4, 5 et 6) est consacrée à l'étude de l'option de rachat partiel ou total dans les contrats GMWB selon deux angles : le point de vue assuré rationnel et le point de vue couvreur appréhendant le pire cas. A ce propos, dans notre cadre général en temps discret avec une volatilité locale et taux d'intérêt à la Hull-White 1 facteur, la stratégie optimale déterminant le coût du contrat dans les deux cas est la solution d'un problème de contrôle stochastique optimal en temps discret. Néanmoins, grâce à une propriété d'homogénéité partielle sur le prix et les flux, on démontre qu'elle est explicite et de type Bang-Bang. Le problème de valorisation étant ainsi ramené à celui d'arrêt optimal, nous avons proposé une méthode de Monte Carlo de type Longstaff Schwartz dont l'étape de régression empirique a été traitée par la méthode de moindres carrés habituels et par une nouvelle méthode appelée VCP (Variables de contrôle préliminaires). Cette dernière consiste dans un premier temps à réduire la variance empirique des flux à régresser à travers une projection L2 sur des variables de contrôle adaptées et centrées, et puis à faire la régression par moindres carrés habituels sur les nouveaux flux à variance réduite. Une étude numérique sur un cas test ainsi qu'une quantification théorique de l'erreur par les techniques de régression non paramétriques ont conclu à son efficacité dans un contexte de faible nombre de simulation (contrainte d'Axa).

Quant au chapitre 3, il est consacré à justifier numériquement et théoriquement l'hypothèse de mutualisation du risque de mortalité souvent supposée par les praticiens dans le cas américain sur un produit simple sensible à ce risque.

Enfin, la dernière partie de la thèse (chapitre 7) est consacrée à la valorisation par formules fermées approchées pour des contrats GMWB simplifiés dans un modèle Black Scholes avec taux d'intérêt de dynamique Hull-White à 1 facteur. En effectuant un développement asymptotique sur le montant des retraits, on obtient des formules approchées du prix du contrat GMWB par un prix Black-Scholes corrigé par une somme explicite de Grecques, le tout étant plus rapide à évaluer. Des estimations d'erreur sont établies lorsque la fonction payoff est régulière. La précision des formules asymptotiques est testée numériquement et montre un excellent comportement de ces approximations, même pour des contrats à longue maturité (20 ans).

Abstract

This thesis deals with the problem of pricing GMWB contracts by efficient numerical methods using closed formulas or Monte Carlo method under the constraint of few simulations. GMWB products are highly complex products that have in recent years experienced a great success due to the guarantee given to the insured people on future withdrawals with an upside effect depending on the performance of the underlying fund contract. In addition, the subscriber has many attractive options that can be exercised at any time such as the option to lapse partially or totally his contract, the possibility of fund switching during the policy life and lastly the option to advance or postpone the withdrawal date. However, such options combined with the complexity of the product, the market risk and the mortality risk expose the insurer who should manage a lot of contracts under several operational constraints (computational time, few simulations, etc..) to a major challenge in terms of valuation and hedging.

A large part of this thesis (Chapters 2, 4, 5 and 6) study the partial and total surrender option in the GMWB contracts from two points of view : the rational client and the hedger worst case. In this regard, in our general discrete time framework with a local volatility and Hull-White interest rates models, the optimal strategy determining the contract cost is in both cases the solution to an optimal stochastic control problem in discrete time. However, thanks to partial homogeneity property of prices and flows, we show that the strategy is Bang-Bang, hence explicit. Thus, the valuation problem is reduced to an optimal stopping problem, we propose a Monte Carlo method using Longstaff-Schwartz methodology whose empirical regression step was treated by the standard least squares method as well as a new method called VCP (preliminary control variates). The latter consists first to reduce the empirical variance of the flows to regress through an L2 projection on adapted and centered control variates, then to perform the standard least square regression methods on the new flows with reduced variance. A numerical study and a theoretical error analysis based on distribution-free nonparametric regression techniques confirm the VCP method efficiency in the context of few simulations.

Chapter 3 gathers mathematical and numerical justifications of the hypothesis considering mortality risk as diversifiable. We achieve this in a simple American exercise style product sensitive to mortality risk.

Finally, the last part (Chapter 7) is devoted to pricing GMWB contracts by closed analytical approximations in a Black Scholes model with Hull-White interest

rates. By performing an asymptotic expansion on the amount of withdrawals, we obtain approximation price formulas as a Black-Scholes price plus a sum of explicit Greeks, faster to evaluate. Error estimates are provided when the payoff function is regular. The accuracy of the asymptotic formulas is experienced numerically and it shows a good behavior even for long maturity contracts (20 years).

Table des matières

1	Introduction	17
1.1	Contrats d'assurance vie	17
1.2	Les variable annuities (annuités variables)	18
1.2.1	Définition des principales composantes d'un produit GMxB	18
1.2.2	Mécanisme de calcul de la base de garantie	19
1.2.3	Les produits GMxB	19
1.2.4	Risques dans les variable annuities	22
1.3	Valorisation des produits GMxB et couverture des risques associés	24
1.4	Synthèse des travaux	25
2	Etude du contrat GMWB	31
2.1	Description et modélisation	31
2.1.1	Notations	31
2.1.2	Dynamique	31
2.1.3	Dates de rachat	34
2.1.4	Flux	34
2.1.5	Variables markoviennes du problème et contrôles	35
2.2	Typologie du client	35
2.2.1	Cas du client rationnel	35
2.2.2	Cas du client non rationnel	35
2.3	Vers la résolution du cas du client rationnel	36
2.3.1	Le problème de contrôle	36
2.3.2	Résultat Principal : Caractérisation du rachat optimal	38
2.3.3	Preuve du théorème 2.3.1	39
2.3.4	Réduction du problème à un problème d'arrêt optimal	44
3	Influence de la mortalité	47
3.1	Tour de la littérature dans le cas européen	47
3.1.1	Exemple de produit financier avec risque de mortalité	47
3.1.2	Quelques résultats sur les stratégies de minimisation du risque globale	48

3.1.3	Application à l'exemple étudié : mutualisation du risque . . .	51
3.1.4	Résultats numériques	52
3.2	Heuristique pour le cas à exercice américain	54
3.2.1	Résultat principal	54
4	Régression empirique	63
4.1	Description du cas test : Modèle et Payoff	63
4.2	Choix de fonctions de base	64
4.2.1	Fonctions de base à 2 variables	64
4.2.2	Réduction de dimension et fonctions à une variable	67
4.2.3	Calcul pratique de la solution du problème des moindres carrés	67
4.3	Régression empirique par moindres carrés standards	71
4.3.1	Rappel de la littérature sur les majorations d'erreur	71
4.3.2	Validation numérique du théorème 4.3.1	71
4.3.3	Validation croisée	74
4.4	Régression par Variables de contrôles préliminaires (VCP)	76
4.4.1	Description et heuristique numérique	76
4.4.2	Quantification de l'erreur	78
4.4.3	Preuve du théorème	81
4.5	Résultats complémentaires en régression non paramétrique	90
4.5.1	Nombres de couverture	90
4.5.2	Coefficients d'éclatement et dimension VC	91
5	Régression empirique à plusieurs périodes	109
5.1	Description du cas test : Modèle et Payoff	109
5.2	Régression par VCP	110
5.2.1	Description	110
5.2.2	Résultats numériques	113
6	Mise en oeuvre numérique sur un contrat GMWB Axa avec le point de vue <i>hedger</i>	125
6.1	Point de vue couvreur	125
6.1.1	Comparaison entre les points de vue client rationnel et <i>hedger</i>	127
6.1.2	Algorithme de programmation dynamique	127
6.2	Résolution numérique : Méthode de Monte Carlo de type Longstaff Schwartz	128
6.2.1	Description	128
6.2.2	Résultats numériques	130
6.2.3	Perspectives	131

7	Valorisation par formules fermées d'un contrat GMWB à exercice européen	137
7.1	Formulation du problème	137
7.2	Formules fermées d'approximation	139
7.2.1	Résultats préliminaires	139
7.2.2	Résultats principaux	141
7.2.3	Estimation de l'erreur dans le cas régulier	148
7.3	Résultats numériques	151
7.4	Annexe	152
7.4.1	Preuve du lemme 7.2.1	152
7.4.2	Preuve du lemme 7.2.2	153
7.4.3	Preuve du lemme 7.2.3	155

Table des figures

1.1	Illustration du calcul de la base.	20
1.2	Illustration du produit GMAB.	21
1.3	Illustration du produit GMDB.	22
1.4	Illustration du produit GMWB.	23
3.1	Double interpolation du prix selon la position de la coordonnée C . . .	58
3.2	Histogramme du coût pour 1 et 10 assurés dans le cas européen. . .	59
3.3	Histogramme du coût pour 50 et 100 assurés dans le cas européen. .	60
3.4	Histogramme du coût pour 1 et 10 assurés dans le cas américain. . .	61
3.5	Histogramme du coût pour 100 et 1000 assurés dans le cas américain.	62
4.1	Découpage en pavés de même surface (voir [GKKW02])	65
4.2	Découpage en pavés de même nombre de données (voir [GKKW02])	66
4.3	Erreur empirique pour les polynômes.	94
4.4	Erreur empirique pour les fonctions locales par morceaux de même aire.	95
4.5	Erreur empirique pour les fonctions locales par morceaux de même nombre de données.	96
4.6	Erreur empirique pour différentes fonctions de base.	97
4.7	Pente de l'erreur pour différentes fonctions de base.	98
4.8	Pente de l'erreur pour différentes fonctions de base et différents K_N .	99
4.9	Erreur empirique pour la régression par VCP pour des polynômes .	100
4.10	Erreur empirique pour la régression par VCP pour le cas <code>1dAffineMemeData</code> .	101
4.11	Erreur empirique pour la régression par VCP pour le cas <code>2dAffineMemeSurf</code> .	102
4.12	Erreur empirique pour la régression par VCP pour le cas <code>1dAffineMemeData</code> avec K_2 grand	103
4.13	Erreur empirique en fonction de N et K pour le cas des polynômes 1D.	103
4.14	Erreur empirique en fonction de N et K pour le cas <code>1dAffineMemeSurf</code> .	104
4.15	Erreur empirique en fonction de N et K pour le cas <code>1dAffineMemeData</code> .	104
4.16	Histogramme de \hat{K} pour les cas polynômes 1D, <code>1dAffineMemeSurf</code> et <code>1dAffineMemeData</code>	105

4.17	Erreur empirique pour la régression par VCP avec échantillon indépendant pour le cas Polynômes 1D.	106
4.18	Erreur empirique pour la régression par VCP avec échantillon indépendant pour le cas <code>1dAffineMemeData</code>	107
4.19	Erreur empirique pour la régression par VCP avec échantillon indépendant pour le cas <code>2dAffineMemeSurf</code>	108
5.1	Erreur empirique en utilisant $K_1 = 2$ polynômes pour la première étape et $K_2 = 10$ polynômes à une variable dans la seconde.	115
5.2	Espérance empirique de la variance conditionnelle associée à la figure 5.1.	116
5.3	Erreur empirique en utilisant $K_1 = 4$ polynômes pour la première étape et $K_2 = 10$ polynômes à une variable dans la seconde.	117
5.4	Espérance empirique de la variance conditionnelle associée à la figure 5.3.	118
5.5	Erreur empirique en utilisant $K_1 = 4$ fonctions affines par morceaux sur $I = [0, 6]$ pour la première étape et $K_2 = 10$ polynômes à une variable dans la seconde.	119
5.6	Espérance empirique de la variance conditionnelle associée à la figure 5.5.	120
5.7	Erreur empirique en utilisant $K_1 = 4$ fonctions affines par morceaux sur $I = [0, 6]$ pour la première étape et $K_2 = 50$ fonctions de l'espace <code>1dAffineMemeSurf</code> sur $[0, 6]$ pour la seconde.	121
5.8	Espérance empirique de la variance conditionnelle associée à la figure 5.7.	122
5.9	Erreur empirique en utilisant $K_1 = 4$ polynômes à une variable pour la première étape et $K_2 = 50$ fonctions de l'espace <code>1dAffineMemeSurf</code> sur $[0, 6]$ pour la seconde.	123
5.10	Espérance empirique de la variance conditionnelle associée à la figure 5.9.	124
6.1	Courbe du Zero Coupon et courbe des taux utilisées.	132
6.2	Comparaison entre le cas homogène et non homogène des fonctions de bases dans le cadre de la méthode standard avec $T^{def} = 5$	133
6.3	Intervalle de confiance du coût de couverture pour les différentes méthodes avec $T^{def} = 5$	134
6.4	Intervalle de confiance du coût de couverture pour les différentes méthodes avec $T^{def} = 9$	135
7.1	Valeur de $GMWB_0$ pour $\sigma=10\%$ et $\delta=9.5$ (retrait annuel à partir de l'année 6 jusqu'à maturité (pas de retrait avant)).	152

7.2	Erreur relative du prix DAS5 pour $\delta=9.5$ (retrait annuel à partir de l'année 6 jusqu'à maturité (pas de retrait avant)).	161
7.3	Erreur relative en échelle log du prix DAS5 en fonction de δ (retrait annuel) pour $\sigma = 15\%$, $\sigma^r = 1\%$ et $a=4\%$	162
7.4	Probabilité pour que l' <i>account value</i> devienne négative pour $\sigma = 15\%$ et $a=4\%$	162

Chapitre 1

Introduction

Il y a plus de vingt ans, les compagnies d'assurance aux Etats Unis proposaient à leurs clients des contrats d'assurance vie à annuités constantes ou des contrats en unité de compte très risqués pour l'assuré car son épargne était investie dans des actifs comme les fonds FCP. Cependant, pendant ces deux dernières décennies, le marché a connu des changements qui ont facilité l'accès à des produits d'investissement sur le marché (fonds mutuel) et accru l'exigence des clients en terme de rendement. Ceci a conduit ces compagnies à innover ses contrats pour proposer de nouveaux produits (*variable annuities*) comportant de nouvelles annuités garanties avec un effet *upside* (rehaussement) via l'investissement dans des fonds. Récemment, ces contrats se sont répandus ailleurs dans le monde avec des succès différents selon les pays.

1.1 Contrats d'assurance vie

Un contrat d'assurance-vie est une convention par laquelle le souscripteur s'engage à verser une ou des primes à l'assureur en échange de quoi l'assureur versera au bénéficiaire du contrat une ou des prestations sous forme de capital ou de rente lors de la réalisation du risque porté par l'assuré. En assurance-vie, on distingue généralement deux types de contrats :

- 1) Les contrats en euros : ils offrent à l'assuré à chaque moment le choix entre un capital garanti ou une rente sous forme d'annuités fixes.
- 2) Les contrats en unités de compte ("UC") : ils garantissent des parts d'actifs. L'épargne de l'assuré est investie sur des actifs, appelés supports (actions de SICAV ou de FCP par exemple) et l'engagement de l'assureur se limite au versement de la contrevaletur de ces unités de placements. La valorisation de ces contrats dépend donc de l'évolution des marchés financiers.

Contrairement aux contrats en euros qui peuvent garantir contractuellement un rendement minimum, les contrats en unités de compte font supporter à l'assuré l'intégralité du risque lié aux fluctuations des marchés financiers.

1.2 Les variable annuities (annuités variables)

Les variable annuities sont des produits hybrides qui offrent une garantie moins forte que les fonds en euros mais aussi la possibilité d'*upside* via l'investissement dans des sous-jacents risqués appelés supports. Ces différents types de variable annuities sont de la forme GMxB (Guaranteed Minimum x Benefit) où x est une caractéristique spécifique du produit. Nous les détaillons dans la suite.

1.2.1 Définition des principales composantes d'un produit GMxB

- **Valeur comptable ou *account value*** : elle représente le montant de l'épargne à une date précise. La valeur comptable est (souvent) une moyenne pondérée entre plusieurs fonds communs de placement. Afin d'assurer la couverture de ces produits, on doit être capable de répliquer la valeur comptable en investissant sur les marchés. Les fonds n'étant généralement pas liquides, ils sont modélisés à l'aide d'indices boursiers (SP500, CAC40, etc.) et donc la valeur comptable est exprimée en fonction de différents indices.
- **Base de garantie ou *Benefit Base*** : c'est une garantie sur l'épargne faite à l'assuré qui lui assure une valeur minimum de son épargne. La manière de calculer cette quantité est expliquée par la suite. La garantie peut être de différentes natures (rentes, retraits, capital, etc.).
- **Frais de gestion ou *management fees*** : c'est l'ensemble des charges prélevées sur l'*account value* et exprimées le plus souvent en pourcentage de l'*account value*. Elles peuvent parfois être exprimées en pourcentage de la base. Ces frais sont prélevés tout au long de la vie du contrat.
- **Frais de couverture** : elles servent à couvrir le risque engendré par les garanties offertes au client.
- **Période d'attente ou *deferral period*** : cette période est le plus souvent fixée lors de la signature du contrat. Pendant cette durée, l'assuré ne reçoit pas d'argent de son contrat.

- **Période de paiements** : c'est la période succédant à la *defferal period* où l'assuré commence à recevoir de l'argent. Elle peut avoir une durée déterminée ou, au contraire, s'étendre jusqu'au décès de l'assuré.
- **Roll up (bonus d'attente)** : c'est le pourcentage annuel d'augmentation de la base durant la période d'attente.
- **Ratchet (cliquet)** : c'est un mécanisme s'appliquant à la base et qui permet de bloquer les performances du marché avec une fréquence qui dépend de la nature du cliquet (annuel ou mensuel).

1.2.2 Mécanisme de calcul de la base de garantie

A la signature du contrat, la *Benefit Base* est égale au montant de la prime nette du contrat (valeur de la prime à laquelle on a retranché des frais d'entrée, par exemple égaux à 5% de la prime mais qui dépendent du distributeur du produit). Cette garantie peut augmenter suivant trois types de mécanisme :

- 1) **L'effet roll-up** : la base est capitalisée au taux de *roll-up* qui peut être linéaire ou composé.
- 2) **L'effet cliquet ou ratchet** : la base est mise à jour à une fréquence donnée (souvent annuelle) durant la *defferal period* à la valeur de l'*account value* si cette dernière est plus grande. Ainsi, ce mécanisme permet à l'assuré de bénéficier des hausses des marchés financiers pendant la période d'attente.
- 3) **L'effet combo** : la base est égale au maximum entre la base *roll-up* et la base *ratchet*.

La figure 1.1 permet d'illustrer ces mécanismes sur un contrat avec une prime de départ de 100 000 et un taux de *roll-up* de 6%.

La courbe verte représente la base *ratchet*, la violette la base *roll-up* et la rouge correspond à la valeur du portefeuille sous-jacent (*account value*).

1.2.3 Les produits GMxB

- 1) **GMAB : Guaranteed Minimum Accumulation Benefit**
Ce produit se comporte comme un put vis-à-vis de l'assuré. A la fin de la période d'attente, l'assuré récupère le maximum entre l'*account value* et sa base de garantie appelée *Accumulation Base*. La maturité de ces produits est

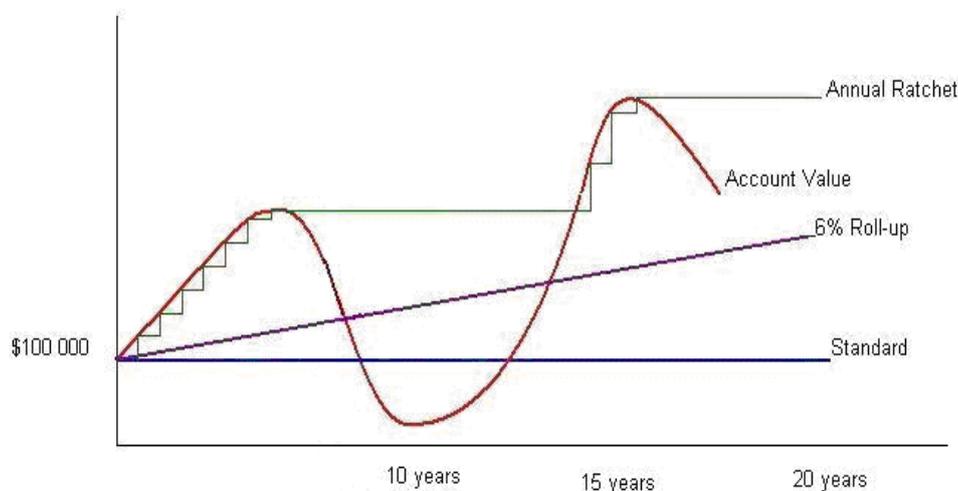


FIGURE 1.1 – Illustration du calcul de la base.

généralement d'une dizaine d'années (voir figure 1.2).

– 2) **GMDB : Guaranteed Minimum Death Benefit**

Ce produit offre une garantie sur l'épargne qui s'exerce au décès de l'assuré. L'assureur verse au bénéficiaire le maximum entre la base appelée *Death Base* (qui a éventuellement été augmentée tout au long de la période d'attente par les mécanismes de *roll-up*, de *ratchet* ou combo) et l'*account value* à la mort de l'assuré.

Sur le graphe 1.3, ces trois mécanismes sont représentés. La garantie que récupère le bénéficiaire est le cliquet annuel en supposant que l'assuré décède au bout de 8 ans.

– 3) **GMSB : Guaranteed Minimum Surrender Benefit**

La GMSB est une garantie qui offre à l'assuré la possibilité de racheter son contrat (sans payer de frais supplémentaires) au bout d'un certain temps (*surrender period*). A la fin de cette période, l'assuré peut choisir de quitter son contrat et repartir avec le maximum entre la *Surrender Base* et l'*account value*.

– 4) **GMWB : Guaranteed Minimum Withdrawal Benefit**

La GMWB est une garantie sur l'épargne qui offre au souscripteur la possibi-

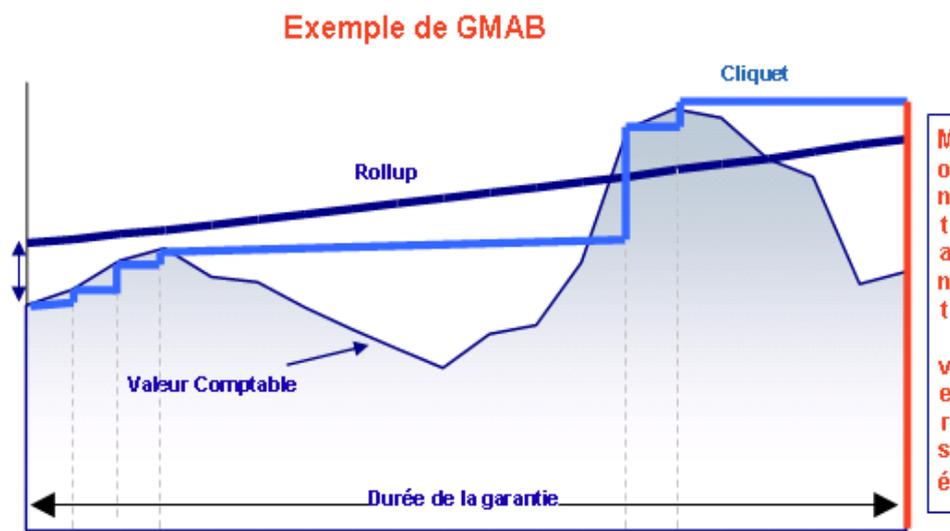
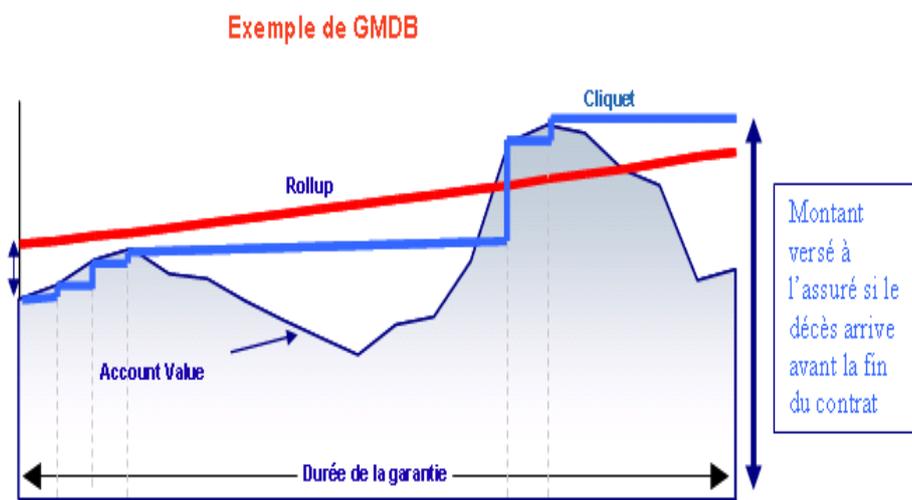


FIGURE 1.2 – Illustration du produit GMAB.

lité de retirer un certain nombre de coupons à l'avance et à des dates données. Quand l'assuré souscrit au produit, commence alors une période d'attente (*defferal period*), dont la durée est fixée à la souscription (à moins que l'assuré ait l'option de la modifier durant le contrat), pendant laquelle l'assuré ne touche rien. La base appelée *Withdrawal Base* peut être augmentée grâce aux mécanismes de *roll-up* et de *ratchet*. Lorsque la *defferal period* se termine, débute la période de paiements où le souscripteur peut retirer à intervalles réguliers (par exemple annuels) des coupons qui sont un pourcentage fixe de la *Withdrawal Base* qui a été atteinte à la fin de la période d'attente. Ils sont soustraits de l'épargne du souscripteur (son *account value*). Durant la période de paiements, l'assuré a toujours la possibilité de partir à n'importe quel moment avec une partie ou la totalité de son *account value* (rachat partiel ou total du contrat) et la *Withdrawal Base* est diminuée dans ce cas par le pourcentage racheté dans le cas d'un rachat partiel. Si l'*account value* atteint 0, elle reste nulle et l'assureur continue à verser le reste des coupons en prélevant dans ses fonds propres. Si l'épargne n'atteint pas 0, le coût de cette garantie est nul pour l'assureur et l'assuré récupère la valeur de l'épargne restante à la fin du contrat. Il est important de noter que la valeur de la *Withdrawal Base* n'est pas réduite au fur et à mesure que les coupons sont versés. Les coupons ont donc une valeur constante. La durée de la période de paiement peut être soit fixée (terme certain) par avance soit "à vie".

La figure 1.4 illustre l'évolution temporelle possible de l'*account value* d'un client qui possède une garantie GMWB. On remarque que l'épargne atteint 0



alors que 4 coupons doivent encore être versés. Ainsi, l'assureur doit lui-même puiser dans ses réserves pour payer les 4 derniers coupons.

– 5) **Produits GMxB à garanties cumulées :**

Les garanties GMxB peuvent aussi être cumulées. Ainsi, on peut avoir un produit GMWB assorti d'une garantie de type GMDB en cas de décès de l'assuré.

Notons que dans tous les produits ci-dessus le souscripteur paie les frais (frais de gestion et de couverture) périodiquement durant toute la vie du contrat.

1.2.4 Risques dans les variable annuities

Les produits GMxB contiennent des options cachées offertes à l'assuré mais difficiles à tarifier et très risquées pour l'assureur. On peut citer :

- L'option de rachat (*lapse*) partiel ou total durant la vie du contrat (commun à tous les produits).
- La possibilité de changement de l'allocation de la prime ou *fund switching* pendant la vie de la police.
- L'option de modifier la *Deferral period* pour les produits GMWB.

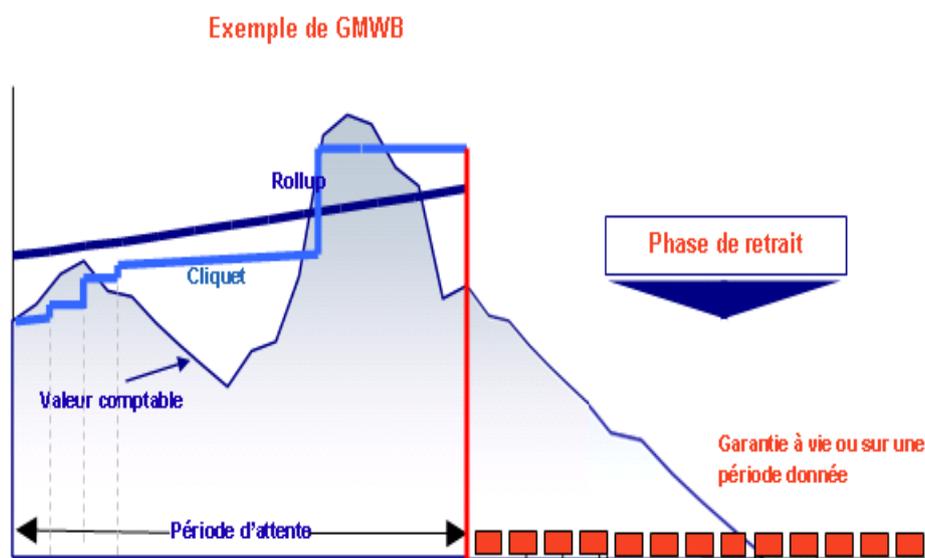


FIGURE 1.4 – Illustration du produit GMWB.

En sus des risques ci-dessus, l'assureur s'expose aussi aux risques actuariels de mortalité et de longévité ainsi qu'aux risques financiers dont le risque de base (dû à la régression des fonds sur certains indices).

La gestion efficace d'un seul des contrats GMxB s'apparente souvent à celle des contrats complexes de type exercice américain (optionalité de la date d'exercice, voir [LS01] [BW12]) ou à des option swing [JRT04] [BBD⁺06] sur les marchés énergétiques (optionalité de la date d'exercice et flexibilité en volumes). Une difficulté supplémentaire pour les assureurs est que le nombre de polices à gérer est considérable (plusieurs dizaines de milliers) et que chaque police est différente ; cela rend le problème de gestion globale très difficile. Ceci constitue un des grands enjeux de l'assurance, qui ne peut être résolu seulement par l'utilisation massive d'ordinateurs plus puissants et plus nombreux. La réussite passe aussi par la dérivation de méthodes numériques performantes et adéquates à cette situation. Ce travail de thèse contribue à apporter des solutions et réponses dans cette direction, en particulier pour les GMWB qui sont les contrats les plus complexes. Notamment, un des enjeux consiste à proposer des méthodes efficaces à base d'un faible nombre de trajectoires Monte Carlo (de l'ordre de 1000 à 5000), ce qui est une contrainte opérationnelle réaliste pour Axa.

1.3 Valorisation des produits GMxB et couverture des risques associés

Nous avons vu dans la section précédente que les variable annuities sont des produits très complexes et comportent pour l'assureur plusieurs risques rendant leur valorisation ainsi que leur couverture très compliquées. Ainsi, les techniques habituelles de valorisation dans les marchés financiers doivent être adaptées pour tenir compte des grandes différences qui existent entre les produits financiers et les produits d'assurance. En effet, les maturités des produits GMxB sont assez grandes pouvant aller jusqu'à 60 ans ou jusqu'à la date du décès de l'assuré. En outre, ces produits contiennent un risque actuariel comme la mortalité et le comportement des assurés qui ne peut être couvert sur le marché. De surcroît, l'assureur joue souvent sur l'effet mutualisation en vendant un même contrat à plusieurs assurés. Enfin, la tarification des variable annuities se traduit par des frais (souvent un pourcentage de l'*account value* ou de la *Benefit Base*) prélevées à une certaine fréquence durant toute la vie du contrat alors que souvent le coût des produits financier est connu et payé à la date initiale.

Boyle, Bremann et Schwartz [BS76] [BS77] sont les premiers à avoir proposé d'utiliser les méthodes classiques de valorisation et de couverture des options dans les marchés financiers à des produits d'assurance assortis avec une garantie en supposant un taux d'intérêt et une volatilité constantes. Grosen et Jorgensen [GJ97], [GJ00] se sont intéressés à certains cas de produits d'assurance avec garantie dans la cas européen (sans rachat) aussi bien qu'américain (avec rachat). Ils utilisent la méthode d'arbre binomial pour valoriser le cas de rachat anticipé. La même méthode est exploitée par Bacinello [Bac03] [Bac05] en incluant le risque de mortalité. L'application de l'approche Longstaff Schwartz par simulations Monte Carlo et régressions empiriques a été faite par Bacinello [BBM09] [BBM10]. Concernant le rachat partiel ou total à travers une flexibilité dans les retraits des contrats, Milevsky et Salisbury [MS06] ont proposé une valorisation du produit GMWB sous l'hypothèse de comportement statique aussi bien que dynamique de l'assuré. Ils arrivent en temps continu à déterminer la stratégie de retrait optimale mais sous des hypothèses simplificatrices de taux d'intérêt et volatilité constantes. Sous ces mêmes hypothèses, Dai, Kwok et Zong [DKZ08] ont proposé pour la valorisation de ce produit un modèle de contrôle stochastique singulier en temps continu pour les *fees* et en temps continu et discret pour les retraits. Ils retrouvent les mêmes résultats de Milevsky et Salisbury [MS06]. Un cadre général pour les variable annuities tenant compte des stratégies de rachat partiel ou total (en tolérant le rachat du contrat sous forme d'un retrait supplémentaire autre que le coupon du contrat) pour des assurés irrationnels ou rationnels a été développé par Bauer [BKR08] en temps

discret. Bacinello, Millosovich, Olivieri et Pitacco [BMOP11] ont étendu le cadre de Bauer en incorporant la caractéristique stochastique des taux et de la volatilité des sous-jacents. Kling, Ruez et Russ [KRR11] ainsi que Holz, Kling et Russ [HKR08] développent un cadre de travail pour des contrats GMWB à vie assortis de l'option de rachat. Le point commun entre ces différentes approches est qu'elles ne donnent aucune information sur la nature de la stratégie optimale et proposent uniquement des méthodes numériques. Peng, Leung et Kwok [PLK12] se sont intéressés à la valorisation en temps continu d'un produit GMWB sans rachat sous l'hypothèse d'un retrait déterministe et de taux et volatilité stochastiques. Ils dérivent dans ce cadre une formule analytique d'approximation sous forme de bornes inférieure et supérieure du prix.

Concernant le risque de mortalité, Bacinello et Ortu [BO93], Nielsen et Sandmann [NS95] ainsi que Milevsky et Posner [MP01] supposent une mutualisation de ce risque à travers les assurés et valorisent dans un monde risque neutre des contrats contenant ce risque. Sous l'hypothèse d'indépendance du risque de mortalité avec le risque financier, Moller [Mol98] [Mol01b] [Mol01a] explore ce risque dans des contrats GMDB et justifie l'hypothèse de mutualisation en appliquant les résultats de la théorie de minimisation du risque en marché incomplet développée par Föllmer et Sondermann [FS86] ainsi que Schweizer [Sch91]. Il arrive ainsi à déterminer des stratégies de couverture minimisantes pour différents types de variable annuités sans option de rachat anticipé du contrat.

1.4 Synthèse des travaux

Dans cette thèse, on s'est concentré principalement sur les contrats GMWB assortis de l'option de rachat partiel ou total du contrat. Cette option est similaire à celle du contrat à exercice américain ou bermudéen ou celle du contrat swing dans les marchés énergétiques. Il est bien connu que l'évaluation par la méthode de Monte Carlo de ces deux contrats peut être réalisée par l'algorithme de Longstaff-Schwartz en utilisant des projections sur des fonctions de bases adaptées [LS01] [BBD⁺06] et dont l'étude théorique de convergence a été faite sous certaines hypothèses [CLP02] [GY04] [Egl05]. L'objectif de cette thèse est d'allier résultats théoriques et méthodes numériques dans un cadre général de temps discret (ce qui correspond à la réalité vu que les *fees* et les coupons tombent à des dates discrètes) pour conduire à une gestion efficace (valorisation et couverture) des contrats GMWB.

A ce propos dans le chapitre 2, on se place du point de vue de l'assuré rationnel

et on propose un cadre général de valorisation des produits GMWB en supposant le taux d'intérêt aussi bien que la volatilité stochastiques et que l'assuré peut racheter partiellement ou totalement son *account value*. La *Withdrawal Base* peut être augmentée par les trois effets *roll up*, *rachet* ou *combo*. On a fait l'hypothèse que le rachat se fait à des dates discrètes qui sont différentes des dates relatives aux prélèvements de frais, d'actualisation de la base ou retrait. Le prix du contrat est alors la solution d'un problème de contrôle stochastique optimal à temps discret où les contrôles sont les pourcentages rachetés à différentes dates d'exercice. Néanmoins, en utilisant la théorie de contrôle stochastique optimal en temps discret développée par Bertsekas et Shreve [BS78], on a pu démontrer l'existence d'une stratégie optimale markovienne. Dans notre modèle (volatilité locale, taux stochastique à la Hull-White), les variables markoviennes sont au nombre de 6. Pour être plus précis, nous démontrons une propriété d'homogénéité partielle, dont les conséquences sont doubles :

- 1) On peut réduire le nombre de variables à 5.
- 1) La stratégie optimale est de type bang-bang (tout ou rien).

La première propriété de réduction de dimension est très utile pour améliorer l'efficacité des méthodes de régression intervenant dans les algorithmes de type Longstaff-Schwartz. La seconde propriété indique que le rachat optimal ne peut être partiel mais seulement total. On s'est ramené ainsi à un problème d'arrêt optimal (option américaine) plus simple à résoudre numériquement. Ainsi, on peut aboutir à la conclusion très importante suivante : l'assureur doit couvrir uniquement le rachat total sans soucier des stratégies de rachats partiels. Notre contribution par rapport aux travaux déjà faits (Milevsky et Salisbury [MS06] et Dai, Kwok et Zong [DKZ08]) est qu'ici la stratégie optimale est explicitement déterminée (bang-bang) dans un cadre plus général (caractère stochastique de la volatilité et du taux d'intérêt) et plus représentatif de la réalité (temps discret). Nous justifions aussi la réduction de dimension qui est souvent utilisée par les praticiens. D'un autre côté Bauer [BKR08], Bacinello [BMOP11], Kling, Ruez et Russ [KRR11] [KRR12] et Holz et al., Kling et Russ [HKR08] même s'ils proposent un cadre de travail général en temps discret, ils ne déterminent pas théoriquement la stratégie optimale ou se limitent dans leur analyse numérique (sans justification théorique) aux stratégies bang-bang. Comme perspectives de travaux ultérieurs, il serait intéressant d'étendre notre résultat de stratégie optimale à davantage de produits GMxB.

Le chapitre 3 a été consacré à l'étude du risque de mortalité surtout lorsqu'il est combiné avec celui de rachat anticipé. D'abord, dans le cas de contrats d'assurance vie dépendant de la durée de vie du client mais sans rachat anticipé (cas européen), les travaux de Moller [Mol98] [Mol01b] montrent que ce risque peut être mutualisé pour un grand nombre d'individus. Pour notre part, on a considéré le

cas d'un contrat d'assurance assez simple dont la payoff dépend de la durée de vie des assurés et qui offre la possibilité d'un rachat anticipé à tout moment avant la maturité ou le décès. En suivant l'approche de Karatzas [Kar88] pour le pricing des options américaines, nous nous sommes ramenés au cas américain habituel ce qui nous a permis d'établir que le risque de mortalité peut être aussi mutualisé pour un nombre significatif de clients malgré le caractère américain de l'exercice. En effet, on exhibe un portefeuille autofinçant dont la variance de l'erreur liée à la mortalité tend vers 0 quand le nombre d'individus tend vers l'infini. De surcroît, on a effectué une comparaison numérique sur le comportement de la mutualisation pour les deux cas (européen et américain) en traçant les histogrammes du coût de couverture associés. Ces tests montrent que dans le cas américain, l'hypothèse de mutualisation peut exiger plus d'assurés que dans le cas européen. En perspective, il reste à déterminer un critère (éventuellement similaire à celui de la minimisation du risque) selon lequel on peut calculer un prix et une stratégie de couverture optimaux pour le cas de rachat anticipé.

Dans le chapitre 4, nous donnons les fondements théoriques pour la partie numérique du chapitre 6 consacrée à la valorisation d'un contrat GMWB en suivant une approche Monte Carlo de type Longstaff-Schwartz. Sachant que l'algorithme est très sensible à la méthode de régression utilisée, nous avons d'abord étudié différents choix de fonctions de bases (polynômes, fonctions constantes ou affines par morceaux sur des blocs contenant un même nombre de données ou de même taille) pour la régression et différents réglages de paramètres (nombre de trajectoires, nombre de fonctions de base), sur un produit financier simple (un portefeuille d'options d'échange) ayant des propriétés théoriques similaires au produit GMWB (dont l'homogénéité). Cette étude numérique a montré que l'erreur entre le prix exact et le prix estimé par régression dépend des fonctions de base utilisées pour la projection, de la dimension de l'espace ainsi que du nombre de simulations. Les travaux de Györfi et autres [GKKW02] montrent que cette erreur peut être décomposée en la somme d'un terme d'erreur statistique dû au nombre fini de simulations et d'un terme d'erreur d'approximation lié à la qualité de l'espace de fonctions de base choisi. En pratique, il y a une balance optimale à faire entre ces deux termes car une augmentation de la dimension engendre une diminution du terme d'approximation mais en même temps un accroissement du terme statistique. Nous avons testé la méthode de validation croisée pour sélectionner le nombre de fonctions de base optimale pour un nombre de simulations donné. De plus, nous mettons en évidence que l'erreur d'approximation est meilleure quand on choisit des fonctions de base qui préservent la propriété d'homogénéité du prix. D'un autre côté, vu qu'on se place dans un contexte de faible nombre de simulations, nous avons proposé une nouvelle méthode de régression avec VCP (Variables de Contrôle Préliminaires) se

basant sur la réduction de variance dont le but est d'accélérer la convergence de l'erreur de régression vers l'erreur d'approximation. Cette méthode se fait en deux étapes.

- 1) Dans la première, on fait une réduction de la variance empirique des flux de régression par une projection L^2 sur des variables de contrôle bien choisies : à la différence de Bouchaud, Potters et Sestovic [BPS01] qui utilisent des idées semblables avec des intégrales stochastiques finement discrétisées, nous recommandons plutôt d'utiliser des variables de contrôle de faible cardinalité dépendant des variables aléatoires d'intérêt aux temps de régression (qui sont assez espacées, c.à.d 1 an ou 1 mois par exemple).
- 2) Dans la deuxième, on fait la régression habituelle par moindres carrées des flux auxquels on a soustrait les variables de contrôle.

Les résultats obtenus sont très satisfaisants car ils montrent qu'on peut gagner un facteur 100 dans le nombre de simulations à effectuer par rapport à la méthode standard. Une preuve théorique de cette amélioration a été apportée à l'aide des techniques de régression non paramétrique. En effet, on montre que dans le cas de fonctions constantes par morceaux sur des blocs de même nombre de données ou équiprobables (des extensions partielles au cas de fonctions générales sont aussi données), l'erreur peut se décomposer en la somme de trois termes :

- Erreur liée à l'estimation des coefficients de réduction de variance par un nombre fini de simulations dans la première étape. C'est ici que le choix d'un nombre relativement faible de variables de contrôle assure une erreur petite.
- Erreur statistique dû au nombre fini de simulations dans l'étape de régression.
- Erreur d'approximation habituelle.

Une extension de la méthode VCP pour le cas d'un produit à plusieurs périodes a été proposée dans le chapitre 5 sous forme de trois variantes possibles. Notre analyse s'est limitée à des tests numériques qui se sont révélés concluants pour une des variantes. Ainsi, une quantification théorique de l'erreur justifiant nos observations numériques est envisageable pour une recherche ultérieure.

Le chapitre 6 a été consacré à l'étude du contrat GMWB selon le point de vue couvreur sur les deux plans théoriques et numériques en se basant sur les résultats des chapitres précédents. D'abord, nous avons pu montrer en se basant sur l'analyse faite au chapitre 2 que la stratégie correspondante au pire cas pour le couvreur est de type bang-bang. Ainsi, le problème de valorisation du coût de couverture étant réduit à celui d'un problème d'arrêt optimal, un algorithme de valorisation numé-

rique par méthode de Monte Carlo de type Longstaff Schwartz est possible. On a adapté la méthode VCP en proposant des variantes à l'algorithme avec différents choix de fonctions de base pour l'étape de régression. La première conclusion satisfaisante est que la conservation de la propriété d'homogénéité dans les fonctions de bases choisies permet d'obtenir un meilleur intervalle de confiance pour le coût de couverture. La deuxième conclusion non satisfaisante constatée est que les variantes de la méthodes VCP proposée n'apportent pas d'amélioration à notre problème de valorisation. Ainsi, nous proposons une quatrième variante de la méthode VCP, que nous pensons être mieux adaptée à l'itération sur les temps d'arrêts à la Longstaff-Schwartz : les tests numériques sont en cours.

Dans le dernier chapitre, nous avons proposé une valorisation par formules fermées approchées pour des contrats GMWB sans rachat et sans ratchet dans la cadre d'un modèle Black Scholes avec taux d'intérêt de dynamique Hull-White à 1 facteur. En utilisant l'idée de développement asymptotique dans [BGM09] [BGM12] et en se basant sur un travail récent d'Etore et Gobet [EG12] à propos de cette méthode pour les dividendes discrets, nous avons approché le prix du contrat GMWB par un prix Black-Scholes corrigé par une somme explicite de Grecques, le tout étant plus rapide à évaluer. Nous sommes ainsi arrivés à valoriser des contrats de maturité 20 ans avec une erreur relative de l'ordre de 0.1%, en utilisant un développement de l'ordre 5. Les tests numériques ont exhibé que l'erreur est plus sensible aux paramètres de la dynamique du taux d'intérêt (*mean reversion* et volatilité) qu'à la volatilité du sous-jacent. Ces conclusions numériques ont été confortées par une analyse mathématique de l'erreur. Notre apport par rapport à Peng, Leung et Kwok [PLK12] qui proposent deux bornes inférieure et supérieure pour le prix est double : d'abord on donne une quantification de l'erreur en fonction des volatilités, de la *mean reversion* et de la maturité. Deuxièmement, notre approche offre toujours la possibilité d'augmenter la précision en calculant les termes correctifs pour des ordres supérieurs, et ce, de manière générique. Pour plustard, il serait intéressant d'étendre cette méthode en intégrant un critère de rationalité du client selon lequel il peut sortir prématurément de son contrat.

Chapitre 2

Etude du contrat GMWB

2.1 Description et modélisation

2.1.1 Notations

Considérons un contrat GMWB avec les notations suivantes :

- T : la maturité du contrat.
- S : le titre sous-jacent du contrat. Sa valeur en t est notée S_t .
- $T_0 = 0$: la date de valorisation.
- T^{def} : la date de la *defferal period* (période de report des retraits).
- $r^{rol} \geq 0$: le taux de *rollup* (roulement).
- A_t^v : l'*account value* (valeur du compte) en t .
- $P_0 \geq 0$: la prime initiale.
- β : le pourcentage de A^v pris par Axa sous forme de charges. Les dates de déduction de charges (*fees*) sont notées $T_i^f, 1 \leq i \leq n_f$. On suppose $0 \leq \beta \leq 1$.
- $\gamma \geq 0$: le taux de retrait. Les dates de retrait sont notées $T_i^w, 1 \leq i \leq n_w$. Elles sont postérieures à T^{def} . On suppose aussi $0 \leq \gamma \leq 1$.
- r_t : le taux d'intérêt instantané à la date t . Il est a priori stochastique.

2.1.2 Dynamique

On suppose que S a une dynamique de diffusion avec volatilité locale $\sigma_t = \sigma(t, S_t)$ bornée :

$$\frac{dS_t}{S_t} = r_t dt + \sigma_t dW_t, \quad S_0 \text{ donné,}$$

où W est un mouvement brownien standard (directement écrit sous la probabilité risque neutre).

Le taux $(r_t)_t$ suit un modèle gaussien à 1 facteur (Hull et White) de corrélation déterministe avec S . On suppose que (S, r) forme un processus de diffusion. On

rappelle la notation : $A_{t-}^v = \lim_{s \rightarrow t, s < t} A_s^v$, c'est à dire la valeur de A^v juste avant l'instant t (qui peut être différente de A_t^v du fait des rachats, des retraits et des fees). Le saut de A^v est noté : $\Delta A_t^v = A_t^v - A_{t-}^v$.

La dynamique de A^v s'écrit sous la forme :

$$\begin{aligned} A_0^v &= P_0, \\ dA_t^v &= (A_{t-}^v)_+ \frac{dS_t}{S_t} - \underbrace{dC_t}_{\text{rachat}} - \underbrace{\beta(A_{t-}^v)_+ dZ_t^f}_{\text{fees}} - \underbrace{\gamma B_{t-}^w dZ_t^w}_{\text{retrait}} \end{aligned}$$

où l'on note $(x)_+$ la partie positive de x . Autrement dit, entre deux instants de rachat/fees/retrait, et avant que A^v ne devienne négative, les rendements de A^v et de S sont égaux :

$$\frac{dA_t^v}{A_t^v} = \frac{dS_t}{S_t}.$$

Sinon, aux instants de rachat/fees/retrait on a :

$$A_t^v = A_{t-}^v - (C_t - C_{t-}) - \beta(A_{t-}^v)_+ \mathbf{1}_{\{t \text{ est un instant de fees}\}} - \gamma B_{t-}^w \mathbf{1}_{\{t \text{ est un instant de retrait}\}},$$

où les différentes quantités sont définies comme suit.

- $Z_t^f = \sum_{i=1}^{n_f} \mathbf{1}_{\{T_i^f \leq t\}}$ (le processus de comptage des fees),
- $Z_t^w = \sum_{i=1}^{n_w} \mathbf{1}_{\{T_i^w \leq t\}}$ (le processus de comptage des retraits (*withdrawals*)),
- C_t représente le montant de consommation de A^v que le client peut obtenir à n'importe quel moment en demandant de racheter partiellement ou totalement son A^v . **On suppose que lors des moments de fee ou retrait, le rachat ne peut avoir lieu, sinon il y aurait une ambiguïté sur le calcul des fees et du retrait.** Le client peut retirer au maximum son A^v . En notant α_t le pourcentage racheté de A^v à la date t , on récupère ainsi la contrainte simple :

$$\boxed{0 \leq \alpha_t \leq 1}$$

(On suppose aussi que le client ne rajoute pas d'argent sur son compte, ce qui est traduit par la positivité de α_t). Ainsi, la dynamique de C aura la forme :

$$\begin{aligned} C_0 &= 0, \\ dC_t &= \alpha_t (A_{t-}^v)_+ dZ_t^c \text{ avec } 0 \leq \alpha_t \leq 1, 0 \leq t \leq T. \end{aligned}$$

Ici Z^c est le processus de comptage associé aux dates de rachats du client.

Sous forme intégrale, on écrit : $C_t = \int_0^t \alpha_s (A_{s-}^v)_+ dZ_s^c, 0 \leq t \leq T$, et sous forme de variation entre deux temps : $C_t = C_{t-} + \alpha_t (A_{t-}^v)_+ \mathbf{1}_{\{\Delta Z_t^c \neq 0\}}$.

– B^w est la *Withdrawal Base* (base de retrait) qui sert à calculer les retraits des clients aux dates $T_i^w, 1 \leq i \leq n_w$. Pour la définir, on introduit deux bases auxiliaires : B^{rat} (*Ratchet Base*) et B^{rol} (*Rollup Base*).

1. Pour calculer B^{rat} , on note $T_i^r, 0 \leq i \leq n_r$, les dates quand a lieu le ratchet. On suppose $T_0^r = T_0$ et $\exists p \in \mathbb{N}$ tel que $T_p^r = T^{def}$. **On suppose aussi que le rachat ne peut avoir lieu au moment du ratchet.** On définit B_i^{rat} , la valeur de B^{rat} en t par :

$$\begin{aligned} B_0^{rat} &= A_0^v, \\ dB_t^{rat} &= -\alpha_t B_{t-}^{rat} dZ_t^c \quad \text{si } T_i^r < t < T_{i+1}^r \quad \text{avec } 0 \leq i \leq n_r - 1, \\ B_{T_i^r}^{rat} &= \max(B_{T_i^r-}^{rat}, A_{T_i^r-}^v) \quad \text{pour } 1 \leq i \leq n_r. \end{aligned}$$

Il est à noter que $(B_t^{rat})_t$ reste positif au cours du temps car $\alpha_t \leq 1$.

2. La B^{rol} est calculée en prenant en compte les rachats comme suit :

$$\begin{aligned} B_0^{rol} &= A_0^v, \\ dB_t^{rol} &= r^{rol} B_{t-}^{rol} dt - \alpha_t B_{t-}^{rol} dZ_t^c \quad \text{pour } t \leq T. \end{aligned}$$

Comme $\alpha_t \leq 1$, $(B_t^{rol})_t$ reste aussi positif.

Ainsi, B^w est définie comme suit :

$$\begin{aligned} B_t^w &= B_t^{rat} \quad \text{pour } t < T_p^r = T^{def}, \\ B_t^w &= \max(B_t^{rat}, B_t^{rol}) \quad \text{pour } t = T_p^r = T^{def}, \\ dB_t^w &= -\alpha_t B_{t-}^w dZ_t^c \quad \text{si } T_i^r < t < T_{i+1}^r \quad \text{avec } p \leq i \leq n_r - 1, \\ B_{T_i^r}^w &= \max(B_{T_i^r-}^w, A_{T_i^r-}^v) \quad \text{pour } p + 1 \leq i \leq n_r. \end{aligned}$$

La withdrawal base reste également positive.

On remarque que dès que A^v passe négative, seuls les retraits sont conservés parmi les trois opérations rachat/fees/retrait. Par suite, A^v devient constante par morceaux.

En cas de décès du client, l'assuré reçoit le maximum entre A^v et une garantie décès qu'on note B^d (*Death Base*). Cette garantie est égale au début à la prime initiale et diminue au cours du temps au prorata de rachat partiel, et aussi soit par le montant que perçoit le client à la date de retrait, soit au prorata du pourcentage que représente le montant par rapport à A^v . Ainsi, sa dynamique est la suivante :

$$\begin{cases} B_0^d = A_0^v, \\ dB_t^d = -\alpha_t (B_{t-}^d)_+ dZ_t^c - \gamma B_{t-}^w dZ_t^w \quad \text{pour } t \leq T, \end{cases}$$

ou

$$dB_t^d = -\alpha_t(B_{t-}^d)_+ dZ_t^c - \frac{\gamma B_{t-}^w}{A_{t-}^v} B_{t-}^d dZ_t^w \quad \text{pour } t \leq T,$$

ou

$$\begin{cases} B_t^r = r^d B_{t-}^r dt - \alpha_t B_{t-}^r dZ_t^c, \\ dB_t^d = -\alpha_t(B_{t-}^d)_+ dZ_t^c - \gamma B_{t-}^w dZ_t^w + \mathbf{1}_{\{t \neq T^d\}}(A_{t-}^v - B_{t-}^d)_+ dZ_t^r \\ + \mathbf{1}_{\{t=T^d\}} \max((B_{t-}^r - B_{t-}^d)_+, (A_{t-}^v - B_{t-}^d)_+) dZ_t^r. \end{cases}$$

2.1.3 Dates de rachat

Les décisions du client ne concernent que le rachat, en date et en pourcentage. On suppose que **les décisions de rachat prises par le client surviennent de manière discrète aux dates** $T_i^c, 1 \leq i \leq n_c$ (par exemple une fois par semaine). La nature de T_i^c (aléatoire ou déterministe) sera discutée après. Donc Z_t^c a la forme :

$$Z_t^c = \sum_{i=1}^{n_c} \mathbf{1}_{\{T_i^c \leq t\}}.$$

Discutons rapidement de la nature des temps T_i^c (aléatoire ou déterministe). En fait, **on peut considérer ces dates déterministes** (avec une période à fixer), quitte à transférer l'incertain des rachats sur les quantités α_t . Notamment, en absence de rachat en T_i^c , on a $\alpha_{T_i^c} = 0$.

2.1.4 Flux

On identifie les flux du produit :

– Axa reçoit à chaque date $T_i^f, 1 \leq i \leq n_f$, un flux de fees :

$$\beta(A_{T_i^f-}^v)_+.$$

– Axa s'engage à verser au client à la date $T_i^w, 1 \leq i \leq n_w$ le flux associé aux retraits :

$$\gamma B_{T_i^w-}^w.$$

– Axa s'engage aussi aux dates de rachats $T_i^c, 1 \leq i \leq n_c$ à verser :

$$\alpha_{T_i^c}(A_{T_i^c-}^v)_+.$$

– A maturité T , Axa restitue au client son A^v si elle est positive, soit :

$$(A_T^v)_+.$$

2.1.5 Variables markoviennes du problème et contrôles

D'après la description précédente, on peut résumer les variables du problème comme suit :

- pour $t \leq T^{def}$, les variables markoviennes à considérer sont : S , A^v , B^{rat} , B^{rol} et r .
- pour $t > T^{def}$, les variables markoviennes à considérer sont : S , A^v , B^w , r .

L'identification de ces variables permet d'obtenir la dépendance des fonctions de base à utiliser pour les régressions empiriques dans un algorithme de valorisation de type Longstaff-Schwartz (Monte Carlo américain).

Les processus de contrôles sont les pourcentages $(\alpha_t)_t$ de rachat du client.

2.2 Typologie du client

On peut diviser les acheteurs de contrats en deux catégories :

1. Le client rationnel qui exerce son droit au rachat de manière optimale (pire cas pour Axa).
2. Le client irrationnel qui prend ses décisions de rachat d'une manière aléatoire.

2.2.1 Cas du client rationnel

Dans ce cas, les pourcentages rachetés $(\alpha_t)_t$ sont des contrôles stochastiques optimaux. Du point de vue du couvreur du contrat, la valeur résultante du contrat est donc :

$$\Pi_0 = \sup_{\alpha_t} \left\{ \mathbb{E} \left[\gamma \sum_{i=1}^{n_w} D(0, T_i^w) B_{T_i^w-}^w + \sum_{i=1}^{n_c} \alpha_{T_i^c} D(0, T_i^c) (A_{T_i^c-}^v)_+ - \beta \sum_{i=1}^{n_f} D(0, T_i^f) (A_{T_i^f-}^v)_+ + D(0, T) (A_T^v)_+ \right] \right\}$$

où $D(0, t) = e^{-\int_0^t r_s ds}$ est le facteur d'actualisation (*discount factor*) stochastique.

Ici la valeur du contrat doit s'entendre comme étant le prix de couverture parfaite : l'espérance \mathbb{E} est donc l'espérance sous la probabilité risque neutre. C'est analogue à [Kar88] pour les options américaines.

2.2.2 Cas du client non rationnel

Dans ce cas, les $\alpha_{T_i^c}$ sont aléatoires et non optimaux. Donc, il y a des choix à faire sur la loi des $(\alpha_t)_t$ (par exemple, indépendant des conditions de marché). Nous

écartons ce cas de notre analyse numérique car cela se ramène à l'utilisation d'une méthode de Monte Carlo standard pour un produit à exercice européen.

2.3 Vers la résolution du cas du client rationnel

2.3.1 Le problème de contrôle

Par souci de simplification, on introduit les notations suivantes pour regrouper toutes les dates de rachat/fees/rachet en un seul ensemble :

$$\{T_i\}_{0 \leq i \leq n} = \{T_i^r\}_{0 \leq i \leq n_r} \cup \{T_i^c\}_{1 \leq i \leq n_c} \cup \{T_i^f\}_{1 \leq i \leq n_f} \cup \{T_i^w\}_{1 \leq i \leq n_w} \cup \{T\}.$$

On définit pour $0 \leq i \leq n-1$, α_i , β_i et γ_i comme suit :

- $\alpha_i = \alpha_{T_j^c}$ si T_i correspond à une certaine date de rachat T_j^c , sinon on pose $\alpha_i = 0$.
- $\beta_i = \alpha_{T_j^f}$ si T_i correspond à une certaine date de fees T_j^f , sinon on pose $\beta_i = 0$.
- $\gamma_i = \alpha_{T_j^w}$ si T_i correspond à une certaine date de retrait T_j^w , sinon on pose $\gamma_i = 0$.
- \mathcal{U}_i est le domaine du choix du contrôle α_i en T_i .
Donc $\mathcal{U}_i = [0, 1]$ si T_i est une date de rachat, et $\mathcal{U}_i = \{0\}$ sinon.

La dynamique de notre problème peut être décrite par la chaîne de Markov suivante :

$$(i, S_i, A_i^v, B_i^w, B_i^{rol}, r_i, D_i)_{0 \leq i \leq n} = (i, S_{T_i-}, A_{T_i-}^v, B_{T_i-}^w, B_{T_i-}^{rol}, r_{T_i-}, D(0, T_i-))_{0 \leq i \leq n}.$$

Cette dynamique est contrôlée par la stratégie de rachat $(\alpha_i)_{1 \leq i \leq n-1}$.

A l'état (i, S, a, b, c, r, D) , le prix actualisé du contrat est :

$$\begin{aligned} & P(i, S, a, b, c, r, D) \\ &= \sup_{(\alpha_j)_{i \leq j \leq n-1}} \mathbb{E} \left\{ \sum_{j=i}^{n-1} D_j [(\alpha_j - \beta_j)(A_j^v)_+ + \gamma_j B_j^w] + D_n (A_n^v)_+ \right. \\ & \quad \left. | S_i = S, A_i^v = a, B_i^w = b, B_i^{rol} = c, r_i = r, D_i = D \right\}. \end{aligned}$$

Cette dernière équation peut être écrite sous la forme :

$$\begin{aligned} & P(i, S, a, b, c, r, D) \\ &= \sup_{(\alpha_j)_{i \leq j \leq n-1}} \mathbb{E} \left\{ \sum_{j=i}^{n-1} \Psi_j(\alpha_j, A_j^v, B_j^w, D_j) + D_n (A_n^v)_+ \right. \\ & \quad \left. | S_i = S, A_i^v = a, B_i^w = b, B_i^{rol} = c, r_i = r, D_i = D \right\}, \end{aligned}$$

avec :

$$\Psi_j(\alpha, a, b, D) = \begin{cases} D[(\alpha - \beta_j)(a)_+ + \gamma_j b] & \text{pour } j < n \\ D(a)_+ & \text{pour } j = n. \end{cases} \quad (2.3.1)$$

Démontrons maintenant qu'il existe une stratégie optimale de rachat $(\alpha_i^*)_{0 \leq i \leq n-1}$. On utilise pour cela la théorie de contrôle stochastique optimal en temps discret et principalement la proposition 8.5 de [BS78]. D'abord, introduisons la notion de stratégies admissibles markoviennes.

Définition 2.3.1 (voir définition 7.12 et proposition 7.26 de [BS78])

Notons $E = \mathbb{R}^+ \times \mathbb{R} \times \mathbb{R}^+ \times \mathbb{R}^+ \times \mathbb{R} \times \mathbb{R}^+$.

Une famille de mesures de probabilité sur $[0, 1]$ $(p_e)_{e \in E}$ est dite noyau stochastique mesurable sur $[0, 1]$ si l'application $f_B : E \rightarrow [0, 1]$ définie par $f_B(e) = p_e(B)$ est mesurable pour tout B ouvert de E .

Définition 2.3.2 Une suite de variables aléatoires $(\alpha_j)_{i \leq j \leq n-1}$ est dite stratégie admissible markovienne à l'état $(i, x = (S, a, b, c, r, D))$ pour le problème de contrôle considéré si

1. Pour tout $i \leq j \leq n-1$, la famille des mesures de probabilité conditionnelle de α_j sachant $(S_j, A_j^v, B_j^w, B_j^{rol}, r_j, D_j, \alpha_{j-1}, S_{j-1}, A_{j-1}^v, B_{j-1}^w, B_{j-1}^{rol}, r_{j-1}, D_{j-1}, \dots, \alpha_i, S_i, A_i^v, B_i^w, B_i^{rol}, r_i, D_i)$ ne dépend que de $(S_j, A_j^v, B_j^w, B_j^{rol}, r_j, D_j)$, et c'est un noyau stochastique mesurable sur $[0, 1]$ vérifiant

$$\mathbb{P}(\alpha_j \in \mathcal{U}_i | (S_j, A_j^v, B_j^w, B_j^{rol}, r_j, D_j) = (S_1, r_1, D_1, a_1, b_1, c_1)) = 1$$

pour tout $(S_1, a_1, b_1, c_1, r_1, D_1) \in E$.

L'ensemble de ce type de stratégies est noté \mathcal{C}_x^i .

La stratégie admissible est non randomisée si la loi conditionnelle de α_j est une mesure de Dirac (ne charge qu'un élément).

On remarque que selon la définition ci dessus \mathcal{C}_x^i est indépendant de x et peut être noté uniquement par \mathcal{C}^i .

Proposition 2.3.1 (Voir proposition 8.5 de [BS78])

Supposons

- a) Pour toute stratégie $(\alpha_j)_{0 \leq j \leq n-1}$ et pour tout $0 \leq i \leq n-1$ on a :

$$\mathbb{E}(|\Psi_i(\alpha_i, A_i^v, B_i^w, D_i)|) < \infty,$$

- b) et pour tout $0 \leq i \leq n-1$ et $(S, a, b, c, r, D) \in E$

$$\sup_{\alpha \in \mathcal{U}_i} \{\Psi_i(\alpha, a, b, D) + \mathbb{E}^\alpha [P(i+1, S_{i+1}, A_{i+1}^v, B_{i+1}^w, B_{i+1}^{rol}, r_{i+1}, D_{i+1}) | S_i = S, A_i^v = a, B_i^w = b, B_i^{rol} = c, r_i = r, D_i = D]\}$$

est atteint (\mathbb{E}^α est l'espérance associée à la chaîne de Markov à laquelle on applique le contrôle α au temps i).

Alors il existe une stratégie markovienne optimale non randomisée, c'est à dire une suite de contrôles $(\alpha_0^*, \alpha_1^*, \dots, \alpha_{n-1}^*)$ tel que chaque α_i^* dépend uniquement de $(S_i, A_i^v, B_i^w, B_i^{rol}, r_i, D_i)$, i.e. l'état de la chaîne de Markov en i . Cette stratégie est générée par l'algorithme de programmation dynamique, i.e. en sélectionnant pour chaque (S, a, b, c, r, D) le contrôle α pour lequel le sup est atteint dans l'équation de la programmation dynamique suivante :

$$\begin{aligned} P(i, S, a, b, c, r, D) &= \sup_{\alpha \in \mathcal{U}_i} \{ \Psi_i(\alpha, a, b, D) \\ &\quad + \mathbb{E}^\alpha(P(i+1, S_{i+1}, A_{i+1}^v, B_{i+1}^w, B_{i+1}^{rol}, r_{i+1}, D_{i+1}) \\ &\quad | S_i = S, A_i^v = a, B_i^w = b, B_i^{rol} = c, r_i = r, D_i = D) \} \\ &\quad (0 \leq i \leq n-1), \\ P(n, S, a, b, c, r, D) &= (a)_+ D. \end{aligned}$$

Notons :

$$\begin{aligned} J_i(\alpha, S, a, b, c, r, d) &= \Psi_i(\alpha, a, b, D) + \mathbb{E}^\alpha[P(i+1, S_{i+1}, A_{i+1}^v, B_{i+1}^w, B_{i+1}^{rol}, r_{i+1}, D_{i+1}) \\ &\quad | S_i = S, A_i^v = a, B_i^w = b, B_i^{rol} = c, r_i = r, D_i = D]. \end{aligned}$$

Puisque α prend ses valeurs dans la compact $[0, 1]$ ou $\{0\}$, il suffit de vérifier la continuité de J_i par rapport à α .

2.3.2 Résultat Principal : Caractérisation du rachat optimal

Avant d'énoncer le théorème 2.3.1, pour chaque $0 \leq i \leq n$, notons par :

- $x = (S, a, b, c, r, D) \in E$.
- $(S_t^{i,x}, A_t^{v,i,x}, B_t^{w,i,x}, B_t^{r,i,x}, r_t^{i,x}, D_t^{i,x})_{T_i \leq t \leq T}$, le processus à temps continu $(S_t, A_t^v, B_t^w, B_t^{rol}, r_t, D_t)_{T_i \leq t \leq T}$ partant en T_i de x mais pour un processus $(\alpha_t)_t$ donné.
- $(S_j^{i,x}, A_j^{v,i,x}, B_j^{w,i,x}, B_j^{r,i,x}, r_j^{i,x}, D_j^{i,x})_{i+1 \leq j \leq n}$, le processus à temps discret $(S_{T_j-}, A_{T_j-}^v, B_{T_j-}^w, B_{T_j-}^{rol}, r_{T_j-}, D_{T_j-})_{i+1 \leq j \leq n}$ partant en T_i de x (i.e. $(S_{T_i}, A_{T_i}^v, B_{T_i}^w, B_{T_i}^{rol}, r_{T_i}, D_{T_i}) = x$) mais pour un processus $(\alpha_t)_t$ donné.

Théorème 2.3.1 *Il existe une stratégie markovienne optimale non randomisée $(\alpha_i^*)_{0 \leq i \leq n-1}$. De plus, en chaque date i qui correspond à une date de rachat, le contrôle optimal est de type bang-bang. En notant par $x = (S, a, b, c, r, D)$ l'état de la chaîne de Markov en T_i , on a :*

$$\alpha_i^*(S, a, b, c, r, D) = \begin{cases} \mathbf{1}_{\{\mathbb{E}(P(i+1, S_{i+1}, A_{i+1}^{v,i,x}, B_{i+1}^{w,i,x}, B_{i+1}^{r,i,x}, r_{i+1}^{i,x}, D_{i+1}^{i,x})) \leq (a)+D\}} \\ \text{(si } T_i \text{ est une date de rachat),} \\ 0 \text{ (sinon).} \end{cases}$$

Ici, en exploitant une propriété remarquable d'homogénéité (voir lemmes qui suivent), on montre que J_i est une fonction affine en α , donc continue. De plus, le supremum de cette fonction est forcément atteint au borne de l'intervalle $[0,1]$, donc la stratégie optimale est bang-bang (on rachète tout ou rien).

L'intérêt de savoir que le contrôle optimal est bang-bang est d'avoir juste à déterminer la date optimale du rachat complet du contrat, ce qui se ramène à un problème d'arrêt optimal (option américaine).

2.3.3 Preuve du théorème 2.3.1

Pour démontrer le théorème nous utiliserons la proposition 2.3.1 et les lemmes qui suivent.

Lemme 2.3.2 *(Majorsations a priori sur les processus) Pour $0 \leq t \leq T$:*

$$\begin{aligned} (A_t^v)_+ &\leq A_0^v \frac{S_t}{S_0}, \\ 0 \leq B_t^w &\leq e^{r \cdot \text{rol} T^{def}} A_0^v \sup_{0 \leq T_i \leq t} \frac{S_{T_i}}{S_0}, \\ 0 \leq B_t^{\text{rol}} &\leq e^{r \cdot \text{rol} t} A_0^v. \end{aligned}$$

Lemme 2.3.3 *(Homogénéité de degré 1 par rapport à A^v , B^w et B^{rol}) Pour tout $\lambda > 0$, $0 \leq i \leq n$, $x = (S, a, b, c, r, D) \in E$ et un contrôle $(\alpha_j)_{i \leq j \leq n} \in \mathcal{C}^i$, on a :*

$$\begin{aligned} (A_t^{v,i,x^\lambda}, B_t^{w,i,x^\lambda}, B_t^{r,i,x^\lambda}, r_t^{i,x^\lambda}, D_t^{i,x^\lambda})_{T_i \leq t \leq T} &= (\lambda A_t^{v,i,x}, \lambda B_t^{w,i,x}, \lambda B_t^{r,i,x}, r_t^{i,x}, D_t^{i,x})_{T_i \leq t \leq T} \text{ p.s.}, \\ (A_j^{v,i,x^\lambda}, B_j^{w,i,x^\lambda}, B_j^{r,i,x^\lambda}, r_j^{i,x^\lambda}, D_j^{i,x^\lambda})_{i+1 \leq j \leq n} &= (\lambda A_j^{v,i,x}, \lambda B_j^{w,i,x}, \lambda B_j^{r,i,x}, r_j^{i,x}, D_j^{i,x})_{i+1 \leq j \leq n} \text{ p.s.}, \\ P(i, S, \lambda a, \lambda b, \lambda c, r, D) &= \lambda P(i, S, a, b, c, r, D) \end{aligned}$$

où $x^\lambda = (S, \lambda a, \lambda b, \lambda c, r, D)$.

Preuve du Lemme 2.3.2

Pour cela, nous allons démontrer par récurrence sur $0 \leq i \leq n$ que les trois inégalités sont vérifiées sur $[T_i, T_{i+1}[$.

Pour $i = 0$, on a $T_0 = 0$ et donc $:A_{T_0}^v = A_0^v$. Soit $T_0 \leq t < T_1$, donc :

$$dA_t^v = (A_t^v)_+ \frac{dS_t}{S_t}.$$

On note que $(A_0^v \frac{S_t}{S_0})_t$ est bien solution sur $[T_0, T_1[$. Par unicité à l'EDS dirigée par $dZ_t = \frac{dS_t}{S_t}$ (voir [Pro04], chapitre 5, section 3), on déduit

$$A_t^v = A_0^v \frac{S_t}{S_0}.$$

On a aussi :

$$\begin{aligned} 0 \leq B_t^w &= B_0^w = A_0^v \leq e^{r^{rol} T^{def}} A_0^v \sup_{0 \leq T_i \leq t} \frac{S_{T_i}}{S_0}, \\ 0 \leq B_t^{rol} &= e^{r^{rol} t} A_0^v. \end{aligned}$$

Ainsi pour $i = 0$, les trois propriétés sont vérifiées sur $[T_i, T_{i+1}[$.

Supposons alors qu'elles sont vérifiées pour $0 \leq i \leq n - 2$ sur $[T_i, T_{i+1}[$ et démontrons les sur $[T_{i+1}, T_{i+2}[$. Un passage à la limite pour $t \uparrow T_{i+1-}$ dans l'hypothèse de récurrence permet d'avoir :

$$\begin{aligned} (A_{T_{i+1-}}^v)_+ &\leq A_0^v \frac{S_{T_{i+1}}}{S_0}, \\ 0 \leq B_{T_{i+1-}}^w &\leq e^{r^{rol} T^{def}} A_0^v \sup_{0 \leq j \leq i} \frac{S_{T_j}}{S_0}, \\ 0 \leq B_{T_{i+1-}}^{rol} &\leq e^{r^{rol} T_{i+1}} A_0^v. \end{aligned}$$

D'autre part on a :

$$A_{T_{i+1}}^v = A_{T_{i+1-}}^v - (\alpha_i + \beta_i)(A_{T_{i+1-}}^v)_+ - \gamma_i B_{T_{i+1-}}^w \leq A_{T_{i+1-}}^v \leq A_0^v \frac{S_{T_{i+1}}}{S_0},$$

et

$$dA_t^v = (A_t^v)_+ \frac{dS_t}{S_t} \text{ sur } [T_{i+1}, T_{i+2}[\text{ .}$$

Notons que $A_{T_{i+1}}^v \frac{S_t}{S_{T_{i+1}}}$ est solution si $A_{T_{i+1}}^v \geq 0$, sinon le processus constant $A_{T_{i+1}}^v$ est solution. Par unicité de la solution, on a pour $T_{i+1} \leq t < T_{i+2}$:

$$A_t^v = \begin{cases} A_{T_{i+1}}^v \frac{S_t}{S_{T_{i+1}}} & \text{si } A_{T_{i+1}}^v \geq 0, \\ A_{T_{i+1}}^v & \text{sinon.} \end{cases}$$

D'où :

$$(A_t^v)_+ \leq (A_{T_{i+1}}^v)_+ \frac{S_t}{S_{T_{i+1}}} \leq A_0^v \frac{S_{T_{i+1}}}{S_0} \frac{S_t}{S_{T_{i+1}}} = A_0^v \frac{S_t}{S_0}.$$

Concernant B^w , on a :

$$B_{T_{i+1}}^w = \begin{cases} \max(B_{T_{i+1}-}^{rol}, \max(B_{T_{i+1}-}^w, A_{T_{i+1}-}^v)) & \text{si } T_{i+1} = T^{def} \quad , \\ \max(B_{T_{i+1}-}^w, A_{T_{i+1}-}^v) & \text{si } T_{i+1} \neq T^{def} \text{ et } T_{i+1} \text{ est une date de ratchet,} \\ (1 - \alpha_i)B_{T_{i+1}-}^w & \text{sinon .} \end{cases}$$

En utilisant les trois comparaisons démontrées en T_{i+1-} , on obtient :

– si $T_{i+1} = T^{def}$ alors :

$$\begin{aligned} B_{T_{i+1}}^w &\leq \max(e^{r^{rol}T_{i+1}} A_0^v, e^{r^{rol}T^{def}} A_0^v \sup_{0 \leq j \leq i} \frac{S_{T_j}}{S_0}, A_0^v \frac{S_{T_{i+1}}}{S_0}) \\ &\leq e^{r^{rol}T^{def}} A_0^v \sup_{0 \leq j \leq i+1} \frac{S_{T_j}}{S_0}. \end{aligned}$$

– si $T_{i+1} \neq T^{def}$ et T_{i+1} est une date de ratchet :

$$\begin{aligned} B_{T_{i+1}}^w &\leq \max(e^{r^{rol}T^{def}} A_0^v \sup_{0 \leq j \leq i} \frac{S_{T_j}}{S_0}, A_0^v \frac{S_{T_{i+1}}}{S_0}) \\ &\leq e^{r^{rol}T^{def}} A_0^v \sup_{0 \leq j \leq i+1} \frac{S_{T_j}}{S_0}. \end{aligned}$$

– Sinon :

$$\begin{aligned} B_{T_{i+1}}^w &\leq B_{T_{i+1}-}^w \leq e^{r^{rol}T^{def}} A_0^v \sup_{0 \leq j \leq i} \frac{S_{T_j}}{S_0} \\ &\leq e^{r^{rol}T^{def}} A_0^v \sup_{0 \leq j \leq i+1} \frac{S_{T_j}}{S_0}. \end{aligned}$$

En résumé, on montré que $B_{T_{i+1}}^w \leq e^{r^{rol}T^{def}} A_0^v \sup_{0 \leq j \leq i+1} \frac{S_{T_j}}{S_0}$ dans tous les cas, ce qui implique pour $T_{i+1} \leq t < T_{i+2}$:

$$0 \leq B_t^w = B_{T_{i+1}}^w = e^{r^{rol}T^{def}} A_0^v \sup_{0 \leq T_j \leq t} \frac{S_{T_j}}{S_0}.$$

Enfin, pour la base rollup on a :

$$\begin{aligned} B_{T_{i+1}}^{rol} &= (1 - \alpha_i)B_{T_{i+1}-}^{rol} \leq e^{r^{rol}T_{i+1}} A_0^v, \\ B_t^{rol} &= e^{(r^{rol}(t-T_{i+1}))} B_{T_{i+1}}^{rol} \leq e^{r^{rol}t} A_0^v. \end{aligned}$$

Ainsi l'hypothèse de récurrence est vérifiée pour $i + 1$ sur $[T_{i+1}, T_{i+2}[$.

En conclusion, pour $0 \leq t < T_n = T$, on a

$$\begin{aligned} (A_t^v)_+ &\leq A_0^v \frac{S_t}{S_0}, \\ 0 \leq B_t^w &\leq e^{r^{rol} T^{def}} A_0^v \sup_{0 \leq T_i \leq t} \frac{S_{T_i}}{S_0}, \\ 0 \leq B_t^{rol} &\leq e^{r^{rol} t} A_0^v. \end{aligned}$$

Puisqu'aucune opération de rachat/fees/retrait/ratchet n'a lieu en T , les mêmes inégalités s'étendent pour T . \square

Preuve du Lemme 2.3.3

Pour un contrôle $(\alpha_k)_{i \leq k \leq n-1}$ donné, on a pour $i \leq j \leq n-1$

$$\begin{aligned} A_{T_j}^v &= A_{T_j-}^v - (\alpha_j + \beta_j)(A_{T_j-}^v)_+ - \gamma_j B_{T_j-}^w \\ dA_t^v &= (A_{t-}^v)_+ \frac{dS_t}{S_t} \text{ sur } [T_j, T_{j+1}[\quad . \end{aligned}$$

Donc pour $T_j \leq t < T_{j+1}$

$$A_t^v = \begin{cases} (A_{T_j-}^v - (\alpha_j + \beta_j)(A_{T_j-}^v)_+ - \gamma_j B_{T_j-}^w) \frac{S_t}{S_{T_j}} \\ \text{si } (A_{T_j-}^v - (\alpha_j + \beta_j)(A_{T_j-}^v)_+ - \gamma_j B_{T_j-}^w) \geq 0, \\ A_{T_j-}^v - (\alpha_j + \beta_j)(A_{T_j-}^v)_+ - \gamma_j B_{T_j-}^w \text{ sinon.} \end{cases} \quad (2.3.2)$$

On a aussi pour $T_j \leq t < T_{j+1}$

$$B_t^w = \begin{cases} \max(B_{T_j-}^{rol}, \max(B_{T_j-}^w, A_{T_j-}^v)) & \text{si } T_j = T^{def} \quad , \\ \max(B_{T_j-}^w, A_{T_j-}^v) & \text{si } T_{j+1} \neq T^{def} \\ \text{et } T_j \text{ est une date de ratchet,} \\ (1 - \alpha_j) B_{T_j-}^w & \text{sinon,} \end{cases} \quad (2.3.3)$$

$$B_t^{rol} = (1 - \alpha_j) e^{r^{rol}(t-T_j)} B_{T_j-}^{rol}. \quad (2.3.4)$$

Au vu de (2.3.2), (2.3.3) et (2.3.4) et en faisant une récurrence sur j , il est facile de voir que le premier et le deuxième résultat du lemme sont vérifiés.

Concernant le troisième résultat, posons pour $\alpha = (\alpha)_{i \leq j \leq n-1} \in \mathcal{C}^i$:

$$\begin{aligned} P(i, S, a, b, c, r, D, \alpha) &= \mathbb{E} \left\{ \sum_{j=i}^{n-1} \Psi_j(\alpha_j, A_j^v, B_j^w, D_j) + D_n(A_T^v)_+ \right. \\ &\quad \left. | S_i = S, A_i^v = a, B_i^w = b, B_i^{rol} = c, r_i = r, D_i = D \right\}. \end{aligned}$$

On a,

$$\begin{aligned}
P(i, S, \lambda a, \lambda b, \lambda c, r, D, \alpha) &= \mathbb{E} \left\{ \sum_{j=i}^{n-1} \Psi_j(\alpha_j, A_j^v, B_j^w, D_j) + D_n(A_T^v)_+ \right. \\
&\quad \left. | S_i = S, A_i^v = \lambda a, B_i^w = \lambda b, B_i^{rol} = \lambda c, r_i = r, D_i = D \right\} \\
&= \mathbb{E} \left\{ \sum_{j=i}^{n-1} \Psi_j(\alpha_j, A_j^{v,i,x^\lambda}, B_j^{w,i,x^\lambda}, D_j^{i,x^\lambda}) + D_n(A_T^{v,i,x^\lambda})_+ \right\} \\
&= \mathbb{E} \left\{ \sum_{j=i}^{n-1} \Psi_j(\alpha_j, \lambda A_j^{v,i,x}, \lambda B_j^{w,i,x}, D_j^{i,x}) + D_n(\lambda A_T^{v,i,x})_+ \right\} \\
&= \lambda \mathbb{E} \left\{ \sum_{j=i}^{n-1} \Psi_j(\alpha_j, A_j^{v,i,x}, B_j^{w,i,x}, D_j^{i,x}) + D_n(A_T^{v,i,x})_+ \right\} \\
&= \lambda P(i, S, a, b, c, r, D, \alpha),
\end{aligned}$$

en utilisant à la 4ème égalité que les flux sont homogènes de degré 1 en A et B (voir 2.3.1).

D'où,

$$\begin{aligned}
P(i, S, \lambda a, \lambda b, \lambda c, r, D) &= \sup_{\alpha \in \mathcal{C}^i} P(i, S, \lambda a, \lambda b, \lambda c, r, D, \lambda) \\
&= \sup_{\alpha \in \mathcal{C}^i} \lambda P(i, S, a, b, c, r, D, \alpha) \\
&= \lambda P(i, S, a, b, c, r, D). \quad \square
\end{aligned}$$

Preuve du Théorème 2.3.1.

Nous vérifierons les hypothèses de la proposition 2.3.1. Le lemme 7.2.1 permet de vérifier facilement la première hypothèse de la proposition car $\mathbb{E}[\sup_{t \leq T} S_t] < \infty$ (σ bornée et $(r_t)_t$ gaussien). Pour vérifier la seconde hypothèse rappelons que :

$$\begin{aligned}
J_i(\alpha, S, r, D, a, b, c) &= \Psi_i(\alpha, a, b, D) \\
&\quad + \mathbb{E}^\alpha \left\{ P(i+1, S_{i+1}, A_{i+1}^v, B_{i+1}^w, B_{i+1}^{rol}, r_{i+1}, D_{i+1}) \right. \\
&\quad \left. | S_i = S, A_i^v = a, B_i^w = b, B_i^{rol} = c, r_i = r, D_i = D \right\}.
\end{aligned}$$

Si T_i n'est pas une date de rachat, $\mathcal{U}_i = \{0\}$ et évidemment $J_i(\cdot, S, r, D, a, b, c)$ atteint son maximum.

Si T_i est une date de rachat, alors $\mathcal{U}_i = [0, 1]$. Puisque les fees, les retraits et le rachat ne peuvent avoir lieu à cette date, on déduit que :

$$\begin{aligned}
J_i(\alpha, S, r, D, a, b, c) &= (a)_+ D \alpha \\
&\quad + \mathbb{E}(P(i+1, S_{i+1}, A_{i+1}^{v,i,x^{1-\alpha}}, B_{i+1}^{w,i,x^{1-\alpha}}, B_{i+1}^{r,i,x^{1-\alpha}}, r_{i+1}^{i,x^{1-\alpha}}, D_{i+1}^{i,x^{1-\alpha}})).
\end{aligned}$$

En utilisant le lemme 2.3.3, on a :

$$\begin{aligned}
J_i(\alpha, S, r, D, a, b, c) &= (a)_+ D\alpha + \mathbb{E}(P(i+1, S_{i+1}, (1-\alpha)A_{i+1}^{v,i,x}, (1-\alpha)B_{i+1}^{w,i,x}, \\
&\quad (1-\alpha)B_{i+1}^{r,i,x}, r_{i+1}^{i,x^{1-\alpha}}, D_{i+1}^{i,x^{1-\alpha}})) \\
&= (a)_+ D\alpha \\
&\quad + (1-\alpha)\mathbb{E}(P(i+1, S_{i+1}, A_{i+1}^{v,i,x}, B_{i+1}^{w,i,x}, B_{i+1}^{r,i,x}, r_{i+1}^{i,x}, D_{i+1}^{i,x})).
\end{aligned}$$

Ce qui prouve la continuité en α , par conséquent le maximum en α est atteint. Comme $J_i(\cdot, S, r, D, a, b, c)$ est affine, le rachat optimal α_i^* en T_i est atteint au bord de l'intervalle $[0, 1]$, donc de type bang-bang :

$$\alpha_i^*(S, a, b, c, r, D) = \mathbf{1}_{\{\mathbb{E}(P(i+1, S_{i+1}, A_{i+1}^{v,i,x}, B_{i+1}^{w,i,x}, B_{i+1}^{r,i,x}, r_{i+1}^{i,x}, D_{i+1}^{i,x})) \leq (a)_+ D\}}. \quad \square$$

2.3.4 Réduction du problème à un problème d'arrêt optimal

Ayant démontré que la stratégie optimale est bang-bang, la dynamique du problème peut être mise sous la forme suivante :

$$\begin{aligned}
A_0^v &= P_0, \\
dA_t^v &= (A_{t-}^v)_+ \frac{dS_t}{S_t} - \beta(A_{t-}^v)_+ dZ_t^f - \gamma B_{t-}^w dZ_t^w \quad \text{avant rachat,} \\
A_t^v &= 0 \quad \text{après rachat.}
\end{aligned} \tag{2.3.5}$$

B^w devient comme suit :

$$\begin{aligned}
B_t^w &= A_0^v \quad \text{pour } T_0^r \leq t < T_1^r, \\
B_t^w &= \max(B_{T_i^r-}^w, A_{T_i^r-}^v) \quad \text{si } T_i^r \leq t < T_{i+1}^r \quad \text{avec } 1 \leq i \leq p-1, \\
B_t^w &= \max(A_0^v e^{r \cdot \text{rol} T_p^r}, \max(B_{T_p^r-}^w, A_{T_p^r-}^v)) \quad \text{pour } T_p^r \leq t < T_{p+1}^r, \\
B_t^w &= \max(B_{T_i^r-}^w, A_{T_i^r-}^v) \quad \text{si } T_i^r \leq t < T_{i+1}^r \quad \text{avec } i \geq p+1.
\end{aligned}$$

Nous constatons que la base Roll-up a été éliminée et donc le prix optimal actualisé ne dépend que de $(S_i, A_i^v, B_i^w, r_i, D_i)$, $0 \leq i \leq n$. Par souci de simplification, pour $0 \leq i \leq n$ introduisons les notations suivantes :

$$\begin{aligned}
X_i &= (S_i, A_i^v, B_i^w, r_i, D_i), \\
P_i(X_i) &= P(i, X_i), \\
Q_i(X_i) &= \mathbb{E}[P_{i+1}(X_{i+1}) | X_i], \quad \text{si } i < n, \\
f_i(X_i) &= D_i(A_i^v)_+, \\
\Psi_i(X_i) &= D_i[\gamma_i B_i^w - \beta_i(A_i^v)_+].
\end{aligned}$$

L'algorithme de programmation dynamique devient :

$$P_n(X_n) = D_n(A_n^v)_+,$$

$$P_i(X_i) = \begin{cases} \max(f_i(X_i), E[P_{i+1}(X_{i+1})|X_i]) & \text{si } T_i \text{ est une date de rachat } (i < n) \quad , \\ \Psi_i(X_i) + E[P_{i+1}(X_{i+1})|X_i] & \text{sinon} \quad . \end{cases}$$

C'est une équation sur les prix, on peut aussi écrire une équation sur les valeurs de continuation :

$$Q_i(X_i) = \begin{cases} E[\max(f_{i+1}(X_{i+1}), Q_{i+1}(X_{i+1}))|X_i] & \text{si } T_{i+1} \text{ est une date de rachat} \quad , \\ E[\Psi_{i+1}(X_{i+1}) + Q_{i+1}(X_{i+1})|X_i] & \text{sinon.} \end{cases}$$

Chapitre 3

Influence de la mortalité

A la différence des contrats purement financiers, les produits d'assurance-vie intègrent en sus du risque de marché habituel le risque lié à la mortalité des souscripteurs. Ne disposant pas d'instruments pour couvrir ce dernier, l'assureur espère que ce risque soit mutualisé dans son portefeuille grâce au grand nombre d'assurés dont il dispose. Le but de ce chapitre est de mettre en exergue d'abord cet effet de mutualisation dans un cas de produit de type européen avant de donner une heuristique concernant le cas américain.

3.1 Tour de la littérature dans le cas européen

3.1.1 Exemple de produit financier avec risque de mortalité

Nous suivons la présentation de [Mol98]. Soit $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t \geq 0})$ un espace de probabilité filtré, satisfaisant aux conditions habituelles. Considérons un actif sans risque rémunéré au taux déterministe $(r_t)_{t \geq 0}$ et un actif risqué S dont la dynamique sous \mathbb{P} (supposée risque neutre) est :

$$dS_t = r_t S_t dt + \sigma_t S_t dW_t$$

avec une volatilité $(\sigma_t)_{t \geq 0}$ déterministe, $S_0 > 0$ et $W = (W_t)_{0 \leq t \leq T}$ est un mouvement brownien standard. Posons $\mathbb{G} = (\mathcal{G}_t)_t$ où $\mathcal{G}_t = \sigma\{S_u, u \leq t\} \subset \mathcal{F}_t$. Considérons l_x assurés de même âge x dont les durées de vie sont des variables aléatoires indépendantes et de même loi T_1, T_2, \dots, T_{l_x} tels que :

$$P(T_i > t) = e^{-\int_0^t \mu_{x+s} ds}$$

où μ est le taux de hasard (fonction déterministe du temps) associé à la fonction de survie.

Et soit $(N_t)_{0 \leq t \leq T}$ le processus de comptage associé aux décès, i.e, $N_t = \sum_{i=1}^{l_x} \mathbf{1}_{\{T_i \leq t\}}$.

Notons $\mathcal{H}_t = \sigma\{N_u, u \leq t\}$, $\mathbb{H} = (\mathcal{H}_t)_{0 \leq t \leq T}$ et $\mathcal{F}_t = \mathcal{G}_t \vee \mathcal{H}_t$.

On sait que

$$M_t = N_t - \int_0^t \lambda_u du$$

est \mathbb{H} -martingale avec $\lambda_t = (l_x - N_{t-})\mu_{x+t}$ et $\tilde{S}_t = \exp(-\int_0^t r_s ds)S_t$ est \mathbb{G} -martingale.

On supposera que \mathcal{G}_T et \mathcal{H}_T sont indépendants.

On considère un contrat qui verse à chaque assuré au temps T un flux de type $g(S_T)$ uniquement s'il est encore en vie. Ainsi pour chaque souscripteur, l'assureur verse le flux (actualisé) suivant :

$$H_i = \mathbf{1}_{\{T_i > T\}} g(S_T) e^{-\int_0^T r_s ds}.$$

En conséquence, le portefeuille globale détenu par l'assureur verse le flux :

$$H = g(S_T) e^{-\int_0^T r_s ds} \sum_{i=1}^{l_x} \mathbf{1}_{\{T_i > T\}} = g(S_T) e^{-\int_0^T r_s ds} (l_x - N_T).$$

Pour pouvoir valoriser ce portefeuille et le couvrir, on introduit d'abord la théorie de minimisation du risque dans la section qui suit.

3.1.2 Quelques résultats sur les stratégies de minimisation du risque globale

Dans la section précédente, on a présenté un modèle combinant le risque du marché financier via le sous-jacent S et le risque actuariel lié à la durée de vie des assurés. Cependant, dans ce cadre le marché est incomplet et une nouvelle notion de couverture s'impose. C'est pourquoi, on résume ici quelques résultats importants de la théorie de minimisation du risque globale utiles pour donner un sens à la couverture dans le cadre de marchés incomplets. Dans ce cadre, Föllmer et Sondermann (voir [FS86]) ont déterminé des stratégies optimales dans le sens de la minimisation de l'erreur quadratique d'un certain processus du risque. Par contre, ils se sont restreints à une mesure de probabilité de type martingale. Schweizer (voir [Sch91]) a étendu ce travail dans le cadre de prix qui sont semimartingales en introduisant la notion de minimisation locale du risque.

On se replace dans le cadre probabiliste du paragraphe 4.2.3. Une stratégie va être de la forme $\phi = (\xi, \eta)$ où $\xi = (\xi_t)_{0 \leq t \leq T}$ et $\eta = (\eta_t)_{0 \leq t \leq T}$ sont respectivement les quantités investies dans le sous-jacent et l'actif sans risque (cash). Ainsi,

$$V_t^\phi = \xi_t \tilde{S}_t + \eta_t$$

est la valeur actualisée du portefeuille à l'instant t , avec $S_t^* = e^{-\int_0^t r_s ds} S_t$. On introduit maintenant la définition rigoureuse d'une stratégie ([FS86]).

Définition 3.1.1 $\phi = (\xi, \eta)$ est appelée une stratégie si

1. ξ est un processus prévisible tel que $\mathbb{E} \left[\int_0^T \xi_t^2 d\langle \tilde{S} \rangle_t \right] < \infty$.
2. η est adapté.
3. V^ϕ est continu à droite avec $V_t^\phi \in L^2(\mathbb{P}), 0 \leq t \leq T$.

Le processus coût C^ϕ associé à la stratégie ϕ est défini par

$$C_t^\phi = V_t^\phi - \int_0^t \xi_u d\tilde{S}_u,$$

et le processus risque résiduel R^ϕ de ϕ s'écrit

$$R_t^\phi = \mathbb{E} \left[(C_T^\phi - C_t^\phi)^2 | \mathcal{F}_t \right].$$

Une stratégie est dite *autofinancante en moyenne* si le processus coût C^ϕ est \mathcal{F} -martingale. Enfin, une stratégie est *autofinancante* si son coût est constant, i.e.

$$V_t^\phi = V_0^\phi + \int_0^t \xi_u d\tilde{S}_u.$$

Maintenant on s'intéresse à caractériser les stratégies optimales, associées à un produit. On considère un produit versant un flux $H \in L^2(\mathbb{P})$ \mathcal{F}_T -adapté. On se restreint dans la suite aux stratégies admissibles vérifiant :

$$V_T^\phi = H \quad p.s.$$

où H est de carré intégrable.

Ainsi, par une stratégie admissible, on peut générer le flux H mais avec un certain coût C_T^ϕ .

D'abord, on s'intéresse aux stratégies admissibles ϕ minimisant R_0^ϕ . On a

$$R_0^\phi = \mathbb{E} \left[(C_T^\phi - C_0^\phi)^2 \right].$$

Puisque

$$C_T^\phi = V_T^\phi - \int_0^T \xi_u d\tilde{S}_u = H - \int_0^T \xi_u d\tilde{S}_u$$

alors

$$R_0^\phi = \mathbb{E} \left[\left(H - \int_0^T \xi_u d\tilde{S}_u - C_0^\phi \right)^2 \right].$$

Pour minimiser R_0^ϕ , il faut choisir

$$C_0^\phi = \mathbb{E} \left[H - \int_0^T \xi_u d\tilde{S}_u \right] = \mathbb{E}[H] = \mathbb{E}[C_T^\phi].$$

Ainsi, le coût initial est déterminé. Enfin, il reste à trouver ξ minimisant $\mathbb{E} \left[(C_T^\phi - \mathbb{E}[H])^2 \right]$. Vu qu'on minimise la variance de C_T^ϕ , ce type de stratégie est appelé *stratégie de variance minimale*.

Notons V^* le processus défini par

$$V_t^* = \mathbb{E}[H | \mathcal{F}_t].$$

D'après la décomposition de Galtchouk-Kunita-Watanabe, V_t^* s'écrit de façon unique sous la forme

$$V_t^* = \mathbb{E}[H] + \int_0^t \xi_u^H d\tilde{S}_u + L_t^H,$$

avec $L^H = (L_t^H)_{0 \leq t \leq T}$ est une \mathbb{F} -martingale orthogonale à \tilde{S} de moyenne nulle et ξ^H est un processus prévisible vérifiant $\mathbb{E} \left[\int_0^T (\xi_t^H)^2 d\langle \tilde{S} \rangle_t \right] < \infty$. En se basant sur cette décomposition, Follmer et Sondermann (voir [FS86]) obtiennent le résultat suivant :

Théorème 3.1.1 *Une stratégie admissible $\phi = (\xi, \eta)$ est de variance minimale*

$$\mathbb{E} \left[(C_T^\phi - \mathbb{E}[C_T^\phi])^2 \right] = \mathbb{E}[(L_T^H)^2]$$

si et seulement si $\xi = \xi^H$.

Le théorème ci-dessus n'assure pas l'unicité de la stratégie vu que η_t est arbitraire pour $t < T$ (η_T est unique grâce à la condition d'admissibilité $V_T = H$). Pour déterminer de façon unique la stratégie optimale, on introduit la notion de stratégie risque-minimisante. Dans la suite, on appellera une prolongation admissible d'une stratégie ϕ au temps t toute stratégie $\tilde{\phi}$ qui coïncide avec ϕ aux temps $u < t$ et ayant la même valeur terminale V_T que ϕ .

Définition 3.1.2 *ϕ est appelée une stratégie risque-minimisante si elle minimise à tout instant $0 \leq t \leq T$ le risque résiduel, i.e.*

$$R_t^\phi \leq R_t^{\tilde{\phi}}$$

pour toute prolongation admissible $\tilde{\phi}$ de ϕ au temps t .

Remarques :

1. Il est facile de voir que toute stratégie autofinancante est risque-minimisante car $R_t^\phi = 0$.
2. Une stratégie risque-minimisante est une stratégie de variance minimale (voir [FS86]).

On a le théorème d'unicité suivant (voir [FS86]) :

Théorème 3.1.2 *Il existe une unique stratégie admissible risque-minimisante $\phi = (\xi, \eta)$ donnée par*

$$(\xi_t, \eta_t) = (\xi_t^H, V_t^* - \xi_t^H \tilde{S}_t), 0 \leq t \leq T.$$

Le processus de risque associé est donné par $R_t^\phi = \mathbb{E} [(L_T^H - L_t^H)^2 | \mathcal{F}_t]$ et est appelé aussi le processus du risque intrinsèque.

3.1.3 Application à l'exemple étudié : mutualisation du risque

Moller (voir [Mol98]) applique les résultats rappelés dans la section précédente pour déterminer des stratégies minimisantes de risque ainsi que leurs risques intrinsèques associés pour différents types de contrats d'assurance en unité de compte. Il étend aussi l'approche de minimisation de risque pour des contrats avec un flux plus général que celui décrit dans la section précédente (voir [Mol01b]). Dans cette section, on énoncera (sans preuve) les principaux résultats dans [Mol98] liés à notre portefeuille initial de flux H afin de les exploiter dans la partie suivante pour visualiser numériquement la mutualisation du risque pour un grand nombre d'assurés.

Dans notre cas, on a

$$H = g(S_T) e^{-\int_0^T r_s ds} (l_x - N_T),$$

et

$$V_t^* = \mathbb{E}[g(S_T) e^{-\int_0^T r_s ds} (l_x - N_T) | \mathcal{F}_t].$$

Notons $F^g(t, S_t) = \mathbb{E} \left[e^{-\int_t^T r_s ds} g(S_T) | \mathcal{G}_t \right]$ et par F_s^g la dérivée de F^g par rapport au sous-jacent.

On a les résultats suivants (voir [Mol98]) :

Théorème 3.1.3 *Le processus V^* admet la décomposition suivante :*

$$V_t^* = V_0^* + \int_0^t \xi_u^H d\tilde{S}_u + \int_0^t \nu_u^H dM_u,$$

avec (ξ^H, ν^H) sont donnés par

$$\begin{aligned} \xi_t^H &= (l_x - N_{t-}) \mathbb{P}(T_1 > T | T_1 > t) F_s^g(t, S_t), \\ \nu_t^H &= -e^{-\int_0^t r_s ds} F^g(t, S_t) \mathbb{P}(T_1 > T | T_1 > t), 0 \leq t \leq T. \end{aligned}$$

Théorème 3.1.4 *Considérons le portefeuille de flux H . Les stratégies admissibles de variance minimale ϕ sont déterminées par*

$$\begin{aligned} \xi_t &= (l_x - N_{t-}) \mathbb{P}(T_1 > T | T_1 > t) F_s^g(t, S_t), 0 \leq t \leq T, \\ \eta_T &= H - \xi_T^* \tilde{S}_T. \end{aligned}$$

La variance minimale est :

$$R_0^\phi = \mathbb{E} \left[(C_T^\phi - \mathbb{E} [C_T^\phi])^2 \right] = l_x \mathbb{P}(T_1 > T) \int_0^T \mathbb{E} \left[\left(e^{-\int_0^u r_s ds} F^g(u, S_u) \right)^2 \right] \mathbb{P}(T_1 > T | T_1 > u) \mu_{x+u} du.$$

On remarque alors que $\frac{\sqrt{R_0^\phi}}{l_x}$ tend vers 0 comme $\frac{1}{\sqrt{l_x}}$ quand l_x tend vers ∞ . Ainsi, il y a bien une mutualisation du risque de mortalité pour un grand nombre d'assurés.

Théorème 3.1.5 *Considérons le portefeuille de flux H . L'unique stratégie risque-minimisante est donnée par*

$$\begin{aligned}\xi_t^* &= (l_x - N_{t-})\mathbb{P}(T_1 > T | T_1 > t)F_s^g(t, S_t), \\ \eta_t^* &= (l_x - N_t)\mathbb{P}(T_1 > T | T_1 > t)e^{-\int_0^t r_s ds}F^g(t, S_t) - \xi_t^* \tilde{S}_t, 0 \leq t \leq T.\end{aligned}$$

La variance minimale est :

$$\begin{aligned}R_0^\phi &= \mathbb{E} \left[(C_T^\phi - \mathbb{E} [C_T^\phi])^2 \right] \\ &= l_x \mathbb{P}(T_1 > T) \int_0^T \mathbb{E} \left[\left(e^{-\int_0^u r_s ds} F^g(u, S_u) \right)^2 \right] \mathbb{P}(T_1 > T | T_1 > u) \mu_{x+u} du.\end{aligned}$$

Ainsi la valeur en t du portefeuille risque-minimisant, calculée par assuré restant en vie, est

$$\frac{V_t^*}{l_x - N_{t-}} = \mathbb{E}(e^{-\int_t^T (r_s + \mu_{x+s}) ds} g(S_T) | \mathcal{G}_t). \quad (3.1.1)$$

C'est la règle de valorisation risque-neutre habituelle en modifiant le facteur d'actualisation avec le taux de hasard.

3.1.4 Résultats numériques

Pour les tests numériques, on se place dans le cadre Black-Scholes suivant :

$$S_0 = 1, \sigma_t = \sigma = 25\%, T = 15, r_t = r = 6\%, g(s) = \max(s, K), K = 2.45.$$

Concernant la mortalité on opte pour un taux constant :

$$\mu_{x+s} = \mu = 0.00439.$$

Cela entraîne que le premier décès parmi l_x assurés a lieu en moyenne au bout de $\frac{1}{l_x \mu}$ années (car de loi exponentielle de paramètre $l_x \mu$). On a

$$\begin{aligned}F^g(t, S_t) &= \mathbb{E}[e^{-r(T-t)}(K + (S_T - K)_+) | \mathcal{F}_t] \\ &= Ke^{-r(T-t)} + (S_t \mathcal{N}(z_t) - Ke^{-r(T-t)} \mathcal{N}(z_t - \sigma\sqrt{T-t})) \\ &= Ke^{-r(T-t)} \mathcal{N}(-z_t + \sigma\sqrt{T-t}) + S_t \mathcal{N}(z_t),\end{aligned}$$

où \mathcal{N} est la fonction de répartition de la loi normale centrée réduite et

$$z_t = \frac{\log(S_t/K) + (r + \frac{\sigma^2}{2})(T-t)}{\sigma\sqrt{T-t}}.$$

Ainsi, la stratégie risque-minimisante est donnée par

$$\begin{aligned}\xi_t &= (l_x - N_{t-})\mathbb{P}(T_1 > T|T_1 > t)\mathcal{N}(z_t), \\ \eta_t &= (l_x - N_t)\mathbb{P}(T_1 > T|T_1 > t)Ke^{-rT}\mathcal{N}(-z_t + \sigma\sqrt{T-t}) \\ &\quad - \Delta N_t\mathbb{P}(T_1 > T|T_1 > t)\mathcal{N}(z_t)\tilde{S}_t.\end{aligned}$$

Le coût de couverture lié au risque de mortalité est :

$$C_T^\phi - C_0^\phi = \int_0^T \nu_u^H dM_u,$$

avec

$$\nu_t^H = -e^{-rt}F^g(t, S_t)\mathbb{P}(T_1 > T|T_1 > t).$$

Le théorème 3.1.4 assure que $\sqrt{\mathbb{E}[(C_T^\phi - C_0^\phi)^2]}/l_x$ tend vers 0 pour un l_x assez grand. Pour voir plus précisément la mutualisation du risque de mortalité, on se propose de visualiser dans cette section le comportement de la quantité aléatoire suivante

$$\frac{C_T^\phi - C_0^\phi}{l_x},$$

qui s'interprète comme un risque par assuré. Pour cela, on génère 20000 scénarii indépendants pour (N, S) avec une discrétisation de pas $\Delta t = 0.01$ pour S . On considère les cas $l_x = 1, 10, 50$ et 100 assurés, induisant un premier décès en moyenne au bout de respectivement 228 ans, 22.8 ans, 4.55 ans et 2.28 ans environ.

On trace l'histogramme (couleur bleue) de

$$\frac{C_T^\phi - C_0^\phi}{l_x}$$

ainsi que celui de la loi normale centrée réduite (couleur violette) de paramètres égaux à la moyenne et l'écart type estimés de l'échantillon considéré. On obtient les graphes suivants pour les différents cas considérés.

Pour un seul assuré (Figure 3.2), les valeurs positives correspondent à des pertes dues au fait que l'assuré en question ne décède pas. Elles ne sont pas très grandes mais leur probabilité est très grande ce qui colle avec la loi de décès supposée. Les gains sont assez grands mais de faible probabilité. Quand il s'agit de 10 assurés (Figure 3.2), on constate deux pics : celui de droite correspond a priori au cas où tous les assurés sont en vie. Quant à celui de gauche, il représente le cas où un seul assuré décède parmi les 10. Quand le nombre d'assurés est grand (Figure 3.3), on constate la mutualisation du risque de mortalité qui se traduit par la concentration des histogrammes comparés à ceux des graphes précédents et par leurs rapprochements de ceux de la loi normale. De surcroît, l'histogramme devient nettement

symétrique, et laisse deviner la convergence en loi vers une loi gaussienne centrée (ce qui n'est pas difficile à montrer).

Ainsi, on peut conclure que la mutualisation des risques a lieu à partir de 100 assurés et que la valorisation du produit se fait en modifiant le facteur d'actualisation $e^{-\int_0^T r_s ds}$ par l'ajout au sein de l'intégrale du taux de mortalité μ_{x+s} (voir 4.4.2).

3.2 Heuristique pour le cas à exercice américain

On considère maintenant un ensemble de l_x contrats qui offre à chacun des assurés la possibilité d'exercer une seule fois à n'importe quel instant t entre 0 et T et de toucher une flux $g(S_t) \geq 0$ s'il est en vie, et rien sinon. Notons $\tilde{g}_t = e^{-\int_0^t r_s ds} g(S_t)$, supposons que $\mathbb{E}(\sup_{0 \leq t \leq T} \tilde{g}_t^2) < +\infty$ et que \tilde{g} est continu.

3.2.1 Résultat principal

Un candidat pour le prix actualisé de ce contrat en t est :

$$\tilde{V}_t^* = \sum_{i=1}^{l_x} \text{esssup}_{\tau \in \tau_{t,T}^i} \mathbb{E}[\mathbf{1}_{\{T_i > \tau\}} \tilde{g}_\tau | \mathcal{F}_t]$$

avec $\tau_{t,T}^i$ est l'ensemble des $\mathbb{G} \vee \sigma\{\mathbf{1}_{\{T_i \leq u\}}, u \leq T\}$ -temps d'arrêts compris entre t et T . On prend \tilde{V}^* de cette forme car

- En raisonnant par assuré, on s'inspire du cas américain habituel : ce prix domine bien les flux à toutes les dates.
- pour chaque assuré, la stratégie d'arrêt est prise dans l'ensemble $\tau_{t,T}^i$ car il est plus naturel de considérer que le souscripteur exerce son droit en tenant compte de l'état du marché et de sa mortalité mais indépendamment des durées de vie des autres assurés auxquelles il n'a même pas accès a priori.

Dans la suite, notre but est de montrer qu'on se ramène au cas américain habituel (sans mortalité) (voir [KS98] pour un traitement détaillé sur les options américaines) en actualisant avec un taux supplémentaire et d'exhiber une stratégie de couverture qui mutualise le risque de mortalité pour un grand nombre d'individus. Notons $\tau_{t,T}^{\mathbb{G}}$ l'ensemble des \mathbb{G} -temps d'arrêts compris entre t et T .

Théorème 3.2.1 *On a*

$$\tilde{V}_t^* = (l_x - N_t)Y_t,$$

où

$$Y_t := \text{esssup}_{\tau \in \tau_{t,T}^{\mathbb{G}}} \mathbb{E}[e^{-\int_t^\tau \mu_{x+s} ds} \tilde{g}_\tau | \mathcal{G}_t] = e^{\int_0^t \mu_{x+s} ds} \text{esssup}_{\tau \in \tau_{t,T}^{\mathbb{G}}} \mathbb{E}[e^{-\int_0^\tau \mu_{x+s} ds} \tilde{g}_\tau | \mathcal{G}_t]$$

peut se décomposer comme :

$$Y_t = e^{\int_0^t \mu_{x+s} ds} \left(Y_0 + \int_0^t \delta_u d\tilde{S}_u - \Lambda_t \right),$$

avec $(\Lambda_t)_t$ est un processus croissant adapté continu de valeur nulle pour $t = 0$.

Corollaire 3.2.2 La stratégie de couverture avec $(l_x - N_{t-})e^{\int_0^t \mu_{x+s} ds} \delta_t$ parts investis dans S mutualise bien les risques de mortalité :

$$\min \left(0, \frac{\tilde{V}_0^* + \int_0^t e^{\int_0^u \mu_{x+s} ds} \delta_u d\tilde{S}_u - (l_x - N_t) \tilde{g}_t}{l_x} \right) \rightarrow 0$$

en probabilité quand $l_x \rightarrow +\infty$.

Le corollaire nous assure qu'en suivant une stratégie auto-finançante consistant à investir $(l_x - N_{t-})e^{\int_0^t \mu_{x+s} ds} \delta_t$ parts dans l'actif risqué, la perte éventuelle par assuré, liée à l'exercice en t est négligeable lorsque le nombre de souscripteurs l_x est suffisamment grand.

Preuve du théorème 3.2.1

On a :

$$\begin{aligned} \tilde{V}_t^* &= \sum_{i=1}^{l_x} \text{esssup}_{\tau \in \tau_{t,T}^i} \mathbb{E}[\mathbf{1}_{\{T_i > \tau\}} \tilde{g}_\tau | \mathcal{F}_t] \\ &= \sum_{i=1}^{l_x} \text{esssup}_{\tau \in \tau_{t,T}^i} \mathbf{1}_{\{T_i > t\}} \mathbb{E}[\mathbf{1}_{\{T_i > \tau\}} \tilde{g}_\tau | \mathcal{F}_t] \\ &= \sum_{i=1}^{l_x} \text{esssup}_{\tau \in \tau_{t,T}^i} \mathbf{1}_{\{T_i > t\}} \mathbb{E}[\mathbf{1}_{\{T_i > \tau\}} \tilde{g}_\tau | \mathcal{G}_t \vee \sigma\{\mathbf{1}_{\{T_i \leq u\}}, u \leq t\}]. \end{aligned}$$

On sait que pour tout $\tau \in \tau_{t,T}^i$, il existe $\tau' \in \tau_{t,T}^{\mathbb{G}}$ tel que :

$$\tau \wedge T_i = \tau' \wedge T_i$$

où $\tau_{t,T}^{\mathbb{G}}$ désigne l'ensemble des \mathbb{G} -temps d'arrêts compris entre t et T (voir [Pro04], page 370).

Notons $\mathcal{F}_t^i = \mathcal{G}_t \vee \sigma\{\mathbf{1}_{\{T_i \leq u\}}, u \leq t\}$. Donc

$$\begin{aligned} \text{esssup}_{\tau \in \tau_{t,T}^i} \mathbb{E}[\mathbf{1}_{\{T_i > \tau\}} \tilde{g}_\tau | \mathcal{F}_t^i] &= \text{esssup}_{\tau \in \tau_{t,T}^i} \mathbb{E}[\mathbf{1}_{\{T_i > T_i \wedge \tau\}} \tilde{g}_{T_i \wedge \tau} | \mathcal{F}_t^i] \\ &= \text{esssup}_{\tau \in \tau_{t,T}^{\mathbb{G}}} \mathbb{E}[\mathbf{1}_{\{T_i > T_i \wedge \tau\}} \tilde{g}_{T_i \wedge \tau} | \mathcal{F}_t^i] \\ &= \text{esssup}_{\tau \in \tau_{t,T}^{\mathbb{G}}} \mathbb{E}[\mathbf{1}_{\{T_i > \tau\}} \tilde{g}_\tau | \mathcal{F}_t^i]. \end{aligned}$$

La dernière expression se développe comme suit

$$\begin{aligned} \text{esssup}_{\tau \in \tau_{t,T}^{\mathbb{G}}} \mathbb{E}[\mathbf{1}_{\{T_i > \tau\}} \tilde{g}_\tau | \mathcal{F}_t^i] &= \frac{\mathbb{P}(T_i > t | \mathcal{F}_t^i)}{\mathbb{P}(T_i > t | \mathcal{G}_t)} \text{esssup}_{\tau \in \tau_{t,T}^{\mathbb{G}}} \mathbb{E}[\mathbf{1}_{\{T_i > \tau\}} \tilde{g}_\tau | \mathcal{G}_t] \\ &= \mathbf{1}_{\{T_i > t\}} e^{\int_0^t \mu_{x+s} ds} \text{esssup}_{\tau \in \tau_{t,T}^{\mathbb{G}}} \mathbb{E}[\mathbf{1}_{\{T_i > \tau\}} \tilde{g}_\tau | \mathcal{G}_t] \\ &\quad (\text{voir lemme 5.1.2 p.143 dans [BR02]}). \end{aligned}$$

Ainsi :

$$\tilde{V}_t^* = (l_x - N_t) Y_t$$

avec

$$\begin{aligned} Y_t &= e^{\int_0^t \mu_{x+s} ds} \text{esssup}_{\tau \in \tau_{t,T}^{\mathbb{G}}} \mathbb{E}[\mathbf{1}_{\{T_1 > \tau\}} \tilde{g}_\tau | \mathcal{G}_t] \\ &= e^{\int_0^t \mu_{x+s} ds} \text{esssup}_{\tau \in \tau_{t,T}^{\mathbb{G}}} \mathbb{E}[\tilde{g}_\tau \mathbb{E}[\mathbf{1}_{\{T_1 > \tau\}} | \mathcal{G}_\tau] | \mathcal{G}_t] \\ &= \text{esssup}_{\tau \in \tau_{t,T}^{\mathbb{G}}} \mathbb{E}[e^{-\int_t^\tau \mu_{x+s} ds} \tilde{g}_\tau | \mathcal{G}_t]. \end{aligned}$$

On sait que $(e^{-\int_0^t \mu_{x+s} ds} Y_t)_t$ est l'enveloppe de Snell de $(e^{-\int_0^t \mu_{x+s} ds} \tilde{g}_t)_{t \geq 0}$: dans notre cadre de marché brownien, Y est une \mathbb{G} -supermartingale continue qui admet une décomposition de Doob-Meyer :

$$e^{-\int_0^t \mu_{x+s} ds} Y_t = R_t - \Lambda_t$$

où R est une \mathbb{G} -martingale continue et Λ est un processus \mathbb{G} -adapté croissant continu avec $\Lambda_0 = 0$ (voir [Kar88]). R étant une martingale brownienne de carré intégrable, $R_t = R_0 + \int_0^t \delta_u d\tilde{S}_u$ avec $(\delta_t)_{0 \leq t \leq T}$ est un processus prévisible tel que $\mathbb{E} \left[\int_0^T \delta_u^2 d\langle \tilde{S} \rangle_u \right] < \infty$. Par suite

$$Y_t = e^{\int_0^t \mu_{x+s} ds} \left(Y_0 + \int_0^t \delta_u d\tilde{S}_u - \Lambda_t \right). \quad \square$$

Preuve du corollaire 3.2.2

Une application de la formule d'Itô à l'expression $\tilde{V}_t^* = (l_x - N_t) Y_t$ (théorème 3.2.1) nous permet d'écrire :

$$\begin{aligned} \tilde{V}_t^* &= \tilde{V}_0^* + \int_0^t (l_x - N_{u-}) dY_u - \int_0^t Y_{u-} dN_u \\ &= \tilde{V}_0^* + \int_0^t (l_x - N_{u-}) e^{\int_0^u \mu_{x+s} ds} d(e^{-\int_0^u \mu_{x+s} ds} Y_u) \\ &\quad + \int_0^t (l_x - N_{u-}) \mu_{x+u} Y_{u-} du - \int_0^t Y_{u-} dN_u \\ &= \tilde{V}_0^* + \int_0^t (l_x - N_{u-}) e^{\int_0^u \mu_{x+s} ds} dR_u - \int_0^t (l_x - N_{u-}) e^{\int_0^u \mu_{x+s} ds} d\Lambda_u \\ &\quad - \int_0^t Y_{u-} d \left(N_u - \int_0^u (l_x - N_{s-}) \mu_{x+s} ds \right). \end{aligned}$$

Finalement, on obtient

$$\tilde{V}_t^* = \tilde{V}_0^* + \int_0^t (l_x - N_{u-}) e^{\int_0^u \mu_{x+s} ds} \delta_u d\tilde{S}_u - \int_0^t (l_x - N_{u-}) e^{\int_0^u \mu_{x+s} ds} d\Lambda_u - \int_0^t Y_{u-} dM_u.$$

Ainsi, en suivant la stratégie auto-finançante avec $\xi_t = (l_x - N_{t-}) e^{\int_0^t \mu_{x+s} ds} \delta_t$ parts de S en t , si l'exercice se fait en t , la perte actualisée par assuré est :

$$\begin{aligned} C_t &= \frac{\tilde{V}_0^* + \int_0^t e^{\int_0^u \mu_{x+s} ds} \xi_u d\tilde{S}_u - (l_x - N_t) \tilde{g}_t}{l_x} \\ &= \frac{\tilde{V}_t^* - (l_x - N_t) \tilde{g}_t + \int_0^t (l_x - N_{u-}) e^{\int_0^u \mu_{x+s} ds} d\Lambda_u + \int_0^t Y_{u-} dM_u}{l_x} \\ &\geq \frac{\int_0^t Y_{u-} dM_u}{l_x}, \end{aligned}$$

car $\tilde{V}_t^* \geq (l_x - N_t) \tilde{g}_t$ (évident à partir du théorème 3.2.1) et $\int_0^t (l_x - N_{u-}) e^{\int_0^u \mu_{x+s} ds} d\Lambda_u \geq 0$ (Λ est un processus croissant).

Or $\mathbb{E} \left[\left(\frac{\int_0^t Y_{u-} dM_u}{l_x} \right)^2 \right] = \frac{1}{l_x} C \rightarrow 0$ quand $l_x \rightarrow +\infty$ avec C est une constante indépendante de l_x . D'après l'inégalité de Markov, pour tout $\epsilon > 0$

$$\begin{aligned} P(|\min(C_t, 0)| > \epsilon) &\leq \frac{E[\min(C_t, 0)^2]}{\epsilon^2} \leq \frac{E \left[\min\left(\frac{\int_0^t Y_{u-} dM_u}{l_x}, 0\right)^2 \right]}{\epsilon^2} \\ &\leq \frac{\mathbb{E} \left[\left(\frac{\int_0^t Y_{u-} dM_u}{l_x} \right)^2 \right]}{\epsilon^2}. \end{aligned}$$

D'où le résultat du corollaire. \square Pour les tests numériques, on se place dans le même cadre que dans la section précédente avec $S_0 = 1, \sigma_t = \sigma = 25\%, T = 15, r_t = r = 6\%, g(s) = \max(s, K), K = 2.45$. Concernant la mortalité on opte toujours pour un taux constant :

$$\mu_{x+s} = \mu = 0.00439.$$

On a dans cas

$$C_T = \frac{-\int_0^T Y_u \lambda_u du + \int_0^T Y_u dN_u}{l_x}.$$

Le calcul de Y s'apparente à celui d'un call américain avec dividende, donc différent d'un call européen (voir théorème 3.1 dans [Vil99]). On génère $M = 20000$ simulations. Concernant $\int_0^T Y_u \lambda_u du$ qu'on approche par la formule des rectangles, la difficulté de la simulation ici est de devoir calculer Y à chaque date de discrétisation $j\Delta t$ et pour chaque valeur simulée S_j^i (correspondant à la valeur du sous-jacent de

la simulation i à la date $j\Delta t$) alors qu'on ne dispose pas de formule fermée pour Y (le prix américain). Deux choix sont possibles ; Soit d'utiliser pour chaque simulation et à chaque date de discrétisation une méthode de Monte-Carlo pour calculer Y ce qui est, malgré la bonne précision, très coûteux en temps de calcul et n'est pas réaliste en pratique, soit d'avoir recours (ce que nous utilisons dans notre cas) à une méthode d'arbre binomiale construite avec les mêmes dates de discrétisation en temps $j\Delta t$ pour calculer Y . On interpole linéairement entre les deux valeurs des noeuds correspondant au segment contenant la coordonnée $(j\Delta t, S_j^i)$ ou l'on prend la valeur du noeud extrême le plus proche de cette coordonnée si ce segment n'existe pas. Quant à la quantité $\int_0^T Y_u dN_u$, elle est égale à $\sum_{k=1}^{l_x} Y_{T_k^*} \mathbf{1}_{\{T_k^* \leq T\}}$ avec $(T_k^*)_k$ est la permutation en ordre croissant de la suite $(T_k)_k$. On simule ici M réalisations pour $(T_k^*)_k$ qu'on note $(T_k^*)_{i,k}$ et ensuite M réalisations pour $(S_{T_k^*})_k$ notées $(S_k^i)_{i,k}$. Pour calculer Y pour un temps $T_k^* < T$ et pour une valeur de sous-jacent égale à S_k^i , on se sert de l'arbre binomiale construit précédemment en faisant cette fois deux interpolations linéaires. Pour ce faire, on considère dans l'arbre le triangle contenant la coordonnée (T_k^i, S_k^i) et le point P d'intersection de la droite déterminée par le noeud N de coordonnée temps différente parmi les trois formant le triangle et le point (T_k^i, S_k^i) avec la droite contenant les deux noeuds restants. On fait une première interpolation en S entre ces deux derniers noeuds pour calculer le prix au point P . Ensuite on procède à une seconde interpolation en temps entre le noeud N et le point P pour calculer la valeur de Y au point (T_k^i, S_k^i) (voir figure 3.1).

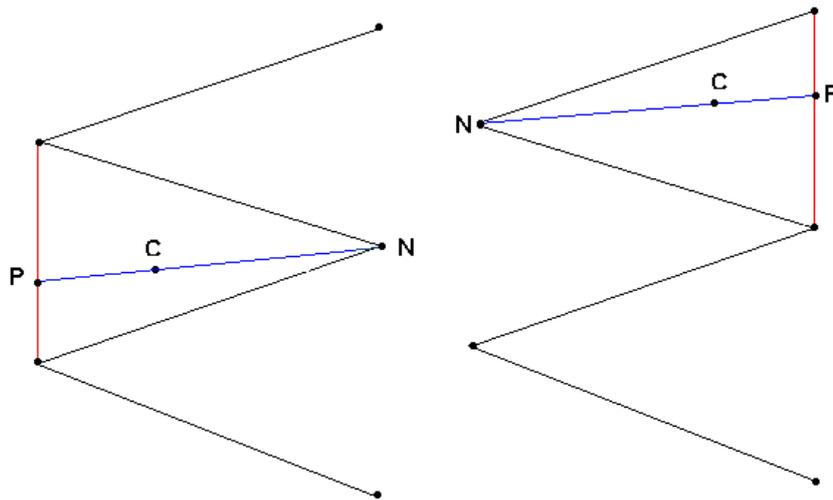


FIGURE 3.1 – Double interpolation du prix selon la position de la coordonnée C .

On trace l'histogramme du coût par assuré $-\frac{C_T}{l_x} = -\frac{\int_0^T Y_u dM_u}{l_x}$ ainsi que celui de la loi normale centrée réduite correspondante aux moyenne et écart-type estimés.

Concernant le cas à 1 et 10 assurés (Figure 4.3), les mêmes remarques que dans le cas européen restent valables à une nuance près due au fait que dans ce cas les valeurs extrêmes de l'histogramme du coût sont plus grandes que dans le cas européen. D'un autre côté, la visualisation de la figure 3.5 exhibe que la mutualisation du risque de mortalité dans le cas américain peut nécessiter plus d'assurés (1000 assurés) que dans la cas Européen (100 assurés).

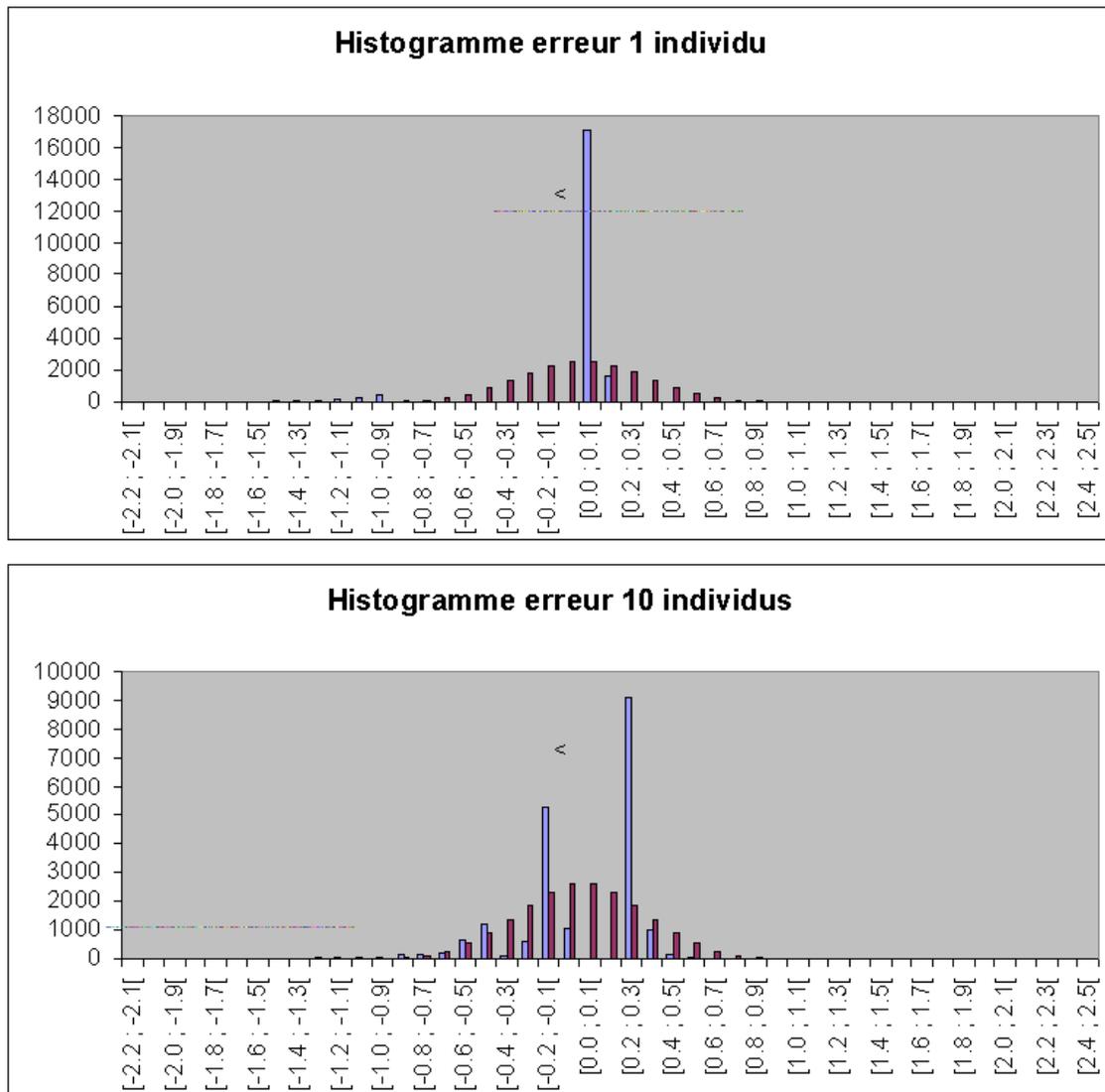


FIGURE 3.2 – Histogramme du coût pour 1 et 10 assurés dans le cas européen.

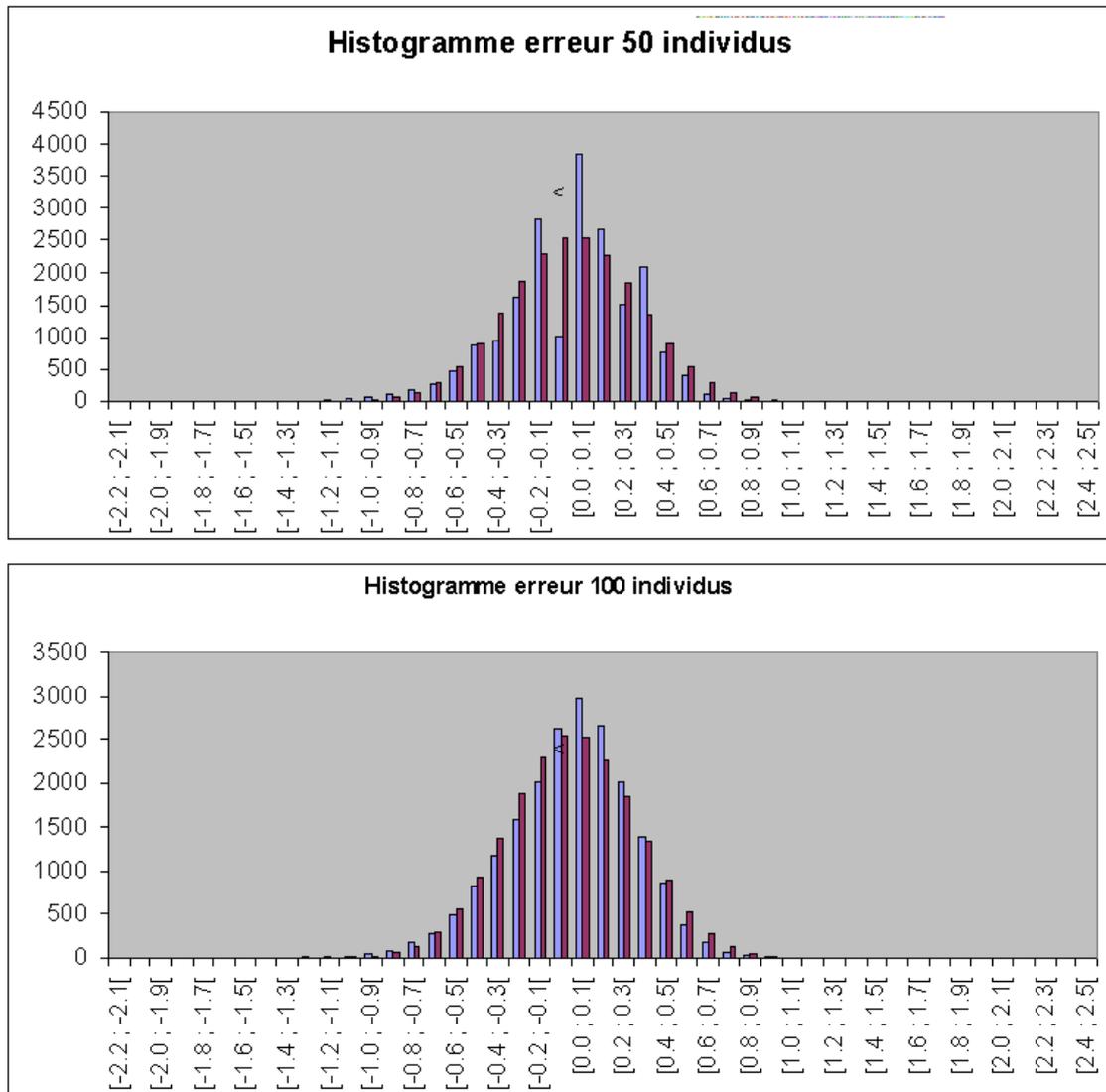


FIGURE 3.3 – Histogramme du coût pour 50 et 100 assurés dans le cas européen.

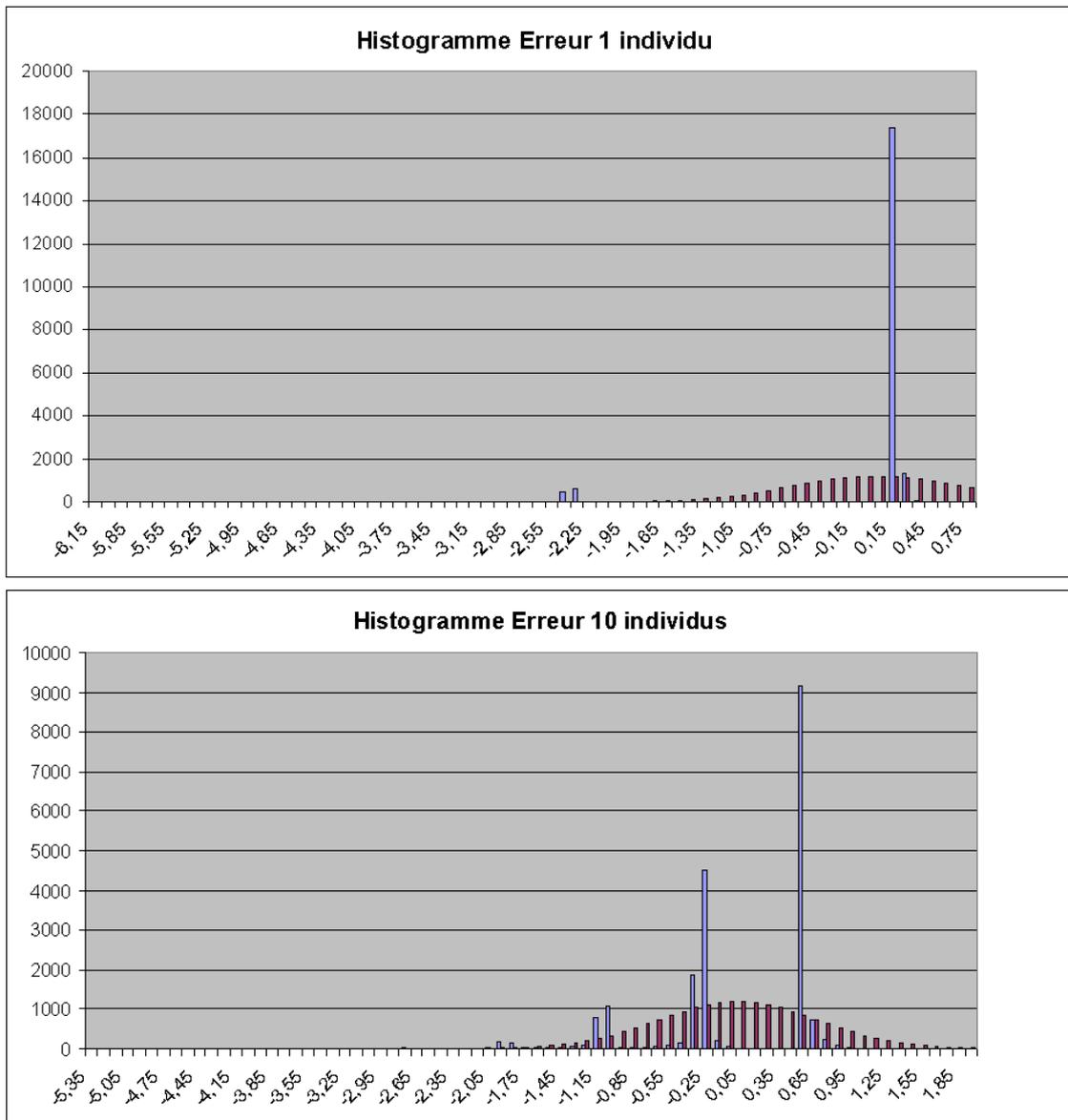


FIGURE 3.4 – Histogramme du coût pour 1 et 10 assurés dans le cas américain.

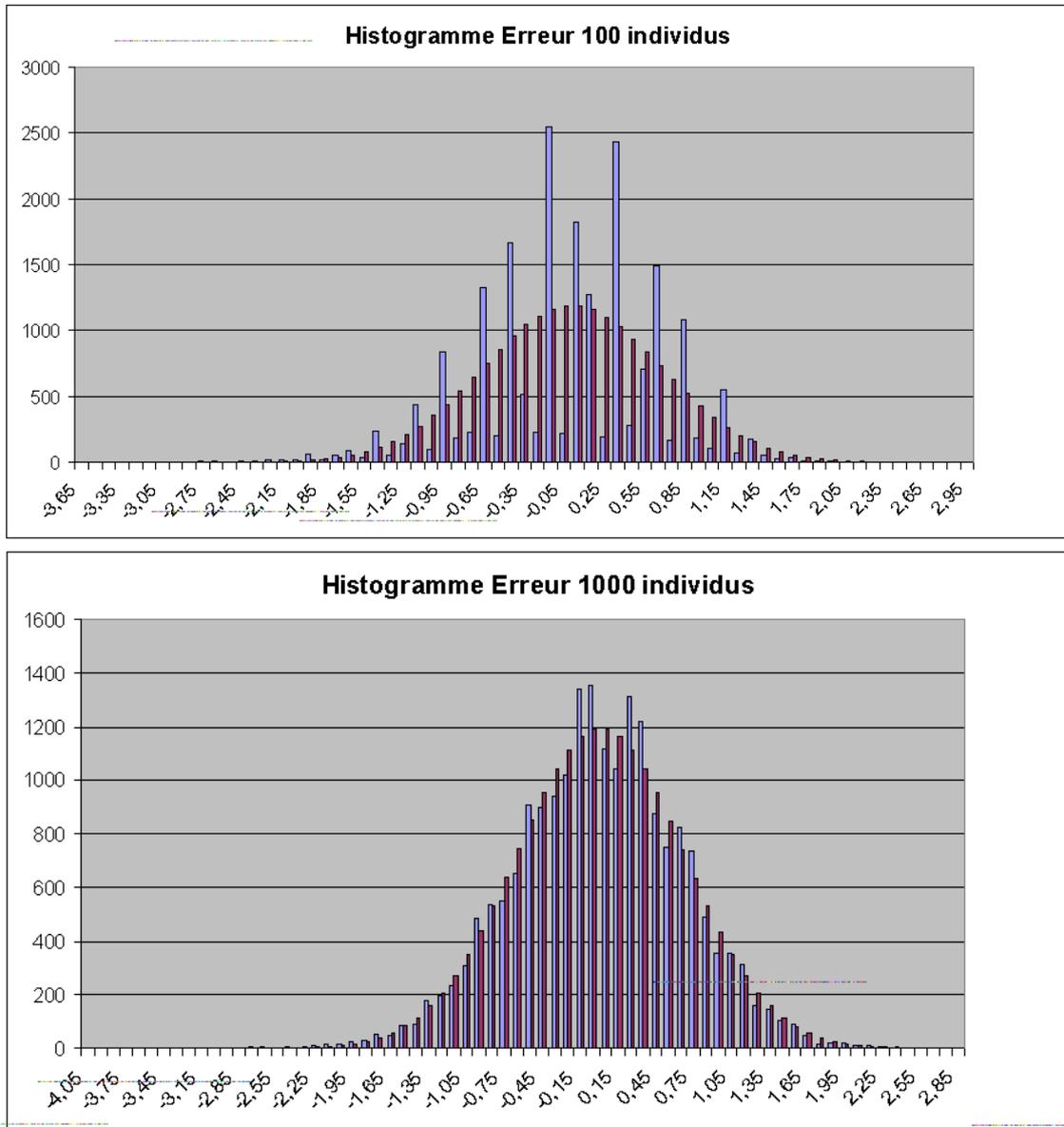


FIGURE 3.5 – Histogramme du coût pour 100 et 1000 assurés dans le cas américain.

Chapitre 4

Régression empirique

En utilisant la régression par moindres carrés pour calculer le prix d'un produit européen, le but dans ce chapitre est de mettre en évidence l'impact du choix de fonctions de base, de leur nombre ainsi que du nombre de simulations, sur la convergence de l'estimateur par moindres carrés.

4.1 Description du cas test : Modèle et Payoff

On considère deux sous-jacents S^1 et S^2 de dynamique :

$$\begin{aligned}\frac{dS_t^1}{S_t^1} &= \sigma^1 dW_t^1, \\ \frac{dS_t^2}{S_t^2} &= \sigma^2 dW_t^2\end{aligned}$$

où W^1, W^2 sont deux mouvements browniens standards indépendants. On néglige le taux d'intérêt.

On considère un produit qui verse à maturité le flux :

$$H = (S_T^2 - (1 - \alpha)S_T^1)_+ - (\alpha + \beta)(S_T^2 - S_T^1)_+ - (S_T^2 - (1 + \beta)S_T^1)_+.$$

On choisit ce type de payoff pour avoir une formule explicite de prix au temps $u < T$ qu'on veut homogène en S^1 et S^2 et différent des options d'échange habituelles où le prix est proche d'une forme quadratique. On prendra pour les tests numériques : $T = 2, u = 1, \sigma^1 = \sigma^2 = 0.15, S_0^1 = S_0^2 = 10, \alpha = 0.1, \beta = 0.15$.

Par la formule de Margrabe ([Mar78]) pour l'option d'échange, $P(S_u^1, S_u^2)$ le prix du produit au temps u s'écrit sous la forme :

$$P(S_u^1, S_u^2) = EO(T-u, S_u^2, (1-\alpha)S_u^1) - (\alpha+\beta)EO(T-u, S_u^2, S_u^1) - EO(T-u, S_u^2, (1+\beta)S_u^1)$$

avec :

$$EO(\tau, s^2, s^1) = s^2 \mathcal{N}(d_+) - s^1 \mathcal{N}(d_-)$$

où

$$d_{\pm} = \frac{\log(s^2/s^1) \pm (\sigma^2/2)\tau}{\sigma\sqrt{\tau}}.$$

$\sigma = \sqrt{\sigma_1^2 + \sigma_2^2}$ et \mathcal{N} est la fonction de répartition standard de la loi normale centrée réduite.

Le but est de calculer $P(S_u^1, S_u^2) = \mathbb{E}[H|S_u^1, S_u^2]$ en utilisant la régression par moindres carrés. On note par \tilde{P}_N le prix estimé pour le jeu de données $(S^{1,i}, S^{2,i})_{1 \leq i \leq N}$ et par F_N l'espace d'approximation contenant les fonctions de base (peut dépendre des données). Ainsi, on a :

$$\tilde{P}_N = \arg \min_{f \in F_N} \frac{1}{N} \sum_{i=1}^N (f(S_u^{1,i}, S_u^{2,i}) - H_i)^2.$$

4.2 Choix de fonctions de base

4.2.1 Fonctions de base à 2 variables

Dans nos tests numériques, on va considérer dans un premier temps que F_N est un espace vectoriel engendré par les fonctions de base à deux variables $(\psi_{ij})_{1 \leq i, j \leq K}$. On s'intéresse aux cas suivants des fonctions ψ_{ij} :

1. Polynômes à 2 variables :

$$\psi_{i,j}(X, Y) = X^i Y^j.$$

2. Espace (noté `2dAffineMemeSurf`) de *fonctions* à 2 variables affines par morceaux sur des pavés A_{ij} de même surface formant une partition d'un pavé $[\min_1, \max_1] \times [\min_2, \max_2]$ contenant les données (voir Figure 4.2) :

$$\psi_{ij}(X, Y) = (a_{ij}X + b_{ij}Y + c_{ij})1_{A_{ij}}$$

avec :

$A_{ij} = [\min_1 + \delta_1(i-1), \min_1 + \delta_1 i] \times [\min_2 + \delta_2(j-1), \min_2 + \delta_2 j]$ (avec intervalle fermé pour $i = K$ ou $j = K$), où :

$$\delta_1 = \frac{\max_1 - \min_1}{K}, \delta_2 = \frac{\max_2 - \min_2}{K}$$

On peut choisir

$$\min_1 = \min_n(S_u^{1,n}), \max_1 = \max_n(S_u^{1,n}), \min_2 = \min_n(S_u^{2,n}), \max_2 = \max_n(S_u^{2,n}).$$

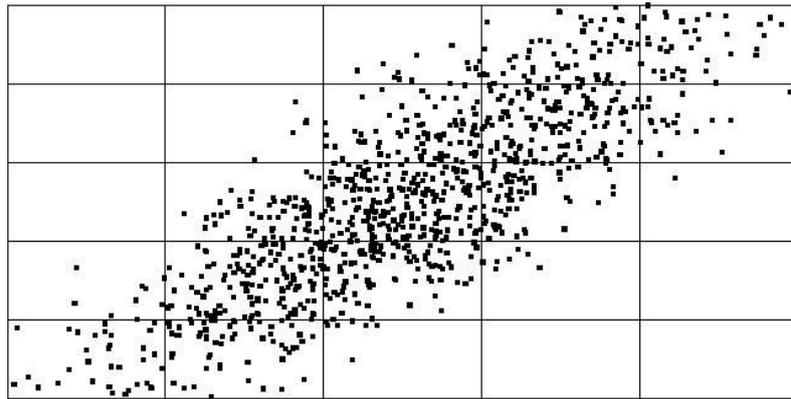


FIGURE 4.1 – Découpage en pavés de même surface (voir [GKKW02])

Sinon, on peut fixer les bords de manière déterministe mais un certain nombre (faible) de données peut se trouver en dehors du pavé global.

3. Espace (noté `2dConstMemeSurf`) de *fonctions* à 2 variables constantes par morceaux sur des pavés A_{ij} de même surface :

$$\psi_{ij}(X, Y) = c_{ij}1_{A_{ij}}$$

avec A_{ij} construit comme dans le cas 2. Notons $\tilde{P}_{3,N}(x, y)$ le prix obtenu par cette régression .

4. On va estimer le prix en procédant à une interpolation bilinéaire entre les centres des pavés A_{ij} (qu'on note (d_j^1, d_j^2)) du prix par moindres carrés obtenu dans le cas 3. Notons ce prix par $\tilde{P}_{4,N}(x, y)$, qui est construit de la manière suivante.

On a $\tilde{P}_{3,N} = \sum_{1 \leq i, j \leq K} c_{ij}1_{A_{ij}}$ avec c_{ij} des constantes.

Le principe est de procéder dans un premier temps à une interpolation linéaire selon une direction et de la refaire selon l'autre direction.

Supposons qu'on veut trouver la valeur de $\tilde{P}_{4,N}$ au point (x, y) entouré des centres $C_{11} = (x_1, y_1)$, $C_{12} = (x_1, y_2)$, $C_{21} = (x_2, y_1)$, $C_{22} = (x_2, y_2)$.

On fait d'abord une interpolation linéaire dans la direction x . On obtient :

$$\tilde{P}_{4,N}(R_1) = \frac{x_2 - x}{x_2 - x_1} \tilde{P}_{4,N}(C_{11}) + \frac{x - x_1}{x_2 - x_1} \tilde{P}_{4,N}(C_{21}),$$

$$\tilde{P}_{4,N}(R_2) = \frac{x_2 - x}{x_2 - x_1} \tilde{P}_{4,N}(C_{12}) + \frac{x - x_1}{x_2 - x_1} \tilde{P}_{4,N}(C_{22})$$

avec $R_1 = (x, y_1)$ et $R_2 = (x, y_2)$.

Ensuite on interpole selon la direction de y :

$$\tilde{P}_{4,N}(x, y) = \frac{y_2 - y}{y_2 - y_1} \tilde{P}_{4,N}(R_1) + \frac{y - y_1}{y_2 - y_1} \tilde{P}_{4,N}(R_2).$$

Si (x, y) se trouve maintenant près du bord de notre partition, l'interpolation reste linéaire parallèlement au bord et constante perpendiculairement. Au coin, l'approximation est constante. On remarque que même si chaque étape est linéaire, l'interpolation résultante globale est une fonction quadratique.

5. Espace (noté `2dAffineMemeData`) de *fonctions* à 2 variables affines par morceaux sur des pavés A_{ij} de même nombre de données qui forment une partition de $[\min_n(S_u^{1,n}), \max_n(S_u^{1,n})] \times [\min_n(S_u^{2,n}), \max_n(S_u^{2,n})]$:

$$\psi_{ij}(X, Y) = (a_{ij}X + b_{ij}Y + c_{ij})1_{A_{ij}}.$$

Pour construire les A_{ij} (voir Figure 4.2), on procède comme suit : selon l'axe de S_u^1 , on découpe les données $(S_u^{1,i}, S_u^{2,i})_{1 \leq i \leq N}$ en K paquets tels que les premières coordonnées forment des blocs de même nombre de données. La frontière selon cette direction entre deux blocs adjacents est la coordonnée milieu entre la dernière coordonnée du premier bloc et la première coordonnée du second. On obtient ainsi K rectangles. De la même manière, on découpe chaque rectangle selon le deuxième axe de S_u^2 en K blocs de même nombre de données. Ainsi on obtient K^2 blocs contenant $\frac{N}{K^2}$ réalisations (sauf éventuellement ceux de la frontière). Une autre manière de procéder est de découper d'abord selon S_u^2 et puis selon S_u^1 . Cependant, vu que dans nos tests on prend S^1 et S^2 de même volatilité, alors les deux façons de découper sont équivalentes (problème symétrique) et par suite on utilisera uniquement le premier découpage.

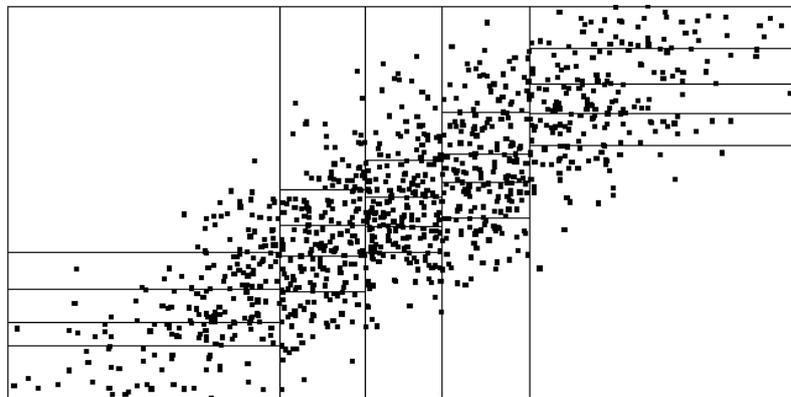


FIGURE 4.2 – Découpage en pavés de même nombre de données (voir [GKKW02])

6. Espace (noté `2dConstMemeData`) de *fonctions* à 2 variables constantes par morceaux sur des pavés A_{ij} de même nombre de données qui forment une partition de $[\min_n(S_u^{1,n}), \max_n(S_u^{1,n})] \times [\min_n(S_u^{2,n}), \max_n(S_u^{2,n})]$:

$$\psi_{ij}(X, Y) = c_{ij}1_{A_{ij}}$$

avec A_{ij} construit comme dans le cas `2dAffineMemeData`.

7. On va estimer le prix en procédant à une interpolation bilinéaire entre les centres des pavés A_{ij} du prix par moindres carrés obtenu dans le cas 6. Notons ce prix par \tilde{P}_7 . L'interpolation est similaire à celle du cas 4.

4.2.2 Réduction de dimension et fonctions à une variable

Puisque le payoff H est homogène et le processus (S_u^1, S^2) est log-normal, alors : $P(S_u^1, S_u^2) = S_u^2 P(\frac{S_u^1}{S_u^2}, 1)$. Pour préserver cette propriété d'homogénéité dans le prix estimé par moindres carrés, il est judicieux d'utiliser des fonctions de base de la forme $(S^2 \phi_k(\frac{S^1}{S^2}))_{1 \leq k \leq K}$. De manière similaire, on s'intéressera dans nos tests aux cas suivants :

1. Polynômes à 1 variable :

$$\phi_k(X) = X^k.$$

2. Espace (noté `1dAffineMemeSurf`) de *fonctions* à 1 variable affines par morceaux sur des intervalles I_k de même longueur formant une partition d'un intervalle $[min, max]$:

$$\phi_k(X) = (a_k X + b_k) 1_{I_k}$$

avec

$$I_k = [min + \delta(k-1), min + \delta k] \quad \text{où} \quad \delta = \frac{max - min}{K}.$$

On peut choisir

$$min = \min_n \left(\frac{S_u^{1,n}}{S_u^{2,n}} \right), \quad max = \max_n \left(\frac{S_u^{1,n}}{S_u^{2,n}} \right).$$

Sinon on peut fixer les bords de manière déterministe.

3. Espace (noté `1dAffineMemeData`) de *fonctions* à 1 variable affines par morceaux sur des intervalles I_k de même nombre de données qui forment une partition de $[\min_n(\frac{S_u^{1,n}}{S_u^{2,n}}), \max_n(\frac{S_u^{1,n}}{S_u^{2,n}})]$:

$$\phi_k(X) = (a_k X + b_k) 1_{I_k}.$$

4.2.3 Calcul pratique de la solution du problème des moindres carrés

On a vu que le prix estimé vérifie :

$$\tilde{P}_N = \arg \min_{f \in F_N} \frac{1}{N} \sum_{i=1}^N (f(X_i) - H_i)^2 \quad \text{avec} \quad X_i = (S_u^{1,i}, S_u^{2,i}).$$

Si F_N est g n r  par p_1, p_2, \dots, p_K , alors on se ram ne au probl me de moindres carr es suivant :

$$\inf_{\alpha} \frac{1}{N} \sum_{i=1}^N (H_i - \alpha \cdot p(X_i))^2$$

avec $p = (p_1, \dots, p_K)$ et $\alpha = (\alpha_1, \dots, \alpha_K) \in \mathbf{R}^K$. Si on pose $B = (p(X_i))_{1 \leq i \leq N}$ (matrice de taille $N \times K$) et $H = (H_1, \dots, H_N)^t$, alors notre probl me s' crit sous forme :

$$\inf_{\alpha} \|B\alpha - H\|_2^2.$$

Quand B est de rang K , il existe une solution unique au probl me. Cependant, lorsqu'il y a des colin arit s entre les colonnes de B , il y a une infinit  de solutions α et il est essentiel de choisir une fa on de r soudre le probl me de moindres carr es. On peut proc der par deux m thodes.

M thode 1(SVD).

On effectue une d composition en valeurs singuli res (not e SVD, voir chapitre 5 de [GVL96]) de B sous la forme : $B = U\Sigma^B V^t$ avec U et V deux matrices orthogonales respectivement de tailles $N \times N$ et $K \times K$ et Σ^B une matrice diagonale de taille $N \times K$ de la forme :

$$\Sigma^B = \begin{pmatrix} \sigma_1^B & 0 & \dots & 0 \\ 0 & \sigma_2^B & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \sigma_K^B \\ 0 & \dots & \dots & 0 \\ \vdots & & & \vdots \\ 0 & \dots & \dots & 0 \end{pmatrix}$$

o  $\sigma_1^B \geq \sigma_2^B \geq \dots \geq \sigma_K^B$.

Si B n'est pas de rang plein, alors il existe $r < K$ tel que $\sigma_{r+1}^B = \dots = \sigma_K^B = 0$.

On a :

$$\begin{aligned} \|B\alpha - H\|_2^2 &= \|U\Sigma^B V^t \alpha - H\|_2^2 \\ &= \|\Sigma^B V^t \alpha - U^t H\|_2^2 \\ &= \sum_{i=1}^N ((\Sigma^B V^t \alpha)_i - (U^t H)_i)^2 \\ &= \sum_{i=1}^r (\sigma_i^B (V^t \alpha)_i - (U^t H)_i)^2 + \sum_{i=r+1}^N ((U^t H)_i)^2 \end{aligned}$$

L'ensemble des solutions \mathcal{S} est :

$$\begin{aligned}\mathcal{S} &= \left\{ \alpha \in \mathbb{R}^K / (V^t \alpha)_i = \frac{(U^t H)_i}{\sigma_i^B}, 1 \leq i \leq r \right\} \\ &= \left\{ V \left(\frac{(U^t H)_1}{\sigma_1^B}, \dots, \frac{(U^t H)_r}{\sigma_r^B}, \lambda_{r+1}, \dots, \lambda_K \right)^t, \lambda_{r+1}, \dots, \lambda_K \in \mathbb{R} \right\}\end{aligned}$$

On choisit la solution de norme euclidienne minimale qui correspond à :

$$\begin{aligned}\alpha^{SVD} &= V \left(\frac{(U^t H)_1}{\sigma_1^B}, \dots, \frac{(U^t H)_r}{\sigma_r^B}, 0, \dots, 0 \right)^t \\ &= V \Sigma^{B,-1} U^t H\end{aligned}$$

avec $\Sigma^{B,-1}$ de taille $K \times N$ définie comme suit :

$$\Sigma^{B,-1} = \begin{pmatrix} \frac{1}{\sigma_1^B} & 0 & \cdots & \cdots & \cdots & \cdots & 0 & \cdots & \cdots & 0 \\ 0 & \frac{1}{\sigma_2^B} & \ddots & \cdots & \cdots & \cdots & \vdots & & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \cdots & \cdots & \vdots & & & \vdots \\ \vdots & \cdots & \ddots & \frac{1}{\sigma_r^B} & \ddots & \cdots & \vdots & & & \vdots \\ \vdots & \cdots & \cdots & \ddots & 0 & \ddots & \vdots & & & \vdots \\ \vdots & \cdots & \cdots & \cdots & \ddots & \ddots & 0 & & & \vdots \\ 0 & \cdots & \cdots & \cdots & \cdots & 0 & 0 & \cdots & \cdots & 0 \end{pmatrix}.$$

Méthode 2 (Equation normale).

La fonction de α à minimiser étant convexe et différentiable, le problème est équivalent à :

$$B^t B \alpha = B^t H.$$

On effectue une décomposition en valeurs singulières de $B^t B$ qui s'écrit $B^t B = V_1 \Sigma V_1^t$ avec V_1 une matrice orthogonale de taille $K \times K$, Σ une matrice de la forme :

$$\Sigma = \begin{pmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_K \end{pmatrix}$$

où $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_K$ sont les valeurs propres de $B^t B$. Si B n'est pas de rang plein alors $B^t B$ n'est pas inversible et par suite il existe $r < K$ tel que $\sigma_{r+1} = \cdots = \sigma_K = 0$.

Il est facile de voir que $V_1 = V$ et $\sigma_i = (\sigma_i^B)^2, 1 \leq i \leq K$. On introduit le pseudo inverse A de $B^t B$ défini par :

$$A = V \Sigma^{-1} V^t$$

où

$$\Sigma^{-1} = \begin{pmatrix} \frac{1}{\sigma_1} & 0 & \dots & \dots & \dots & \dots & 0 \\ 0 & \frac{1}{\sigma_2} & \ddots & \dots & \dots & \dots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \dots & \dots & \vdots \\ \vdots & \dots & \ddots & \frac{1}{\sigma_r} & \ddots & \dots & \vdots \\ \vdots & \dots & \dots & \ddots & 0 & \ddots & \vdots \\ \vdots & \dots & \dots & \dots & \ddots & \ddots & 0 \\ 0 & \dots & \dots & \dots & \dots & 0 & 0 \end{pmatrix}.$$

La solution qu'on choisit au problème de moindres carrées est $\alpha^{EN} = AB^tH$.

Comparaison des deux méthodes.

On a :

$$\begin{aligned} \alpha^{EN} &= V\Sigma^{-1}U^tB^tH = V\Sigma^{-1}V^tV(\Sigma^B)^tU^tH \\ &= V\Sigma^{-1}(\Sigma^B)^tU^tH = V\Sigma^{B,-1}U^tH = \alpha^{SVD}. \end{aligned}$$

Donc d'un point de vue théorique les deux méthodes aboutissent au même résultat. Par contre, il est intéressant de les comparer d'un point de vue pratique. D'abord, avec la méthode 2 il peut y avoir une perte de précision numérique lors du calcul de B^tB , mais cet impact peut être négligé vu que les machines actuelles ont une double précision. Quant au conditionnement de la matrice B^tB , il est égal à $\left(\frac{\sigma_1^B}{\sigma_r^B}\right)^2$ c'est à dire égal au carré de celui de la matrice B . Ainsi, la méthode 1 est plus stable. Enfin, d'un point de vue de coût algorithmique, la deuxième méthode est plus rapide car elle exige $\frac{NK^2}{2} + 12K^3$ opérations tandis que la première demande $2NK^2 + 11K^3$ (Voir chapitre 5 de [GVL96]).

Il est à noter que dans le cas où B est de rang plein, le problème peut être aussi résolu par la méthode d'équations normales via une décomposition de Cholesky sur B^tB , ce qui requiert $(N + K/3)K^2$ opérations ou par un méthode de décomposition QR sur B qui exige $2K^2(N - K/3)$ opérations. Ainsi, d'un point de vue temps de calcul et stockage, l'approche de l'équation normale est mieux adaptée (car $N \gg K$). Cependant, la solution par la deuxième approche reste assez stable seulement lorsque le conditionnement de B n'est pas trop grand et l'erreur résiduelle $\rho = \inf_{\alpha} \|B\alpha - H\|_2$ du problème est faible (Voir chapitre 5 de [GVL96]).

4.3 Régression empirique par moindres carrés standards

4.3.1 Rappel de la littérature sur les majorations d'erreur

Pour discuter de la qualité du prix estimé on s'intéresse à l'erreur quadratique empirique $\|\tilde{P}_N - P\|_N$ définie par :

$$\|\tilde{P}_N - P\|_N^2 = \frac{1}{N} \sum_{i=1}^N (\tilde{P}_N(S_u^{1,i}, S_u^{2,i}) - P(S_u^{1,i}, S_u^{2,i}))^2.$$

Avant de présenter les résultats numériques, on énonce le théorème suivant (voir théorème 11.1 de [GKKW02]) :

Théorème 4.3.1 *Supposons que $\Sigma^2 = \sup_{s_1, s_2 \in \mathbb{R}} \mathbf{Var}\{H | S_u^1 = s_1, S_u^2 = s_2\} < \infty$. Si F_N est un espace vectoriel linéaire de fonctions $f : \mathbb{R}^2 \mapsto \mathbb{R}$ de dimension K_N alors :*

$$\mathbb{E} \left[\|\tilde{P}_N - P\|_N^2 \right] \leq \Sigma^2 \frac{K_N}{N} + \mathbb{E} \left[\inf_{f \in F_N} \|f - P\|_N^2 \right].$$

L'erreur globale $E = \mathbb{E} \left[\|\tilde{P}_N - P\|_N^2 \right]$ se majore par la somme de deux termes connue sous le nom *décomposition biais/variance*. Le premier terme représente l'erreur statistique E^s qui est inversement proportionnelle au nombre de simulations et proportionnelle à la dimension de F_N . Quant au deuxième, il représente l'erreur d'approximation E^{app} . Ainsi, même si l'espace F_N a une bonne qualité d'approximation, l'erreur statistique peut être assez considérable et par conséquent il faut augmenter beaucoup le nombre de simulations, surtout si K_N est grand, pour la rendre comparable à E^{app} .

Commentaires :

1. On a :

$$E^{app} \leq \inf_{f \in F_N} \mathbb{E} \left[\|f - P\|_N^2 \right].$$

2. Si F_N est déterministe :

$$E^{app} \leq \inf_{f \in F_N} \mathbb{E} \left[|f(S_u^1, S_u^2) - P(S_u^1, S_u^2)|^2 \right].$$

4.3.2 Validation numérique du théorème 4.3.1

Dans nos tests, on s'intéressera à la quantité $\mathbb{E} \left[\|\tilde{P}_N - P\|_N^2 \right]$. Vu qu'on ne dispose pas de formule fermée pour $\mathbb{E} \left[\|\tilde{P}_N - P\|_N^2 \right]$, on va donc l'approcher par sa

moyenne empirique qu'on notera \tilde{E} en générant M jeux de scénarii de N trajectoires de (S^1, S^2) aux temps $u = 1$ et $T = 2$ qu'on note $(S_{u,j}^{1,i}, S_{u,j}^{2,i})_{1 \leq i \leq N}$ (avec $1 \leq j \leq M$). On note par \tilde{P}_N^j le prix estimé par le jeu j . On a donc :

$$\tilde{E} = \frac{1}{M} \sum_{j=1}^M \frac{1}{N} \sum_{i=1}^N (\tilde{P}_N^j(S_{u,j}^{1,i}, S_{u,j}^{2,i}) - P(S_{u,j}^{1,i}, S_{u,j}^{2,i}))^2.$$

Par la loi des grands nombres, \tilde{E} converge presque sûrement vers E quand $M \rightarrow \infty$. D'abord, on visualise $\log(\tilde{E})$ en fonction de N pour différents choix de fonctions de base. Un des objectifs des tests suivants est de valider expérimentalement le théorème 4.3.1 et voir dans quelles mesures la majoration donnée est plutôt fine ou grossière, dans les exemples de fonctions de base considérées.

On peut tirer les conclusions suivantes :

- Concernant la régression par polynômes (voir Figures 3) où $K_N = 10$ (le degré de polynôme deg est inférieur ou égal à 3 pour le cas 2 variables et inférieur ou égale à 9 pour le cas 1 variable), les polynômes à une variable sont meilleures que ceux à deux variables. Ceci est cohérent avec l'effet de réduction de la dimension et avec le fait que le cas à 1 variable, au contraire de celui à 2 variables, préserve la propriété d'homogénéité du prix. Mais l'erreur globale est importante pour un nombre de simulations inférieur à 100 000. Ceci s'explique par la dominance de E^s qui décroît en $\frac{1}{N}$ (voir théorème 4.3.1) par rapport à l'erreur d'approximation. Pour le cas à 1 seule variable, il faut beaucoup plus de simulations pour que l'erreur globale converge vers E^{app} ce qui est attendu vu que l'erreur d'approximation pour ce cas est plus petite (effet réduction de dimension et préservation de l'homogénéité). Cependant, l'augmentation du nombre de fonctions de base (voir Figure 3) et par conséquent le degré de polynômes pour $K_N = 45$ ($deg \leq 8$ pour le cas 2 variables et $deg \leq 44$ pour le cas à 1 variable) et $K_N = 105$ ($deg \leq 13$ pour le cas 2 variables et $deg \leq 104$ pour le cas à 1 variable), engendre beaucoup d'instabilité dans la régression. Ceci est dû au fait que la matrice d'inversion du système de moindres carrés est mal conditionnée à cause de ses valeurs très grandes et très petites.
- La visualisation des résultats de fonctions locales par morceaux sur des parties de même aire (Figure 4) permet de conclure que la qualité d'approximation de l'espace `1dAffineMemeSurf` est meilleure. L'erreur statistique est considérable lorsqu'on se limite à un petit nombre de simulations. Cette observation s'accroît encore davantage si on augmente le nombre de fonctions de base ce qui est en cohérence avec l'expression de E^s dans le Théorème 3.1. En d'autres termes, il vaut mieux utiliser un espace de faible dimension lorsqu'en pratique

on est limité en nombre de simulations.

- Pour le cas de fonctions locales par morceaux sur des parties de même nombre de données (Figure 5), il en sort que l'espace `1dAffineMemeData` approche mieux le prix (Effet réduction de dimension et homogénéité) à condition que l'erreur statistique soit écrasée ce qui se constate pour un nombre de simulations assez grand et pour un faible nombre de fonctions de base.
- L'interpolation de l'estimateur par fonctions locales constantes par morceaux est moins avantageuse qu'une régression sur des fonctions locales affines par morceaux, que ce soit sur des pavés de même surface ou de même nombre de données (Figures 4 et 5).
- Enfin, la figure 6 montre que la comparaison des différentes méthodes demeure plus subtile. En effet, les polynômes sont les mieux adaptés mais à condition de se limiter à un faible degré. Si on veut augmenter plus la dimension de l'espace d'approximation, les polynômes cèdent la première place aux fonctions affines par morceaux à une variable. On remarque aussi que pour ces fonctions, l'utilisation d'intervalles avec le même nombre de données est mieux pour un petit nombre de fonction de bases. Par contre pour un grand nombre, il n'est pas possible de trancher au vu de nos simulations limitées à $N = 200000$.

Pour fonder une idée sur l'optimalité de la majoration de l'erreur dans le théorème 3.1, on trace \tilde{E} en fonction de $\frac{1}{N}$. On obtient les graphes suivants des figures 7 et 8.

Concernant le premier graphe de la figure 7 consacré aux fonctions à 2 variables avec à peu près le même K , on remarque que la dépendance de l'erreur est linéaire et que les pentes sont assez proches. Ceci est en concordance avec la majoration de l'erreur qui est linéaire en $\frac{1}{N}$ à nombre de fonctions de base égal. Le deuxième graphe confirme cette remarque pour les fonctions à une seule variable. Il est intéressant de noter que les rapports des pentes des courbes de fonctions constantes par morceaux pour $K_N = 49$ avec celles pour $K_N = 16$ valent $\frac{240}{61}$ et $\frac{142}{49}$ c'est à dire proche de $\frac{49}{16} = 3.06$. Ceci est cohérent avec la croissance linéaire de la majoration de l'erreur en fonction du nombre de fonctions de base. Cette remarque est aussi valable pour les fonctions affines par morceaux (Figure 8).

Enfin, la comparaison des différentes figures précédentes permet de conclure que les pentes des courbes demeurent du même ordre de grandeur à nombre de fonctions

de base égal. En effet, le facteur de majoration de E^s dans le théorème 3.1 dépend de l'espace F_N uniquement via sa dimension.

4.3.3 Validation croisée

On a vu que l'erreur globale dépend de N et de K . En pratique, on dispose d'un nombre de simulations limité et par conséquent augmenter la dimension K , certes permet d'améliorer l'erreur d'approximation, mais d'un autre côté ceci agrandit l'erreur statistique E^s . Ainsi, on peut s'intéresser à la dimension optimale \hat{K} permettant de faire une meilleure balance entre E^{app} et E^s afin de minimiser l'erreur globale. Dans un premier temps, on visualisera le comportement de l'erreur globale en fonction de N et de K et dans un second temps, on présentera comment on peut estimer la dimension optimale en utilisant la méthode de validation croisée.

Dans les figures qui suivent (Figures 13, 14 et 15), on trace la surface représentant l'erreur globale en fonction de N et de K pour l'espace des polynômes à 1 variable, l'espace `1dAffineMemeSurf`, et l'espace `1dAffineMemeData`. On remarque que les surfaces ont des allures convexes et que pour chaque N correspond une dimension optimale \hat{K} . On note aussi que lorsque N est petit (proche de 1000), l'erreur globale s'écarte brutalement de l'erreur globale optimale dès que K bouge un peu autour de \hat{K} . Ceci est dû essentiellement à la valeur considérable de E^s pour N petit. En effet, l'erreur empirique est plus plate autour de \hat{K} pour N proche de 10000.

En pratique, on dispose d'un seul jeu de scénarii et il est donc impossible de connaître la valeur exacte de \hat{K} . Ainsi, on propose ci-dessous la méthode de validation croisée pour estimer \hat{K} .

On note $X_i = (S_u^{1,i}, S_u^{2,i})$, $Y_i = H_i$, $D_n = \{(X_1, Y_1), \dots, (X_n, Y_n)\}$, et Q un ensemble fini contenant les dimensions possibles de l'espace de fonctions de base considéré parmi lesquels on désire déterminer la dimension optimale \hat{K} . Pour estimer \hat{K} , on divise notre échantillon D_n en deux ($n = n_l + n_t$) :

$$D_{n_l} = \{(X_1, Y_1), \dots, (X_{n_l}, Y_{n_l})\} \quad \text{et} \quad D_{n_t} = \{(X_{n_l+1}, Y_{n_l+1}), \dots, (X_n, Y_n)\}.$$

Pour chaque $K \in Q$ on note $\tilde{P}_{n_l}^K$ le prix estimé en utilisant l'échantillon D_{n_l} (*learning set*). \hat{K} est définie comme étant la dimension minimisant l'espérance de l'erreur quadratique calculée calculée sous la vraie loi μ de (S_u^1, S_u^2) , i.e.

$$\int |\tilde{P}_{n_l}^{\hat{K}}(x) - P(x)|^2 \mu(dx) = \min_{K \in Q} \int |\tilde{P}_{n_l}^K(x) - P(x)|^2 \mu(dx).$$

Pour estimer \hat{K} (pour n_l simulations), on utilise ensuite l'échantillon de test D_{n_t} par \tilde{K} construit comme suit :

$$\frac{1}{n_t} \sum_{i=n_l+1}^n |\tilde{P}_{n_l}^{\tilde{K}}(X_i) - Y_i|^2 = \min_{K \in Q} \frac{1}{n_t} \sum_{i=n_l+1}^n |\tilde{P}_{n_l}^K(X_i) - Y_i|^2.$$

On a le résultat suivant (voir théorème 7.1 de [GKKW02]) :

Théorème 4.3.2 *Soit $0 < L < +\infty$. Supposons que $|H| \leq L$ et $\max_{K \in Q} \|P_{n_l}^K\|_\infty \leq L$.*

Alors, pour tout $\delta > 0$,

$$\begin{aligned} & \mathbb{E} \left[\int |\tilde{P}_{n_l}^{\tilde{K}}(x) - P(x)|^2 \mu(dx) \right] \\ & \leq (1 + \delta) \mathbb{E} \left[\int |\tilde{P}_{n_l}^{\hat{K}}(x) - P(x)|^2 \mu(dx) \right] + c \frac{1 + \log(\text{Card}(Q))}{n_t}, \\ & \text{avec } c = L^2(16/\delta + 35 + 19\delta). \end{aligned}$$

Dans nos tests numériques, on traite le cas de l'espace des polynômes à 1 variable, l'espace `1dAffineMemeSurf` et l'espace `1dAffineMemeData`. On choisit $Q = \{2, 4, 6, 8, 12, 14, 16, 18, 20\}$ pour le premier espace et $Q = \{4, 8, 12, 16, 20, 24, 28, 32, 36, 40\}$ pour le deuxième et le troisième. Enfin on s'intéresse au cas de $n_l \in \{1000, 4000, 7000, 10000\}$ avec $n_l = n_t$. En se basant sur les surfaces tracées précédemment, on reporte dans le tableau suivant \hat{K} pour chaque espace et chaque $n_l \in \{1000, 4000, 7000, 10000\}$.

	$n_l = 1000$	$n_l = 4000$	$n_l = 7000$	$n_l = 10000$
Espace Polynômes 1D	4	4	6	6
Espace 1AffMemeSurf	8	16	16	20
Espace 1AffMemeData	8	8	8	12

Pour vérifier la méthode de validation croisée, on l'applique 100 fois pour chaque $n_l \in \{1000, 4000, 7000, 10000\}$ et chaque espace en utilisant à chaque fois un jeu de scénario indépendant de ceux utilisés avant. On obtient ainsi 100 réalisations de \tilde{K} et par suite on trace le pourcentage de réalisations des valeurs possibles de \tilde{K} pour les différents n_l et les différents espaces considérés. On obtient les graphes de la figure 1.

On remarque que pour le premier et le troisième graphe de la figure 16, \tilde{K} ne s'écarte pas trop de la dimension optimale \hat{K} et que son pourcentage de réalisation (histogramme empirique) a toujours un pic pour $\tilde{K} = \hat{K}$ pour tout n_l . Quant au graphe du milieu, les queues de distribution des courbes sont plus épaisses avec des pics se produisant pour des \tilde{K} légèrement différents de \hat{K} . Ceci est dû essentiellement à la grande stabilité de l'erreur globale autour de \hat{K} pour l'espace `1dAffineMemeSurf` (voir la figure 14).

4.4 Régression par Variables de contrôles préliminaires (VCP)

4.4.1 Description et heuristique numérique

Sachant que l'erreur statistique est liée à la variance conditionnelle du payoff par rapport aux deux variables sous-jacentes, on tente ici la bonne qualité de réduction de cette variance par une méthode de régression qu'on appelle régression par VCP (variables de contrôles préliminaires) exploitant la qualité d'approximation par polynômes à faible degré. On approche dans la première étape notre payoff par de tels polynômes et dans la seconde on régresse sur des fonctions de base (polynômes, locales par morceaux) le payoff auquel on a retranché les polynômes d'approximation de la première étape.

Notons : $X_i = (S_u^{1,i}, S_u^{2,i})$ et $Z_i = (S_u^{1,i}, S_u^{2,i}, S_T^{1,i}, S_T^{2,i})$.

Pour cela on calcule des coefficients $(\tilde{\alpha}_k)_{1 \leq k \leq K_1}$ qui vérifient :

$$(\tilde{\alpha}_k)_k = \arg \inf_{(\alpha_k)_k} \frac{1}{N} \sum_{i=1}^N \left| H_i - \sum_{k=1}^{K_1} \alpha_k P_k(Z_i) \right|^2$$

avec $(P_k)_{1 \leq k \leq K_1}$ est une famille de polynômes de faible degré.

Ensuite on calcule les coefficients de régression par moindres carrées $(\tilde{\beta}_k)_{1 \leq k \leq K_2}$ vérifiant :

$$(\tilde{\beta}_k)_k = \arg \inf_{(\beta_k)_k} \frac{1}{N} \sum_{i=1}^N \left| H_i - \sum_{k=1}^{K_1} \tilde{\alpha}_k P_k(Z_i) - \sum_{k=1}^{K_2} \beta_k \Phi_k(X_i) \right|^2$$

avec $(\Phi_k)_{1 \leq k \leq K_2}$ est une famille de fonctions locales par morceaux ou d'autres fonctions de base.

On choisit :

$$\begin{aligned} P_k(Z_i) &:= S_T^{2,i} \left(\frac{S_T^{1,i}}{S_T^{2,i}} \right)^k - \mathbb{E} \left[S_T^{2,i} \left(\frac{S_T^{1,i}}{S_T^{2,i}} \right)^k \mid S_u^{1,i}, S_u^{2,i} \right] \\ &= S_T^{2,i} \left(\frac{S_T^{1,i}}{S_T^{2,i}} \right)^k - \exp \left(\frac{\sigma^2}{2} (k^2 + (k-1)^2 - 1) \right) S_u^{2,i} \left(\frac{S_u^{1,i}}{S_u^{2,i}} \right)^k \end{aligned}$$

en exploitant que S^1 et S^2 sont log-normaux, indépendants, de même volatilité σ et que $T - u = 1$.

Puisque $\mathbb{E}[P_k(Z_i) | X_i] = 0$, retrancher les $P_k(Z_i)$ du payoff ne change pas la fonction de régression, par contre on peut espérer que la variance de $H - \sum_{k=1}^{K_1} \alpha_k P_k(Z)$ soit bien plus petite. Par suite, il est légitime de penser pouvoir ainsi diminuer

l'erreur statistique (terme E^s dans le théorème 3.1). On définit alors P par $\tilde{P}_N^h = \sum_{k=1}^{K_2} \tilde{\beta}_k \Phi_k$.

Dans nos tests numériques on prendra $K_1 \in \{0, 4, 8, 10, 12\}$. Les figures 9-10-11-12 visualisent en échelle log l'erreur empirique globale pour différents choix de K_1 et de $(\Phi_k)_{1 \leq k \leq K_2}$ pour une plage de faible nombre de simulations ($< 10\ 000$) et une autre de grand nombre de simulations (entre $10\ 000$ et $100\ 000$). Dans un souci de comparaison, nous avons aussi ajouté sur ces mêmes graphes l'erreur pour la régression standard calculée sur un échantillon de variables antithétiques.

La figure 9 visualise pour différents K_1 l'erreur globale pour $K_2 = 10$ polynômes à une variable. On remarque que la réduction de variance par la régression par VCP, comparée à une régression classique, améliore la convergence de l'erreur globale (à part pour $K_2 = 12$). Il est intéressant de noter que l'erreur statistique est considérable pour N très petit surtout lorsque K_1 est grand. Ainsi, on peut noter que pour un très faible nombre de simulations, l'utilisation de $K_1 = 4$ polynômes suffit, tandis que lorsque N est suffisamment grand le choix $K_1 = 8$ devient optimal.

L'étude de l'espace `1dAffineMemeData` pour $K_2 = 12$ (Figure 10) permet d'aboutir aux mêmes conclusions que précédemment. En outre, les courbes pour $K_1 \in \{4, 8, 10\}$ convergent vers une même erreur globale (erreur d'approximation) dès que $N \geq 10000$. Ceci s'explique par la réduction de variance qui permet de rendre l'erreur statistique rapidement comparable au biais. Quant au cas $K_1 = 12$, il n'est pas stable vu le degré élevé des polynômes.

Concernant l'espace `2dAffineMemeSurf` (Figure 11), outre ce qui a été dit, on voit que la convergence de l'erreur globale est plus rapide que précédemment surtout pour $K_1 \in \{4, 8, 10\}$. Ceci est dû à la mauvaise qualité d'approximation de l'espace `2dAffineMemeSurf` par rapport aux espaces d'avant.

Il est important de noter que, sur ce cas de test, la méthode par variables antithétiques n'apporte aucune amélioration par rapport à la régression standard.

Enfin, la comparaison de la figure 12 à la figure 10 montre que lorsque K_2 est grand (par exemple 50), l'effet de réduction de variance est encore plus visible vu que le terme de l'erreur statistique initiale est plus important. C'est ce qui explique le grand écart des erreurs globales entre le cas sans réduction et les cas avec réduction de variance.

Avant de conclure quant à la pertinence de la méthode proposée, il est légitime

de se demander comment se comporte la régression par VCP si le calcul des coefficients à la première étape est fait avec un échantillon indépendant. Les figures 17, 18 et 19 apportent une réponse à cette question. En effet, on constate que dans ce cas la méthode est moins bonne et moins stable surtout lorsque N est petit. Ceci peut s'expliquer par le fait que la minimisation de la variance empirique lors de la première étape n'assure pas forcément que cette quantité soit petite en changeant d'échantillon dans la seconde étape, surtout avec un petit nombre de simulation.

En guise de conclusion, la régression par VCP accélère nettement la convergence de l'erreur globale de l'estimateur des moindres carrés vers l'erreur d'approximation indépendamment de l'espace de projection choisi. Néanmoins, à la lumière de ces tests, il est vraisemblable que le choix optimal de K_1 obéit encore à une balance équilibrée entre biais et variance :

- Choisir un K_1 trop petit va peu réduire la variance mais l'estimation des coefficients $\tilde{\alpha}_k$ est meilleure (moindre erreur statistique dans leurs calculs).
- Opter pour un K_1 trop grand va mieux réduire la variance, mais l'erreur statistique dans l'estimation des coefficients $\tilde{\alpha}_k$ va être très importante et peut faire perdre l'intérêt d'une régression par VCP.

La quantification de l'erreur globale en fonction de K_1 , K_2 et N fait l'objet des parties suivantes. On peut aussi envisager ici une validation croisée pour le choix de optimal K_1 et K_2 , problème que nous laissons pour des recherches ultérieures.

4.4.2 Quantification de l'erreur

Soient d_1, d_2, K_1 et K_2 des entiers strictement positifs.

Soient h une fonction de $\mathbb{R}^{d_1} \rightarrow \mathbb{R}$ et $(P_k)_{1 \leq k \leq K_1}$ une famille de fonctions de $\mathbb{R}^{d_1} \rightarrow \mathbb{R}$ tel que $\|P_k\|_\infty \leq 1$ pour $1 \leq k \leq K_1$.

Soient Z et X des variables aléatoires à valeurs respectivement dans \mathbb{R}^{d_1} et \mathbb{R}^{d_2} .

On suppose que :

$$\forall 1 \leq k \leq K_1 : \mathbb{E}[P_k(Z)|X] = 0 \quad \text{et} \quad \exists L > 1 : \|h\|_\infty \leq L.$$

On se donne un échantillon $(Z_i, X_i)_{1 \leq i \leq N}$ de variables aléatoires indépendantes et de même que loi que (Z, X) . Soit \mathcal{F}_N un espace vectoriel de fonctions $f : \mathbb{R}^{d_1} \rightarrow \mathbb{R}$ de dimension K_2 engendré par une base de fonctions qu'on note $(\Phi_k)_{1 \leq k \leq K_2}$ qui peuvent éventuellement dépendre des données (X_1, \dots, X_N) .

Le but est de calculer par une méthode de moindres carrés avec réduction de variance un estimateur \tilde{m}_N de la fonction de régression m définie par :

$$m(x) = \mathbb{E}[H|X = x] \quad \text{avec} \quad H = h(Z).$$

Voici les notations :

- $H_i = h(Z_i)$ et $D_N = \{(Z_1, X_1), \dots, (Z_N, X_N)\}$.
- μ la mesure de loi associée à (Z, X) et μ_N sa version empirique calculée sur l'échantillon D_N :

$$\mu_N(dz, dx) := \frac{1}{N} \sum_{i=1}^N \delta_{(Z_i, X_i)}(dz, dx)$$

où $\delta_{(z,x)}$ est la masse de dirac au point $(z, x) \in \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$.

- pour deux fonctions g_1, g_2 de $\mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \rightarrow \mathbb{R}$, on définit les produits scalaires $\langle \cdot, \cdot \rangle_\mu$ et $\langle \cdot, \cdot \rangle_{\mu_N}$ ainsi que les deux normes associées $\|\cdot\|_\mu$ et $\|\cdot\|_{\mu_N}$ par :

$$\langle g_1, g_2 \rangle_\mu = \int g_1(z, x) g_2(z, x) \mu(dz, dx),$$

$$\langle g_1, g_2 \rangle_{\mu_N} = \frac{1}{N} \sum_{i=1}^N g_1(Z_i, X_i) g_2(Z_i, X_i),$$

$$\|g_1\|_\mu^2 = \int g_1(z, x)^2 \mu(dz, dx) \quad \text{et} \quad \|g_1\|_{\mu_N}^2 = \frac{1}{N} \sum_{i=1}^N |g_1(Z_i, X_i)|^2.$$

- On introduit un espace de coefficients utilisés pour la réduction de variance :

$$\mathcal{A} = \left\{ \alpha \in \mathbb{R}^{K_1} : \|\alpha\|_1 = \sum_{i=1}^{K_1} |\alpha_i| \leq L \right\}.$$

Par l'inégalité triangulaire, on remarque que \mathcal{A} est convexe fermé non vide. On fait ce choix pour \mathcal{A} car on aura besoin pour la preuve du théorème 4.1 de divers inégalités de probabilité nécessitant des hypothèses de bornitude.

- T_L l'opérateur de troncature des fonctions au niveau L défini comme suit :

$$T_L f(x) = \begin{cases} f(x), & \text{si } |f(x)| \leq L, \\ L \text{sign}(f(x)), & \text{sinon.} \end{cases}$$

- $\mathcal{G} = \left\{ \sum_{k=1}^{K_1} \alpha_k P_k : \alpha \in \mathbb{R}^{K_1} \right\}$, $\mathcal{G}^{\mathcal{A}} = \left\{ \sum_{k=1}^{K_1} \alpha_k P_k : \alpha \in \mathcal{A} \right\}$,

$T_L \mathcal{G} = \{T_L g : g \in \mathcal{G}\}$. On remarque que si $f \in \mathcal{G}^{\mathcal{A}}$, alors $|f(x)| \leq |\alpha|_1 \leq L$ et $f \in \mathcal{G}$. Ainsi $f \in T_L \mathcal{G}$.

Maintenant, on introduit la notion de projection sur $\mathcal{G}^{\mathcal{A}}$.

Définition 4.4.1 (Voir théorème V.2, p.79 dans [Bre99])

Comme $\mathcal{G}^{\mathcal{A}}$ un convexe fermé non vide de $\mathbb{L}^2(\mathbb{R}^{d_1}, \mathbb{P} \circ Z^{-1})$, pour toute fonction $u \in \mathbb{L}^2(\mathbb{R}^{d_1}, \mathbb{P} \circ Z^{-1})$, on peut associer une fonction $\mathcal{P}(u) \in \mathcal{G}^{\mathcal{A}}$ qui est l'unique projection de u dans $\mathcal{G}^{\mathcal{A}}$ qui vérifie : $\mathbb{E}[|u(Z) - \mathcal{P}(u)(Z)|^2] = \inf_{g \in \mathcal{G}^{\mathcal{A}}} \mathbb{E}[|u(Z) - g(Z)|^2]$. On a aussi :

$$\forall g \in \mathcal{G}^{\mathcal{A}} : \mathbb{E}[(u(Z) - \mathcal{P}(u)(Z))(g(Z) - \mathcal{P}(u)(Z))] \leq 0. \quad (4.4.1)$$

Il n'est pas difficile de noter que de (4.4.1) découle que \mathcal{P} est 1-Lipschitz :

$$\mathbb{E}[|\mathcal{P}(u)(Z) - \mathcal{P}(v)(Z)|^2] \leq \mathbb{E}[|u(Z) - v(Z)|^2] \text{ pour } u, v \in \mathbb{L}^2(\mathbb{R}^{d_1}, \mathbb{P} \circ Z^{-1})$$

On propose l'algorithme suivant pour calculer \tilde{m}_N :

1. Etape 1 : Réduction de variance.

On calcule les coefficients $(\tilde{\alpha}_k)_{1 \leq k \leq K_1}$ qui vérifient :

$$(\tilde{\alpha}_k)_{1 \leq k \leq K_1} = \arg \inf_{\alpha \in \mathcal{A}} \frac{1}{N} \sum_{i=1}^N \left| H_i - \sum_{k=1}^{K_1} \alpha_k P_k(Z_i) \right|^2.$$

C'est résoudre le problème de projection sur le convexe fermé non vide $\mathcal{G}^{\mathcal{A}}$ en utilisant le produit scalaire associé à la norme empirique des $(Z_i)_{1 \leq i \leq N}$.

2. Etape 2 : Régression par moindres carrés.

On calcule les coefficients de régression par moindres carrés $(\tilde{\beta}_k)_{1 \leq k \leq K_2}$ vérifiant :

$$(\tilde{\beta}_k)_{1 \leq k \leq K_2} = \arg \inf_{(\beta_k)_k} \frac{1}{N} \sum_{i=1}^N \left| H_i - \sum_{k=1}^{K_1} \tilde{\alpha}_k P_k(Z_i) - \sum_{k=1}^{K_2} \beta_k \Phi_k(X_i) \right|^2.$$

\tilde{m}_N est calculé par :

$$\tilde{m}_N = \sum_{k=1}^{K_2} \tilde{\beta}_k \Phi_k.$$

On a le résultat suivant :

Théorème 4.4.1 Posons $H^\alpha = H - \sum_{k=1}^{K_1} \alpha_k P_k$ pour $\alpha \in \mathcal{A}$.

1. Si les Φ_k sont constants par morceaux sur des pavés disjoints de même nombre de données ou équiprobables, alors

$$\begin{aligned} \mathbb{E}[\|\tilde{m}_N - m\|_{\mu_N}^2] \leq & 2 \left\{ \frac{c_1}{N} + \frac{(c_2 + c_3 \log(N))(K_1 + 1)}{N} \right\} \\ & + 2 \frac{K_2}{N} \inf_{\alpha \in \mathcal{A}} \mathbb{E}[\mathbf{Var}(H^\alpha | X)] + 2 \mathbb{E} \left\{ \inf_{\Phi \in \mathcal{F}_N} \|\Phi - m\|_{\mu_N}^2 \right\} \end{aligned}$$

avec c_1, c_2 et c_3 des constantes indépendantes de K_1, K_2 et N (mais dépendantes de L).

2. Si $\Sigma^2(\hat{\alpha}) = \sup_{x \in \mathbb{R}^{d_2}} \mathbf{Var}\{H^{\hat{\alpha}} | X = x\} < \infty$. Alors

$$\begin{aligned} \mathbb{E}[\|\tilde{m}_N - m\|_{\mu_N}^2] \leq & 2 \left\{ \frac{c_1}{N} + \frac{(c_2 + c_3 \log(N))(K_1 + 1)}{N} \right\} \\ & + 2 \frac{K_2}{N} \Sigma^2(\hat{\alpha}) + 2 \mathbb{E} \left\{ \inf_{\Phi \in \mathcal{F}_N} \|\Phi - m\|_{\mu_N}^2 \right\} \end{aligned}$$

avec c_1, c_2 et c_3 des constantes indépendantes de K_1, K_2 et N (mais dépendantes de L).

4.4.3 Preuve du théorème

On introduit \hat{m}_N défini par :

$$\hat{m}_N = \sum_{k=1}^{K_2} \hat{\beta}_k \Phi_k$$

avec :

$$(\hat{\alpha}_k)_{1 \leq k \leq K_1} = \arg \inf_{\alpha \in \mathcal{A}} \mathbb{E} \left[\left| H - \sum_{k=1}^{K_1} \alpha_k P_k(Z) \right|^2 \right],$$

$$(\hat{\beta}_k)_{1 \leq k \leq K_2} = \arg \inf_{(\beta_k)_k} \frac{1}{N} \sum_{i=1}^N \left| H_i - \sum_{k=1}^{K_1} \hat{\alpha}_k P_k(Z_i) - \sum_{k=1}^{K_2} \beta_k \Phi_k(X_i) \right|^2.$$

On a :

$$\mathbb{E}[\|\tilde{m}_N - m\|_{\mu_N}^2] \leq 2\mathbb{E}[\|\tilde{m}_N - \hat{m}_N\|_{\mu_N}^2] + 2\mathbb{E}[\|\hat{m}_N - m\|_{\mu_N}^2]. \quad (4.4.2)$$

Etape 1 : Majoration de $\mathbb{E}[\|\tilde{m}_N - \hat{m}_N\|_{\mu_N}^2]$.

\tilde{m}_N et \hat{m}_N sont respectivement les projections orthogonales sur $\text{vect}((\Phi_k)_{1 \leq k \leq K_2})$ par rapport au produit scalaire $\langle \cdot, \cdot \rangle_{\mu_N}$ de $h - \sum_{k=1}^{K_1} \tilde{\alpha}_k P_k$ et $h - \sum_{k=1}^{K_1} \hat{\alpha}_k P_k$. Par contraction de la projection on a :

$$\|\tilde{m}_N - \hat{m}_N\|_{\mu_N}^2 \leq \left\| \sum_{k=1}^{K_1} \tilde{\alpha}_k P_k - \sum_{k=1}^{K_1} \hat{\alpha}_k P_k \right\|_{\mu_N}^2.$$

Et ainsi :

$$\mathbb{E}[\|\tilde{m}_N - \hat{m}_N\|_{\mu_N}^2] \leq \mathbb{E} \left[\left\| \sum_{k=1}^{K_1} \tilde{\alpha}_k P_k - \sum_{k=1}^{K_1} \hat{\alpha}_k P_k \right\|_{\mu_N}^2 \right].$$

On définit r et r_N par : $r_N = \sum_{k=1}^{K_1} \tilde{\alpha}_k P_k$ et $r = \sum_{k=1}^{K_1} \hat{\alpha}_k P_k$. Notons que r et r_N sont dans $\mathcal{G}^{\mathcal{A}}$ et que $r = \mathcal{P}(h)$.

On a :

$$\begin{aligned} & \int |r_N(z) - h(z)|^2 \mu(dz, dx) \\ &= \int |r_N(z) - r(z) + r(z) - h(z)|^2 \mu(dz, dx) \\ &= \int |r_N(z) - r(z)|^2 \mu(dz, dx) + 2 \int (r_N(z) - r(z))(r(z) - h(z)) \mu(dz, dx) \\ & \quad + \int |r(z) - h(z)|^2 \mu(dz, dx) \\ &\geq \int |r_N(z) - r(z)|^2 \mu(dz, dx) + \int |r(z) - h(z)|^2 \mu(dz, dx) \quad (\text{utilisation de (4.4.1)}). \end{aligned}$$

Ainsi,

$$\begin{aligned}
\|r_N - r\|_\mu^2 &= \int |r_N(z) - r(z)|^2 \mu(dz, dx) \\
&\leq \int |r_N(z) - h(z)|^2 \mu(dz, dx) - \int |r(z) - h(z)|^2 \mu(dz, dx) \\
&= \int |r_N(z) - h(z)|^2 \mu(dz, dx) - \mathbb{E}[|r(Z) - H|^2].
\end{aligned}$$

Et puisque par définition de r_N , $\frac{1}{N} \sum_{i=1}^N |r_N(Z_i) - H_i|^2 - \frac{1}{N} \sum_{i=1}^N |r(Z_i) - H_i|^2 \leq 0$, alors on déduit :

$$\begin{aligned}
\int |r_N(z) - r(z)|^2 \mu(dz, dx) &\leq \int |r_N(z) - h(z)|^2 \mu(dz, dx) - \mathbb{E}[|r(Z) - H|^2] \\
&\quad - 2 \left(\frac{1}{N} \sum_{i=1}^N |r_N(Z_i) - H_i|^2 - \frac{1}{N} \sum_{i=1}^N |r(Z_i) - H_i|^2 \right) \\
&:= T_N.
\end{aligned}$$

Soit $t \geq \frac{1}{N}$, On a :

$$\begin{aligned}
\mathbb{P}(T_N > t) &= \mathbb{P} \left\{ \int |r_N(z) - h(z)|^2 \mu(dz, dx) - \mathbb{E}[|r(Z) - H|^2] \right. \\
&\quad \left. - \left(\frac{1}{N} \sum_{i=1}^N |r_N(Z_i) - H_i|^2 - \frac{1}{N} \sum_{i=1}^N |r(Z_i) - H_i|^2 \right) \right. \\
&\quad \left. > \frac{1}{2} \left(t + \int |r_N(z) - h(z)|^2 \mu(dz, dx) - \mathbb{E}[|r(Z) - H|^2] \right) \right\} \\
&\leq \mathbb{P} \left\{ \exists f \in T_L \mathcal{G} : \mathbb{E}[|f(Z) - H|^2] - \mathbb{E}[|r(Z) - H|^2] \right. \\
&\quad \left. - \left(\frac{1}{N} \sum_{i=1}^N |f(Z_i) - H_i|^2 - \frac{1}{N} \sum_{i=1}^N |r(Z_i) - H_i|^2 \right) \right. \\
&\quad \left. > \frac{1}{2} \left(\frac{t}{2} + \frac{t}{2} + \mathbb{E}[|f(Z) - H|^2] - \mathbb{E}[|r(Z) - H|^2] \right) \right\} \\
&\quad (\text{car } \mathcal{G}^A \subset T_L \mathcal{G}).
\end{aligned}$$

En utilisant le théorème 11.4 de [GKKW02], on a

$$\begin{aligned}
\mathbb{P}(T_N > t) &\leq 14 \sup_{z_1^N} \mathcal{N}_1 \left(\frac{t}{80L}, T_L \mathcal{G}, z_1^N \right) \exp \left(-\frac{N}{24.214L^4} t \right) \\
&\leq 14 \sup_{z_1^N} \mathcal{N}_1 \left(\frac{1}{80L.N}, T_L \mathcal{G}, z_1^N \right) \exp \left(-\frac{N}{24.214L^4} t \right) (\text{car } t \geq \frac{1}{N}).
\end{aligned}$$

Par le théorème 4.5.1 on obtient (pour $\frac{1}{80LN} < \frac{L}{2}$) :

$$\begin{aligned} \mathcal{N}_1 \left(\frac{t}{80L}, T_L \mathcal{G}, z_1^N \right) &\leq 3 \left(\frac{2e(2L)}{\frac{1}{80L \cdot N}} \log \left(\frac{2e(2L)}{\frac{1}{80L \cdot N}} \right) \right)^{V_{T_L \mathcal{G}^+}} \\ &\leq 3(480eL^2N)^{2V_{T_L \mathcal{G}^+}}. \end{aligned}$$

En utilisant l'inégalité ci-dessus avec le fait que $V_{T_L \mathcal{G}^+} \leq V_{\mathcal{G}^+} \leq K_1 + 1$ (voir théorème 4.5.2), on déduit pour $\epsilon \geq \frac{1}{N}$:

$$\begin{aligned} \mathbb{E}[T_N] &= \int_0^\infty \mathbb{P}(T_N > t) dt \leq \epsilon + \int_\epsilon^\infty \mathbb{P}(T_N > t) dt \\ &\leq \epsilon + \int_\epsilon^\infty 42(480eL^2N)^{2(K_1+1)} \exp\left(-\frac{N}{24 \cdot 214L^4} t\right) dt \\ &\leq \epsilon + 42(480eL^2N)^{2(K_1+1)} \frac{24 \cdot 214L^4}{N} \exp\left(-\frac{N}{24 \cdot 214L^4} \epsilon\right). \end{aligned}$$

L'expression ci-dessus est minimale pour $\epsilon = \frac{24 \cdot 214L^4}{N} \log\left(\left(42(480eL^2N)^{2(K_1+1)}\right)\right)$ (on vérifie que $\epsilon \geq \frac{1}{N}$ car $L \geq 1$). En injectant cette valeur on aboutit à :

$$\mathbb{E}[T_N] \leq \frac{24 \cdot 214L^4}{N} \left((\log(42) + 2(K_1 + 1) \log(480eL^2N)) + \frac{24 \cdot 214L^4}{N} \right).$$

Ainsi,

$$\mathbb{E}[\|r_N - r\|_\mu^2] \leq \frac{a_1}{N} + \frac{(a_2 + a_3 \log(N))(K_1 + 1)}{N}, \text{ pour } N \geq \frac{1}{40L^2}.$$

Pour déduire une majoration de $\mathbb{E}[\|r_N - r\|_{\mu_N}^2]$ à partir de la majoration ci-dessus, on a besoin de la décomposition suivante :

$$\begin{aligned} \|r_N - r\|_{\mu_N}^2 &= (\|r_N - r\|_{\mu_N} - 2\|r_N - r\|_\mu + 2\|r_N - r\|_\mu)^2 \\ &\leq (\max\{\|r_N - r\|_{\mu_N} - 2\|r_N - r\|_\mu, 0\} + 2\|r_N - r\|_\mu)^2 \\ &\leq 2(\max\{\|r_N - r\|_{\mu_N} - 2\|r_N - r\|_\mu, 0\})^2 + 8\|r_N - r\|_\mu^2 \\ &\leq T_{1,N} + 8T_N. \end{aligned}$$

Soit $u > \frac{576L^2}{N}$. Vu que $\mathcal{G}^A \subset T_L \mathcal{G}$, par le lemme 4.5.3, on obtient :

$$\begin{aligned} \mathbb{P}(T_{1,N} > u) &= \mathbb{P}(2(\max\{\|r_N - r\|_{\mu_N} - 2\|r_N - r\|_\mu, 0\})^2 > u) \\ &\leq \mathbb{P}(\exists f \in T_L \mathcal{G} : \|f - r\|_{\mu_N} - 2\|f - r\|_\mu > \sqrt{u/2}) \\ &\leq 3\mathbb{E} \left[\mathcal{N}_2 \left(\frac{\sqrt{u}}{24}, T_L \mathcal{G}, Z_1^{2N} \right) \right] \exp\left(-\frac{N \cdot u}{576(2L)^2}\right) \\ &\leq 3\mathbb{E} \left[\mathcal{N}_2 \left(\frac{L}{\sqrt{N}}, T_L \mathcal{G}, Z_1^{2N} \right) \right] \exp\left(-\frac{N \cdot u}{576(2L)^2}\right). \end{aligned}$$

En utilisant le théorème 4.5.1 et $V_{T_L \mathcal{G}^+} \leq V_{\mathcal{G}^+}$, on obtient

$$\mathbb{E} \left[\mathcal{N}_2 \left(\frac{L}{\sqrt{N}}, T_L \mathcal{G}, Z_1^{2N} \right) \right] \leq 3 \left(\frac{3e(2L)^2}{(L/\sqrt{N})^2} \right)^{2(K_1+1)} = 3(12eN)^{2(K_1+1)}$$

pour $\frac{L}{\sqrt{N}} \leq 2L$ (ce qui est toujours vérifié).

Donc pour $u > \frac{576L^2}{N}$,

$$\mathbb{P}(T_{1,N} > u) \leq 9(12eN)^{2(K_1+1)} \exp \left(-\frac{Nu}{2304L^2} \right).$$

Ainsi, pour $v > \frac{576L^2}{N}$,

$$\begin{aligned} \mathbb{E}[T_{1,N}] &\leq v + \int_v^\infty \mathbb{P}(T_{1,N} > t) dt \\ &\leq v + 9(12eN)^{2(K_1+1)} \int_v^\infty \exp \left(-\frac{N.t}{2304L^2} \right) dt \\ &\leq v + 9(12eN)^{2(K_1+1)} \frac{2304L^2}{N} \exp \left(-\frac{Nv}{2304L^2} \right). \end{aligned}$$

En prenant

$$v = \frac{2304L^2}{N} \log(9.(12eN)^{2(K_1+1)}) > \frac{576L^2}{N},$$

on aboutit à :

$$\mathbb{E}[T_{1,N}] \leq \frac{b_1}{N} + \frac{(b_2 + b_3 \log(N))(K_1 + 1)}{N}.$$

Finalement :

$$\begin{aligned} \mathbb{E} [\|\tilde{m}_N - \hat{m}_N\|_{\mu_N}^2] &\leq \mathbb{E} [\|r_N - r\|_{\mu_N}^2] \\ &\leq T_{1,N} + T_N \\ &\leq \frac{b_1}{N} + \frac{(b_2 + b_3 \log(N))(K_1 + 1)}{N} \\ &\quad + 8 \left\{ \frac{a_1}{N} + \frac{(a_2 + a_3 \log(N))(K_1 + 1)}{N} \right\} \end{aligned}$$

Par conséquent :

$$\mathbb{E} [\|\tilde{m}_N - \hat{m}_N\|_{\mu_N}^2] \leq \frac{c_1}{N} + \frac{(c_2 + c_3 \log(N))(K_1 + 1)}{N}. \quad (4.4.3)$$

Etape 2 : Majoration de $\mathbb{E}[\|\hat{m}_N - m\|_{\mu_N}^2]$.

Pour simplifier, on introduit la notation $\mathbb{E}^*[\cdot] = \mathbb{E}[\cdot | X_1, \dots, X_N]$ et $P(Z) = (P_1(Z), \dots, P_{K_1}(Z))$.

D'abord on va montrer que $\|\mathbb{E}^*[\hat{m}_N] - m\|_{\mu_N}^2 = \inf_{\Phi \in \mathcal{F}_N} \|\Phi - m\|_{\mu_N}^2$.

Pour cela, considérons une base orthonormale complète (f_1, \dots, f_K) dans \mathcal{F}_N par rapport au produit scalaire empirique $\langle \cdot, \cdot \rangle_{\mu_N}$, défini par :

$$\langle f, g \rangle_{\mu_N} = \frac{1}{N} \sum_{i=1}^N f(X_i)g(X_i).$$

Remarquons que $K \leq K_2$ (il peut y avoir des colinéarités entre les fonctions de base) et que les fonctions f_k sont toujours aléatoires car elles dépendent de X_1, \dots, X_N (même si les Φ_k ne dépendent pas des données). Nous voici ramenés à un problème de moindres carrés sur la base $(f_k)_{1 \leq k \leq K}$ orthonormale par rapport au produit scalaire empirique $\langle \cdot, \cdot \rangle_{\mu_N}$. D'où, pour tout x , on a :

$$\hat{m}_N(x) = f(x)^t \frac{1}{N} B^t (Y - A\hat{\alpha}),$$

avec $Y = (H_1, \dots, H_N)^t$, $A = (P_j(Z_i))_{1 \leq i \leq N, 1 \leq j \leq K_1}$, $\hat{\alpha} = (\hat{\alpha}_1, \dots, \hat{\alpha}_{K_1})^t$, $f(x) = (f_1(x), \dots, f_K(x))^t$ et $B = (f_j(X_i))_{1 \leq i \leq N, 1 \leq j \leq K}$, puisque :

$$\frac{1}{N} B^t B = (\langle f_j, f_k \rangle_{\mu_N})_{1 \leq j, k \leq K} = Id_K \text{ (voir paragraphe 4.2.3).}$$

En prenant l'espérance conditionnelle par rapport à X_1, \dots, X_N on obtient que :

$$\mathbb{E}^*[\hat{m}_N(x)] = f(x)^t \frac{1}{N} B^t (m(X_1), \dots, m(X_N))^t$$

car :

$$\begin{aligned} \mathbb{E}^* \left[\frac{1}{N} B^t (Y - A\hat{\alpha}) \right] &= \frac{1}{N} B^t \mathbb{E}^*[(Y - A\hat{\alpha})] \\ &= \frac{1}{N} B^t (\mathbb{E}^*[H_1], \dots, \mathbb{E}^*[H_N])^t \\ &\quad - \frac{1}{N} B^t (\mathbb{E}^*[\hat{\alpha}^t P(Z_1)], \dots, \mathbb{E}^*[\hat{\alpha}^t P(Z_N)])^t \\ &= \frac{1}{N} B^t (m(X_1), \dots, m(X_N))^t \text{ (car } \mathbb{E}^*[P(Z_i)] = 0). \end{aligned}$$

Par suite $\mathbb{E}^*[\hat{m}_N]$ est l'estimateur de moindres carrés dans \mathcal{F}_N sur les données $(X_1, m(X_1)), \dots, (X_N, m(X_N))$. D'où :

$$\|\mathbb{E}^*[\hat{m}_N] - m\|_{\mu_N}^2 = \inf_{\Phi \in \mathcal{F}_N} \|\Phi - m\|_{\mu_N}^2.$$

Remarquons que :

$$\begin{aligned}
\mathbb{E}^*[\|\hat{m}_N - m\|_{\mu_N}^2] &= \mathbb{E}^* \left[\frac{1}{N} \sum_{i=1}^N |\hat{m}_N(X_i) - m(X_i)|^2 \right] \\
&= \mathbb{E}^* \left[\frac{1}{N} \sum_{i=1}^N |\hat{m}_N(X_i) - \mathbb{E}^*[\hat{m}_N(X_i)]|^2 \right] \\
&\quad + \mathbb{E}^* \left[\frac{1}{N} \sum_{i=1}^N |\mathbb{E}^*[\hat{m}_N(X_i)] - m(X_i)|^2 \right] \\
&= \mathbb{E}^*[\|\hat{m}_N - \mathbb{E}^*[\hat{m}_N]\|_{\mu_N}^2] + \|\mathbb{E}^*[\hat{m}_N] - m\|_{\mu_N}^2
\end{aligned}$$

puisque :

$$\begin{aligned}
&\mathbb{E}^* \left[\frac{1}{N} \sum_{i=1}^N (\hat{m}_N(X_i) - \mathbb{E}^*[\hat{m}_N(X_i)])(\mathbb{E}^*[\hat{m}_N(X_i)] - m(X_i)) \right] \\
&= \frac{1}{N} \sum_{i=1}^N (\mathbb{E}^*[\hat{m}_N(X_i)] - m(X_i)) \mathbb{E}^*[\hat{m}_N(X_i) - \mathbb{E}^*[\hat{m}_N(X_i)]] = 0.
\end{aligned}$$

En conclusion on a :

$$\mathbb{E}^*[\|\hat{m}_N - m\|_{\mu_N}^2] = \mathbb{E}^*[\|\hat{m}_N - \mathbb{E}^*[\hat{m}_N]\|_{\mu_N}^2] + \inf_{\Phi \in \mathcal{F}_N} \|\Phi - m\|_{\mu_N}^2.$$

Et par suite :

$$\begin{aligned}
\mathbb{E}[\|\hat{m}_N - m\|_{\mu_N}^2] &= \mathbb{E} \left\{ \mathbb{E}^*[\|\hat{m}_N - m\|_{\mu_N}^2] \right\} \\
&= \mathbb{E} \left\{ \|\hat{m}_N - \mathbb{E}^*[\hat{m}_N]\|_{\mu_N}^2 \right\} + \mathbb{E} \left\{ \inf_{\Phi \in \mathcal{F}_N} \|\Phi - m\|_{\mu_N}^2 \right\} \quad (4.4.4)
\end{aligned}$$

Ainsi il nous reste à majorer $\mathbb{E}[\|\hat{m}_N - \mathbb{E}^*[\hat{m}_N]\|_{\mu_N}^2]$. On a

$$\begin{aligned}
&\mathbb{E}^* [|\hat{m}_N(x) - \mathbb{E}^*[\hat{m}_N(x)]|^2] \\
&= \mathbb{E}^* \left[\left| f(x)^t \frac{1}{N} B^t (Y - A\hat{\alpha} - \mathbb{E}^*[Y - A\hat{\alpha}]) \right|^2 \right] \\
&= \mathbb{E}^* \left[f(x)^t \frac{1}{N} B^t \right. \\
&\quad \left. ((H_i - P(Z_i)\hat{\alpha} - \mathbb{E}^*[H_i - P(Z_i)\hat{\alpha}])(H_j - P(Z_j)\hat{\alpha} - \mathbb{E}^*[H_j - P(Z_j)\hat{\alpha}]))_{1 \leq i, j \leq N} \right. \\
&\quad \left. \frac{1}{N} B f(x) \right] \\
&= f(x)^t \frac{1}{N} B^t \\
&\quad (\mathbb{E}^* \{ (H_i - P(Z_i)\hat{\alpha} - \mathbb{E}^*[H_i - P(Z_i)\hat{\alpha}])(H_j - P(Z_j)\hat{\alpha} - \mathbb{E}^*[H_j - P(Z_j)\hat{\alpha}]) \})_{1 \leq i, j \leq N} \\
&\quad \frac{1}{N} B f(x)
\end{aligned}$$

où x est point (X_1, \dots, X_N) -mesurable de l'espace.

Notons $H^{\hat{\alpha}} = H - P(Z)\hat{\alpha}$ et $H_i^{\hat{\alpha}} = H_i - P(Z_i)\hat{\alpha}$. On a :

$$\mathbb{E}^* [(H_i^{\hat{\alpha}} - \mathbb{E}^*[H_i^{\hat{\alpha}}])(H_j^{\hat{\alpha}} - \mathbb{E}^*[H_j^{\hat{\alpha}}])] = \delta_{ij} \mathbf{Var}(H_i^{\hat{\alpha}}|X_i).$$

Ainsi,

$$\mathbb{E}^* [|\hat{m}_N(x) - \mathbb{E}^*[\hat{m}_N(x)]|^2] = \sum_{i=1}^N \mathbf{Var}(H_i^{\hat{\alpha}}|X_i) \left(\frac{\sum_{j=1}^K f_j(X_i) f_j(x)}{N} \right)^2.$$

On déduit

$$\begin{aligned} \mathbb{E}^* [\|\hat{m}_N - \mathbb{E}^*[\hat{m}_N]\|_{\mu_N}^2] &= \frac{1}{N} \sum_{l=1}^N \left\{ \sum_{i=1}^N \mathbf{Var}(H_i^{\hat{\alpha}}|X_i) \left(\frac{\sum_{j=1}^K f_j(X_i) f_j(X_l)}{N} \right)^2 \right\} \\ &= \frac{1}{N} \sum_{i=1}^N \mathbf{Var}(H_i^{\hat{\alpha}}|X_i) \sum_{l=1}^N \left(\frac{\sum_{j=1}^K f_j(X_i) f_j(X_l)}{N} \right)^2 \\ &= \frac{1}{N^3} \sum_{i=1}^N \mathbf{Var}(H_i^{\hat{\alpha}}|X_i) \sum_{l=1}^N \left(\sum_{j_1, j_2=1}^K f_{j_1}(X_i) f_{j_1}(X_l) f_{j_2}(X_i) f_{j_2}(X_l) \right) \\ &= \frac{1}{N^3} \sum_{i=1}^N \mathbf{Var}(H_i^{\hat{\alpha}}|X_i) \left(\sum_{j_1, j_2=1}^K f_{j_1}(X_i) f_{j_2}(X_i) \sum_{l=1}^N f_{j_1}(X_l) f_{j_2}(X_l) \right). \end{aligned}$$

Puisque $(f_k)_k$ est une base orthonormale, on a

$$\sum_{l=1}^N f_{j_1}(X_l) f_{j_2}(X_l) = \delta_{j_1 j_2} N.$$

Et donc,

$$\begin{aligned} \mathbb{E}^* [\|\hat{m}_N - \mathbb{E}^*[\hat{m}_N]\|_{\mu_N}^2] &= \frac{1}{N^3} \sum_{i=1}^N \mathbf{Var}(H_i^{\hat{\alpha}}|X_i) \left(\sum_{j=1}^K (f_j(X_i))^2 N \right) \\ &= \frac{1}{N^2} \sum_{i=1}^N \mathbf{Var}(H_i^{\hat{\alpha}}|X_i) \sum_{j=1}^K (f_j(X_i))^2. \end{aligned}$$

Ecrivant $\mathbb{E} [\|\hat{m}_N - \mathbb{E}^*[\hat{m}_N]\|_{\mu_N}^2] = \mathbb{E} \{ \mathbb{E}^* [\|\hat{m}_N - \mathbb{E}^*[\hat{m}_N]\|_{\mu_N}^2] \}$, on a démontré

$$\mathbb{E} [\|\hat{m}_N - \mathbb{E}^*[\hat{m}_N]\|_{\mu_N}^2] = \frac{1}{N^2} \sum_{i=1}^N \mathbb{E} \left[\mathbf{Var}(H_i^{\hat{\alpha}}|X_i) \sum_{j=1}^K (f_j(X_i))^2 \right]. \quad (4.4.5)$$

La majoration finale de (4.4.5) dépend des fonctions de base $(\Phi_k)_{1 \leq k \leq K_2}$ (et donc $(f_j)_{1 \leq j \leq K}$). On va maintenant utiliser ou non des hypothèses supplémentaires sur les fonctions de base $(\Phi_k)_{1 \leq k \leq K_2}$ (donc $(f_j)_{1 \leq j \leq K}$) pour majorer précisément (4.4.5).

- Dans le cas général, posons $\Sigma^2(\hat{\alpha}) = \sup_{x \in \mathbb{R}^{d_2}} \mathbf{Var}\{H^{\hat{\alpha}}|X = x\}$ ($\Sigma^2(\hat{\alpha})$ est toujours fini car H est borné) alors :

$$\begin{aligned}
 \mathbb{E} [\|\hat{m}_N - \mathbb{E}^*[\hat{m}_N]\|_{\mu_N}^2] &\leq \frac{1}{N^2} \sum_{i=1}^N \mathbb{E} \left[\Sigma^2(\hat{\alpha}) \sum_{j=1}^K (f_j(X_i))^2 \right] \\
 &= \Sigma^2(\hat{\alpha}) \frac{1}{N^2} \mathbb{E} \left[\sum_{j=1}^K \sum_{i=1}^N (f_j(X_i))^2 \right] \\
 &= \Sigma^2(\hat{\alpha}) \frac{1}{N^2} \mathbb{E} \left[\sum_{j=1}^K N \right] = \frac{\Sigma^2(\hat{\alpha})}{N} \mathbb{E}[K] \\
 &\leq \frac{K_2}{N} \Sigma^2(\hat{\alpha}).
 \end{aligned}$$

Ainsi,

$$\mathbb{E} [\|\hat{m}_N - \mathbb{E}^*[\hat{m}_N]\|_{\mu_N}^2] \leq \frac{K_2}{N} \Sigma^2(\hat{\alpha}). \quad (4.4.6)$$

La combinaison des résultats (4.4.2), (4.4.3), (4.4.4) et (4.4.6) implique le second résultat du théorème.

- Dans le cas particulier où $\Phi_k(x) = \mathbf{I}_{\{x \in I_k\}}$ avec $(I_k)_{1 \leq k \leq K_2}$ étant une partition contenant $\{X_1, \dots, X_N\}$ avec des pavés contenant le même nombre de données c'est à dire $\frac{N}{K_2}$, alors :

$$K = K_2 \quad \text{et} \quad f_k(x) = \sqrt{N} \frac{\mathbf{I}_{\{x \in I_k\}}}{\sqrt{\sum_{i=1}^N \mathbf{I}_{\{X_i \in I_k\}}}} = \sqrt{K} \mathbf{I}_{\{x \in I_k\}}.$$

A partir de la formule (4.4.5) et par symétrie des (X_1, \dots, X_N) , on déduit :

$$\begin{aligned}
 \mathbb{E} [\|\hat{m}_N - \mathbb{E}^*[\hat{m}_N]\|_{\mu_N}^2] &= \frac{1}{N} \mathbb{E} \left[\mathbf{Var}(H_1^{\hat{\alpha}}|X_1) \sum_{j=1}^K (f_j(X_1))^2 \right] \quad (4.4.7) \\
 &= \frac{1}{N} \mathbb{E} \left[\mathbf{Var}(H_1^{\hat{\alpha}}|X_1) \sum_{j=1}^{K_2} K_2 \mathbf{I}_{\{X_1 \in I_j\}} \right] \\
 &= \frac{K_2}{N} \mathbb{E} [\mathbf{Var}(H_1^{\hat{\alpha}}|X_1)] \\
 &= \frac{K_2}{N} \inf_{\alpha \in \mathbb{R}^{K_1}} \mathbb{E} [\mathbf{Var}(H^\alpha|X)].
 \end{aligned}$$

L'équation ci-dessus combinée avec (4.4.2), (4.4.3) et (4.4.4) implique le deuxième résultat du théorème pour le cas considéré.

- Considérons enfin le cas où $\Phi_k(x) = \mathbf{I}_{\{x \in I_k\}}$ avec $(I_k)_{1 \leq k \leq K_2}$ étant une partition contenant $\{X_1, \dots, X_N\}$ avec des pavés équiprobables c'est à dire : $\forall 1 \leq$

$k \leq K_2 : \mathbb{P}(X \in I_k) = \frac{1}{K_2}$. Concernant l'orthogonalisation des $(\Phi_k)_{1 \leq k \leq K_2}$, seulement $K \leq K_2$ intervalles contiennent au moins une donnée. On note ces intervalles $I_{\sigma(1)}, \dots, I_{\sigma(K)}$ et les autres restant sont notés $I_{\sigma(K+1)}, \dots, I_{\sigma(K_2)}$. Ainsi pour $1 \leq k \leq K$,

$$f_k(x) = \sqrt{N} \frac{\mathbf{I}_{\{x \in I_{\sigma(k)}\}}}{\sqrt{\sum_{i=1}^N \mathbf{I}_{\{X_i \in I_{\sigma(k)}\}}}.$$

On pose :

$$f_k = 0 \text{ pour } K + 1 \leq k \leq K_2.$$

En prenant la convention $\frac{0}{0} = 0$, alors, pour $1 \leq k \leq K_2$

$$f_k(x) = \sqrt{N} \frac{\mathbf{I}_{\{x \in I_{\sigma(k)}\}}}{\sqrt{\sum_{i=1}^N \mathbf{I}_{\{X_i \in I_{\sigma(k)}\}}}.$$

D'après (4.4.7)

$$\begin{aligned} & \mathbb{E} [\|\hat{m}_N - \mathbb{E}^*[\hat{m}_N]\|_{\mu_N}^2] \\ &= \frac{1}{N} \mathbb{E} \left[\mathbf{Var}(H_1^{\hat{\alpha}} | X_1) \sum_{j=1}^{K_2} (f_j(X_1))^2 \right] \\ &= \frac{1}{N} \mathbb{E} \left[\mathbf{Var}(H_1^{\hat{\alpha}} | X_1) \sum_{j=1}^{K_2} N \frac{\mathbf{I}_{\{X_1 \in I_{\sigma(j)}\}}}{\sum_{i=1}^N \mathbf{I}_{\{X_i \in I_{\sigma(j)}\}}} \right] \\ &= \mathbb{E} \left[\mathbf{Var}(H_1^{\hat{\alpha}} | X_1) \sum_{j=1}^{K_2} \frac{\mathbf{I}_{\{X_1 \in I_j\}}}{\sum_{i=1}^N \mathbf{I}_{\{X_i \in I_j\}}} \right] \\ &= \sum_{j=1}^{K_2} \mathbb{E} \left[\mathbf{Var}(H_1^{\hat{\alpha}} | X_1) \frac{\mathbf{I}_{\{X_1 \in I_j\}}}{\sum_{i=1}^N \mathbf{I}_{\{X_i \in I_j\}}} \right] \\ &= \sum_{j=1}^{K_2} \mathbb{E} \left\{ \mathbf{Var}(H_1^{\hat{\alpha}} | X_1) \mathbf{I}_{\{X_1 \in I_j\}} \mathbb{E} \left[\frac{1}{1 + \sum_{i=2}^N \mathbf{I}_{\{X_i \in I_j\}}} \middle| X_1 \right] \right\} \\ &= \sum_{j=1}^{K_2} \mathbb{E} \left\{ \mathbf{Var}(H_1^{\hat{\alpha}} | X_1) \mathbf{I}_{\{X_1 \in I_j\}} \mathbb{E} \left[\frac{1}{1 + \sum_{i=2}^N \mathbf{I}_{\{X_i \in I_j\}}} \right] \right\}. \end{aligned}$$

On sait que $\sum_{i=2}^N \mathbf{I}_{\{X_i \in I_j\}}$ est une variable aléatoire de loi binomiale $\mathcal{B}(N - 1, \frac{1}{K_2})$ car $\mathbb{P}(X \in I_j) = \frac{1}{K_2}$. Donc, d'après le lemme 4.1 de [GKKW02] on a :

$$\mathbb{E} \left[\frac{1}{1 + \sum_{i=2}^N \mathbf{I}_{\{X_i \in I_j\}}} \right] \leq \frac{K_2}{N}.$$

Par suite :

$$\begin{aligned} \mathbb{E} [\|\hat{m}_N - \mathbb{E}^*[\hat{m}_N]\|_{\mu_N}^2] &\leq \frac{K_2}{N} \sum_{j=1}^{K_2} \mathbb{E} [\mathbf{Var}(H_1^{\hat{\alpha}}|X_1)\mathbf{I}_{\{X_1 \in I_j\}}] \\ &= \frac{K_2}{N} \mathbb{E} [\mathbf{Var}(H_1^{\hat{\alpha}}|X_1)] \\ &= \frac{K_2}{N} \inf_{\alpha \in \mathbb{R}^{K_1}} \mathbb{E} [\mathbf{Var}(H^\alpha|X)]. \end{aligned}$$

Ainsi, il en découle pour ce cas le premier résultat du théorème en utilisant (4.4.2), (4.4.3) et (4.4.4). \square

4.5 Résultats complémentaires en régression non paramétrique

Pour plus de détail dans cette section, on réfère le lecteur au chapitre 9 dans [GKKW02].

4.5.1 Nombres de couverture

Définition 4.5.1 Soient $\epsilon > 0$, \mathcal{G} une classe de fonctions $\mathbb{R}^d \rightarrow \mathbb{R}$, $1 \leq p < \infty$ et μ une mesure de probabilité sur \mathbb{R}^d . Pour une fonction $g : \mathbb{R}^d \rightarrow \mathbb{R}$, on définit

$$\|g\|_{L_p(\mu)} := \left(\int |g(z)|^p d\mu \right)^{\frac{1}{p}}.$$

1. Toute famille finie $g_1, \dots, g_N : \mathbb{R}^d \rightarrow \mathbb{R}$ telle que pour toute fonction $g \in \mathcal{G}$, il existe $j = j(g) \in \{1, \dots, N\}$ tel que

$$\|g - g_j\|_{L_p(\mu)} < \epsilon$$

est appelée une ϵ -couverture de \mathcal{G} par rapport à $\|\cdot\|_{L_p(\mu)}$.

2. On note $\mathcal{N}(\epsilon, \mathcal{G}, \|\cdot\|_{L_p(\mu)})$ la taille de la plus petite ϵ -couverture de \mathcal{G} par rapport à $\|\cdot\|_{L_p(\mu)}$. On prend $\mathcal{N}(\epsilon, \mathcal{G}, \|\cdot\|_{L_p(\mu)}) = +\infty$ s'il n'existe de ϵ -couverture. Alors $\mathcal{N}(\epsilon, \mathcal{G}, \|\cdot\|_{L_p(\mu)})$ est appelé nombre de ϵ -couverture de \mathcal{G} par rapport à $\|\cdot\|_{L_p(\mu)}$.
3. Soit $z_1^M = (z_1, \dots, z_M)$ M points fixes de \mathbb{R}^d . Soit μ_M la mesure empirique correspondante, i.e.

$$\mu_M(A) = \frac{1}{M} \sum_{m=1}^M \mathbf{1}_A(z_m), (A \subset \mathbb{R}^d).$$

Alors

$$\|g\|_{L_p(\mu_M)} := \left(\frac{1}{M} \sum_{m=1}^M |g(z_m)|^p d\mu \right)^{\frac{1}{p}},$$

et toute ϵ -couverture de \mathcal{G} par rapport à $\|\cdot\|_{L_p(\mu_M)}$ sera appelée une L_p ϵ -couverture de \mathcal{G} sur z_1^M et le nombre de ϵ -couverture de \mathcal{G} par rapport à $\|\cdot\|_{L_p(\mu_M)}$ sera noté

$$\mathcal{N}_p(\epsilon, \mathcal{G}, z_1^M).$$

4.5.2 Coefficients d'éclatement et dimension VC

Définition 4.5.2 Soit \mathcal{A} une classe de sous-ensembles de \mathbb{R}^d et soit $M \in \mathbb{N}$.

1. Pour $z_1, \dots, z_M \in \mathbb{R}^d$, on définit

$$s(\mathcal{A}, \{z_1, \dots, z_M\}) = |\{A \cap \{z_1, \dots, z_M\} : A \in \mathcal{A}\}|,$$

c'est-à-dire que $s(\mathcal{A}, \{z_1, \dots, z_M\})$ est le nombre de sous-ensembles distincts de $\{z_1, \dots, z_M\}$ en intersectant avec des éléments de \mathcal{A} .

2. Soit G un sous-ensemble de \mathbb{R}^d de cardinal M . On dit que \mathcal{A} fragmente G si $s(\mathcal{A}, G) = 2^M$, i.e. si tout sous-ensemble de G peut être représenté sous la forme $A \cap G$ pour un élément $A \in \mathcal{A}$.

3. Le M -ème coefficient d'éclatement $\mathcal{S}(\mathcal{A}, M)$ de \mathcal{A} est défini comme

$$\mathcal{S}(\mathcal{A}, M) = \max_{\{z_1, \dots, z_M\} \subset \mathbb{R}^d} s(\mathcal{A}, \{z_1, \dots, z_M\}).$$

Ainsi, le M -ème coefficient d'éclatement est le nombre maximal de sous-ensembles différents d'un ensemble de M points qui peuvent être obtenus par intersection avec les éléments de \mathcal{A} .

Définition 4.5.3 Soit \mathcal{A} une classe de sous-ensembles de \mathbb{R}^d avec $\mathcal{A} \neq \emptyset$. La dimension VC (ou dimension de Vapnik-Chervonenkis) $V_{\mathcal{A}}$ de \mathcal{A} est définie par

$$V_{\mathcal{A}} = \sup\{M \in \mathbb{N} : \mathcal{S}(\mathcal{A}, M) = 2^M\},$$

i.e. $V_{\mathcal{A}}$ est le plus grand entier M tel qu'il existe un sous-ensemble de M points qui peut être éclaté par \mathcal{A} .

Définition 4.5.4 Soit \mathcal{G} une classe de fonctions $\mathbb{R}^d \rightarrow \mathbb{R}$. On note \mathcal{G}^+ le sous-ensemble de \mathbb{R}^{d+1} défini par :

$$\mathcal{G}^+ = \{(z, t) \in \mathbb{R}^d \times \mathbb{R} ; t \leq g(z), g \in \mathcal{G}\}$$

soit l'ensemble des sous-graphes de fonctions de \mathcal{G} .

Théorème 4.5.1 (voir théorème 9.4 dans [GKKW02]) Soit \mathcal{G} une classe de fonctions $g : \mathbb{R}^d \rightarrow [0, L]$ avec $2 \leq V_{\mathcal{G}^+} < \infty$, soit $p \geq 1$, soit μ une mesure de probabilité sur \mathbb{R}^d , et soit $0 < \epsilon < \frac{L}{4}$. Alors

$$\mathcal{N}(\epsilon, \mathcal{G}, \|\cdot\|_{L_p(\mu)}) \leq 3 \left(\frac{2eL^p}{\epsilon^p} \log \left(\frac{3eL^p}{\epsilon^p} \right) \right)^{V_{\mathcal{G}^+}}.$$

Théorème 4.5.2 (voir théorème 9.5 dans [GKKW02]) Soit \mathcal{G} un espace vectoriel de dimension r de fonction de $\mathbb{R}^d \rightarrow \mathbb{R}$ et soit

$$\mathcal{A} = \{z; g(z) \geq 0\} : g \in \mathcal{G}.$$

Alors

$$V_{\mathcal{A}} \leq r.$$

Lemme 4.5.3 Soit \mathcal{F} une classe de fonctions $f : \mathbb{R}^d \rightarrow \mathbb{R}$ bornées par L et $Z_1^{2N} = (Z_{N+1}, \dots, Z_{2N})$ un échantillon de variables indépendantes de même loi que $Z_1^N = (Z_1, \dots, Z_N)$. Notons $Z_1^{2N} = (Z_1, \dots, Z_N, Z_{N+1}, \dots, Z_{2N})$. Soit $\epsilon > 0$, alors

$$\mathbb{P}(\exists f \in \mathcal{F} : \|f\|_{\mu_N} - 2\|f\|_{\mu} > \epsilon) \leq 3\mathbb{E} \left[\mathcal{N}_2 \left(\frac{\sqrt{2}}{24}\epsilon, \mathcal{F}, Z_1^{2N} \right) \right] \exp \left(-\frac{N\epsilon^2}{288L^2} \right).$$

Preuve du lemme : Nous suivons pas à pas la preuve du théorème 11.2 dans [GKKW02]. L'étape 1 est différente, nous la détaillons.

Définissons :

$$\|f\|_{\mu_N}'^2 = \frac{1}{N} \sum_{i=N+1}^{2N} |f(Z_i)|^2.$$

Soit f^* une fonction $f \in \mathcal{F}$ telle que

$$\|f\|_{\mu_N} - 2\|f\|_{\mu} > \epsilon$$

si une telle fonction existe, sinon on prend pour f^* une fonction arbitraire dans \mathcal{F} .

Donc :

$$\begin{aligned} \mathbb{P} \left(2\|f^*\|_{\mu} + \frac{\epsilon}{2} > \|f^*\|_{\mu_N}' \mid Z_1^N \right) &\geq \mathbb{P} \left(4\|f^*\|_{\mu}^2 + \frac{\epsilon^2}{4} > \|f^*\|_{\mu_N}'^2 \mid Z_1^N \right) \\ &= 1 - \mathbb{P} \left(4\|f^*\|_{\mu}^2 + \frac{\epsilon^2}{4} \leq \|f^*\|_{\mu_N}'^2 \mid Z_1^N \right) \\ &= 1 - \mathbb{P} \left(3\|f^*\|_{\mu}^2 + \frac{\epsilon^2}{4} \leq \|f^*\|_{\mu_N}'^2 - \|f^*\|_{\mu}^2 \mid Z_1^N \right). \end{aligned}$$

L'application de l'inégalité de Chebyshev permet d'avoir :

$$\begin{aligned}
\mathbb{P}\left(3\|f^*\|_\mu^2 + \frac{\epsilon^2}{4} \leq \|f^*\|_{\mu_N}'^2 - \|f^*\|_\mu^2 | Z_1^N\right) &\leq \frac{\text{Var}\left(\frac{1}{N} \sum_{i=N+1}^{2N} |f^*(Z_i)|^2 | Z_1^N\right)}{\left(3\|f^*\|_\mu^2 + \frac{\epsilon^2}{4}\right)^2} \\
&\leq \frac{\frac{1}{N} L^2 \|f^*\|_\mu^2}{\left(3\|f^*\|_\mu^2 + \frac{\epsilon^2}{4}\right)^2} \\
&\leq \frac{\frac{L^2}{3N} \left(3\|f^*\|_\mu^2 + \frac{\epsilon^2}{4}\right)}{\left(3\|f^*\|_\mu^2 + \frac{\epsilon^2}{4}\right)^2} \\
&\leq \frac{L^2}{3N} \frac{4}{\epsilon^2} = \frac{4L^2}{3\epsilon^2 N}.
\end{aligned}$$

Par suite, pour $N \geq \frac{4L^2}{\epsilon^2}$

$$\mathbb{P}\left(2\|f^*\|_\mu + \frac{\epsilon}{2} > \|f^*\|_{\mu_N}' | Z_1^N\right) \geq \frac{2}{3}.$$

Pour de tels N , on a

$$\begin{aligned}
&\mathbb{P}\left(\exists f \in \mathcal{F} : \|f\|_{\mu_N} - \|f\|_{\mu_N}' > \frac{\epsilon}{4}\right) \\
&\geq \mathbb{P}\left(\|f^*\|_{\mu_N} - \|f^*\|_{\mu_N}' > \frac{\epsilon}{4}\right) \\
&\geq \mathbb{P}\left(\|f^*\|_{\mu_N} + \frac{3\epsilon}{4} - \|f^*\|_{\mu_N}' > \epsilon, 2\|f^*\|_\mu + \frac{3\epsilon}{4} > \|f^*\|_{\mu_N}'\right) \\
&\geq \mathbb{P}\left(\|f^*\|_{\mu_N} - 2\|f^*\|_\mu > \epsilon, 2\|f^*\|_\mu + \frac{3\epsilon}{4} > \|f^*\|_{\mu_N}'\right) \\
&= \mathbb{E}\left(I_{\{\|f^*\|_{\mu_N} - 2\|f^*\|_\mu > \epsilon\}} \mathbb{P}(2\|f^*\|_\mu + \frac{\epsilon}{2} > \|f^*\|_{\mu_N}' | Z_1^N)\right) \\
&\geq \frac{2}{3} \mathbb{P}(\|f^*\|_{\mu_N} - 2\|f^*\|_\mu > \epsilon) = \frac{2}{3} \mathbb{P}(\exists f \in \mathcal{F} : \|f\|_{\mu_N} - 2\|f\|_\mu > \epsilon).
\end{aligned}$$

En conclusion pour $N \geq \frac{4L^2}{\epsilon^2}$, on a

$$\mathbb{P}(\exists f \in \mathcal{F} : \|f\|_{\mu_N} - 2\|f\|_\mu > \epsilon) \leq \frac{3}{2} \mathbb{P}\left(\exists f \in \mathcal{F} : \|f\|_{\mu_N} - \|f\|_{\mu_N}' > \frac{\epsilon}{4}\right).$$

Remarquons que pour $N < \frac{4L^2}{\epsilon^2}$, le lemme est trivial vu que l'expression de droite est bien supérieure à 1. Les étapes 2, 3 et 4 de la preuve du théorème 11.2 dans [GKKW02] aboutissent à :

$$\mathbb{P}\left(\exists f \in \mathcal{F} : \|f\|_{\mu_N} - \|f\|_{\mu_N}' > \frac{\epsilon}{4}\right) \leq 2\mathbb{E}\left[\mathcal{N}_2\left(\frac{\sqrt{2}}{24}\epsilon, \mathcal{F}, Z_1^{2N}\right)\right] \exp\left(-\frac{N\epsilon^2}{288L^2}\right).$$

Et ainsi notre lemme est démontré. \square

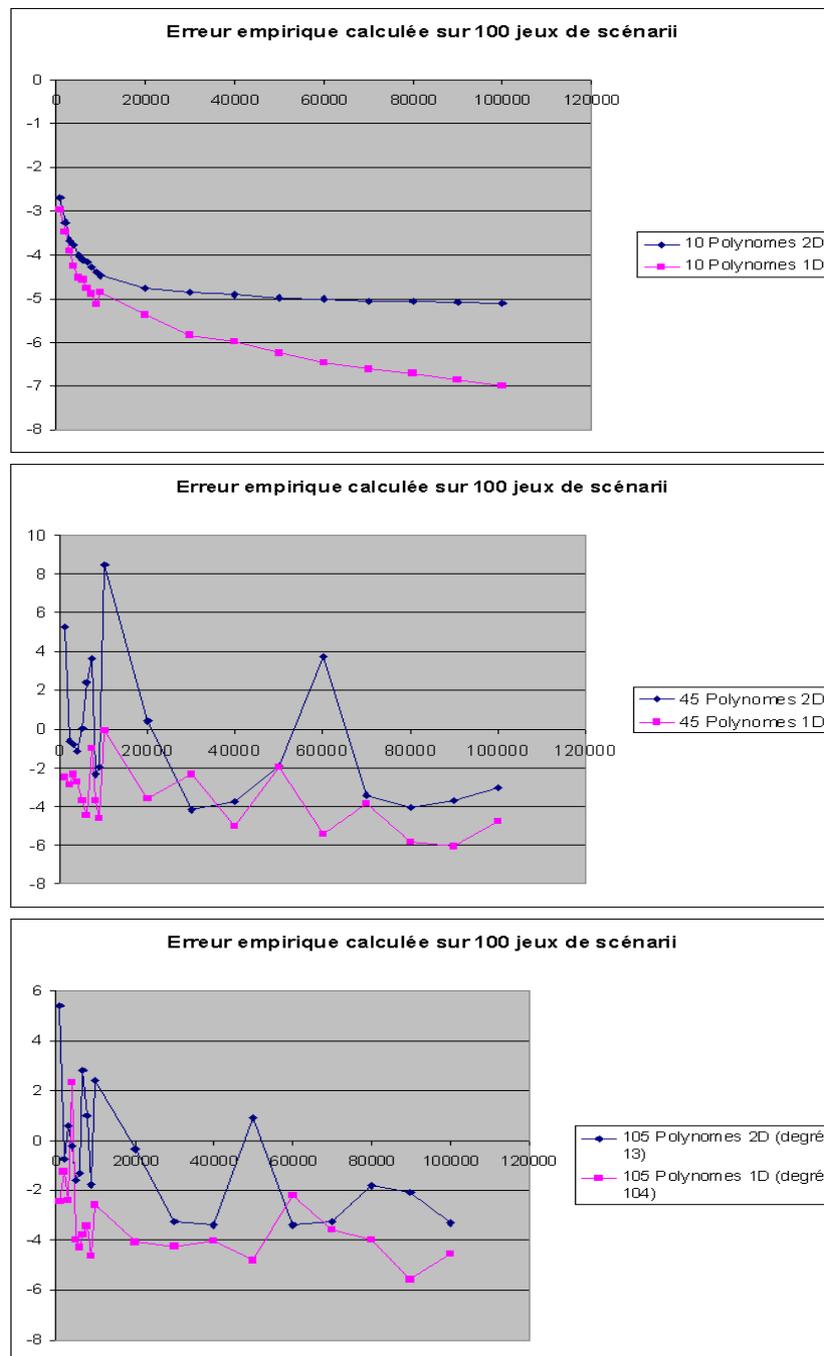


FIGURE 4.3 – Erreur empirique pour les polynômes.

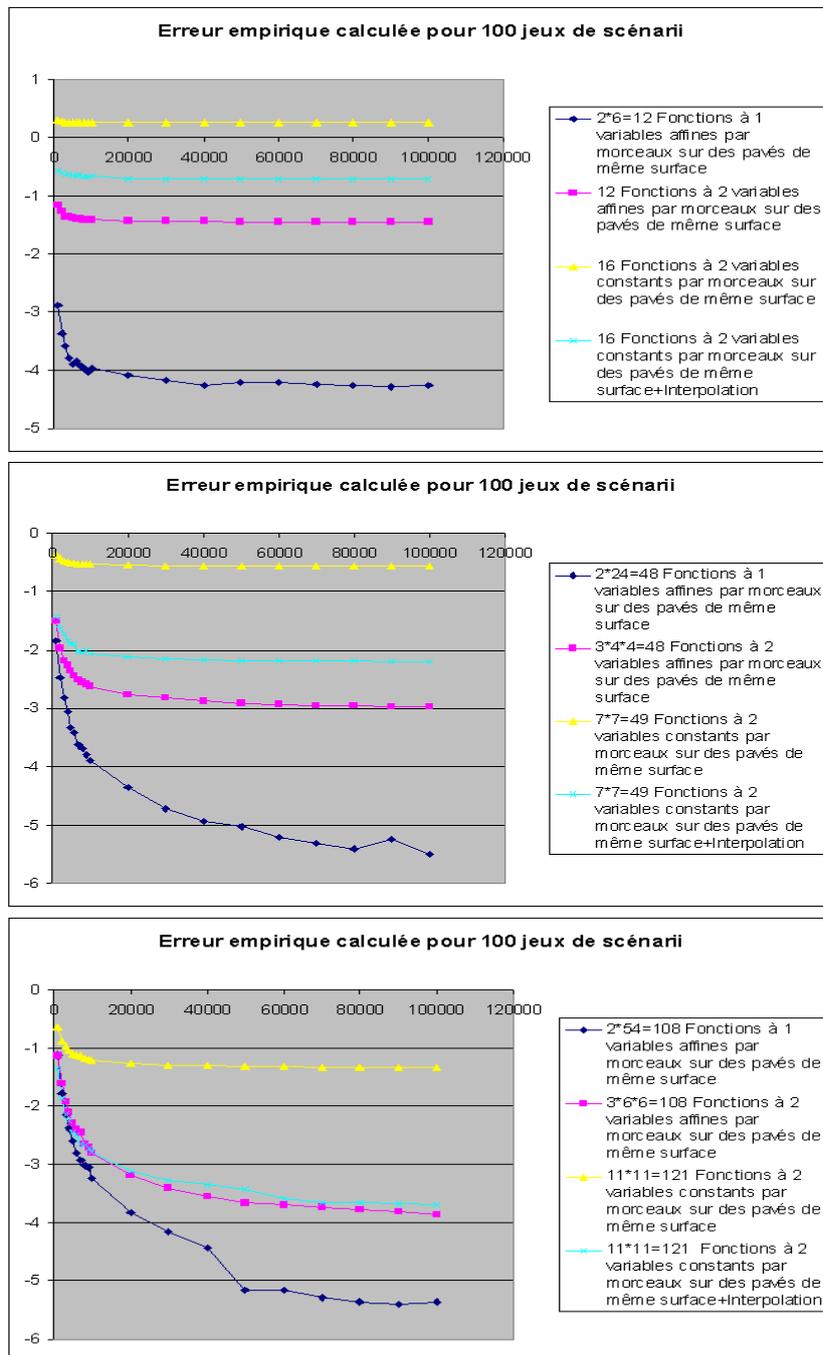


FIGURE 4.4 – Erreur empirique pour les fonctions locales par morceaux de même aire.

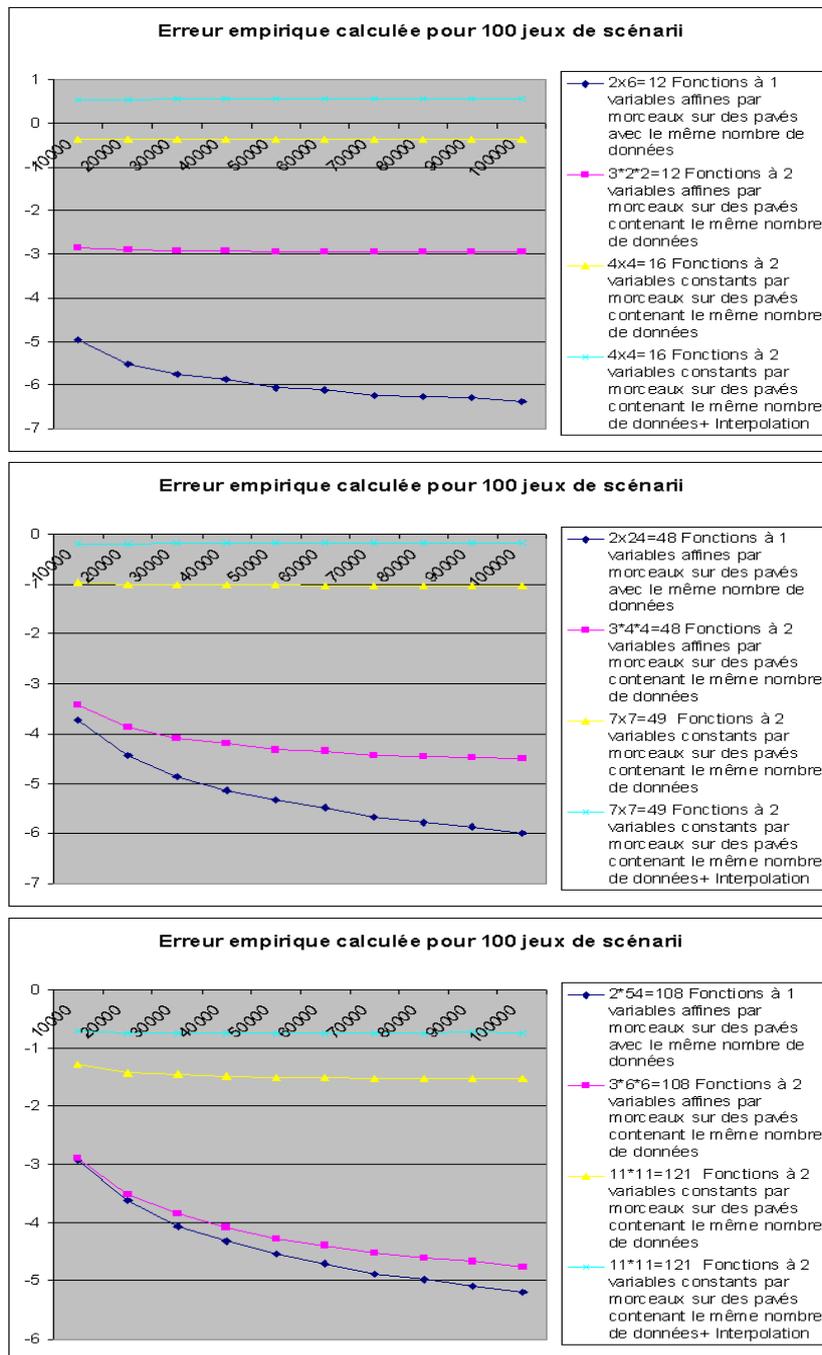


FIGURE 4.5 – Erreur empirique pour les fonctions locales par morceaux de même nombre de données.

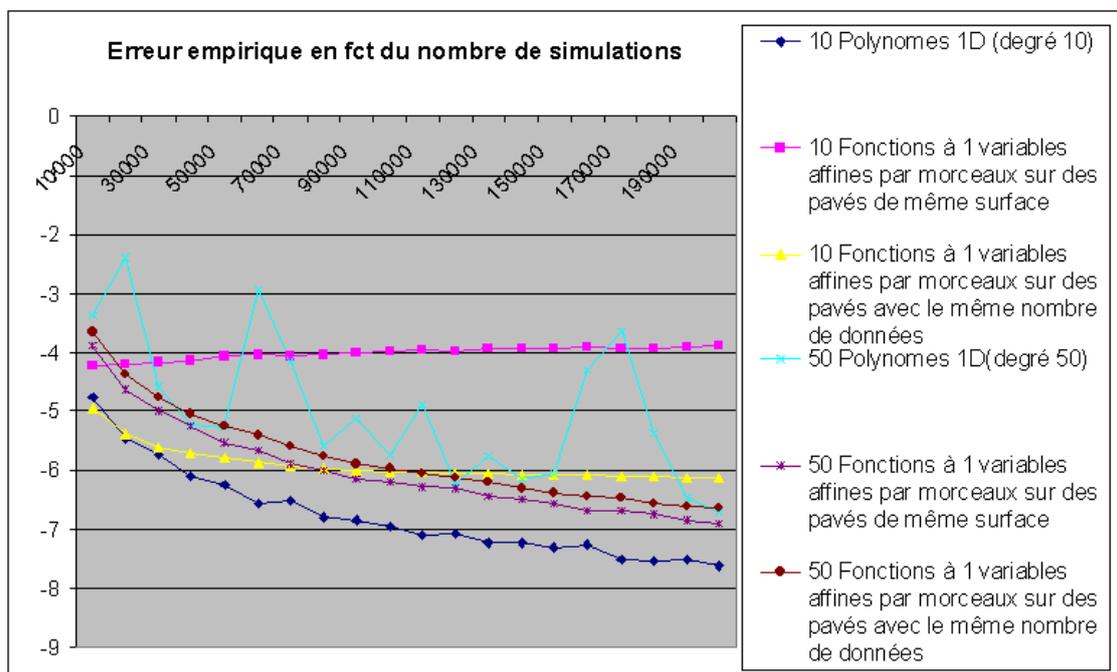


FIGURE 4.6 – Erreur empirique pour différentes fonctions de base.

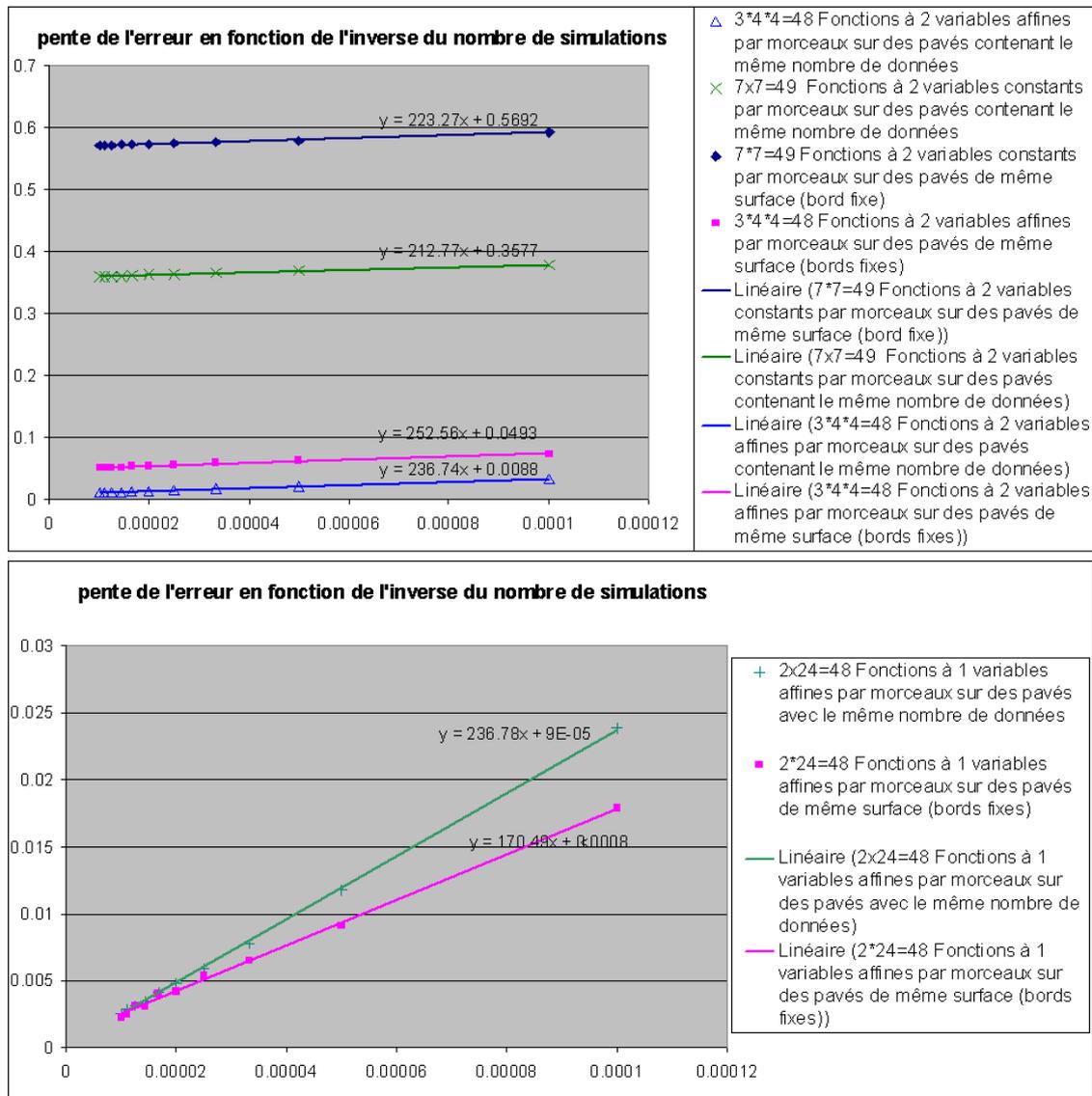


FIGURE 4.7 – Pente de l'erreur pour différentes fonctions de base.

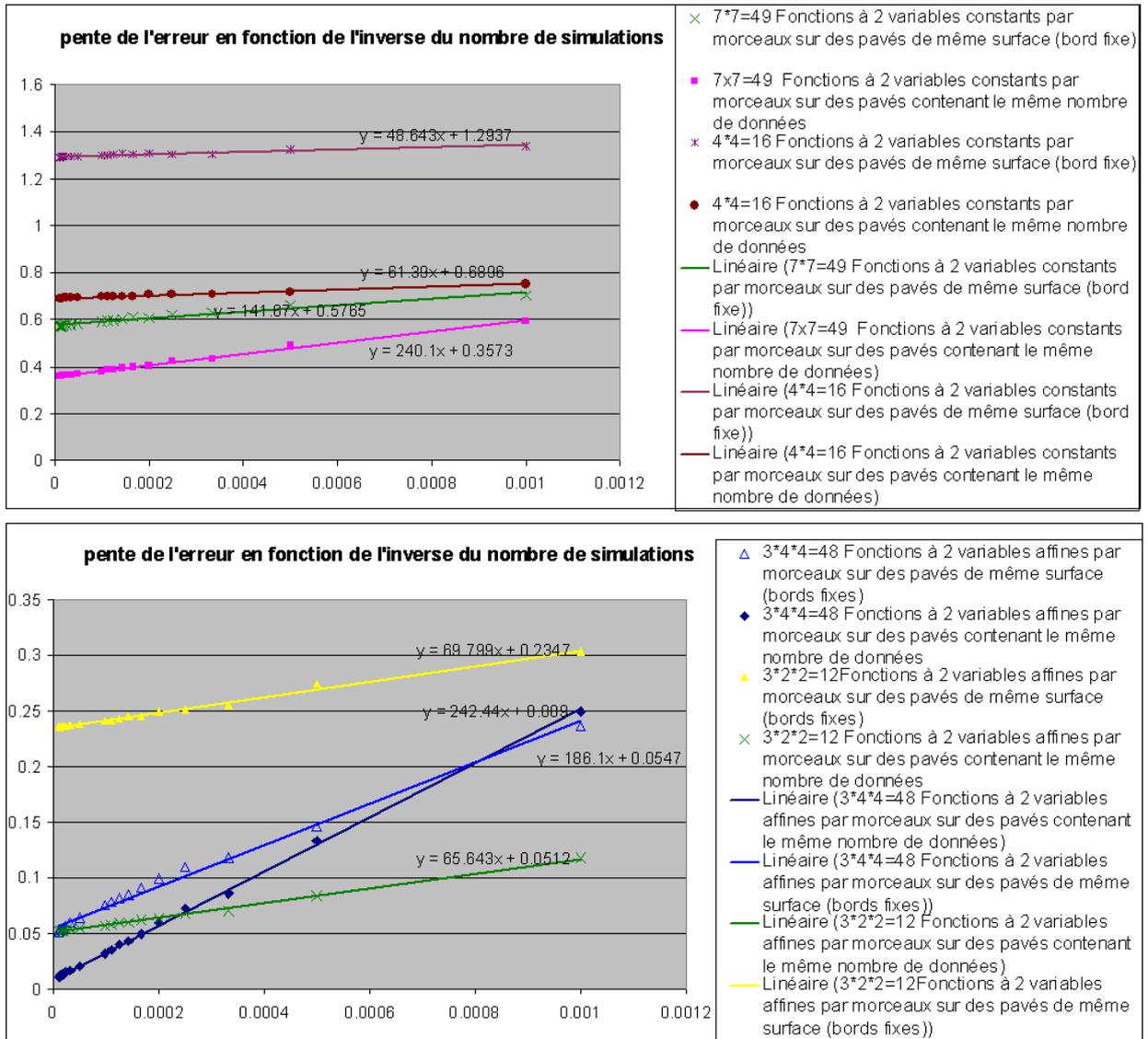


FIGURE 4.8 – Pente de l’erreur pour différentes fonctions de base et différents K_N .

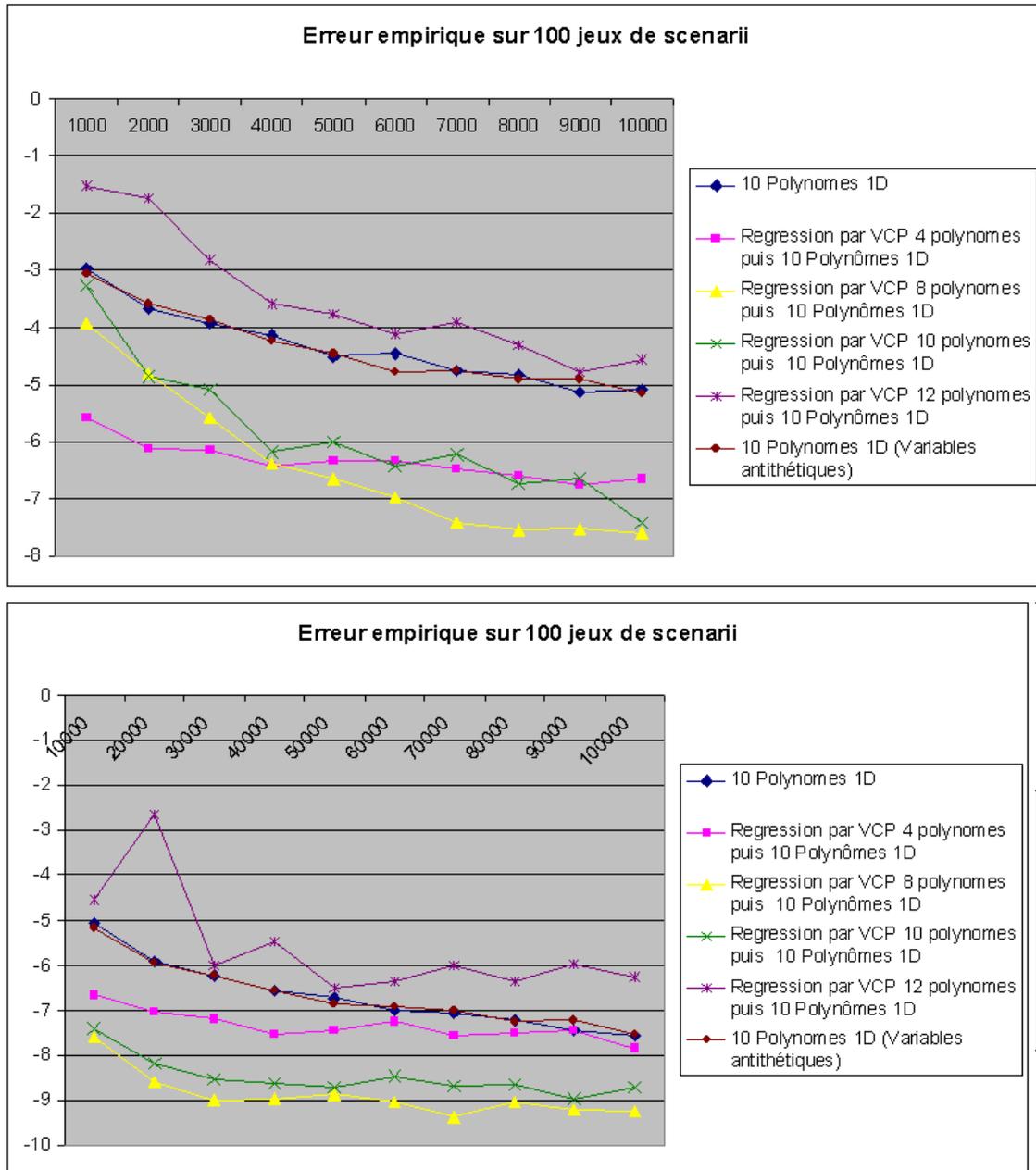


FIGURE 4.9 – Erreur empirique pour la régression par VCP pour des polynômes

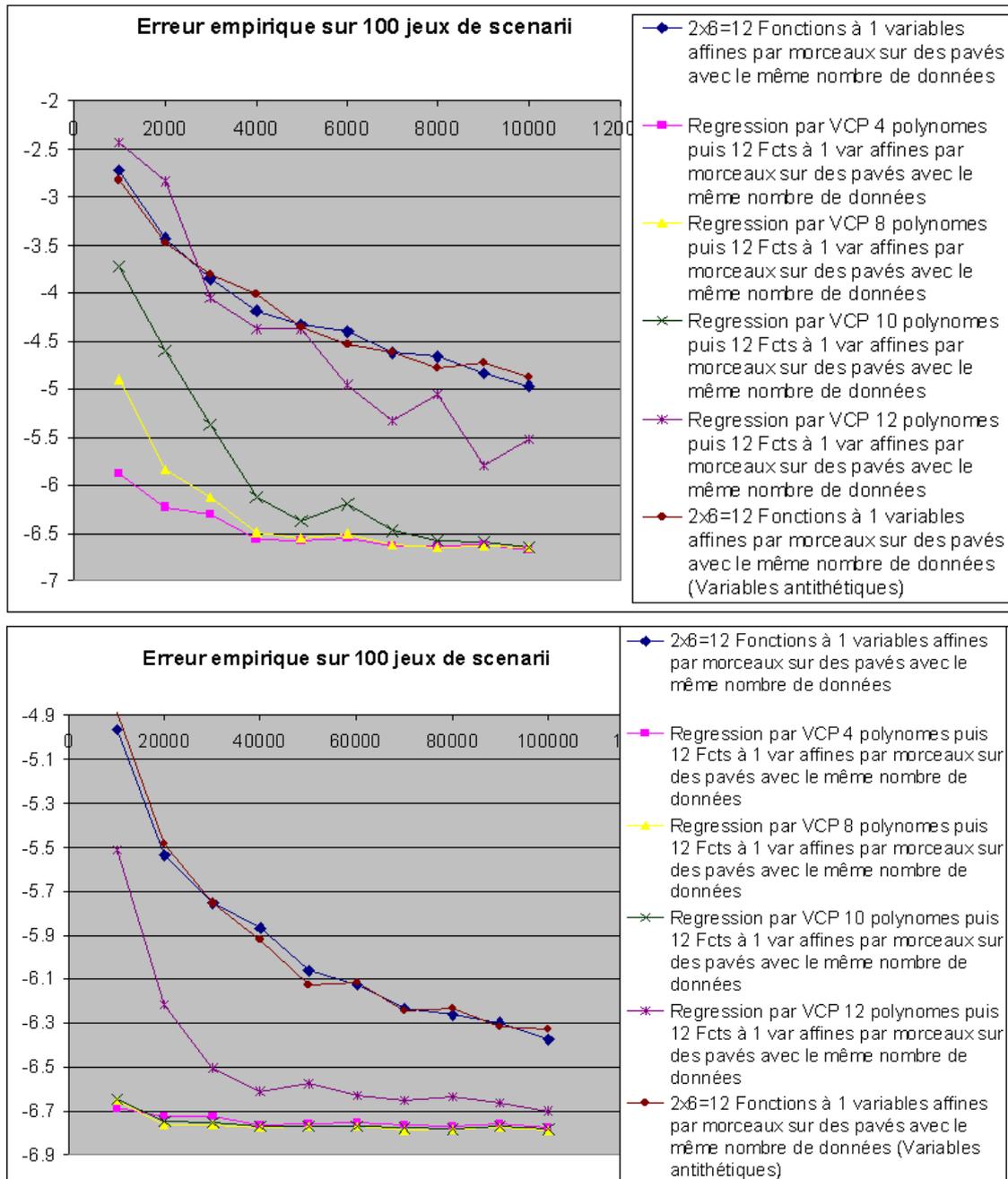


FIGURE 4.10 – Erreur empirique pour la régression par VCP pour le cas 1dAffineMemeData.

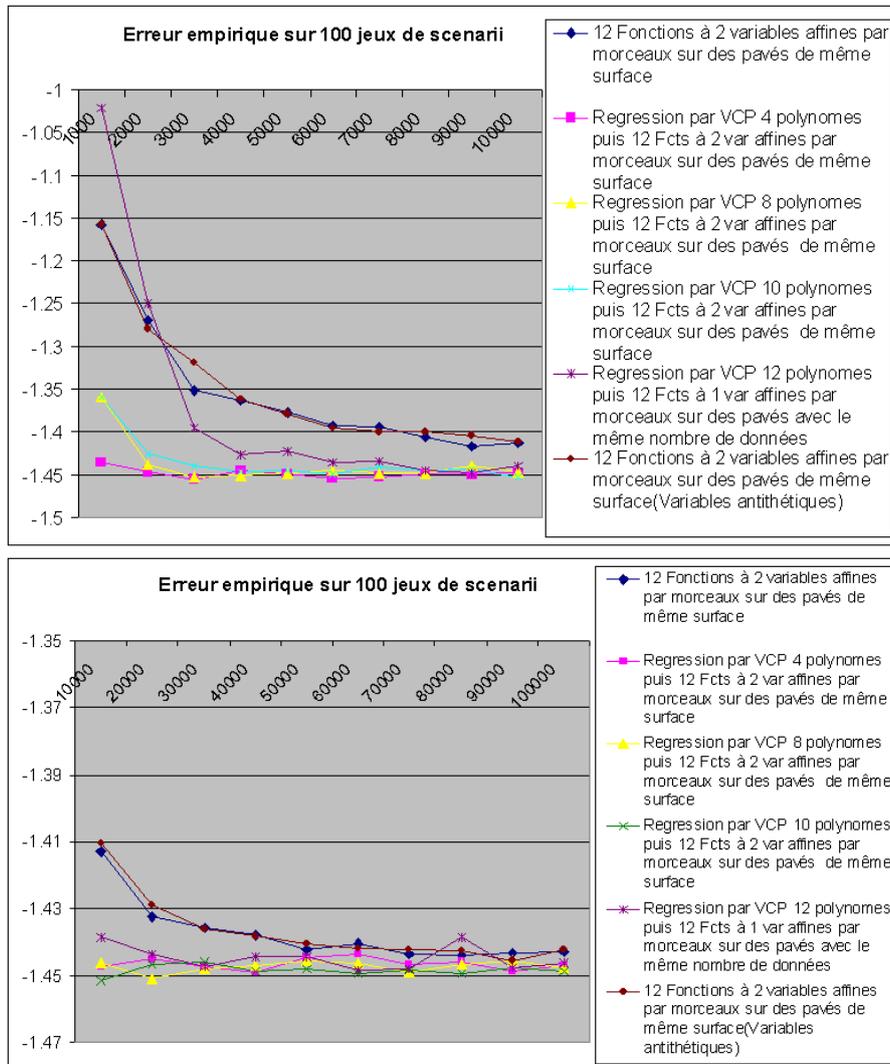


FIGURE 4.11 – Erreur empirique pour la régression par VCP pour le cas 2dAffineMemeSurf.

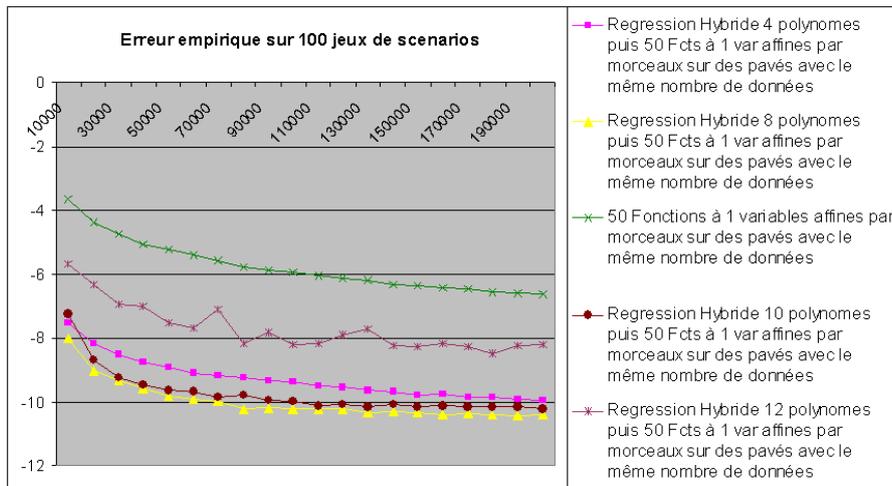


FIGURE 4.12 – Erreur empirique pour la régression par VCP pour le cas 1dAffineMemeData avec K_2 grand

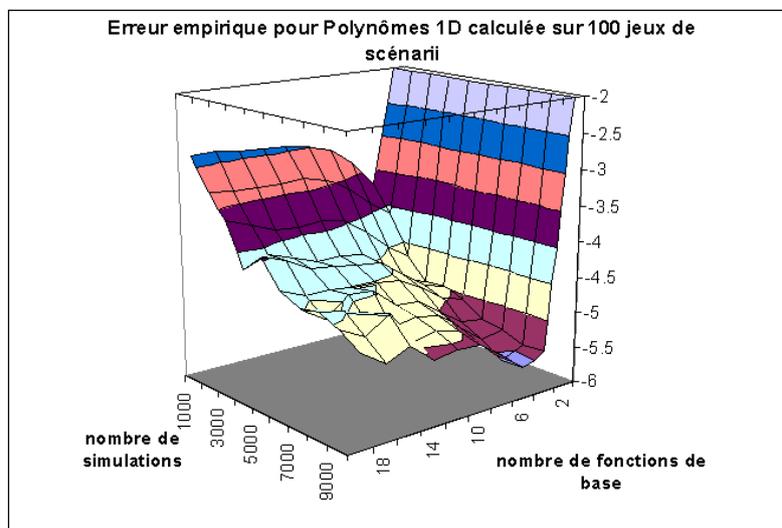


FIGURE 4.13 – Erreur empirique en fonction de N et K pour le cas des polynômes 1D.

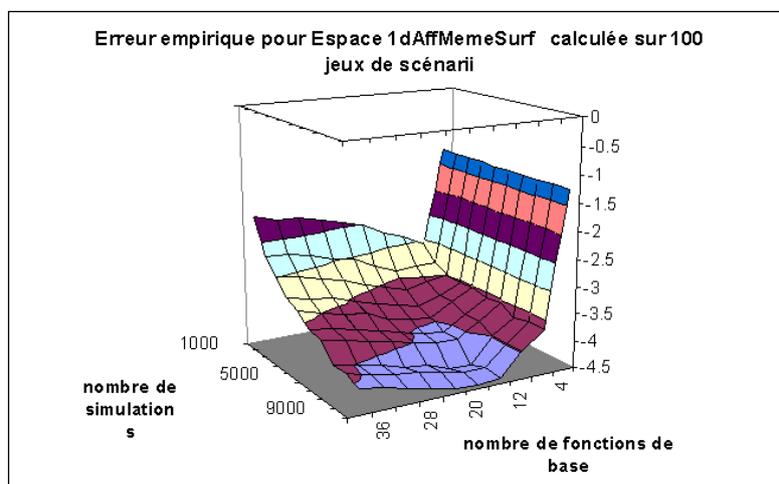


FIGURE 4.14 – Erreur empirique en fonction de N et K pour le cas 1dAffineMemeSurf.

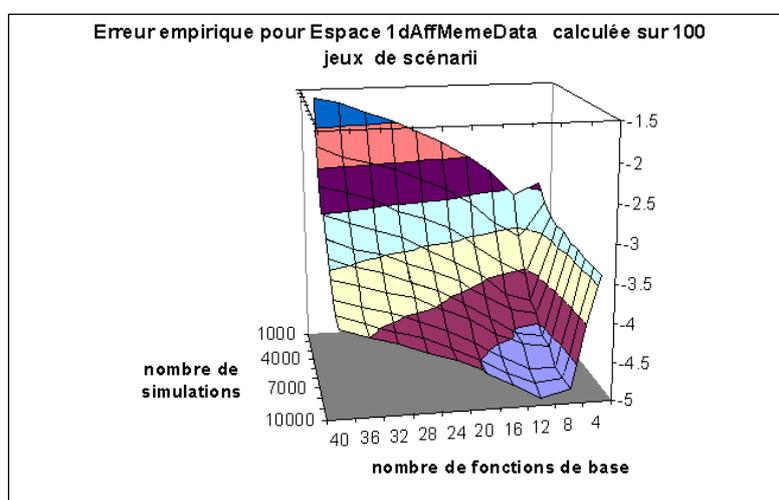


FIGURE 4.15 – Erreur empirique en fonction de N et K pour le cas 1dAffineMemeData.

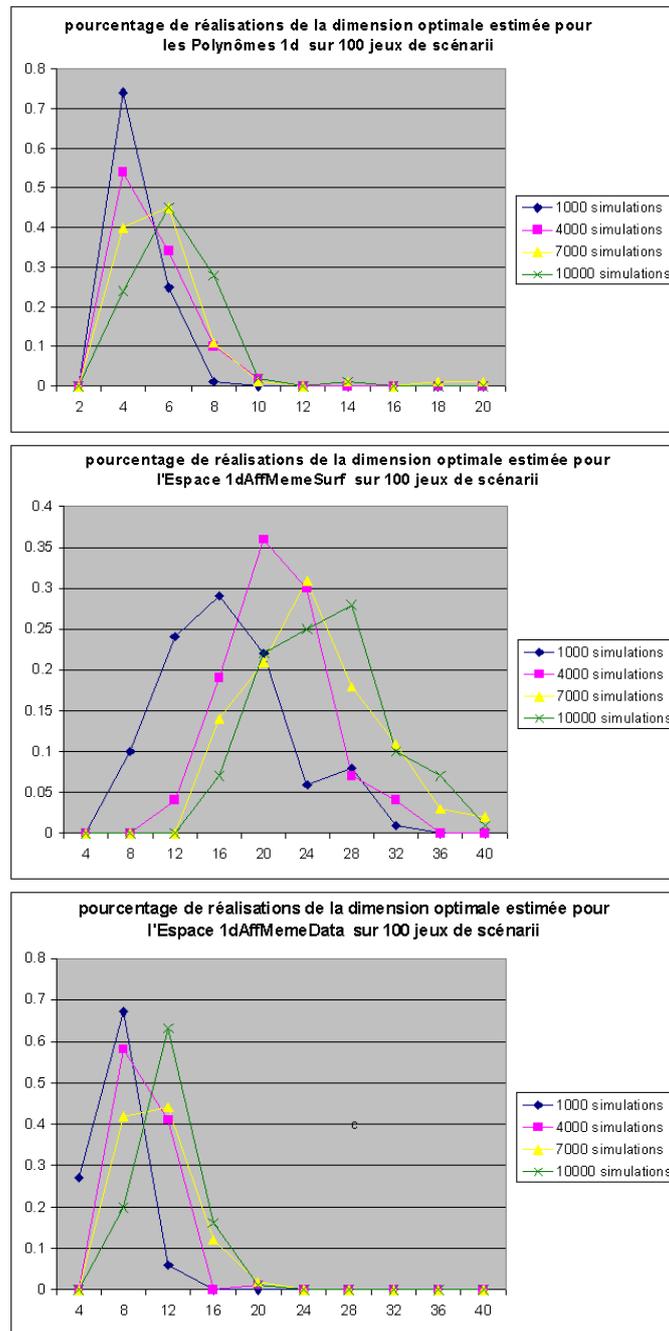


FIGURE 4.16 – Histogramme de \hat{K} pour les cas polynômes 1D, 1dAffineMemeSurf et 1dAffineMemeData.

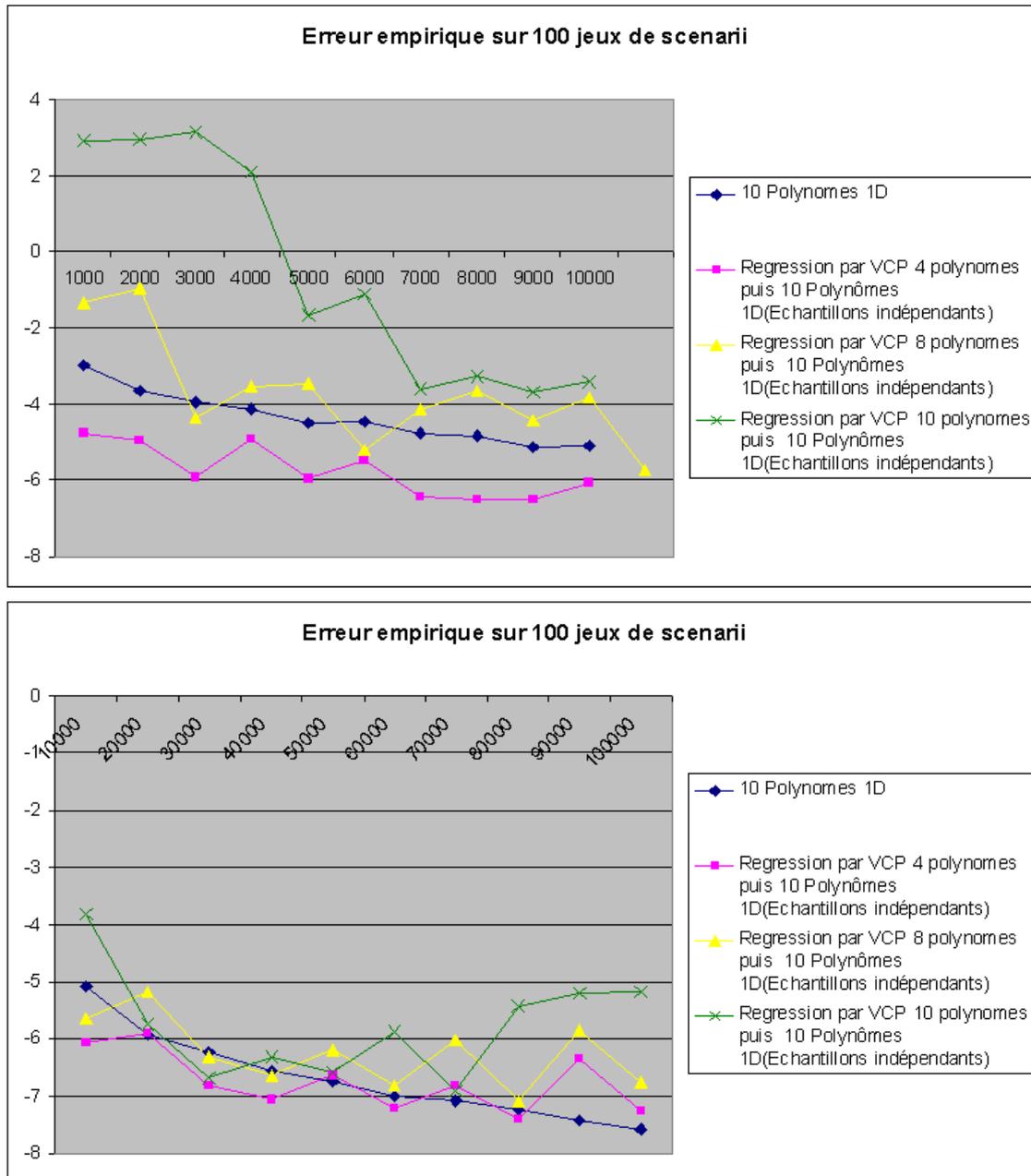


FIGURE 4.17 – Erreur empirique pour la régression par VCP avec échantillon indépendant pour le cas Polynômes 1D.

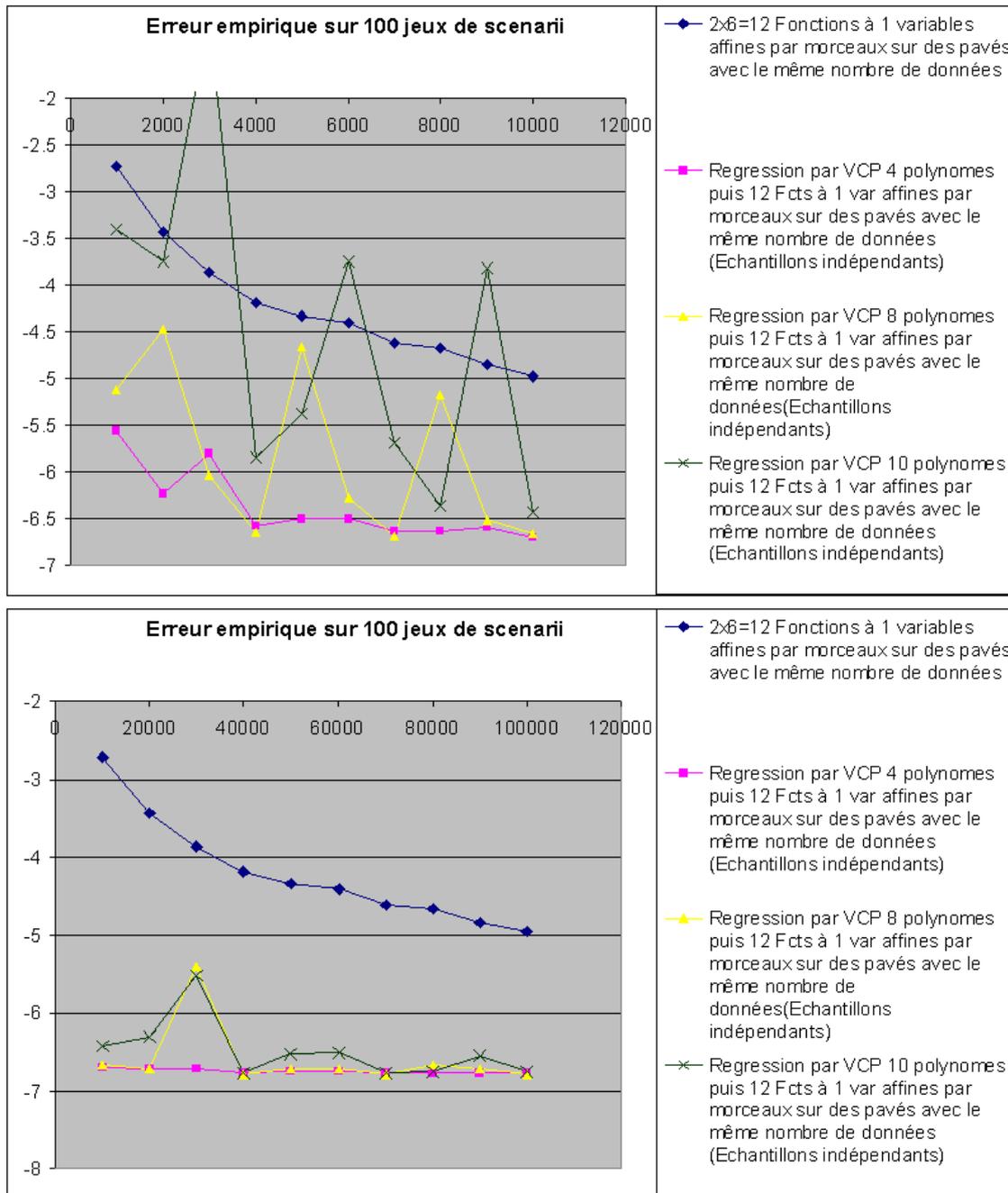


FIGURE 4.18 – Erreur empirique pour la régression par VCP avec échantillon indépendant pour le cas 1dAffineMemeData.

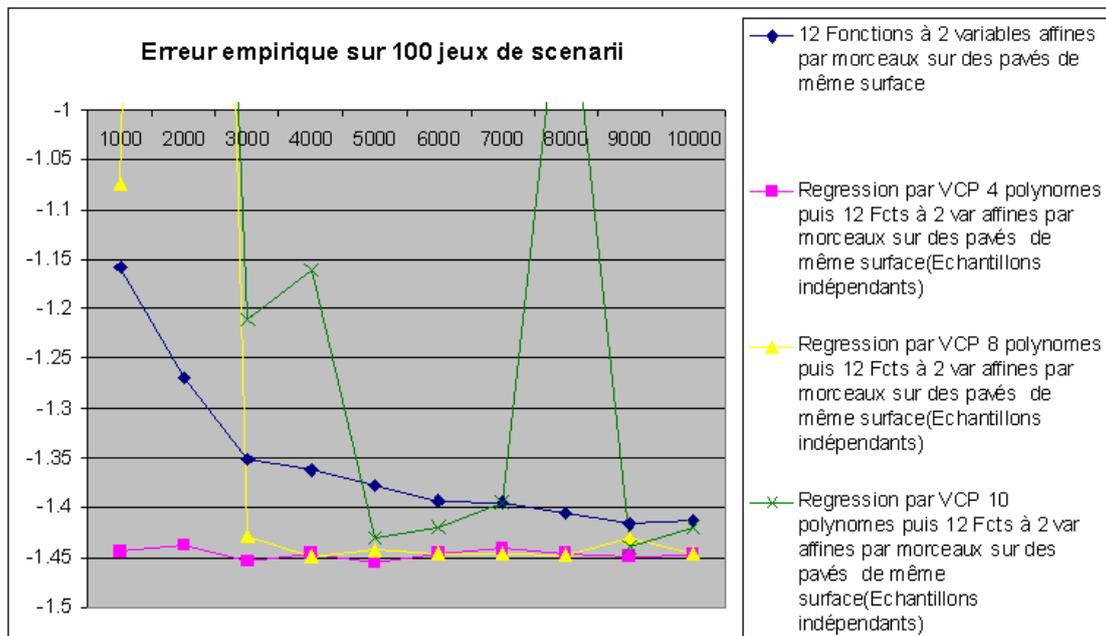


FIGURE 4.19 – Erreur empirique pour la régression par VCP avec échantillon indépendant pour le cas 2dAffineMemeSurf.

Chapitre 5

Régression empirique à plusieurs périodes

Le but ici est d'étendre la méthode de régression par VCP du chapitre précédent au cas d'un produit versant plusieurs flux à différentes dates futures.

5.1 Description du cas test : Modèle et Payoff

Comme au chapitre précédent, on considère deux sous-jacents S^1 et S^2 de dynamique :

$$\begin{aligned}\frac{dS_t^1}{S_t^1} &= \sigma^1 dW_t^1, \\ \frac{dS_t^2}{S_t^2} &= \sigma^2 dW_t^2\end{aligned}$$

où W^1, W^2 sont deux mouvements browniens standards indépendants. On néglige le taux d'intérêt.

On considère un produit qui verse à chaque date t_i , $i \in \{1, \dots, n\}$ (où $(t_i)_{1 \leq i \leq n}$ est équi-repartie de pas Δt et $t_n = T$) le flux :

$$\Psi(S_{t_i}^1, S_{t_i}^2) = (S_{t_i}^2 - (1 - \alpha)S_{t_i}^1)_+ - (\alpha + \beta)(S_{t_i}^2 - S_{t_i}^1)_+ - (S_{t_i}^2 - (1 + \beta)S_{t_i}^1)_+.$$

On prendra pour les tests numériques : $\sigma^1 = \sigma^2 = \sigma = 0.15$, $S_0^1 = S_0^2 = 10$, $\alpha = 0.1$, $\beta = 0.15$. L'enjeu est d'estimer à chaque date t_m , $m < n$ la valeur actuelle des flux futures, en d'autres termes

$$P^m(S_m^1, S_m^2) = \mathbb{E}\left[\sum_{j=m+1}^n \Psi(S_{t_j}^1, S_{t_j}^2) \mid S_{t_m}^1, S_{t_m}^2\right]$$

via la méthode de régression par VCP.

5.2 Régression par VCP

5.2.1 Description

Notons :

$$X^j = (S_{t_j}^1, S_{t_j}^2), 1 \leq j \leq n \quad \text{et} \quad Z^j = (S_{t_j}^1, S_{t_j}^2, S_{t_{j+1}}^1, S_{t_{j+1}}^2), 1 \leq j < n.$$

Pour un jeu de données $(S^{1,i}, S^{2,i})_{1 \leq i \leq N}$, notons :

$$X_i^j = (S_{t_j}^{1,i}, S_{t_j}^{2,i}), 1 \leq j \leq n \quad \text{et} \quad Z_i^j = (S_{t_j}^{1,i}, S_{t_j}^{2,i}, S_{t_{j+1}}^{1,i}, S_{t_{j+1}}^{2,i}), 1 \leq j < n.$$

Soit $(P_k)_{1 \leq k \leq K_1}$ une famille de polynômes de faible degré ou d'autres types de fonctions vérifiant

$$\mathbb{E}[P_k(Z^j)|X^j] = 0, \quad 1 \leq k \leq K_1, 1 \leq j < n.$$

Pour les tests, quand il s'agit de polynômes on prendra :

$$\begin{aligned} P_k(Z^j) &:= S_{t_{j+1}}^2 \left(\frac{S_{t_{j+1}}^1}{S_{t_{j+1}}^2} \right)^k - \mathbb{E} \left[S_{t_{j+1}}^2 \left(\frac{S_{t_{j+1}}^1}{S_{t_{j+1}}^2} \right)^k \mid S_{t_j}^1, S_{t_j}^2 \right] \\ &= S_{t_{j+1}}^2 \left(\frac{S_{t_{j+1}}^1}{S_{t_{j+1}}^2} \right)^k - \exp \left(\frac{\sigma^2}{2} (k^2 + (k-1)^2 - 1) \Delta \right) S_{t_j}^2 \left(\frac{S_{t_j}^1}{S_{t_j}^2} \right)^k. \end{aligned}$$

Et quand il s'agit de fonctions locales par morceaux, on prendra $(P_k)_{1 \leq k \leq K_1}$ de la forme

$$\begin{aligned} P_{2m-1}(Z^j) &= S_{t_{j+1}}^2 \mathbf{1}_{\left\{ \frac{S_{t_{j+1}}^1}{S_{t_{j+1}}^2} \in I_m \right\}} - \mathbb{E} \left[S_{t_{j+1}}^2 \mathbf{1}_{\left\{ \frac{S_{t_{j+1}}^1}{S_{t_{j+1}}^2} \in I_m \right\}} \mid S_{t_j}^1, S_{t_j}^2 \right], \\ P_{2m}(Z^j) &= S_{t_{j+1}}^1 \mathbf{1}_{\left\{ \frac{S_{t_{j+1}}^1}{S_{t_{j+1}}^2} \in I_m \right\}} - \mathbb{E} \left[S_{t_{j+1}}^1 \mathbf{1}_{\left\{ \frac{S_{t_{j+1}}^1}{S_{t_{j+1}}^2} \in I_m \right\}} \mid S_{t_j}^1, S_{t_j}^2 \right] \end{aligned}$$

avec $1 \leq m \leq p$, $K_1 = 2p$, $(I_m = [a_m, b_m])_{1 \leq m \leq p}$ une partition d'intervalles de même longueur d'un intervalle $[min, max]$.

Un calcul simple permet d'obtenir :

$$\begin{aligned}
P_{2m-1}(Z^j) &= S_{t_{j+1}}^2 \mathbf{1}_{\left\{\frac{S_{t_{j+1}}^1}{S_{t_{j+1}}^2} \in I_m\right\}} - S_{t_j}^2 \left[\mathcal{N} \left(\frac{\log(b_m) + \sigma^2 \Delta t - \log\left(\frac{S_{t_j}^1}{S_{t_j}^2}\right)}{\sigma \sqrt{2\Delta t}} \right) \right. \\
&\quad \left. - \mathcal{N} \left(\frac{\log(a_m) + \sigma^2 \Delta t - \log\left(\frac{S_{t_j}^1}{S_{t_j}^2}\right)}{\sigma \sqrt{2\Delta t}} \right) \right] \\
P_{2m}(Z^j) &= S_{t_{j+1}}^1 \mathbf{1}_{\left\{\frac{S_{t_{j+1}}^1}{S_{t_{j+1}}^2} \in I_m\right\}} - S_{t_j}^1 \left[\mathcal{N} \left(\frac{\log(1/a_m) + \sigma^2 \Delta t - \log\left(\frac{S_{t_j}^2}{S_{t_j}^1}\right)}{\sigma \sqrt{2\Delta t}} \right) \right. \\
&\quad \left. - \mathcal{N} \left(\frac{\log(1/b_m) + \sigma^2 \Delta t - \log\left(\frac{S_{t_j}^2}{S_{t_j}^1}\right)}{\sigma \sqrt{2\Delta t}} \right) \right]
\end{aligned}$$

où \mathcal{N} est la fonction de répartition de la loi normale centrée réduite.

Soit $(\Phi_k)_{1 \leq k \leq K_2}$ une famille de fonctions locales par morceaux ou d'autres fonctions de base. L'idée de la régression par VCP peut être exploitée dans le cas de plusieurs flux par les trois façons suivantes :

Méthode VCP 1 :

Pour cette méthode, on calcule des coefficients $(\tilde{\alpha}_k^m)_{1 \leq k \leq K_1}$ qui vérifient :

$$(\tilde{\alpha}_k^m)_k = \arg \inf_{(\alpha_k)_k} \frac{1}{N} \sum_{i=1}^N \left| \sum_{j=m}^{n-1} \Psi(X_i^{j+1}) - \sum_{k=1}^{K_1} \alpha_k P_k(Z_i^m) \right|^2.$$

Ensuite on calcule les coefficients de régression par moindres carrés $(\tilde{\beta}_k^m)_{1 \leq k \leq K_2}$ vérifiant :

$$(\tilde{\beta}_k^m)_k = \arg \inf_{(\beta_k)_k} \frac{1}{N} \sum_{i=1}^N \left| \sum_{j=m}^{n-1} \Psi(X_i^{j+1}) - \sum_{k=1}^{K_1} \tilde{\alpha}_k^m P_k(Z_i^m) - \sum_{k=1}^{K_2} \beta_k \Phi_k(X_i^m) \right|^2.$$

On estime P^m par $\tilde{P}_N^m = \sum_{k=1}^{K_2} \tilde{\beta}_k^m \Phi_k$.

Puisque $\mathbb{E}[P_k(Z^j)|X^j] = 0$, il est facile de voir que pour $(\alpha_k)_{1 \leq k \leq K_1}$ fixés :

$$\begin{aligned} & \mathbb{E}[\text{Var}(\sum_{j=m}^{n-1} \Psi(X^{j+1}) - \sum_{k=1}^{K_1} \alpha_k P_k(Z^m) | X^m)] \\ &= \mathbb{E}[(\sum_{j=m}^{n-1} \Psi(X^{j+1}) - \sum_{k=1}^{K_1} \alpha_k P_k(Z^m))^2] - \mathbb{E}[\mathbb{E}(\sum_{j=m}^{n-1} \Psi(X^{j+1}) | X^m)^2]. \end{aligned}$$

On pose V_1 la version empirique de la quantité ci-dessus, i.e

$$V_1 = \frac{1}{N} \sum_{i=1}^N [(\sum_{j=m}^{n-1} \Psi(X_i^{j+1}) - \sum_{k=1}^{K_1} \tilde{\alpha}_k^m P_k(Z_i^m))^2 - \mathbb{E}(\sum_{j=m}^{n-1} \Psi(X_i^{j+1}) | X_i^m)^2].$$

Pour évaluer V_1 , étant donné les hypothèse du modèle, le terme $\mathbb{E}(\sum_{j=m}^{n-1} \Psi(X_i^{j+1}) | X_i^m)^2$ se calcule par formule fermée.

Méthode VCP 2 :

Cette méthode se base sur une réduction de l'espérance de variance flux par flux et exploite de manière récursive les réductions de l'espérance de variance effectuées aux pas d'avant en sus de celle de l'étape en cours, selon l'algorithme suivant.

Pour $m = n - 1, \dots, 1$,

$$(\tilde{\alpha}_k^m)_k = \arg \inf_{(\alpha_k)_k} \frac{1}{N} \sum_{i=1}^N \left| \Psi(X_i^{m+1}) - \sum_{k=1}^{K_1} \alpha_k P_k(Z_i^m) \right|^2,$$

$$(\tilde{\beta}_k^m)_k = \arg \inf_{(\beta_k)_k} \frac{1}{N} \sum_{i=1}^N \left| \sum_{j=m}^{n-1} \left[\Psi(X_i^{j+1}) - \sum_{k=1}^{K_1} \tilde{\alpha}_k^j P_k(Z_i^j) \right] - \sum_{k=1}^{K_2} \beta_k \Phi_k(X_i^m) \right|^2.$$

On estime P^m par $\tilde{P}_N^m = \sum_{k=1}^{K_2} \tilde{\beta}_k^m \Phi_k$.

Vu que $\mathbb{E}[P_k(Z^j)|X^j] = 0$, pour $(\alpha_k^j)_{1 \leq j \leq n, 1 \leq k \leq K_1}$ fixés on a

$$\begin{aligned} & \mathbb{E}[\text{Var}(\sum_{j=m}^{n-1} \Psi(X^{j+1}) - \sum_{k=1}^{K_1} \alpha_k^j P_k(Z^j) | X^m)] \\ &= \mathbb{E}[(\sum_{j=m}^{n-1} \Psi(X^{j+1}) - \sum_{k=1}^{K_1} \alpha_k^j P_k(Z^j))^2] - \mathbb{E}[\mathbb{E}(\sum_{j=m}^{n-1} \Psi(X^{j+1}) | X^m)^2]. \end{aligned}$$

On pose

$$V_2 = \frac{1}{N} \sum_{i=1}^N [(\sum_{j=m}^{n-1} \Psi(X_i^{j+1}) - \sum_{k=1}^{K_1} \tilde{\alpha}_k^j P_k(Z_i^j))^2 - \mathbb{E}(\sum_{j=m}^{n-1} \Psi(X_i^{j+1}) | X_i^m)^2].$$

Méthode VCP 3 :

Cette dernière version exploite de manière récursive la réduction de l'espérance de

variance de la somme des flux selon l'algorithme suivant :

Pour $m = n - 1, \dots, 1$,

$$(\tilde{\alpha}_k^m)_k = \arg \inf_{(\alpha_k)_k} \frac{1}{N} \sum_{i=1}^N \left| \sum_{j=m+1}^{n-1} \left[\Psi(X_i^{j+1}) - \sum_{k=1}^{K_1} \tilde{\alpha}_k^j P_k(Z_i^j) \right] + \Psi(X_i^{m+1}) - \sum_{k=1}^{K_1} \alpha_k P_k(Z_i^m) \right|^2,$$

$$(\tilde{\beta}_k^m)_k = \arg \inf_{(\beta_k)_k} \frac{1}{N} \sum_{i=1}^N \left| \sum_{j=m}^{n-1} \left[\Psi(X_i^{j+1}) - \sum_{k=1}^{K_1} \tilde{\alpha}_k^j P_k(Z_i^j) \right] - \sum_{k=1}^{K_2} \beta_k \Phi_k(X_i^m) \right|^2.$$

On estime P^m par $\tilde{P}_N^m = \sum_{k=1}^{K_2} \tilde{\beta}_k^m \Phi_k$.

On pose

$$V_3 = \frac{1}{N} \sum_{i=1}^N \left[\left(\sum_{j=m}^{n-1} \Psi(X_i^{j+1}) - \sum_{k=1}^{K_1} \tilde{\alpha}_k^j P_k(Z_i^j) \right)^2 - \mathbb{E} \left(\sum_{j=m}^{n-1} \Psi(X_i^{j+1}) | X_i^m \right)^2 \right].$$

Enfin on note

$$V = \frac{1}{N} \sum_{i=1}^N \left[\left(\sum_{j=m}^{n-1} \Psi(X_i^{j+1}) \right)^2 - \mathbb{E} \left(\sum_{j=m}^{n-1} \Psi(X_i^{j+1}) | X_i^m \right)^2 \right].$$

V_1 , V_2 et V_3 représentent, pour chacune des méthodes associées, la version empirique du facteur du terme d'erreur statistique dans le théorème 4.4.1 ce qui permet ainsi de comparer ces méthodes entre elles. Enfin, V représente la version empirique de ce facteur pour une régression standard (correspondant à des coefficients VCP nuls).

5.2.2 Résultats numériques

Dans nos tests, on prendra $n = 10$, $\Delta t = 1$ et $T = 10$. Comme dans le chapitre précédent, on approche la quantité $E^m = \mathbb{E} \left[\|\tilde{P}_N^m - P\|_N^2 \right]$ par sa version empirique moyennée sur $M = 20$ jeux de scénarii indépendants. Pour différents choix de date t_m et de fonctions $(P_k)_{1 \leq k \leq K_1}$ et $(\Phi_k)_{1 \leq k \leq K_2}$, on visualise $\log(E^m)$ ainsi que la version empirique de la variance conditionnelle pour les trois méthodes aussi bien que pour la régression standard en fonction de N variant de 1000 à 10000. Notre but est de déterminer expérimentalement la méthode la plus performante.

D'abord on teste le cas de $K_2 = 10$ polynômes à 1 variable dans l'étape de régression. La figure 5.1 où on utilise $K_1 = 2$ polynômes pour la première étape des méthodes VCP, montre que pour $T = 10$ la méthode VCP 3 est la plus performante surtout lorsque la date t_m est petite. La deuxième méthode VCP est globalement mieux que la première mais ajoute peu d'amélioration par rapport à la régression

standard. La figure 5.2 visualisant l'espérance empirique de la variance conditionnelle associée à la figure 5.1 appuie ces observations. En effet, V_3 est nettement petite surtout pour $t_m = 1$ alors que V_1 et V_2 restent comparables à V . Par contre si on fixe $K_1 = 4$ (figure 5.3), les trois méthodes VCP ne sont pas satisfaisantes pour $t_m = 9$ et $t_m = 5$ malgré la bonne réduction de l'espérance empirique de variance (Figure 5.4). Par contre à la date $t_m = 1$, la méthode VCP 3 est nettement meilleure tandis que les deux autres améliorent très peu. La différence entre le cas $K_1 = 2$ et $K_1 = 4$ peut s'expliquer par la non vérification de l'hypothèse de bornitude des polynômes d'approximation de la première étape dans le théorème 4.4.1 lorsque K_1 et la maturité sont simultanément grands. En effet, l'utilisation de $K_1 = 4$ fonctions affines par morceaux sur $[0, 6]$ (donc bornées) (figures 5.5 et 5.6) conforte cette explication dans le mesure où on tire les mêmes conclusions que celles de la figure 5.1 .

Dans cette deuxième partie, on étudie l'utilisation de $K_2 = 50$ fonctions de l'espace `1dAffineMemeSurf` sur $[0, 6]$. Concernant la figure 5.7 où la réduction de l'espérance de la variance conditionnelle est réalisée par $K_1 = 4$ fonctions affines par morceaux sur $[0, 6]$, la méthode VCP 3 accélère nettement la convergence surtout pour $t_m = 1$. Quant aux deux autres méthodes VCP, les améliorations de convergence sont limitées avec un avantage pour la VCP 2. Ceci est bien en concordance avec les valeurs des espérances empiriques de variance conditionnelle de la figure 5.8 associée. Quant à la figure 5.9 où on fait la réduction de l'espérance de variance conditionnelle par $K_1 = 4$ polynômes à une variable, on constate que pour $t_m = 9$ et $t_m = 5$ les trois méthodes VCP sont peu ou pas performantes alors que pour $t_m = 1$, c'est la méthode VCP 3 qui l'emporte largement en terme de d'accélération de convergence. L'explication plausible du comportement des méthodes VCP entre ces deux cas réside aussi dans la non bornitude des fonctions polynomiales pour $K_1 = 4$ et des maturités grandes.

En conclusion, la méthode VCP 3 est celle qui est la plus satisfaisante des trois surtout pour des maturités non lointaines.

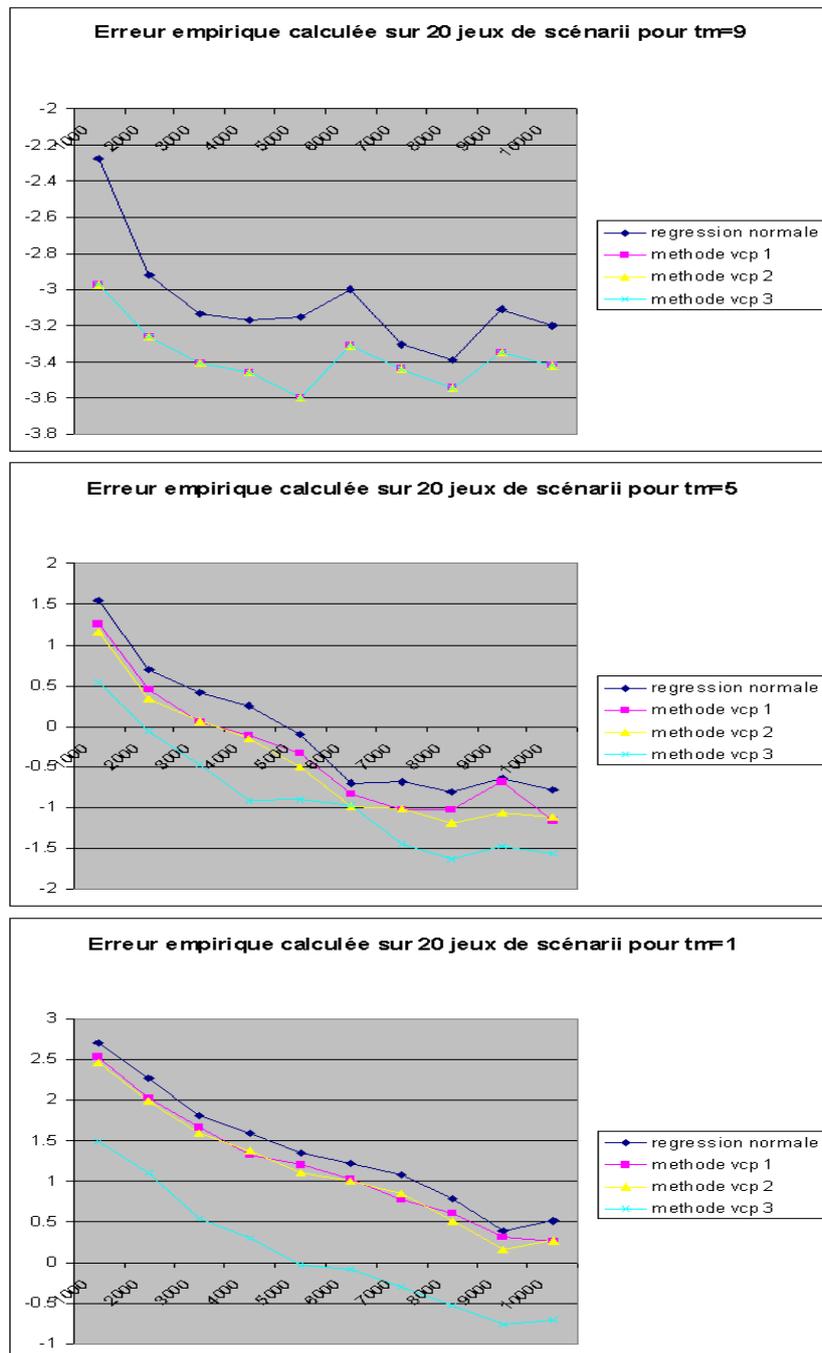


FIGURE 5.1 – Erreur empirique en utilisant $K_1 = 2$ polynômes pour la première étape et $K_2 = 10$ polynômes à une variable dans la seconde.

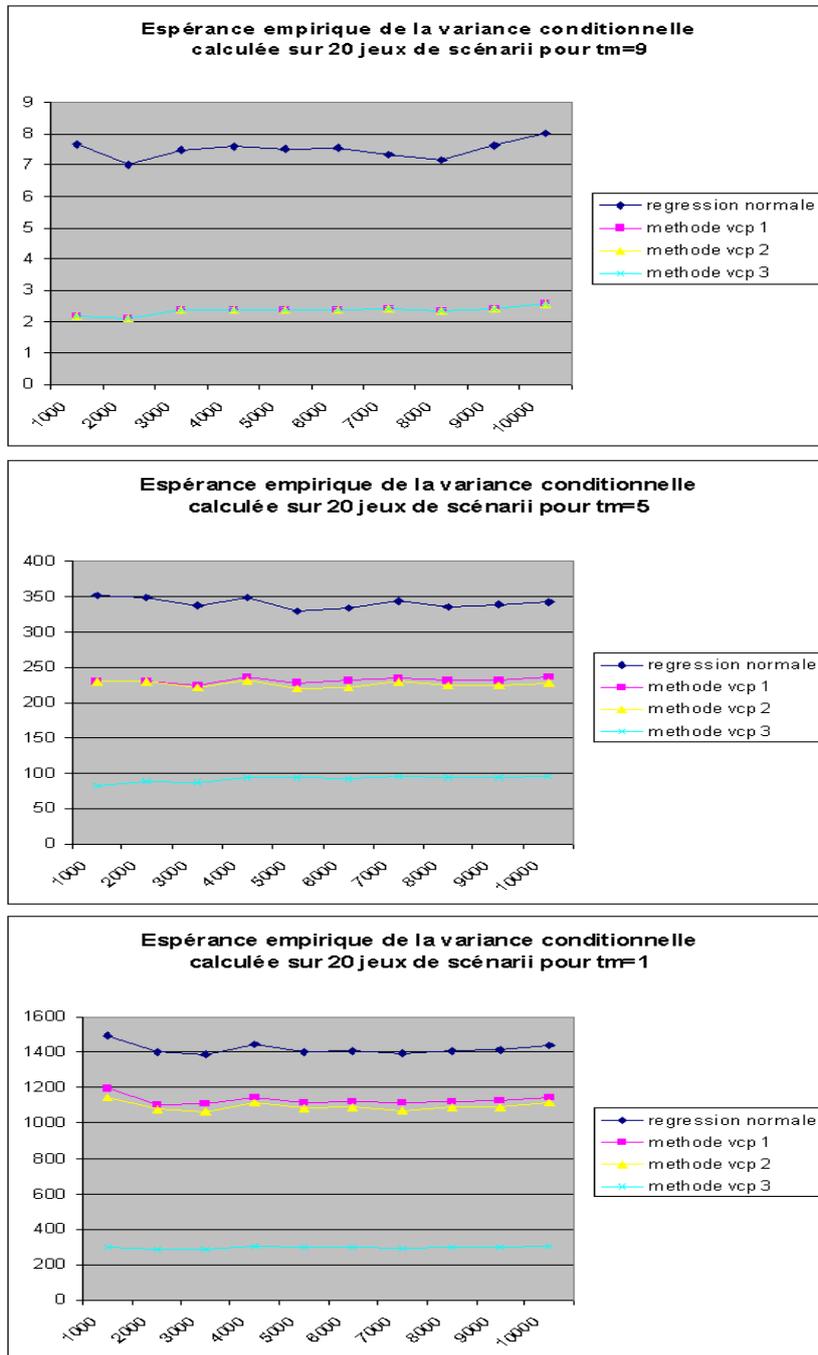


FIGURE 5.2 – Espérance empirique de la variance conditionnelle associée à la figure 5.1.

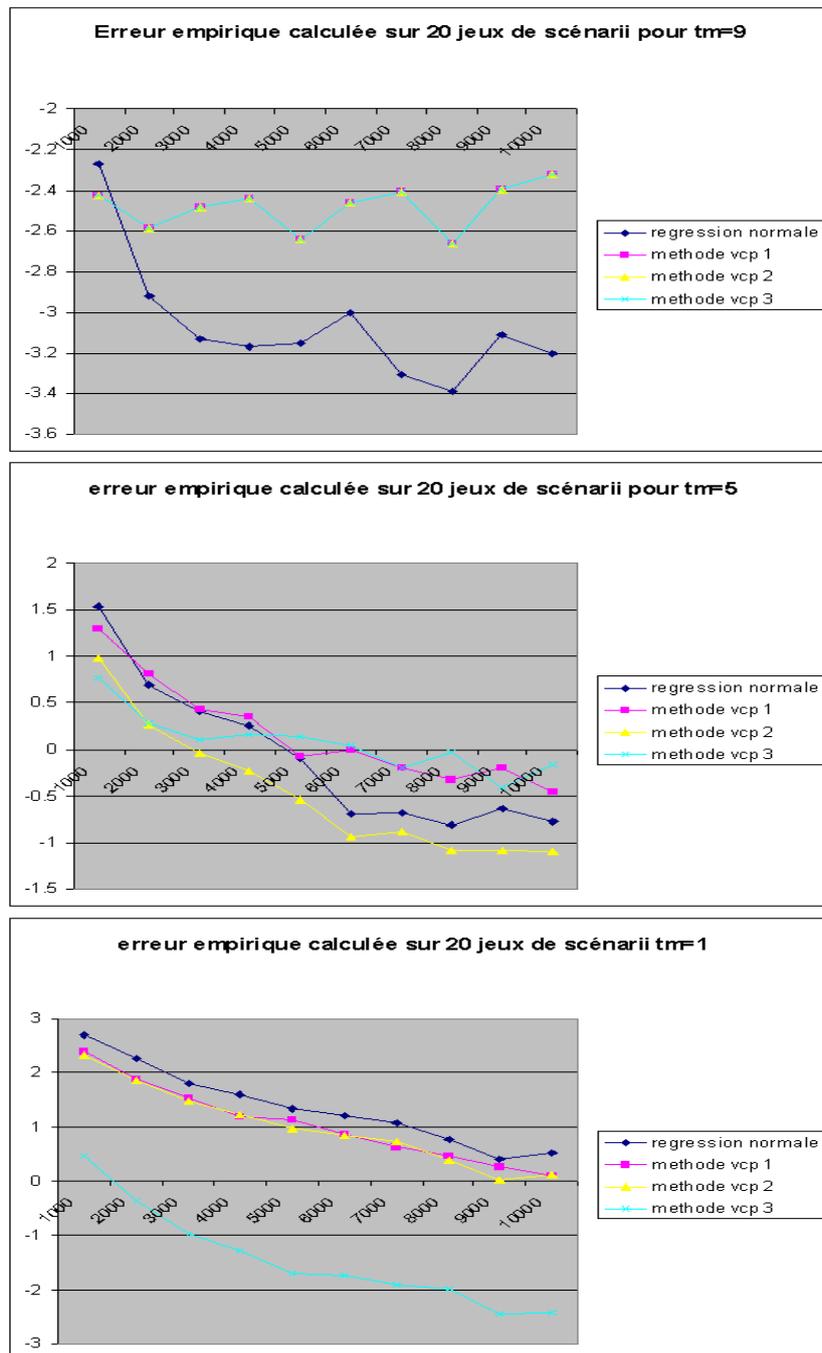


FIGURE 5.3 – Erreur empirique en utilisant $K_1 = 4$ polynômes pour la première étape et $K_2 = 10$ polynômes à une variable dans la seconde.

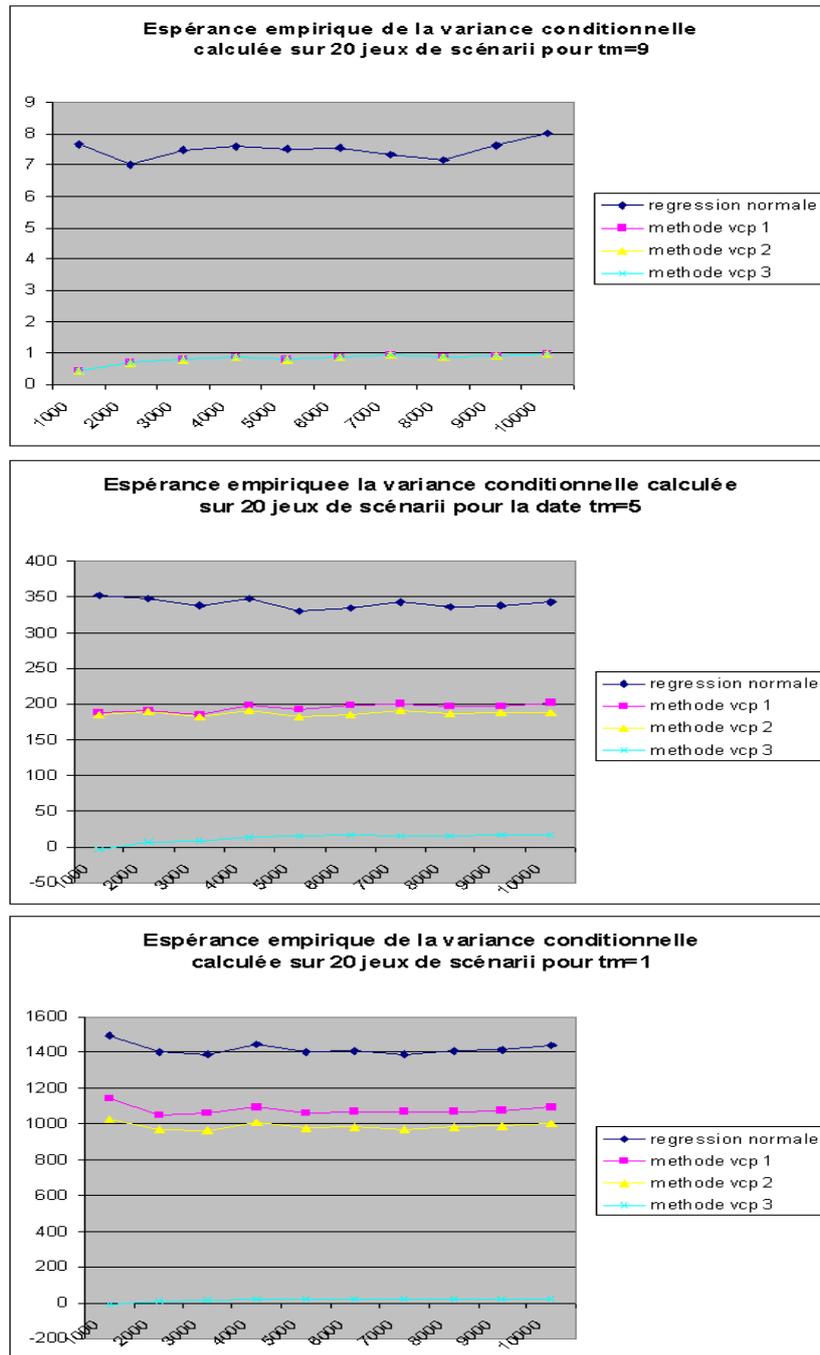


FIGURE 5.4 – Espérance empirique de la variance conditionnelle associée à la figure 5.3.

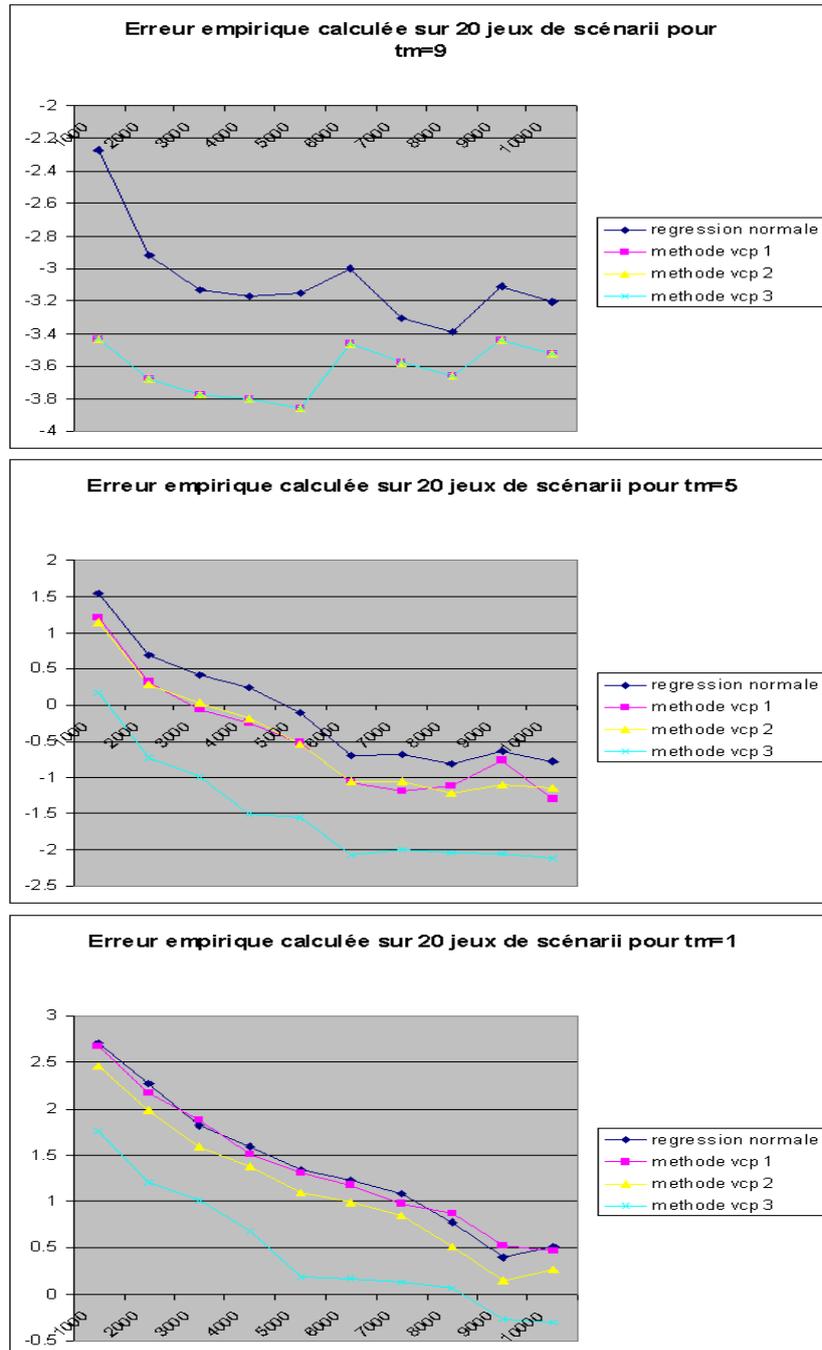


FIGURE 5.5 – Erreur empirique en utilisant $K_1 = 4$ fonctions affines par morceaux sur $I = [0, 6]$ pour la première étape et $K_2 = 10$ polynômes à une variable dans la seconde.

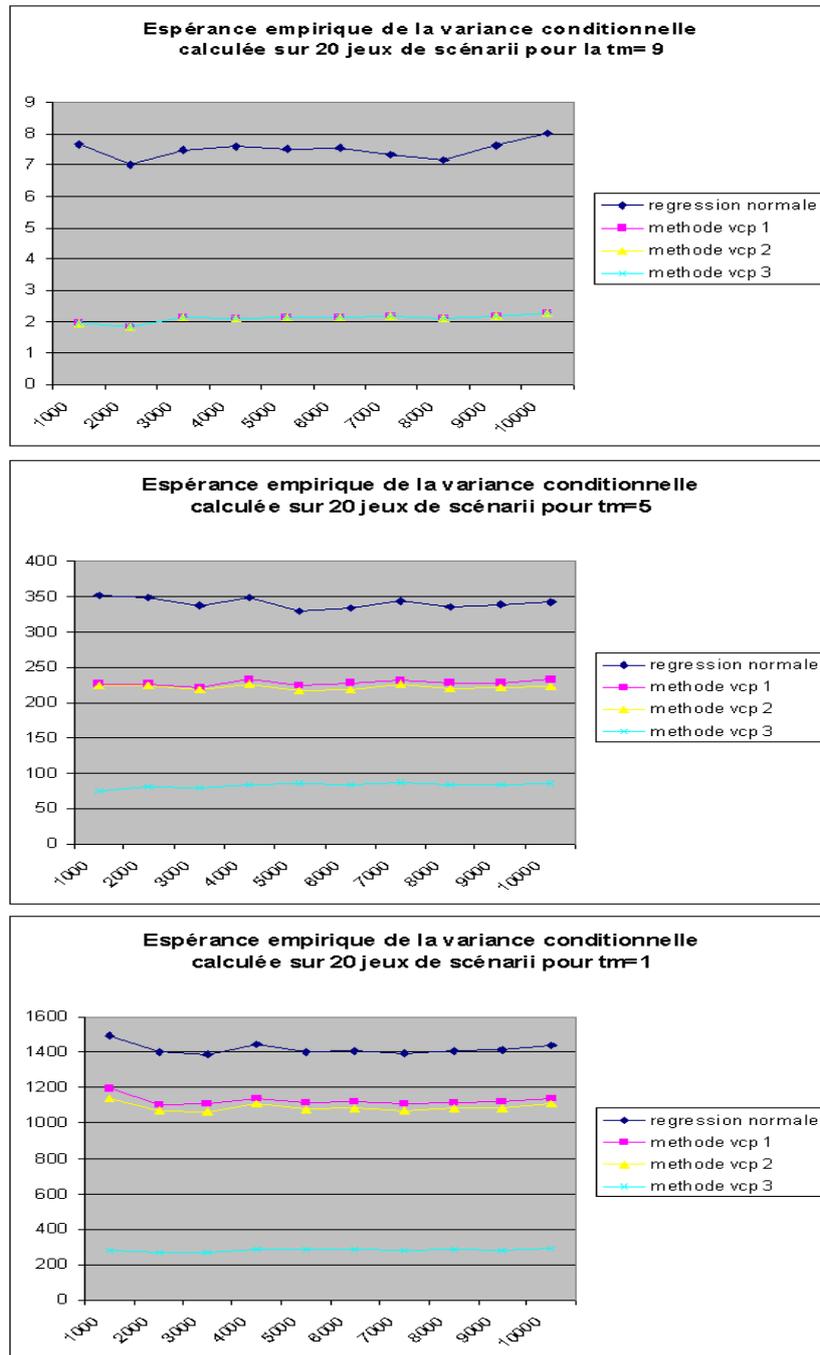


FIGURE 5.6 – Espérance empirique de la variance conditionnelle associée à la figure 5.5.

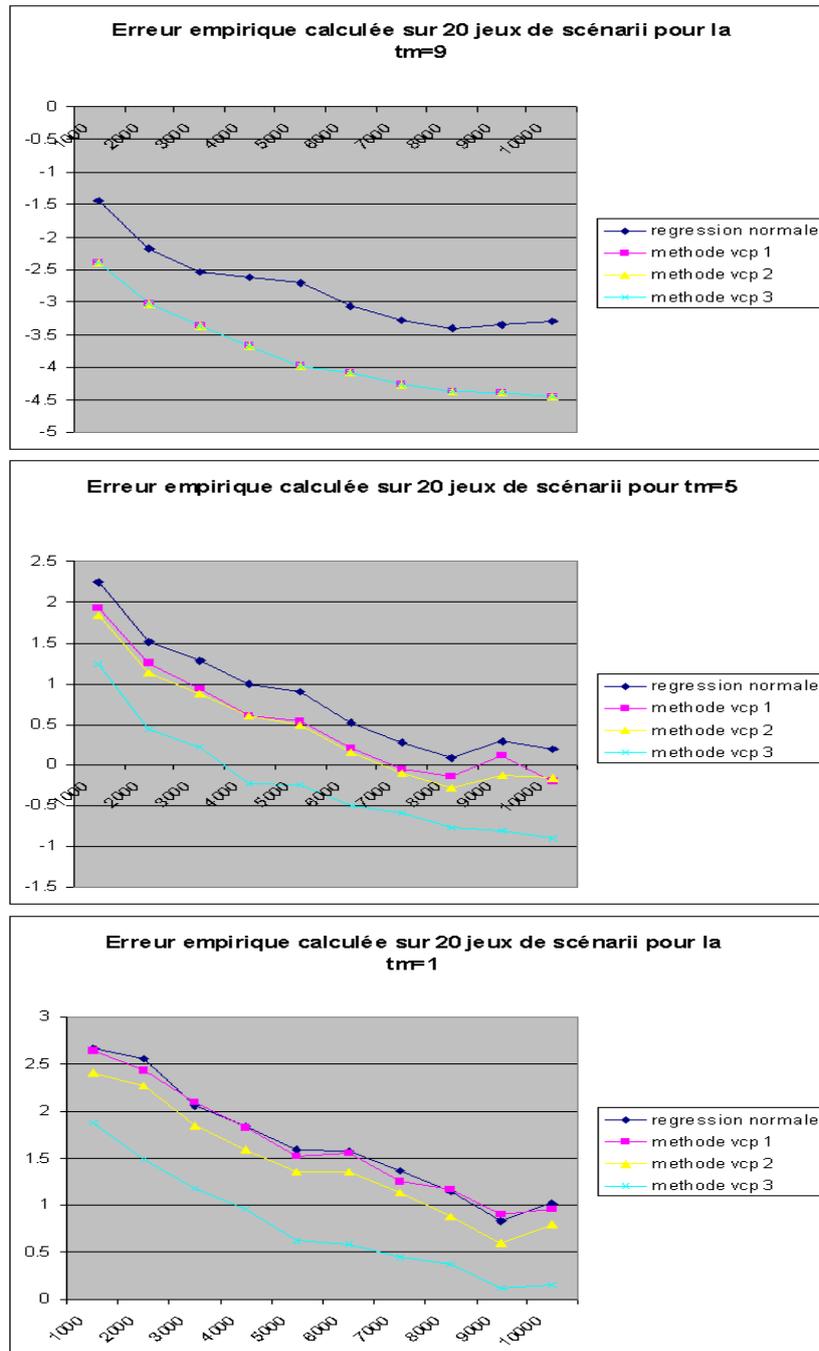


FIGURE 5.7 – Erreur empirique en utilisant $K_1 = 4$ fonctions affines par morceaux sur $I = [0, 6]$ pour la première étape et $K_2 = 50$ fonctions de l'espace $1dAffineMemeSurfsur [0, 6]$ pour la seconde.

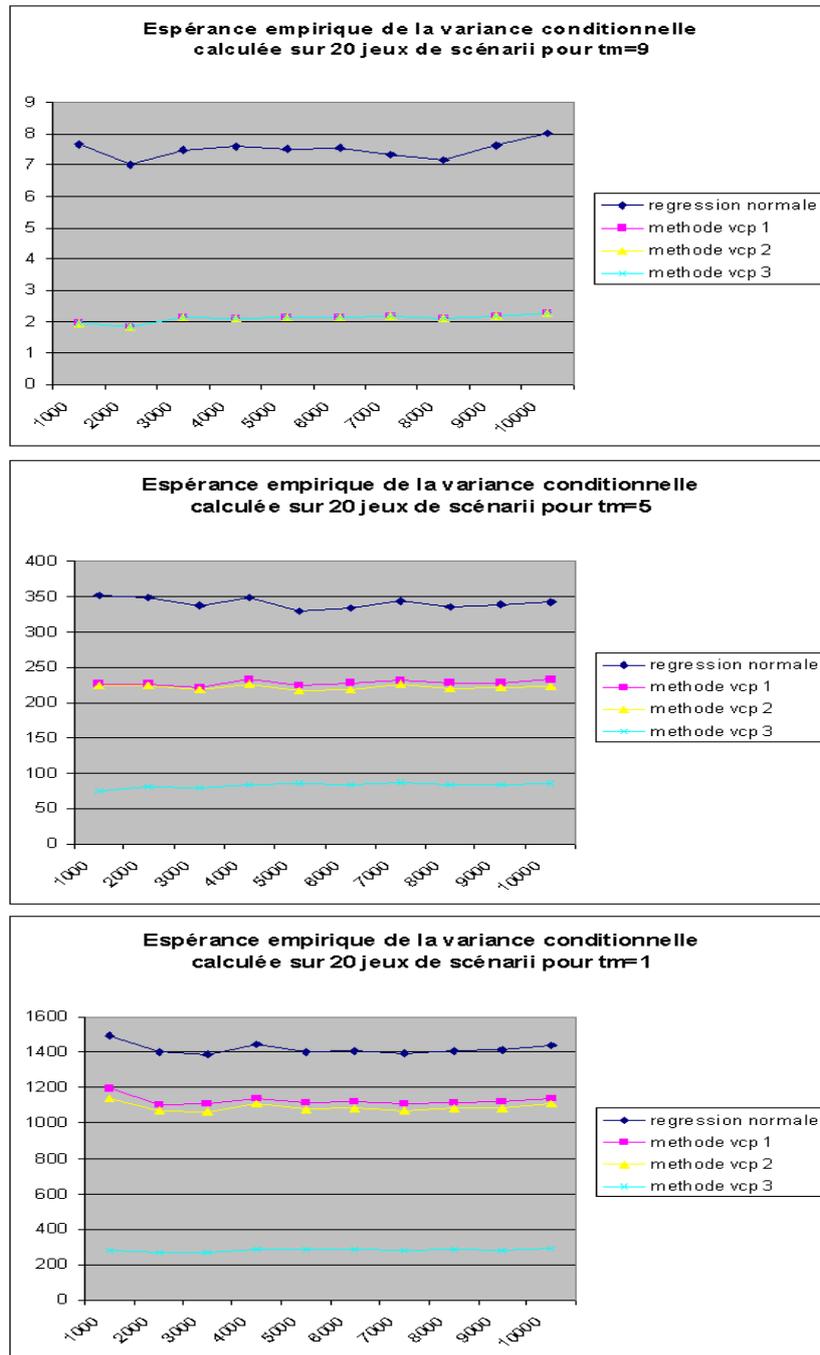


FIGURE 5.8 – Espérance empirique de la variance conditionnelle associée à la figure 5.7.

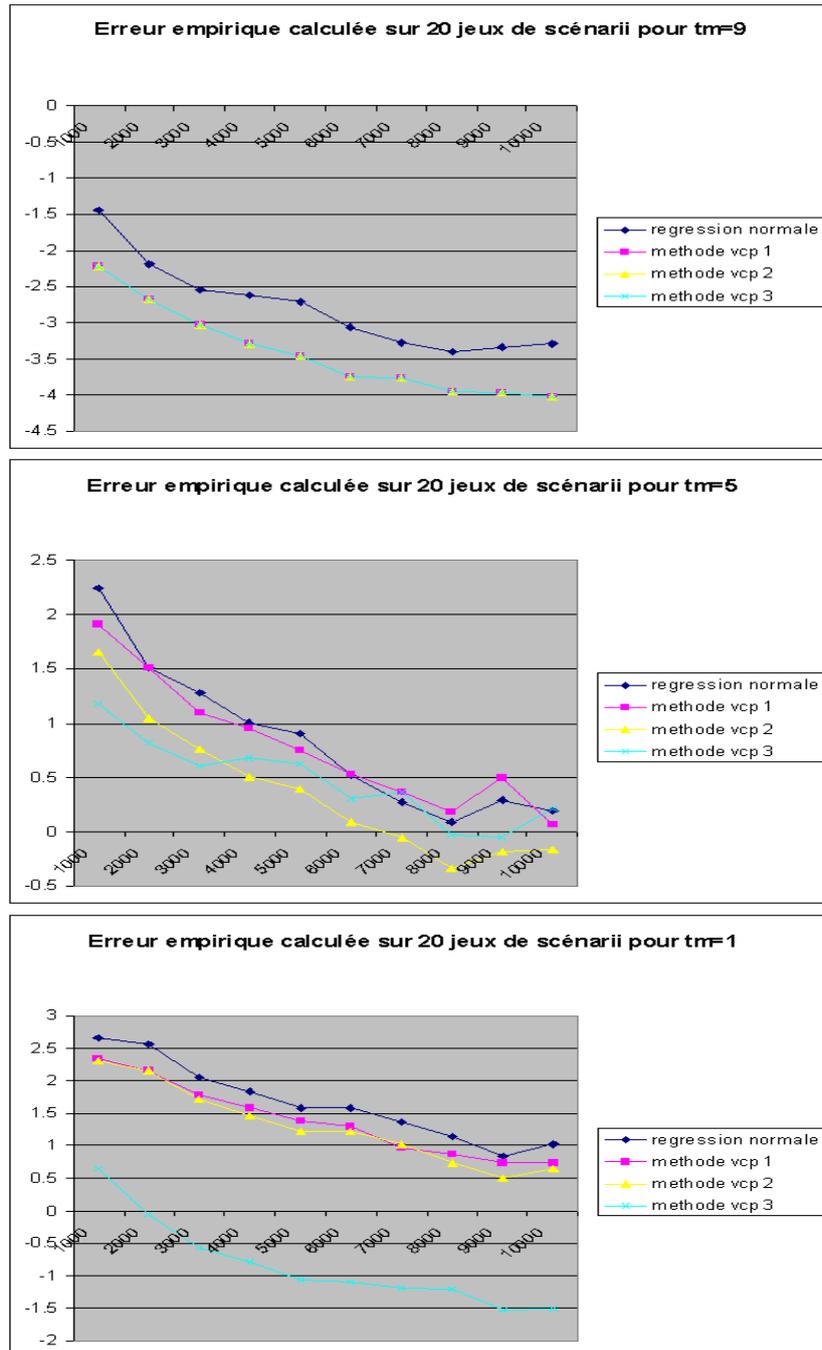


FIGURE 5.9 – Erreur empirique en utilisant $K_1 = 4$ polynômes à une variable pour la première étape et $K_2 = 50$ fonctions de l'espace $1dAffineMemeSurfsur [0, 6]$ pour la seconde.

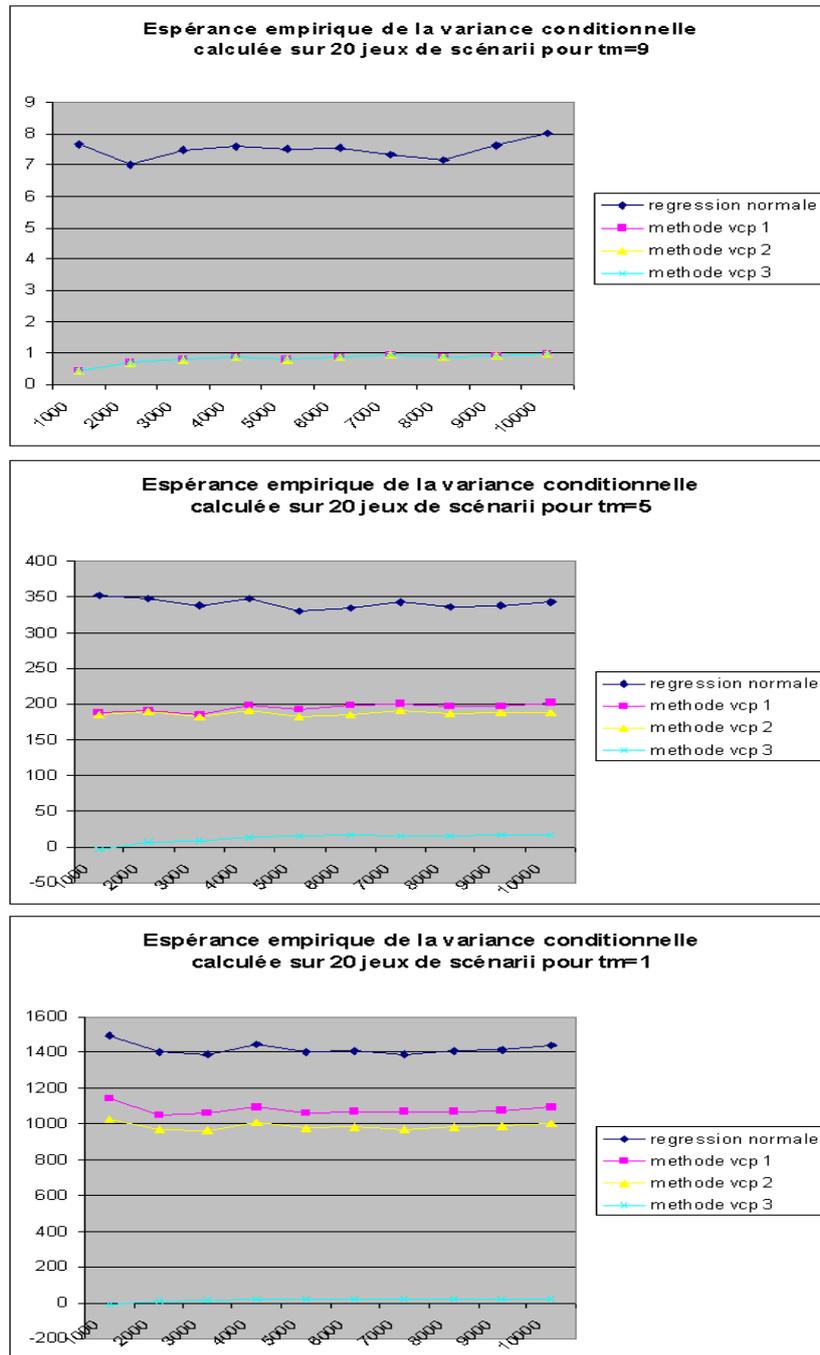


FIGURE 5.10 – Espérance empirique de la variance conditionnelle associée à la figure 5.9.

Chapitre 6

Mise en oeuvre numérique sur un contrat GMWB Axa avec le point de vue *hedger*

Dans cette section on applique les résultats obtenus aux chapitres précédents pour des produits GMWB. La seule différence ici est qu'on s'intéresse au point de vue du couvreur qui n'est pas forcément équivalent à celui du client rationnel étudié au chapitre 2. C'est dans ce contexte qu'on propose plusieurs variantes de l'algorithme de Longstaff Schwartz pour résoudre numériquement le problème de valorisation. Enfin, on expose quelques tests numériques effectués sur des contrats GMWB Axa.

On se place ici dans le même cadre du chapitre 2 en terme de notations et d'hypothèses.

6.1 Point de vue couvreur

En général, l'entité qui couvre les garanties du contrat GMWB est différente de celle qui gère l'*account value* du client. Ainsi, le pourcentage des *fees* β peut être décomposé en deux quantités β^h et β^m que touchent respectivement le couvreur (*hedger*) du contrat et le gérant (*manager*) de la valeur liquidative. Ainsi, les flux du point de vue du *hedger* se déclinent comme suit :

- 1) Il reçoit à chaque date $T_i^f, 1 \leq i \leq n_f$, un flux de fees :

$$\beta^h (A_{T_i^f}^v)_+.$$

- 2) Il s'engage à verser au client à la date T_i^w , $1 \leq i \leq n_w$ le flux manquant à l'*account value* pour honorer les retraits promis :

$$(\gamma B_{T_i^w}^w - (A_{T_i^f}^v)_+)_+.$$

Remarquons que le couvreur est sollicité uniquement si l'*account value* devient insuffisante pour en soustraire les retraits ou carrément nulle. Ainsi, à l'état (i, S, a, b, c, r, D) , le pire cas pour le couvreur correspond au cas où le client effectue une stratégie de rachat $(\alpha_j)_{i \leq j \leq n-1}$ qui maximise en espérance la valeur actualisée de la quantité

$$\sum_{j=i}^{n-1} (\gamma_j B_j^w - (A_j^v)_+)_+ - \beta_j^h (A_j^v)_+.$$

Notons que dans ce point de vue, l'assuré a un contrôle implicite sur les flux futurs en impactant les dynamiques de A^v , B^w et B^{rol} par la stratégie choisie.

Le **coût de couverture** correspondant au pire cas est

$$P^h(i, S, a, b, c, r, D) = \sup_{(\alpha_j)_{i \leq j \leq n-1}} \mathbb{E} \left\{ \sum_{j=i}^{n-1} \Psi_j^h(A_j^v, B_j^w, D_j) \mid S_i = S, A_i^v = a, B_i^w = b, B_i^{rol} = c, r_i = r, D_i = D \right\},$$

avec :

$$\Psi_j^h(\alpha, a, b, D) = D[(\gamma_j b - (a)_+)_+ - \beta_j^h (a)_+].$$

De manière similaire au chapitre 2, en exploitant la propriété d'homogénéité de degré 1 des flux Ψ_j^h en a et b , on peut montrer que P^h est homogène en l'*account value* et la base et qu'il existe une stratégie de rachat de type Bang-Bang correspondant au pire cas.

Théorème 6.1.1 *Pour le point de vue du hedger, il existe une stratégie markovienne non randomisée $(\alpha_i^*)_{0 \leq i \leq n-1}$ correspondant au pire cas. De plus, en chaque date i qui correspond à une date de rachat, le contrôle optimal est de type bang-bang. En notant par $x = (S, a, b, c, r, D)$ l'état de la chaîne de Markov en T_i , on a :*

$$\alpha_i^*(S, a, b, c, r, D) = \begin{cases} \mathbf{1}_{\{\mathbb{E}(P^h(i+1, S_{i+1}, A_{i+1}^{v,i,x}, B_{i+1}^{w,i,x}, B_{i+1}^{rol,i,x}, r_{i+1}^{i,x}, D_{i+1}^{i,x})) \leq 0\}} & (\text{si } T_i \text{ est une date de rachat}), \\ 0 & (\text{sinon}). \end{cases}$$

6.1.1 Comparaison entre les points de vue client rationnel et *hedger*

Il est important de noter que le pire cas du couvreur n'est pas l'équivalent du meilleur cas du client rationnel. En effet, on a vu dans le chapitre 2 que l'assuré rationnel maximise en espérance à l'état (i, S, a, b, c, r, D) la valeur actualisée de la somme des flux réellement perçus ou versés qui s'écrit :

$$\sum_{j=i}^{n-1} [(\alpha_j - \beta_j)(A_j^v)_+ + \gamma_j B_j^w] + (A_n^v)_+.$$

Ainsi, il existe deux stratégies, une correspondant au meilleur cas pour le client rationnel et l'autre au pire cas pour le *hedger*. En pratique, le point de vue intéressant est celui du *hedger* car avec des contrats réels, les *fees* prélevés étant assez élevés (β grand), la stratégie optimale du client rationnel est de ne pas rentrer dans le contrat. C'est pourquoi on s'intéresse dans la suite uniquement au point de vue *hedger*.

6.1.2 Algorithme de programmation dynamique

Notons,

$$\begin{aligned} X_i &= (S_i, A_i^v, B_i^w, r_i, D_i), \\ P_i^h(X_i) &= P^h(i, X_i), \\ \Psi_i^h(X_i) &= D_i[(\gamma_i B_i^w - (A_i^v)_+)_+ - \beta_i^h (A_i^v)_+]. \end{aligned}$$

Grâce au théorème 6.1.1, l'algorithme de programmation dynamique est celui d'un problème d'arrêt optimal qui s'écrit :

$$\begin{aligned} \tau_n &= n, \\ P_n^h(X_n) &= 0, \\ \tau_i &= \begin{cases} i \mathbf{1}_{\{E[P_{i+1}^h(X_{i+1})|X_i] \leq 0\}} + \tau_{i+1} \mathbf{1}_{\{E[P_{i+1}^h(X_{i+1})|X_i] > 0\}} & \text{si } T_i \text{ est une date de rachat } (i < n) \\ \tau_{i+1} & \text{sinon} \end{cases}, \\ P_i^h(X_i) &= \mathbb{E}\left[\sum_{j=i}^{\tau_i} \Psi_j^h(X_j) | X_i\right], i < n. \end{aligned}$$

6.2 Résolution numérique : Méthode de Monte Carlo de type Longstaff Schwartz

6.2.1 Description

On effectue N simulations de $(X_i)_{1 \leq i \leq n}$ qu'on note $(X_i^m)_{1 \leq i \leq n, 1 \leq m \leq N}$. L'idée est de déterminer pour chaque date t_i le temps d'arrêt optimal τ_i^m pour le scénario m . Ensuite, le prix à l'instant initial est calculé par une approximation Monte Carlo comme suit :

$$\frac{1}{N} \sum_{m=1}^N \sum_{i=1}^{\tau_1^m} \Psi_i^h(X_i^m).$$

Il reste maintenant à expliquer comment on obtient $(\tau_i^m)_{1 \leq i \leq n, 1 \leq m \leq N}$. Pour cela on approche la quantité

$$R(i, x) = \mathbb{E}[P_{i+1}^h(X_{i+1}) | X_i = x] = \mathbb{E}\left[\sum_{j=i+1}^{\tau_{i+1}} \Psi_j^h(X_j) | X_i = x\right]$$

qui représente la valeur en t_i du prix du contrat à la date t_{i+1} par une méthode de régression empirique. Notons $\tilde{R}(i, x)$ cette approximation. Ainsi, on calcule $(\tau_i^m)_{1 \leq i \leq n, 1 \leq m \leq N}$ par :

$$\begin{aligned} \tau_n^m &= n, \\ \tau_i^m &= \begin{cases} i \mathbf{1}_{\{\tilde{R}(i, X_i^m) \leq 0\}} + \tau_{i+1}^m \mathbf{1}_{\{E[\tilde{R}(i, X_i^m)] > 0\}} & \text{si } T_i \text{ est une date de rachat } (i < n) \quad , \\ \tau_{i+1}^m & \text{sinon} \quad . \end{cases} \end{aligned}$$

$\tilde{R}(i, x)$ est calculé en utilisant les méthodes de régression empirique (standard, VCP1, VCP2 et VCP3) étudiées dans le chapitre précédent comme suit.

Notons

$$Z_i = (X_i, X_{i+1}), Z_i^m = (X_i^m, X_{i+1}^m), 1 \leq j < n, 1 \leq m < N.$$

Soit $(P_k)_{1 \leq k \leq K_1}$ une famille de fonctions vérifiant

$$\mathbb{E}[P_k(Z_i) | X_i] = 0, \quad 1 \leq k \leq K_1, 1 \leq i < n.$$

Soit $(\Phi_k)_{1 \leq k \leq K_2}$ une famille de fonctions de base.

Alors on estime $R(i, \cdot)$ par $\tilde{R}(i, \cdot) = \sum_{k=1}^{K_2} \tilde{\beta}_k^i \Phi_k$ où les coefficients $(\tilde{\beta}_k^i)$ sont calculés par les quatre façons suivantes.

Méthode standard

On calcule les coefficients de régression par moindres carrés $(\tilde{\beta}_k^i)_{1 \leq k \leq K_2}$ vérifiant :

$$(\tilde{\beta}_k^i)_k = \arg \inf_{(\beta_k)_k} \frac{1}{N} \sum_{m=1}^N \left| \sum_{j=i}^{\tau_{i+1}^m-1} \Psi_{j+1}^h(X_{j+1}^m) - \sum_{k=1}^{K_2} \beta_k \Phi_k(X_i^m) \right|^2.$$

Méthode VCP 1

Ici on détermine les coefficients $(\tilde{\alpha}_k^i)_{1 \leq k \leq K_1}$ tels que :

$$(\tilde{\alpha}_k^i)_k = \arg \inf_{(\alpha_k)_k} \frac{1}{N} \sum_{m=1}^N \left| \sum_{j=i}^{\tau_{i+1}^m-1} \Psi_{j+1}^h(X_{j+1}^m) - \sum_{k=1}^{K_1} \alpha_k P_k(Z_i^m) \right|^2.$$

Ensuite on calcule les coefficients de régression par moindres carrés $(\tilde{\beta}_k^i)_{1 \leq k \leq K_2}$ vérifiant :

$$(\tilde{\beta}_k^i)_k = \arg \inf_{(\beta_k)_k} \frac{1}{N} \sum_{m=1}^N \left| \sum_{j=i}^{\tau_{i+1}^m-1} \Psi_{j+1}^h(X_{j+1}^m) - \sum_{k=1}^{K_1} \alpha_k P_k(Z_i^m) - \sum_{k=1}^{K_2} \beta_k \Phi_k(X_i^m) \right|^2.$$

Méthode VCP 2

Pour $i = n - 1, \dots, 1$,

$$(\tilde{\alpha}_k^i)_k = \arg \inf_{(\alpha_k)_k} \frac{1}{N} \sum_{m=1}^N \left| \Psi_{i+1}^h(X_{i+1}^m) - \sum_{k=1}^{K_1} \alpha_k P_k(Z_i^m) \right|^2,$$

$$(\tilde{\beta}_k^i)_k = \arg \inf_{(\beta_k)_k} \frac{1}{N} \sum_{m=1}^N \left| \sum_{j=i}^{\tau_{i+1}^m-1} \left[\Psi_{j+1}^h(X_{j+1}^m) - \sum_{k=1}^{K_1} \tilde{\alpha}_k^j P_k(Z_j^m) \right] - \sum_{k=1}^{K_2} \beta_k \Phi_k(X_i^m) \right|^2.$$

Méthode VCP 3

Pour $i = n - 1, \dots, 1$,

$$(\tilde{\alpha}_k^i)_k = \arg \inf_{(\alpha_k)_k} \frac{1}{N} \sum_{m=1}^N \left| \sum_{j=i+1}^{\tau_{i+1}^m-1} \left[\Psi_{j+1}^h(X_{j+1}^m) - \sum_{k=1}^{K_1} \tilde{\alpha}_k^j P_k(Z_j^m) \right] + \Psi_{i+1}^h(X_{i+1}^m) - \sum_{k=1}^{K_1} \alpha_k P_k(Z_i^m) \right|^2,$$

$$(\tilde{\beta}_k^i)_k = \arg \inf_{(\beta_k)_k} \frac{1}{N} \sum_{m=1}^N \left| \sum_{j=i}^{\tau_{i+1}^m-1} \left[\Psi_{j+1}^h(X_{j+1}^m) - \sum_{k=1}^{K_1} \tilde{\alpha}_k^j P_k(Z_j^m) \right] - \sum_{k=1}^{K_2} \beta_k \Phi_k(X_i^m) \right|^2.$$

6.2.2 Résultats numériques

Pour les tests numériques, on considère un contrat GMWB avec les caractéristiques suivantes :

- 1/ $P_0 = 20000$, $\beta^h = 1\%$, $\beta^m = 1\%$ et $\gamma = 5\%$.
- 2/ Le rachat du contrat est possible annuellement à partir de l'année 5.
- 3/ Les *fees*, retraits et rachats ont lieu à une fréquence annuelle à chaque date d'anniversaire du contrat.
- 4/ Le ratchet a lieu une seule fois, à la date T^{def} .

On prend $\sigma(t, S_t) = \sigma = 15\%$, $\sigma_t^r = \sigma^r = 0.00583935$, $a = 0.0248547$ et $\rho = 0$ et on varie N dans l'ensemble $\{1000, 5000, 1000, 20000\}$. La courbe du Zéro Coupon et celle des taux utilisées sont tracées dans le graphe 6.1.

Notons que dans ce cas la dynamique de l'*account value* devient :

$$\begin{aligned} A_0^v &= P_0, \\ dA_t^v &= (A_{t-}^v)_+ (r_t dt + \sigma dW_t) - \beta (A_{t-}^v)_+ dZ_t^f - \gamma B_{t-}^w dZ_t^w \quad \text{avant rachat,} \\ A_t^v &= 0 \quad \text{après rachat.} \end{aligned}$$

Par suite la variable sous-jacent S peut être éliminée des variables markoviennes du problème.

a) D'abord on prend $T = 25$ et $T^{def} = 5$.

Dans la figure 6.2, on choisit $K_2 = 3$ et 6 et visualise en fonction de N le coût de couverture du contrat calculé par la méthode standard ainsi que les bornes I_{cmin} et I_{cmax} de l'intervalle de confiance à 95% associé. On choisit Φ_k des deux formes suivantes.

- 1) Φ_k homogènes en la base et l'*account value* : $\Phi_k(X) = B(\frac{A}{B})^k$.
- 2) Φ_k non homogènes : $\Phi_k(X) = A^i B^j$.

On remarque qu'au fur et à mesure que le nombre de simulations augmente, l'intervalle de confiance associé à la forme homogène des Φ_k devient au dessus de celui de la forme non homogène. Ceci est en concordance avec le caractère homogène du coût de couverture en l'*account value* et la base. En plus, la forme 1 permet d'éviter un nouveau calcul si la prime initiale est multipliée par un facteur ce qui n'est pas le cas pour la deuxième forme exigeant un nouveau calcul des coefficients de régression.

Dans la figure 6.3 où on a utilisé $K_1 = 2$ et $K_2 = 8$ fonctions de base de la forme 1, on visualise les intervalles de confiance des coûts de couverture pour les

quatre méthodes pour des taux de *rollup* de 2.5% et de 1%. On remarque que les méthodes VCP1, VCP2 et VCP3 n'apportent pas d'amélioration et peuvent être moins bonnes que la méthode standard (c'est essentiellement le cas pour la VCP3). Ceci peut s'expliquer par le fait que le temps d'arrêt optimal n'est pas forcément le même pour tous les scénarii. Notons aussi que le coût de couverture augmente pour un taux de *rollup* supérieur ce qui est logique car le contrat devient avantageux pour l'assuré dans la mesure où il peut espérer toucher des retraits plus conséquents après T^{def} .

b) Ici on fixe $T = 29$ et $T^{def} = 9$.

Les graphes de la figure 6.4 (utilisation de $K_1 = 2$ et $K_2 = 8$ fonctions de base de la forme 1) tracent les intervalles de confiance des coûts de couverture pour les quatre méthodes respectivement pour des taux de *rollup* de 1% et de 2.5%. Les mêmes conclusions de la figure 6.3 restent valables. D'autre part, le coût de couverture est inférieur à celui de la figure d'avant ce qui est dû au fait que l'assuré doit attendre plus avant de toucher son premier retrait alors que les *fees* débutent toujours dès l'année 1.

6.2.3 Perspectives

Lest tests numériques ont montré que la méthode VCP3, qui a donné des résultats satisfaisants dans le chapitre 5, est moins performante que la méthode standard. Ceci peut s'expliquer par le fait que les coefficients $(\tilde{\alpha}^j)_{i \leq j \leq n}$ calculés à l'itération i de manière à réduire la variance empirique de la somme des flux futurs jusqu'au temps d'arrêt τ_{i+1} sont les mêmes utilisés dans les itérations suivantes : ils peuvent être inadaptées à l'itération $i - 1$ lorsqu'on s'intéresse à réduire la variance empirique des flux jusqu'au temps d'arrêt τ_i . Une manière de résoudre ce problème est d'aménager la méthode VCP3 en recalculant à chaque itération i de nouveaux coefficients $(\tilde{\alpha}^j)_{i \leq j \leq n}$ en ne tenant compte que des flux utiles. Ainsi on propose la variante suivante :

Méthode VCP 3 modifiée

Pour $i = n - 1, \dots, 1$,

Pour $l = n - 1, \dots, i$,

$$(\tilde{\alpha}_k^{i,l})_k = \arg \inf_{(\alpha_k)_k} \frac{1}{N} \sum_{m=1}^N \left| \sum_{j=l+1}^{\tau_{i+1}^m - 1} \left[\Psi_{j+1}^h(X_{j+1}^m) - \sum_{k=1}^{K_1} \tilde{\alpha}_k^{i,j} P_k(Z_j^m) \right] + \Psi_{l+1}^h(X_{l+1}^m) \mathbf{1}_{\{l+1 \leq \tau_{i+1}^m\}} - \sum_{k=1}^{K_1} \alpha_k P_k(Z_l^m) \right|^2,$$

$$(\tilde{\beta}_k^i)_k = \arg \inf_{(\beta_k)_k} \frac{1}{N} \sum_{m=1}^N \left| \sum_{j=i}^{\tau_{i+1}^m - 1} \left[\Psi_{j+1}^h(X_{j+1}^m) - \sum_{k=1}^{K_1} \tilde{\alpha}_k^{i,j} P_k(Z_i^j) \right] - \sum_{k=1}^{K_2} \beta_k \Phi_k(X_i^m) \right|^2.$$

Des tests numériques sont en cours pour voir l'amélioration éventuelle de cette nouvelle méthode.

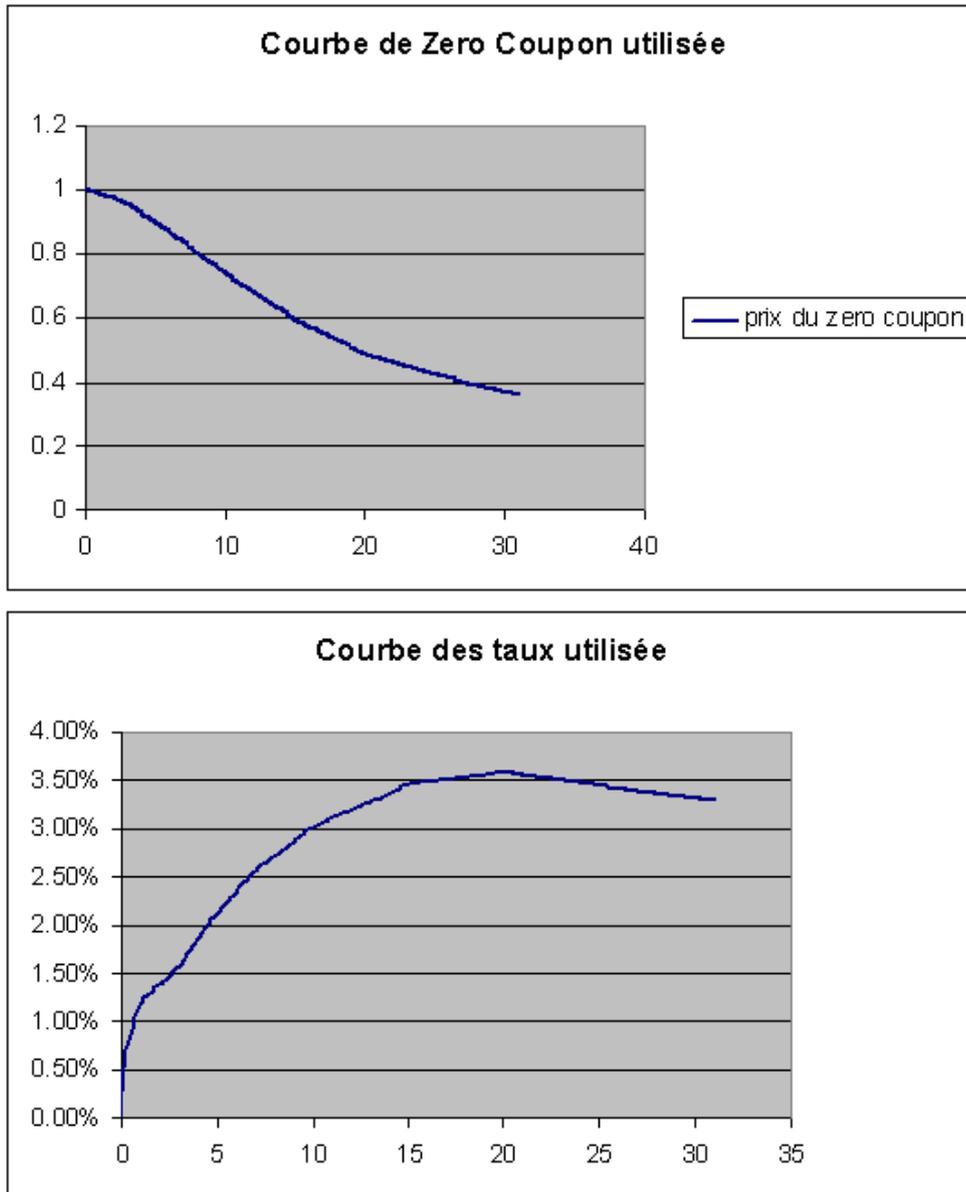


FIGURE 6.1 – Courbe du Zero Coupon et courbe des taux utilisées.

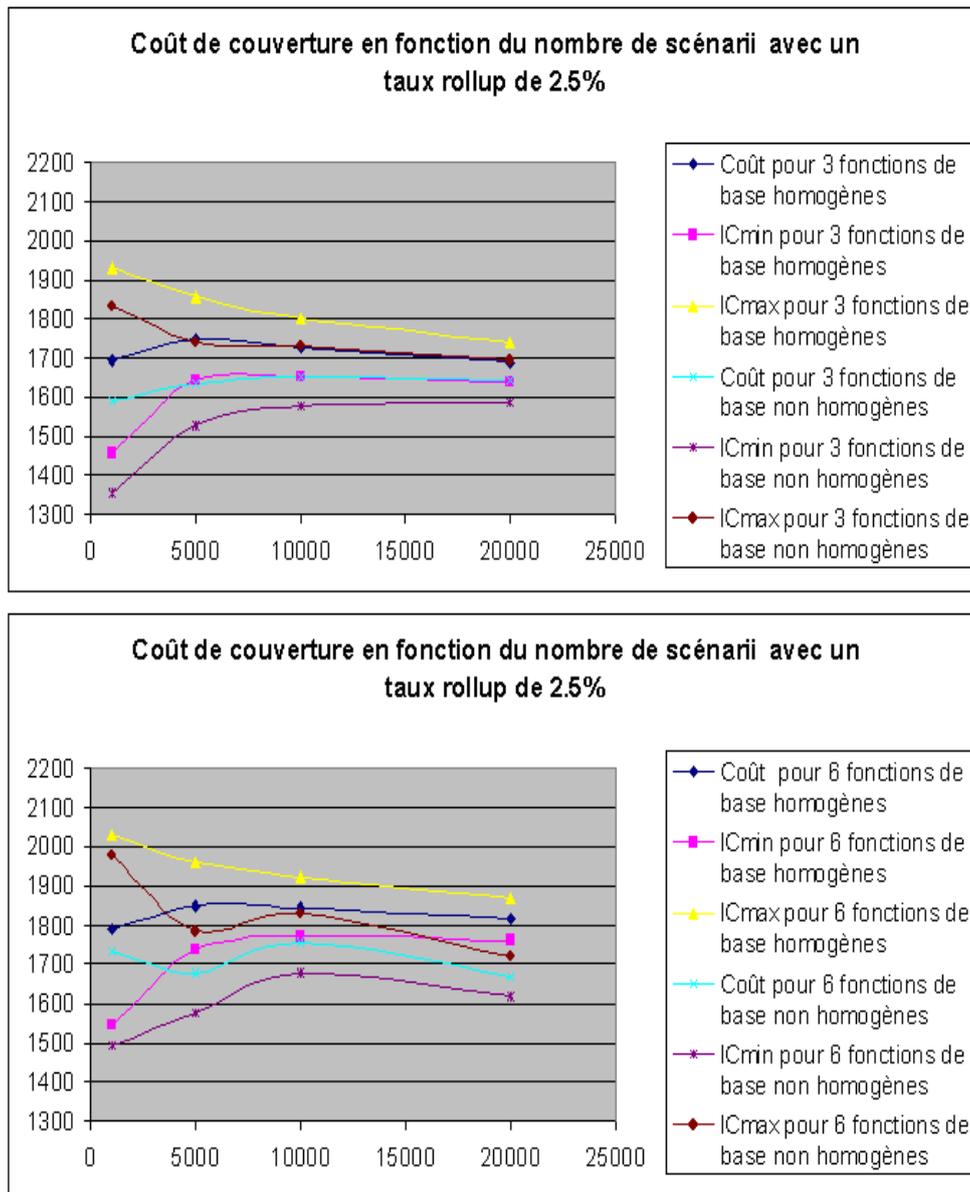


FIGURE 6.2 – Comparaison entre le cas homogène et non homogène des fonctions de bases dans le cadre de la méthode standard avec $T^{def} = 5$.

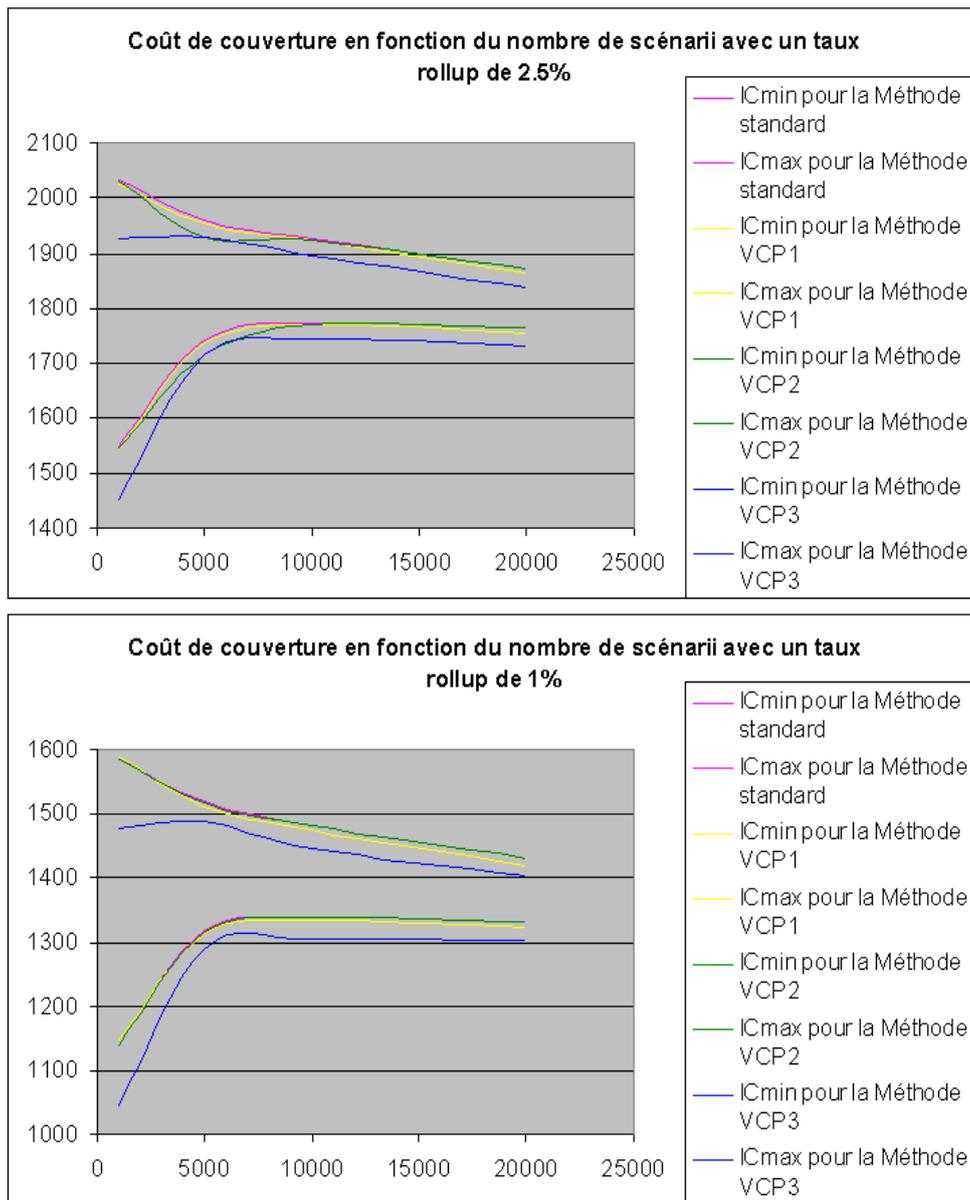


FIGURE 6.3 – Intervalle de confiance du coût de couverture pour les différentes méthodes avec $T^{def} = 5$.

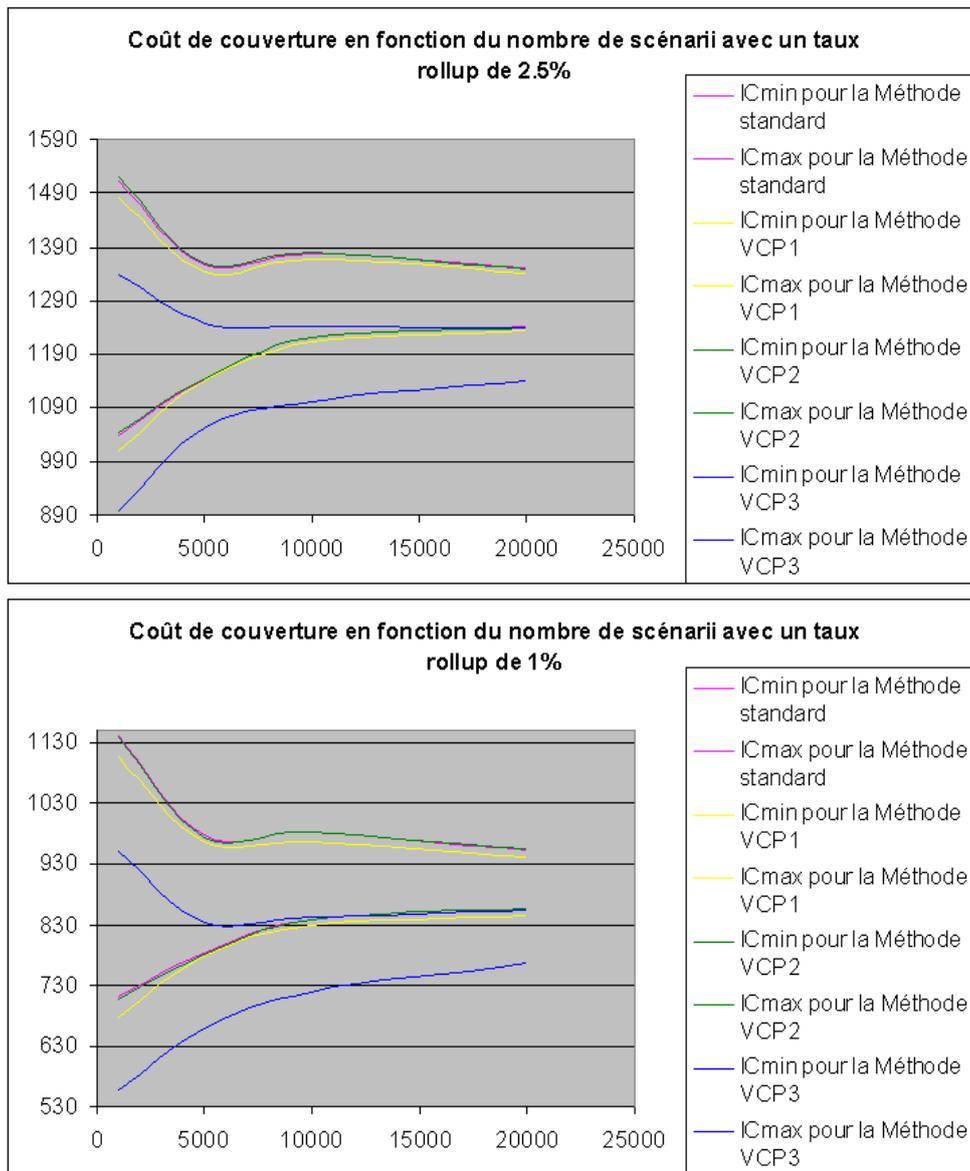


FIGURE 6.4 – Intervalle de confiance du coût de couverture pour les différentes méthodes avec $T^{def} = 9$.

Chapitre 7

Valorisation par formules fermées d'un contrat GMWB à exercice européen

Ce chapitre a pour objectif de proposer des formules fermées pour valoriser des contrats GMWB dans un cadre européen. En effet, les prix calculés par ces formules peuvent être exploités pour calculer des poids optimaux dans un algorithme de type Weighted-Monte Carlo (voir [ABF⁺01]). Ces poids seront ensuite utilisés pour valoriser des contrats GMWB plus compliqués (avec rachat anticipé, ratchet, etc.).

7.1 Formulation du problème

On se place dans le cadre des hypothèses suivantes :

- Absence de coût de transaction et d'opportunité d'arbitrage dans un marché complet .
- La vente à découvert est autorisée.
- Les retraits des souscripteurs et le pourcentage des *fees* (charges) sont supposés déterministes et se produisant à des dates discrètes.
- Le portefeuille sous-jacent de référence a la dynamique d'un mouvement brownien géométrique à volatilité déterministe.
- Le taux d'intérêt est un processus stochastique de type Hull-White à 1 facteur.

Notons par S_t la valeur du portefeuille de référence. Sous la mesure risque neutre \mathbb{Q} , on a :

$$\begin{aligned}\frac{dS_t}{S_t} &= r_t dt + \sigma_t(\rho dW_t + \sqrt{1-\rho^2} dB_t), \\ dr_t &= (\theta_t - ar_t)dt + \sigma_r^t dW_t\end{aligned}$$

où : W et B sont deux mouvements browniens standards, θ_t , a et σ_r^t sont des paramètres déterministes, σ_t est la volatilité déterministe de S et ρ est la corrélation instantanée entre S et r .

Soit $A^{v,y,\delta}$ l'*account value* (valeur du compte) du client définie comme suit :

- Initialement, le client investit la valeur $A_0^{v,y,\delta}$;
- Il reçoit les retraits $(\delta_i)_{1 \leq i \leq n}$ à des dates fixes $(t_i)_{1 \leq i \leq n}$;
- Il reçoit le capital restant $(A_T^{v,y,\delta})_+$ à la maturité $T = t_{n+1}$;
- Entre deux dates, l'*account value* $A^{v,y,\delta}$ et S ont le même rendement.
- L'assureur reçoit les *fees* (prises de l'*account value*) en $(t_i)_{1 \leq i \leq n}$, qui sont les pourcentages $(\alpha_i)_{1 \leq i \leq n}$ de l'*account value* si elle est positive ou zéro sinon. Les *fees* sont fixées de manière à ce que la valeur initiale risque neutre du contrat soit égale à $A_0^{v,y,\delta}$.

La maturité T est grande (10 ans ou plus). Soit $t_0 = 0$. La dynamique de $A^{v,y,\delta}$ s'écrit :

$$\begin{aligned}\frac{dA_t^{v,y,\delta}}{A_t^{v,y,\delta}} &= \frac{dS_t}{S_t} \quad \text{si } t_i \leq t < t_{i+1}, 1 \leq i \leq n, \\ A_{t_i}^{v,y,\delta} &= A_{t_i^-}^{v,y,\delta} - \alpha_i(A_{t_i^-}^{v,y,\delta})_+ - \delta_i \quad \text{pour } 1 \leq i \leq n+1.\end{aligned}$$

Pour être précis dans nos définitions, il faudrait laisser $A_t^{v,y,\delta}$ constant une fois qu'une valeur négative est atteinte : en réalité, puisque les flux ne concernent que les valeurs positives de l'*account value*, la modification proposée ci-dessus est équivalente. En notant $D_t = e^{-\int_0^t r_s ds}$, le prix du contrat est donné par :

$$\text{GMWB}_0 = \mathbb{E} \left[D_T (A_T^{v,y,\delta})_+ + \sum_{i=1}^n D_{t_i} \delta_i - \sum_{i=1}^n D_{t_i} \alpha_i (A_{t_i^-}^{v,y,\delta})_+ \right].$$

7.2 Formules fermées d'approximation

Remarquons qu'une fois $A^{v,y,\delta}$ devienne négative, elle ne peut plus redevenir positive (car $\delta_i \geq 0$ et $\alpha_i \in [0, 1]$). Ainsi, puisque le payoff inclut uniquement les valeurs positives de $A^{v,y,\delta}$, on peut supposer pour le calcul du prix GMWB_0 que :

$$A_{t_i}^{v,y,\delta} = (1 - \alpha_i)A_{t_i^-}^{v,y,\delta} - \delta_i \quad \text{pour } 1 \leq i \leq n.$$

Le prix devient :

$$\text{GMWB}_0 = \sum_{i=1}^n \delta_i B(0, t_i) + \mathbb{E} \left[D_T(A_T^{v,y,\delta})_+ \right] - \sum_{i=1}^n \alpha_i \mathbb{E} \left[D_{t_i}(A_{t_i^-}^{v,y,\delta})_+ \right] \quad (7.2.1)$$

où $B(t, s)$ est le prix en t du zéro-coupon qui commence et de maturité s .

Ainsi, le pricing revient à calculer les termes :

$$\mathbb{E} \left[D_T(A_T^{v,y,\delta})_+ \right] \quad \text{et} \quad \mathbb{E} \left[D_{t_i}(A_{t_i^-}^{v,y,\delta})_+ \right], 1 \leq i \leq n$$

où $A^{v,y,\delta}$ a une dynamique log normale avec des sauts affines en t_i . C'est similaire à un sous-jacent versant des dividendes discrets affines, et nous nous inspirons de [EG12] pour calculer des formules approchées dans une asymptotique de petits retraits. Par rapport à [EG12], la nouveauté vient des taux d'intérêt stochastiques. Pour simplifier, on introduit les notations suivantes :

$$\begin{aligned} \Gamma(t, u) &= \int_t^u \sigma_s^r e^{-a(u-s)} ds, \\ \pi_{j,i} &= (1 - \alpha_i) \dots (1 - \alpha_{j+1}) \quad \text{pour } j \leq i, \\ K^{y,\delta} &= \sum_{j=1}^n \delta_j \pi_{j,n} \mathbb{E} \left[\frac{S_T}{S_{t_j}} \right], \\ \hat{\delta}_i^j &= \delta_i \pi_{i,j}. \end{aligned}$$

Pour alléger les écritures, on suppose $A_0^{v,y,\delta} = S_0$ et on pose $\hat{\delta}_i = \hat{\delta}_i^n$.

Enfin, rappelons ici que la dynamique des zéro-coupons s'écrit (voir chapitres 3 et 5 dans [BM07]) :

$$B(t, s) = B(0, s) \exp \left(\int_0^t r_u du - \frac{1}{2} \int_0^t \Gamma(u, s)^2 du - \int_0^t \Gamma(u, s) dW_u \right), t \leq s.$$

7.2.1 Résultats préliminaires

Avec les notations précédentes, on déduit :

$$A_t^{v,y,\delta} = \begin{cases} S_t & \text{si } t < t_1 \\ A_{t_i}^{v,y,\delta} \frac{S_t}{S_{t_i}} = (1 - \alpha_i) A_{t_i^-}^{v,y,\delta} \frac{S_t}{S_{t_i}} - \delta_i \frac{S_t}{S_{t_i}} & \text{si } t_i \leq t < t_{i+1}, i < n \quad \text{ou } t_n \leq t \leq T. \end{cases}$$

D'où, par récurrence (voir annexe), on a le lemme suivant :

Lemme 7.2.1 *Pour $0 \leq i \leq n$, on a*

$$A_{t_{i+1}-}^{v,y,\delta} = \pi_{0,i} S_{t_{i+1}} - \sum_{j=1}^i \delta_j \pi_{j,i} \frac{S_{t_{i+1}}}{S_{t_j}},$$

$$A_T^{v,y,\delta} = \pi_{0,n} S_T - \sum_{i=1}^n \delta_i \pi_{i,n} \frac{S_T}{S_{t_i}}.$$

Puisque la même méthodologie peut être utilisée pour toutes les dates dans l'expression (7.2.1) de $GMWB_0$, on explique seulement comment calculer $\mathbb{E} \left[D_T(A_T^{v,y,\delta})_+ \right]$.

Le cas associé à l'absence de retraits ($\delta_i = 0$) et pour lequel $A_T^{v,y,\delta} = \pi_{0,n} S_T$ est important. En effet, dans ce cas $A_T^{v,y,\delta}$ étant une variable aléatoire log-normale, la quantité d'intérêt $\mathbb{E} \left[D_T(A_T^{v,y,\delta})_+ \right]$ est explicite est vaut $\pi_{0,n} S_0$.

Ce cas permet de trouver un *proxy* pour notre modèle de départ. Ce *proxy* n'est pas donné directement par le modèle avec $\delta_i = 0$, mais plutôt par ce modèle translaté par l'espérance des retraits déterministes. En d'autres termes, par le lemme 7.2.1, le *proxy* est défini par :

$$\bar{A}_T^{v,y,\delta} = \pi_{0,n} S_T - \mathbb{E} \left[\sum_{i=1}^n \hat{\delta}_i \frac{S_T}{S_{t_i}} \right] = \pi_{0,n} S_T - \sum_{i=1}^n \hat{\delta}_i \mathbb{E} \left[\frac{S_T}{S_{t_i}} \right], \quad (7.2.2)$$

où $\mathbb{E} \left[\frac{S_T}{S_{t_i}} \right]$ peut être calculée explicitement selon le lemme suivant (voir la preuve dans l'annexe) :

Lemme 7.2.2 *Pour $1 \leq i \leq n + 1$, on a :*

$$\mathbb{E} \left[\frac{S_T}{S_{t_i}} \right] = \frac{B(0, t_i)}{B(0, T)} \exp \left(\int_0^{t_i} [\Gamma(s, T)^2 - \Gamma(s, t_i)\Gamma(s, T)] ds + \int_{t_i}^T [\Gamma(s, T)^2 - \rho \sigma_s \Gamma(s, T)] ds \right).$$

Si $\sigma_t = \sigma$ et $\sigma_t^r = \sigma^r$, alors $\mathbb{E} \left[\frac{S_T}{S_{t_i}} \right]$ est égale à :

$$\frac{B(0, t_i)}{B(0, T)} \exp \left(\left(\frac{\sigma_r}{a} \right)^2 \left[T - t_i + \frac{e^{-aT}}{a} - \frac{e^{-at_i}}{a} - \frac{e^{-2aT}}{2a} + \frac{e^{-a(T-t_i)}}{2a} + \frac{e^{-a(T+t_i)}}{2a} - \frac{1}{2a} \right] + \rho \sigma \sigma^r \left[T - t_i + \frac{e^{-a(T-t_i)}}{a} - \frac{1}{a} \right] \right).$$

Ainsi, $\bar{A}_T^{v,y,\delta}$ est une variable aléatoire log-normale translatée, et par suite le calcul de $\mathbb{E}[e^{-\int_0^T r_s ds} (\bar{A}_T^{v,y,\delta})_+]$ demeure explicite en utilisant la variable strike translatée $K^{y,\delta}$ dans la formule de Black :

$$\mathbb{E}[e^{-\int_0^T r_s ds} (\bar{A}_T^{v,y,\delta})_+] = \text{Call}^{BS}(\pi_{0,n} S_0, K^{y,\delta}),$$

où

$$\begin{aligned} \text{Call}^{BS}(x, k) &= x\mathcal{N}\left[d_+\left(\frac{x}{B(0,T)}, k\right)\right] - kB(0,T)\mathcal{N}\left[d_-\left(\frac{x}{B(0,T)}, k\right)\right], \\ d_{\pm}(x, k) &= \frac{\log\left(\frac{x}{k}\right) \pm \frac{1}{2}\Sigma^2 T}{\Sigma\sqrt{T}}, \\ \Sigma^2 &= \frac{1}{T} \int_0^T (\Gamma(s,T)^2 + 2\rho\sigma_s\Gamma(s,T) + \sigma_s^2) ds, \end{aligned}$$

\mathcal{N} étant la fonction de répartition de la loi normale centrée réduite. La quantité ci-dessus représente le terme principal dans notre formule de développement asymptotique de $\mathbb{E}[e^{-\int_0^T r_s ds} (A_T^{v,y,\delta})_+]$.

Notre but maintenant est de fournir des termes correctifs permettant l'obtention d'une bonne précision.

7.2.2 Résultats principaux

Notre méthode est basée sur le développement de Taylor et le calcul astucieux des termes correctifs via le *proxy* $\bar{A}_T^{v,y,\delta}$. Par le lemme 7.2.1 et l'égalité (7.2.2), on obtient la distance au *proxy* :

$$A_T^{v,y,\delta} - \bar{A}_T^{v,y,\delta} = - \sum_{i=1}^n \hat{\delta}_i^n \left(\frac{S_T}{S_{t_i}} - \mathbb{E} \left[\frac{S_T}{S_{t_i}} \right] \right).$$

Nous cherchons à approcher $\mathbb{E}[e^{-\int_0^T r_s ds} h(A_T^{v,y,\delta})]$ pour $h(x) = (x)_+$ par des formules explicites avec une estimation de l'erreur. Ceci est simple quand h est plus régulière que la fonction partie positive. On traite ici uniquement ce cas. L'extension pour $h(x) = (x)_+$ est similaire à [EG12] et nous ne le détaillons pas ici.

Développement de Taylor pour une fonction régulière h

On dit que h a un degré de régularité k ($k \geq 1$) si la propriété (H_k) suivante est vérifiée :

(H_k) : La fonction h est $(k-1)$ fois continuellement différentiable et sa $(k-1)$ -ième dérivée est différentiable presque partout. En plus, les dérivées sont bornées par des polynômes tels que pour des constantes positives C et p on a $|h(x)| + \sum_{j=1}^k |\partial_x^j h(x)| \leq$

$C(1 + |x|^p)$ pour tout $x \in \mathbb{R}$.

Avant d'énoncer le théorème pour la formule d'approximation, on a besoin des lemmes suivants :

Lemme 7.2.3 (Voir preuve dans l'annexe)

Pour $k \in \mathbb{N}$ et $(i_1, i_2, \dots, i_k) \in \{0, 1, \dots, n+1\}^k$, posons

$$c_{i_1, i_2, \dots, i_k} := \mathbb{E} \left[D_T \frac{S_T}{S_{t_{i_1}}} \frac{S_T}{S_{t_{i_2}}} \dots \frac{S_T}{S_{t_{i_k}}} \right].$$

Si $i_1 \leq i_2 \leq \dots \leq i_k$ alors,

$$\begin{aligned} & c_{i_1, i_2, \dots, i_k} \\ &= \frac{\prod_{j=1}^k B(0, t_{i_j})}{B(0, T)^{k-1}} \\ & \exp \left(\frac{1}{2} \sum_{j=0}^k \int_{t_{i_j}}^{t_{i_{j+1}}} \left[(k-1)\Gamma(s, T)^2 - j\sigma_s^2 - \sum_{l=j+1}^k \Gamma(s, t_{i_l})^2 \right] ds \right) \\ & \exp \left(\frac{1}{2} \sum_{j=0}^k \int_{t_{i_j}}^{t_{i_{j+1}}} \left[(k-1)\Gamma(s, T) + j\rho\sigma_s - \sum_{l=j+1}^k \Gamma(s, t_{i_l}) \right]^2 ds + \frac{1}{2} \sum_{j=0}^k \int_{t_{i_j}}^{t_{i_{j+1}}} j^2(1-\rho^2)\sigma_s^2 ds \right) \end{aligned}$$

avec la convention $t_{i_0} = 0$ et $t_{i_{k+1}} = T$.

Si non,

$$c_{i_1, i_2, \dots, i_k} = c_{\zeta(i_1), \zeta(i_2), \dots, \zeta(i_k)}$$

où ζ est une permutation transformant la suite (i_1, i_2, \dots, i_k) en une suite croissante $(\zeta(i_1), \zeta(i_2), \dots, \zeta(i_k))$.

En particulier,

$$c_i = B(0, t_i), 0 \leq i \leq n+1.$$

On attire l'attention que si $\sigma_t = \sigma$ et $\sigma_t^r = \sigma^r$, les intégrales dans c_{i_1, i_2, \dots, i_k} se calculent explicitement par la formule (7.4.3) (voir remarque à la fin de la preuve du lemme 7.2.3 en annexe). C'est cette formule qui a été utilisée lors des tests numériques.

Lemme 7.2.4 Notons pour $k \in \mathbb{N}^*$, $\mathbb{Q}_{t_{i_1}, \dots, t_{i_k}}$ la mesure de probabilité telle que

$$\frac{d\mathbb{Q}_{t_{i_1}, \dots, t_{i_k}}}{d\mathbb{Q}} \Big|_{\mathcal{F}_T} = \frac{D_T \frac{S_T}{S_{t_{i_1}}} \frac{S_T}{S_{t_{i_2}}} \dots \frac{S_T}{S_{t_{i_k}}}}{\mathbb{E} \left[D_T \frac{S_T}{S_{t_{i_1}}} \frac{S_T}{S_{t_{i_2}}} \dots \frac{S_T}{S_{t_{i_k}}} \right]}.$$

Alors, la loi du proxy $\bar{A}_T^{v,y,\delta}$ sous $\mathbb{Q}_{t_{i_1}, \dots, t_{i_k}}$ est la même que celle de la variable log-normale translatée

$$\pi_{0,n} \frac{c_{0,i_1, \dots, i_k}}{c_{i_1, \dots, i_k}} \tilde{S}_T - K^{y,\delta}$$

sous \mathbb{Q} où :

\tilde{S} est un actif fictif log normal sans drift de volatilité Σ , valant S_0 à $t = 0$.

Remarque

- La caractéristique log-normal de $\frac{d\mathbb{Q}_{t_{i_1}, \dots, t_{i_k}}}{d\mathbb{Q}} \Big|_{\mathcal{F}_T}$ est une propriété clé qui induit un changement de probabilité conservant le caractère gaussien des variables aléatoires.
- Pour $t_{i_1} = T$, $\mathbb{Q}_{t_{i_1}}$ coïncide avec la mesure de probabilité habituelle forward neutre de maturité T .
- Pour $t_{i_1} = 0$, $\mathbb{Q}_{t_{i_1}}$ correspond à la mesure de probabilité associée au numéraire S .

Preuve du Lemme 7.2.4

D'après (7.2.2) on a $\bar{A}_T^{v,y,\delta} = \pi_{0,n} S_T - K^{y,\delta}$. Ainsi, sous $\mathbb{Q}_{t_{i_1}, \dots, t_{i_k}}$ la loi de $\bar{A}_T^{v,y,\delta}$ demeure celle d'une variable log-normale translatée de volatilité inchangée et de de moyenne égale à :

$$\begin{aligned} \mathbb{E}_{\mathbb{Q}_{t_{i_1}, \dots, t_{i_k}}} [\bar{A}_T^{v,y,\delta}] &= \pi_{0,n} \mathbb{E}_{\mathbb{Q}_{t_{i_1}, \dots, t_{i_k}}} [S_T] - K^{y,\delta} \\ &= \pi_{0,n} \frac{\mathbb{E} \left[S_T D_T \frac{S_T}{S_{t_{i_1}}} \frac{S_T}{S_{t_{i_2}}} \dots \frac{S_T}{S_{t_{i_k}}} \right]}{\mathbb{E} \left[D_T \frac{S_T}{S_{t_{i_1}}} \frac{S_T}{S_{t_{i_2}}} \dots \frac{S_T}{S_{t_{i_k}}} \right]} - K^{y,\delta} \\ &= \pi_{0,n} S_0 \frac{c_{0,i_1, \dots, i_k}}{c_{i_1, \dots, i_k}} - K^{y,\delta}. \end{aligned}$$

D'où le le résultat du lemme. \square

Théorème 7.2.5 Pour une fonction régulière h vérifiant (\mathbf{H}_2) , on a

$$\begin{aligned} &\mathbb{E} \left[D_T h(A_T^{v,y,\delta}) \right] \\ &= B(0, T) \mathbb{E} \left[h \left(\frac{\pi_{0,n}}{B(0, T)} \tilde{S}_T - K^{y,\delta} \right) \right] + \sum_{i=1}^n \hat{\delta}_i B(0, t_i) \partial_K \mathbb{E} \left[h \left(\pi_{0,n} \frac{c_{0,i}}{B(0, t_i)} \tilde{S}_T - K^{y,\delta} \right) \right] \\ &\quad - \sum_{i=1}^n \hat{\delta}_i \mathbb{E} \left[\frac{S_T}{S_{t_i}} \right] B(0, T) \partial_K \mathbb{E} \left[h \left(\frac{\pi_{0,n}}{B(0, T)} \tilde{S}_T - K^{y,\delta} \right) \right] \\ &\quad + \text{Error}_2(h). \end{aligned}$$

Notons que dans le terme de droite de l'égalité ci-dessus, la fonction h est évaluée en une variable aléatoire log-normale translatée ce qui permet de simplifier le calcul numérique. En outre, $\mathbb{E} \left[\frac{S_T}{S_{t_i}} \right]$, $K^{y,\delta}$ et $c_{0,i}$ sont explicites selon les lemmes 7.2.2 et 7.2.3.

Preuve. Notre objectif est d'approcher $\mathbb{E} \left[D_T h(A_T^{v,y,\delta}) \right]$ pour des fonctions h vérifiant (\mathbf{H}_2) . En utilisant le développement de Taylor au premier ordre, on a

$$\begin{aligned} \mathbb{E} \left[D_T h(A_T^{v,y,\delta}) \right] &= \mathbb{E} \left[e^{-\int_0^T r_s ds} h(\bar{A}_T^{v,y,\delta}) \right] \\ &\quad - \sum_{i=1}^n \hat{\delta}_i \mathbb{E} \left[e^{-\int_0^T r_s ds} \left(\frac{S_T}{S_{t_i}} - \mathbb{E} \left[\frac{S_T}{S_{t_i}} \right] \right) h'(\bar{A}_T^{v,y,\delta}) \right] \quad (7.2.3) \\ &\quad + \text{Error}_2(h). \end{aligned}$$

On a

$$\mathbb{E} \left[e^{-\int_0^T r_s ds} h(\bar{A}_T^{v,y,\delta}) \right] = B(0, T) \mathbb{E}_{\mathbb{Q}_T} \left[h(\bar{A}_T^{v,y,\delta}) \right].$$

Selon le lemme 7.2.4, sous \mathbb{Q}_T la loi de $\bar{A}_T^{v,y,\delta}$ est la même que celle de $\pi_{0,n} \frac{\tilde{S}_T}{B(0,T)} - K^{y,\delta}$ sous \mathbb{Q} . D'où

$$\mathbb{E} \left[e^{-\int_0^T r_s ds} h(\bar{A}_T^{v,y,\delta}) \right] = B(0, T) \mathbb{E} \left[h\left(\pi_{0,n} \frac{\tilde{S}_T}{B(0, T)} - K^{y,\delta}\right) \right].$$

Concernant le deuxième terme de la partie droite de (7.2.3), on a pour tout $1 \leq i \leq n$

$$\begin{aligned} \mathbb{E} \left[e^{-\int_0^T r_s ds} \left(\frac{S_T}{S_{t_i}} - \mathbb{E} \left[\frac{S_T}{S_{t_i}} \right] \right) h'(\bar{A}_T^{v,y,\delta}) \right] &= \mathbb{E} \left[e^{-\int_0^T r_s ds} \frac{S_T}{S_{t_i}} h'(\bar{A}_T^{v,y,\delta}) \right] - \\ &\quad \mathbb{E} \left[\frac{S_T}{S_{t_i}} \right] \mathbb{E} \left[e^{-\int_0^T r_s ds} h'(\bar{A}_T^{v,y,\delta}) \right]. \quad (7.2.4) \end{aligned}$$

Le second terme dans la partie droite peut être écrit sous forme de dérivée (l'hypothèse sur h permet la permutation entre la dérivée et l'espérance) :

$$\mathbb{E} \left[e^{-\int_0^T r_s ds} h'(\bar{A}_T^{v,y,\delta}) \right] = -\partial_K \mathbb{E} \left[e^{-\int_0^T r_s ds} h(\pi_{0,n} S_T - K) \right] \Big|_{K=K^{y,\delta}}.$$

Cette formule est essentielle pour interpréter plus tard les termes d'approximation comme des grecques (donc explicites).

Par le lemme 7.2.4, on transforme le premier terme dans la partie droite de (7.2.4) en interprétant le facteur (à une constante multiplicative près) $e^{-\int_0^T r_s ds} \frac{S_T}{S_{t_i}}$ comme un changement de probabilité dans \mathcal{F}_T . On écrit

$$\begin{aligned}
\mathbb{E} \left[e^{-\int_0^T r_s ds} \frac{S_T}{S_{t_i}} h'(\bar{A}_T^{v,y,\delta}) \right] &= c_i \mathbb{E} \left[\frac{D_T S_T}{c_i S_{t_i}} h'(\bar{A}_T^{v,y,\delta}) \right] \\
&= c_i \mathbb{E}_{\mathbb{Q}_{t_i}} \left[h'(\bar{A}_T^{v,y,\delta}) \right] \\
&= B(0, t_i) \mathbb{E} \left[h' \left(\pi_{0,n} \frac{c_{0,i}}{B(0, t_i)} \tilde{S}_T - \sum_{i=1}^n \hat{\delta}_i^n \mathbb{E} \left[\frac{S_T}{S_{t_i}} \right] \right) \right] \\
&= -B(0, t_i) \partial_K \mathbb{E} \left[h \left(\pi_{0,n} \frac{c_{0,i}}{B(0, t_i)} \tilde{S}_T - K \right) \right] \Big|_{K=K^{y,\delta}}
\end{aligned}$$

en utilisant le lemme 7.2.4 dans la troisième égalité ci-dessus.

En combinant l'égalité ci-dessus avec (7.2.4), et en injectant dans (7.2.3), on aboutit au résultat souhaité. \square

Théorème 7.2.6 *Pour une fonction régulière h vérifiant (\mathbf{H}_3) , on a*

$$\begin{aligned}
&\mathbb{E} \left[D_T h(A_T^{v,y,\delta}) \right] \\
&= B(0, T) \mathbb{E} \left[h \left(\frac{\pi_{0,n}}{B(0, T)} \tilde{S}_T - K^{y,\delta} \right) \right] + \sum_{i=1}^n \hat{\delta}_i c_i \partial_K \mathbb{E} \left[h \left(\pi_{0,n} \frac{c_{0,i}}{c_i} \tilde{S}_T - K^{y,\delta} \right) \right] \\
&\quad - \sum_{i=1}^n \hat{\delta}_i \mathbb{E} \left[\frac{S_T}{S_{t_i}} \right] B(0, T) \partial_K \mathbb{E} \left[h \left(\frac{\pi_{0,n}}{B(0, T)} \tilde{S}_T - K^{y,\delta} \right) \right] \\
&\quad + \frac{1}{2} \left\{ \sum_{i,j=1}^n \hat{\delta}_i \hat{\delta}_j c_{i,j} \partial_K^2 \mathbb{E} \left[h \left(\frac{\pi_{0,n} c_{0,i,j}}{c_{i,j}} \tilde{S}_T - K^{y,\delta} \right) \right] \right. \\
&\quad \quad - 2K^{y,\delta} \sum_{i=1}^n \hat{\delta}_i B(0, t_i) \partial_K^2 \mathbb{E} \left[h \left(\frac{\pi_{0,n} c_{0,i}}{c_i} \tilde{S}_T - K^{y,\delta} \right) \right] \\
&\quad \quad \left. + (K^{y,\delta})^2 B(0, T) \partial_K^2 \mathbb{E} \left[h \left(\frac{\pi_{0,n}}{B(0, T)} \tilde{S}_T - K^{y,\delta} \right) \right] \right\} \\
&\quad + \text{Error}_3(h).
\end{aligned}$$

Preuve. La preuve est similaire à celle du théorème précédent. Cependant, on remplace ici l'égalité (7.2.3) par un développement de Taylor au second ordre. On obtient :

$$\begin{aligned}
\mathbb{E} \left[D_T h(A_T^{v,y,\delta}) \right] &= \mathbb{E} \left[e^{-\int_0^T r_s ds} h(\bar{A}_T^{v,y,\delta}) \right] \\
&\quad - \sum_{i=1}^n \hat{\delta}_i \pi_{i,n} \mathbb{E} \left[e^{-\int_0^T r_s ds} \left(\frac{S_T}{S_{t_i}} - \mathbb{E} \left[\frac{S_T}{S_{t_i}} \right] \right) h'(\bar{A}_T^{v,y,\delta}) \right] \\
&\quad + \frac{1}{2} \mathbb{E} \left[e^{-\int_0^T r_s ds} \left(\sum_{i=1}^n \hat{\delta}_i \pi_{i,n} \left(\frac{S_T}{S_{t_i}} - \mathbb{E} \left[\frac{S_T}{S_{t_i}} \right] \right) \right)^2 h''(\bar{A}_T^{v,y,\delta}) \right] \\
&\quad + \text{Error}_3(h).
\end{aligned}$$

En comparant avec le développement du théorème 7.2.5, il nous reste à calculer le nouveau terme correctif avec le facteur $\frac{1}{2}$. Ce terme est égal à

$$\begin{aligned}
& \mathbb{E} \left[e^{-\int_0^T r_s ds} \left(\sum_{i=1}^n \delta_i \pi_{i,n} \left(\frac{S_T}{S_{t_i}} - E \left[\frac{S_T}{S_{t_i}} \right] \right) \right)^2 h''(\bar{A}_T^{v,y,\delta}) \right] \\
&= \sum_{1 \leq i,j \leq n} \hat{\delta}_i \hat{\delta}_j \mathbb{E} \left[e^{-\int_0^T r_s ds} \frac{S_T}{S_{t_i}} \frac{S_T}{S_{t_j}} h''(\bar{A}_T^{v,y,\delta}) \right] \\
&\quad - 2 \left(\sum_{j=1}^n \hat{\delta}_j \mathbb{E} \left[\frac{S_T}{S_{t_j}} \right] \right) \sum_{i=1}^n \hat{\delta}_i \mathbb{E} \left[e^{-\int_0^T r_s ds} \frac{S_T}{S_{t_i}} h''(\bar{A}_T^{v,y,\delta}) \right] \\
&\quad + \left(\sum_{j=1}^n \hat{\delta}_j \mathbb{E} \left[\frac{S_T}{S_{t_j}} \right] \right)^2 \mathbb{E} \left[e^{-\int_0^T r_s ds} h''(\bar{A}_T^{v,y,\delta}) \right] \\
&= \tau_1 + \tau_2 + \tau_3.
\end{aligned}$$

On traite séparément chacun des trois termes ci-dessus.

- **Le terme τ_1 .** On procède de façon analogue en le transformant par les changements de probabilité \mathbb{Q}_{t_i,t_j} . Donc, d'après le lemme 7.2.4, sous \mathbb{Q}_{t_i,t_j} $\bar{A}_T^{v,y,\delta}$ a la même loi que celle de $\pi_{0,n} \frac{c_{0,i,j}}{c_{i,j}} \tilde{S}_T - K^{y,\delta}$ sous \mathbb{Q} . Ainsi, on obtient

$$\begin{aligned}
\tau_1 &= \sum_{1 \leq i,j \leq n} \hat{\delta}_i \hat{\delta}_j c_{i,j} \mathbb{E}_{\mathbb{Q}_{t_i,t_j}} \left[h''(\bar{A}_T^{v,y,\delta}) \right] \\
&= \sum_{1 \leq i,j \leq n} \hat{\delta}_i \hat{\delta}_j c_{i,j} \mathbb{E} \left[h'' \left(\pi_{0,n} \frac{c_{0,i,j}}{c_{i,j}} \tilde{S}_T - \sum_{i=1}^n \hat{\delta}_i \mathbb{E} \left[\frac{S_T}{S_{t_i}} \right] \right) \right] \\
&= \sum_{1 \leq i,j \leq n} \hat{\delta}_i \hat{\delta}_j c_{i,j} \partial_k^2 \mathbb{E} \left[h'' \left(\pi_{0,n} \frac{c_{0,i,j}}{c_{i,j}} \tilde{S}_T - k \right) \right] \Big|_{k=K^{y,\delta}}.
\end{aligned}$$

- **Le terme τ_2 .** De manière similaire, on a

$$\tau_2 = -2 \left(\sum_{j=1}^n \hat{\delta}_j \mathbb{E} \left[\frac{S_T}{S_{t_j}} \right] \right) \sum_{i=1}^n \hat{\delta}_i B(0, t_i) \partial_k^2 \mathbb{E} \left[h'' \left(\pi_{0,n} \frac{c_{0,i}}{B(0, t_i)} \tilde{S}_T - k \right) \right] \Big|_{k=K^{y,\delta}}.$$

- **Le terme τ_3 .** On obtient facilement

$$\tau_3 = \left(\sum_{j=1}^n \hat{\delta}_j \mathbb{E} \left[\frac{S_T}{S_{t_j}} \right] \right)^2 B(0, T) \partial_k^2 \mathbb{E} \left[h'' \left(\frac{\pi_{0,n}}{B(0, T)} \tilde{S}_T - k \right) \right] \Big|_{k=K^{y,\delta}}.$$

Ainsi le théorème est prouvé. \square

Les théorèmes d'avant peuvent s'étendre à la fonction partie positive par passage à la limite via une suite de fonctions régulières l'approximant. La preuve est similaire à [EG12] et nous ne la détaillons pas. On énonce ci-dessous les théorèmes :

Théorème 7.2.7 *Le développement asymptotique au premier ordre de $\mathbb{E} \left[D_T(A_T^{v,y,\delta})_+ \right]$ est :*

$$\begin{aligned} \mathbb{E} \left[D_T(A_T^{v,y,\delta})_+ \right] = & B(0, T) \text{Call}^{BS} \left(\frac{\pi_{0,n}}{B(0, T)} S_0, K^{y,\delta} \right) \\ & + \sum_{i=1}^n \hat{\delta}_i B(0, t_i) \partial_K \text{Call}^{BS} \left(\frac{\pi_{0,n} c_{0,i}}{B(0, t_i)} S_0, K^{y,\delta} \right) \\ & - K^{y,\delta} B(0, T) \partial_K \text{Call}^{BS} \left(\frac{\pi_{0,n}}{B(0, T)} S_0, K^{y,\delta} \right) \\ & + \text{Error}_2. \end{aligned}$$

Théorème 7.2.8 *Le développement asymptotique au second ordre de $\mathbb{E} \left[D_T(A_T^{v,y,\delta})_+ \right]$ s'écrit :*

$$\begin{aligned} & \mathbb{E} \left[D_T(A_T^{v,y,\delta})_+ \right] \\ & = \text{termes ordre 1} + \\ & + \frac{1}{2} \left\{ \sum_{i,j=1}^n \hat{\delta}_i \hat{\delta}_j c_{i,j} \partial_K^2 \text{Call}^{BS} \left(\frac{\pi_{0,n} c_{0,i,j}}{c_{i,j}} S_0, K^{y,\delta} \right) \right. \\ & \quad - 2K^{y,\delta} \sum_{i=1}^n \hat{\delta}_i B(0, t_i) \partial_K^2 \text{Call}^{BS} \left(\frac{\pi_{0,n} c_{0,i}}{B(0, t_i)} S_0, K^{y,\delta} \right) \\ & \quad \left. + (K^{y,\delta})^2 B(0, T) \partial_K^2 \text{Call}^{BS} \left(\frac{\pi_{0,n}}{B(0, T)} S_0, K^{y,\delta} \right) \right\} \\ & + \text{Error}_3. \end{aligned}$$

Il est important de noter qu'on peut étendre les résultats précédents à n'importe quel ordre k . Dans ce cas, le calcul du k -ième terme du développement asymptotique revient à celui des quantités :

$$\mathbb{E} \left[D_T \frac{S_T}{S_{t_{i_1}}} \frac{S_T}{S_{t_{i_2}}} \dots \frac{S_T}{S_{t_{i_k}}} h^{(k)}(\bar{A}_T^{v,y,\delta}) \right].$$

Le terme ci-dessus correspond à la k -ième dérivée de la formule de Black par rapport au strike à un facteur multiplicatif près. En effet, le facteur

$$\frac{D_T \frac{S_T}{S_{t_{i_1}}} \frac{S_T}{S_{t_{i_2}}} \dots \frac{S_T}{S_{t_{i_k}}}}{\mathbb{E} \left[D_T \frac{S_T}{S_{t_{i_1}}} \frac{S_T}{S_{t_{i_2}}} \dots \frac{S_T}{S_{t_{i_k}}} \right]}$$

induit la mesure de probabilité $\mathbb{Q}_{t_1, \dots, t_k}$ sous laquelle la loi du *proxy* $\bar{A}_T^{v, y, \delta}$ est la même que celle de $\pi_{0, n} \frac{c_{0, i_1, \dots, i_k}}{c_{i_1, \dots, i_k}} \tilde{S}_T - K^{y, \delta}$ sous \mathbb{Q} . Ainsi,

$$\mathbb{E} \left[D_T \frac{S_T}{S_{t_1}} \frac{S_T}{S_{t_2}} \dots \frac{S_T}{S_{t_k}} h^{(k)}(\bar{A}_T^{v, y, \delta}) \right] = c_{i_1, \dots, i_k} \mathbb{E} \left[h^{(k)} \left(\pi_{0, n} \frac{c_{0, i_1, \dots, i_k}}{c_{i_1, \dots, i_k}} \tilde{S}_T - K^{y, \delta} \right) \right].$$

Théorème 7.2.9 *L'ordre 3 s'écrit :*

$$\begin{aligned} & \mathbb{E} \left[D_T (A_T^{v, y, \delta})_+ \right] \\ &= \text{termes ordre 2} + \\ &+ \frac{1}{6} \left\{ \sum_{i, j, l=1}^n \hat{\delta}_i \hat{\delta}_j \hat{\delta}_l c_{i, j, l} \partial_K^3 \text{Call}^{BS} \left(\frac{\pi_{0, n} c_{0, i, j, l}}{c_{i, j, l}} S_0, K^{y, \delta} \right) \right. \\ &\quad - 3K^{y, \delta} \sum_{i, j=1}^n \hat{\delta}_i \hat{\delta}_j c_{i, j} \partial_K^3 \text{Call}^{BS} \left(\frac{\pi_{0, n} c_{0, i, j}}{c_{i, j}} S_0, K^{y, \delta} \right) \\ &\quad + 3(K^{y, \delta})^2 \sum_{i=1}^n \hat{\delta}_i B(0, t_i) \partial_K^3 \text{Call}^{BS} \left(\frac{\pi_{0, n} c_{0, i}}{B(0, t_i)} S_0, K^{y, \delta} \right) \\ &\quad \left. - (K^{y, \delta})^3 B(0, T) \partial_K^3 \text{Call}^{BS} \left(\frac{\pi_{0, n}}{B(0, T)} S_0, K^{y, \delta} \right) \right\} \\ &+ \text{Error}_4. \end{aligned}$$

7.2.3 Estimation de l'erreur dans le cas régulier

On écrit $A \leq_c B$ s'il existe une constante (dont la dépendance est uniformément bornée quand les paramètres tendent vers 0) telle que $A \leq cB$. On énonce le théorème d'estimation d'erreur suivant :

Théorème 7.2.10 *Si $\forall t \in [0, T] : \sigma_t^r \leq \bar{\sigma}^r, \sigma_t \leq \bar{\sigma}$, alors on a les estimations d'erreur suivantes :*

$$\begin{aligned} |Error_2(h)| &\leq_c \left[n \sup_{1 \leq i \leq n} \left(\delta_i \left[\bar{\sigma}^r \sqrt{T} \min \left(T, \frac{1}{a} \right) + \bar{\sigma} \sqrt{T - t_i} \right] \right)^2, \\ |Error_3(h)| &\leq_c \left[n \sup_{1 \leq i \leq n} \left(\delta_i \left[\bar{\sigma}^r \sqrt{T} \min \left(T, \frac{1}{a} \right) + \bar{\sigma} \sqrt{T - t_i} \right] \right)^3. \end{aligned}$$

Preuve du théorème 7.2.10

On fait la preuve juste pour $|Error_2(h)|$. La démonstration concernant $|Error_3(h)|$ est similaire. Notre but est de faire une majoration montrant le comportement de l'erreur en fonction des retraits ainsi que de $\bar{\sigma}$, $\bar{\sigma}^r$, a et T . On a

$$Error_2(h) = \mathbb{E} \left[\int_0^1 D_T \left(\sum_{i=1}^n \hat{\delta}_i \left(\frac{S_T}{S_{t_i}} - \mathbb{E} \left[\frac{S_T}{S_{t_i}} \right] \right) \right)^2 (1 - \lambda) h''(A_T^\lambda) d\lambda \right]$$

en posant

$$A_T^\lambda = \bar{A}^{v,y,\delta} - \lambda \sum_{i=1}^n \hat{\delta}_i \left(\frac{S_T}{S_{t_i}} - \mathbb{E} \left[\frac{S_T}{S_{t_i}} \right] \right).$$

Ainsi

$$\begin{aligned} |Error_2(h)| &\leq_c \int_0^1 \mathbb{E} \left[D_T \left(\sum_{i=1}^n \hat{\delta}_i \left(\frac{S_T}{S_{t_i}} - \mathbb{E} \left[\frac{S_T}{S_{t_i}} \right] \right) \right)^2 (1 + |A_T^\lambda|^p) \right] d\lambda \\ &\leq_c \int_0^1 \|D_T\|_3 \left\| \left(\sum_{i=1}^n \hat{\delta}_i \left(\frac{S_T}{S_{t_i}} - \mathbb{E} \left[\frac{S_T}{S_{t_i}} \right] \right) \right)^2 \right\|_3 \|1 + |A_T^\lambda|^p\|_3 d\lambda \\ &\leq_c \|D_T\|_3 \left(\sum_{i=1}^n \hat{\delta}_i \left\| \frac{S_T}{S_{t_i}} - \mathbb{E} \left[\frac{S_T}{S_{t_i}} \right] \right\|_6 \right)^2 \int_0^1 (1 + \|A_T^\lambda\|_{3p}^p) d\lambda \\ &\leq_c \|D_T\|_3 \left(n \sup_{1 \leq i \leq n} \hat{\delta}_i \left\| \frac{S_T}{S_{t_i}} - \mathbb{E} \left[\frac{S_T}{S_{t_i}} \right] \right\|_6 \right)^2 \\ &\quad \int_0^1 \left[1 + \left(\|\pi_{0,n} S_T\|_{3p} + \lambda \sum_{i=1}^n \hat{\delta}_i \left\| \frac{S_T}{S_{t_i}} \right\|_{3p} + (1 - \lambda) \sum_{i=1}^n \hat{\delta}_i \mathbb{E} \left[\frac{S_T}{S_{t_i}} \right] \right)^p \right] d\lambda \\ &\leq_c \|D_T\|_3 \left[1 + \left(\|\pi_{0,n} S_T\|_{3p} + \sum_{i=1}^n \hat{\delta}_i \left\| \frac{S_T}{S_{t_i}} \right\|_{3p} \right)^p \right] \left(n \sup_{1 \leq i \leq n} \hat{\delta}_i \left\| \frac{S_T}{S_{t_i}} - \mathbb{E} \left[\frac{S_T}{S_{t_i}} \right] \right\|_6 \right)^2. \end{aligned}$$

Donc

$$|Error_2(h)| \leq_c \left(n \sup_{1 \leq i \leq n} \hat{\delta}_i \left\| \frac{S_T}{S_{t_i}} - \mathbb{E} \left[\frac{S_T}{S_{t_i}} \right] \right\|_6 \right)^2. \quad (7.2.5)$$

Notre objectif maintenant est de majorer $\left\| \frac{S_T}{S_{t_i}} - \mathbb{E} \left[\frac{S_T}{S_{t_i}} \right] \right\|_6$ pour $1 \leq i \leq n$. On a selon (7.4.2)

$$\begin{aligned} \frac{S_T}{S_{t_i}} &= \frac{B(0, t_i)}{B(0, T)} \exp \left(\frac{1}{2} \int_0^T \Gamma(s, T)^2 ds - \frac{1}{2} \int_0^{t_i} \Gamma(s, t_i)^2 ds - \frac{1}{2} \int_{t_i}^T \sigma_s^2 ds \right) \\ &\quad \exp \left(\int_0^T \Gamma(s, T) dW_s - \int_0^{t_i} \Gamma(s, t_i) dW_s + \int_{t_i}^T \rho \sigma_s dW_s + \int_{t_i}^T \sqrt{1 - \rho^2} \sigma_s dB_s \right) \\ &= \frac{B(0, t_i)}{B(0, T)} \exp(A) \exp(N_T) \end{aligned}$$

où

$$N_u = \int_0^u \Gamma(s, T) dW_s - \int_0^u \Gamma(s, t_i) \mathbf{1}_{s \leq t_i} dW_s + \int_0^u \rho \sigma_s \mathbf{1}_{s \geq t_i} dW_s + \int_0^u \sqrt{1 - \rho^2} \sigma_s \mathbf{1}_{s \geq t_i} dB_s, \quad 0 \leq u \leq T,$$

et

$$A = \frac{1}{2} \int_0^T \Gamma(s, T)^2 ds - \frac{1}{2} \int_0^{t_i} \Gamma(s, t_i)^2 ds - \frac{1}{2} \int_{t_i}^T \sigma_s^2 ds.$$

D'où

$$\frac{S_T}{S_{t_i}} - \mathbb{E} \left[\frac{S_T}{S_{t_i}} \right] = \frac{B(0, t_i)}{B(0, T)} \exp(A) \exp\left(\frac{1}{2}\langle N \rangle_T\right) \left(\exp(N_T - \frac{1}{2}\langle N \rangle_T) - 1 \right).$$

Par une adaptation du lemme 4.3 dans [EG12], on a pour tout $r \geq 1$:

$$\|e^{J_T} - 1\|_r \leq \|J_T\|_{2r} e^{2r\langle N \rangle_T}$$

où $J_T = N_T - \frac{1}{2}\langle N \rangle_T$.

D'où

$$\left\| \frac{S_T}{S_{t_i}} - \mathbb{E} \left[\frac{S_T}{S_{t_i}} \right] \right\|_6 \leq \frac{B(0, t_i)}{B(0, T)} \exp(A) \exp\left(\frac{1}{2}\langle N \rangle_T\right) \|J_T\|_{12} \exp(12\langle N \rangle_T). \quad (7.2.6)$$

On a

$$\begin{aligned} \langle N \rangle_T &= \int_0^{t_i} (\Gamma(s, T) - \Gamma(s, t_i))^2 ds + \int_{t_i}^T (\Gamma(s, T) + \rho\sigma_s)^2 ds + \int_{t_i}^T (1 - \rho^2)\sigma_s^2 ds \\ &\leq_c \int_0^T \Gamma(s, T)^2 ds + \int_0^{t_i} \Gamma(s, t_i)^2 ds + \bar{\sigma}^2(T - t_i). \end{aligned}$$

D'autre part

$$\begin{aligned} \int_0^T \Gamma(s, T)^2 ds &= \int_0^T \left(\int_s^T \sigma_u^r e^{-a(u-s)} du \right)^2 ds \\ &\leq \int_0^T \left(\bar{\sigma}^r \frac{1 - e^{-a(T-s)}}{a} \right)^2 ds \\ &= \frac{(\bar{\sigma}^r)^2}{a^3} \int_0^{aT} (1 - e^{-u})^2 ds \\ &\leq \frac{(\bar{\sigma}^r)^2}{a^3} \int_0^{aT} (\min(u, 1))^2 ds \\ &\leq \frac{(\bar{\sigma}^r)^2}{a^3} \min\left(\frac{a^3 T^3}{3}, aT\right) \\ &\leq_c (\bar{\sigma}^r)^2 T \min\left(T^2, \frac{1}{a^2}\right). \end{aligned}$$

De même

$$\int_0^{t_i} \Gamma(s, t_i)^2 ds \leq_c (\bar{\sigma}^r)^2 t_i \min\left(t_i^2, \frac{1}{a^2}\right) \leq_c (\bar{\sigma}^r)^2 T \min\left(T^2, \frac{1}{a^2}\right).$$

Par suite

$$\langle N \rangle_T \leq_c (\bar{\sigma}^r)^2 T \min\left(T^2, \frac{1}{a^2}\right) + \bar{\sigma}^2(T - t_i).$$

Ainsi, on déduit par (7.2.6) et l'inégalité de Burkholder-Davis-Gundy que

$$\begin{aligned} \left\| \frac{S_T}{S_{t_i}} - \mathbb{E} \left[\frac{S_T}{S_{t_i}} \right] \right\|_6 &\leq_c \sqrt{(\bar{\sigma}^r)^2 T \min(T^2, \frac{1}{a^2}) + \bar{\sigma}^2 (T - t_i)} \\ &\leq_c \bar{\sigma}^r \sqrt{T} \min(T, \frac{1}{a}) + \bar{\sigma} \sqrt{T - t_i}. \end{aligned}$$

En injectant l'inégalité ci-dessus dans (7.2.5) on aboutit au résultat souhaité. \square

Remarque :

En supposant une hypothèse d'ellipticité sur σ , les mêmes estimations d'erreur restent valables pour la fonction partie positive qu'on peut écrire sous la forme :

$$|Error_k| \leq_c \left[n \sup_{1 \leq i \leq n} \left(\delta_i \left[\bar{\sigma}^r \sqrt{T} \min \left(T, \frac{1}{a} \right) + \bar{\sigma} \sqrt{T - t_i} \right] \right) \right]^k, \quad (7.2.7)$$

pour la méthode de développement asymptotique d'ordre $k - 1$, $k \geq 2$; voir [EG12] pour des arguments détaillés.

Bien que cette estimation croisse quand n et T augmentent laissant supposer une explosion d'erreur à grande maturité, ce n'est probablement pas le cas car le prix d'exercice du proxy augmente avec T , rendant l'option de plus en plus en dehors de la monnaie (prix et grecques tendent vers 0).

7.3 Résultats numériques

Dans cette section, nous discutons la performance de notre méthode d'approximation sur un contrat GMWB en fonction des paramètres du modèle et de l'ordre k du développement asymptotique. Dans tous nos tests, on choisit comme *benchmark* le prix Monte Carlo calculé en utilisant 10^8 simulations et on prendra $T = 20$, $S_0 = 100$, $\rho = 0$ et $y = 0.9\%$ (charges annuels y compris en T).

Le tableau de la figure 7.1 affiche pour un contrat GMWB, sa valeur *benchmark* (colonne MC) ainsi que l'intervalle de confiance à 95% correspondant (colonne IC) et les valeurs calculées par la méthode de développement asymptotique pour les ordres de 1 à 5 (colonnes DAS1, ..., DAS5) en fonction de σ_r et a . On remarque que plus l'ordre de la méthode est grand, meilleure est la précision. Ainsi, le prix DAS5 est bien précis pour toutes les valeurs de σ_r et a . Cependant, quelquefois on peut se limiter à l'ordre 4 pour avoir une bonne précision. Enfin, le gain en temps de calcul est énorme puisque la méthode DAS5 prend environ 2 secondes tandis que le temps de calcul nécessaire pour obtenir un prix Monte Carlo de précision équivalente est proche de 2 heures.

σ_r	a	MC	IC	DAS1	DAS2	DAS3	DAS4	DAS5
0.01	0.001	94.7514	[94.7476,94.7553]	93.5575	94.2223	94.5405	94.674	94.722
0.01	0.041	93.8524	[93.8487,93.856]	92.634	93.383	93.7057	93.8169	93.8434
0.01	0.091	93.1718	[93.1682,93.1753]	91.912	92.7376	93.0658	93.1593	93.1704
0.02	0.001	100.673	100.669,100.678	98.933	99.6428	100.056	100.296	100.437
0.02	0.041	98.3116	98.3071,98.3161	96.9445	97.6249	97.9742	98.1439	98.2236
0.02	0.091	96.3348	96.3306,96.3389	95.0395	95.789	96.1272	96.2593	96.3035

FIGURE 7.1 – Valeur de $GMWB_0$ pour $\sigma=10\%$ et $\delta=9.5$ (retrait annuel à partir de l'année 6 jusqu'à maturité (pas de retrait avant)).

Dans le graphe 7.2, on visualise l'erreur relative du prix de la méthode DAS5 en variant a et σ_r pour trois valeurs de σ . On remarque que a et σ^r ont un impact significatif surtout quand a est inférieur à 2%. En effet l'erreur relative augmente considérablement quand a diminue ou σ^r augmente. Ceci est en concordance avec la majoration d'erreur (7.2.7) qui est proportionnelle à σ^r et $\frac{1}{a}$. D'un autre côté l'effet de la volatilité σ est insignifiant ce qui est en accord avec (7.2.7) montrant que le facteur de σ^r est au moins de l'ordre $T\sqrt{T}$ alors que le facteur de σ est de l'ordre de \sqrt{T} .

La figure 7.3 montre l'impact de l'ordre d'approximation ainsi que du taux de retrait δ sur l'erreur relative. On remarque une stabilité de l'erreur en fonction des retraits ce qui peut s'expliquer par une compensation des erreurs dans la formule du prix (7.2.1). En effet, plus la maturité est grande plus l'erreur est grande mais d'un autre côté plus les retraits augmentent (et aussi la maturité (voir graphe 7.4)), plus la contribution des termes à grandes maturités diminue car dans ce cas l'*account value* a tendance à diminuer rapidement.

Enfin, le graphe 7.4 trace la probabilité que l'*account value* devienne négative en fonction des années.

7.4 Annexe

7.4.1 Preuve du lemme 7.2.1

On prouve seulement la seconde égalité, la première peut être déduite de manière similaire. On le fait par récurrence.

Le résultat est vrai pour $n = 1$. Supposons qu'il est aussi vrai pour $n (\geq 1)$ dates $(t_i)_{1 \leq i \leq n}$ et considérons qu'un nouveau retrait est fait au temps $t_{n+1} \in (t_n, T]$. Ainsi,

on a :

$$\begin{aligned}
A_T^{v,y,\delta} &= (1 - \alpha_{n+1})A_{t_{n+1}}^{v,y,\delta} \frac{S_T}{S_{t_{n+1}}} - \delta_{n+1} \frac{S_T}{S_{t_{n+1}}} \\
&= (1 - \alpha_{n+1}) \frac{S_T}{S_{t_{n+1}}} \left[\left(\prod_{i=1}^n (1 - \alpha_i) \right) S_{t_{n+1}} - \sum_{i=1}^n \left(\delta_i \prod_{j=i+1}^n (1 - \alpha_j) \right) \frac{S_{t_{n+1}}}{S_{t_i}} \right] \\
&\quad - \delta_{n+1} \frac{S_T}{S_{t_{n+1}}} \\
&= \left(\prod_{i=1}^{n+1} (1 - \alpha_i) \right) S_T - \sum_{i=1}^n \left(\delta_i (1 - \alpha_{n+1}) \prod_{j=i+1}^n (1 - \alpha_j) \right) \frac{S_T}{S_{t_i}} - \delta_{n+1} \frac{S_T}{S_{t_{n+1}}} \\
&= \left(\prod_{i=1}^{n+1} (1 - \alpha_i) \right) S_T - \sum_{i=1}^{n+1} \left(\delta_i \prod_{j=i+1}^{n+1} (1 - \alpha_j) \right) \frac{S_T}{S_{t_i}}.
\end{aligned}$$

Par suite le résultat est vrai pour $n + 1$ dates. \square

7.4.2 Preuve du lemme 7.2.2

Sous le modèle Hull-White à 1 facteur, on a :

$$B(t, u) = B(0, u) \exp \left(\int_0^t r_s ds - \frac{1}{2} \int_0^t \Gamma(s, u)^2 ds - \int_0^t \Gamma(s, u) dW_s \right), t \leq u.$$

D'où,

$$1 = B(0, t_i) \exp \left(\int_0^{t_i} r_s ds - \frac{1}{2} \int_0^{t_i} \Gamma(s, t_i)^2 ds - \int_0^{t_i} \Gamma(s, t_i) dW_s \right),$$

$$1 = B(0, T) \exp \left(\int_0^T r_s ds - \frac{1}{2} \int_0^T \Gamma(s, T)^2 ds - \int_0^T \Gamma(s, T) dW_s \right).$$

En divisant les deux égalités ci-dessus, on déduit :

$$\begin{aligned}
\exp \left(\int_{t_i}^T r_s ds \right) &= \frac{B(0, t_i)}{B(0, T)} \exp \left(\int_0^T \Gamma(s, T) dW_s + \frac{1}{2} \int_0^T \Gamma(s, T)^2 ds \right. \\
&\quad \left. - \int_0^{t_i} \Gamma(s, t_i) dW_s - \frac{1}{2} \int_0^{t_i} \Gamma(s, t_i)^2 ds \right). \tag{7.4.1}
\end{aligned}$$

D'autre part, on a :

$$\frac{S_T}{S_{t_i}} = \exp \left(\int_{t_i}^T r_s ds - \frac{1}{2} \int_{t_i}^T \sigma_s^2 ds + \int_{t_i}^T \sigma_s (\rho dW_s + \sqrt{1 - \rho^2} dB_s) \right).$$

En injectant (7.4.1) dans l'égalité ci-dessus, on obtient :

$$\begin{aligned}
& \frac{S_T}{S_{t_i}} \\
&= \frac{B(0, t_i)}{B(0, T)} \exp \left(\int_0^T \Gamma(s, T) dW_s + \frac{1}{2} \int_0^T \Gamma(s, T)^2 ds - \int_0^{t_i} \Gamma(s, t_i) dW_s - \frac{1}{2} \int_0^{t_i} \Gamma(s, t_i)^2 ds \right) \\
& \quad \exp \left(-\frac{1}{2} \int_{t_i}^T \sigma_s^2 ds + \int_{t_i}^T \sigma_s (\rho dW_s + \sqrt{1 - \rho^2} dB_s) \right) \\
&= \frac{B(0, t_i)}{B(0, T)} \exp \left(\frac{1}{2} \int_0^{t_i} (\Gamma(s, T)^2 - \Gamma(s, t_i)^2) ds + \frac{1}{2} \int_{t_i}^T (\Gamma(s, T)^2 - \sigma_s^2) ds \right) \\
& \quad \exp \left(\int_0^{t_i} (\Gamma(s, T) - \Gamma(s, t_i)) dW_s + \int_{t_i}^T (\Gamma(s, T) + \rho \sigma_s) dW_s + \int_{t_i}^T \sqrt{1 - \rho^2} \sigma_s dB_s \right).
\end{aligned} \tag{7.4.2}$$

Donc, en prenant l'espérance, on a :

$$\begin{aligned}
& \mathbb{E} \left[\frac{S_T}{S_{t_i}} \right] \\
&= \frac{B(0, t_i)}{B(0, T)} \exp \left(\frac{1}{2} \int_0^{t_i} (\Gamma(s, T)^2 - \Gamma(s, t_i)^2) ds + \frac{1}{2} \int_{t_i}^T (\Gamma(s, T)^2 - \sigma_s^2) ds \right) \\
& \quad \exp \left(\frac{1}{2} \int_0^{t_i} (\Gamma(s, T) - \Gamma(s, t_i))^2 ds + \frac{1}{2} \int_{t_i}^T (\Gamma(s, T) + \rho \sigma_s)^2 ds + \frac{1}{2} \int_{t_i}^T (1 - \rho^2) \sigma_s^2 ds \right) \\
&= \frac{B(0, t_i)}{B(0, T)} \exp \left(\int_0^{t_i} (\Gamma(s, T)^2 - \Gamma(s, T) \Gamma(s, t_i)) ds + \int_{t_i}^T (\Gamma(s, T)^2 + \rho \sigma_s \Gamma(s, T)) ds \right).
\end{aligned}$$

Ainsi, la première partie du résultat est démontrée.

Supposons maintenant que $\sigma_t = \sigma$ et $\sigma_t^r = \sigma^r$. Dans ce cas particulier, on a :

$$\Gamma(t, s) = \sigma^r \frac{1 - e^{-a(s-t)}}{a}, t \leq s.$$

D'où :

$$\begin{aligned}
& \mathbb{E} \left[\frac{S_T}{S_{t_i}} \right] \\
&= \frac{B(0, t_i)}{B(0, T)} \exp \left(\int_0^{t_i} \left(\sigma^r \frac{1 - e^{-a(T-s)}}{a} \right)^2 ds - \int_0^{t_i} \sigma^r \frac{1 - e^{-a(t_i-s)}}{a} \sigma^r \frac{1 - e^{-a(T-s)}}{a} ds \right. \\
& \quad \left. + \int_{t_i}^T \rho \sigma \sigma^r \frac{1 - e^{-a(T-s)}}{a} ds \right)
\end{aligned}$$

$$\begin{aligned}
&= \frac{B(0, t_i)}{B(0, T)} \exp \left(\left(\frac{\sigma^r}{a} \right)^2 \left[T - 2 \frac{1 - e^{-aT}}{a} + \frac{1 - e^{-2aT}}{2a} \right] \right. \\
&\quad \left. - \left(\frac{\sigma^r}{a} \right)^2 \left[t_i - \frac{e^{-a(T-t_i)} - e^{-aT}}{a} - \frac{1 - e^{-at_i}}{a} + \frac{e^{-a(T-t_i)} - e^{-a(T+t_i)}}{2a} \right] \right. \\
&\quad \left. + \frac{\rho\sigma\sigma^r}{a} \left[T - t_i - \frac{1 - e^{-a(T-t_i)}}{a} \right] \right) \\
&= \frac{B(0, t_i)}{B(0, T)} \exp \left(\left(\frac{\sigma^r}{a} \right)^2 \left[T - t_i + \frac{e^{-aT}}{a} - \frac{e^{-at_i}}{a} - \frac{e^{-2aT}}{2a} + \frac{e^{-a(T-t_i)}}{2a} + \frac{e^{-a(T+t_i)}}{2a} - \frac{1}{2a} \right] \right. \\
&\quad \left. + \rho\sigma\sigma^r \left[T - t_i + \frac{e^{-a(T-t_i)}}{a} - \frac{1}{a} \right] \right).
\end{aligned}$$

D'où la deuxième partie du lemme. \square

7.4.3 Preuve du lemme 7.2.3

Puisque par définition, c_{i_1, i_2, \dots, i_k} est invariant par permutation de (i_1, i_2, \dots, i_k) , le deuxième résultat du lemme est immédiat. Supposons que la suite (i_1, i_2, \dots, i_k) est croissante. En utilisant (7.4.1) et (7.4.2), on peut écrire :

$$\begin{aligned}
& c_{i_1, i_2, \dots, i_k} \\
&= \mathbb{E} \left[B(0, T) \exp \left(- \int_0^T \Gamma(s, T) dW_s - \frac{1}{2} \int_0^T \Gamma(s, T)^2 ds \right) \right. \\
&\quad \prod_{j=1}^k \frac{B(0, t_{i_j})}{B(0, T)} \exp \left(\frac{1}{2} \int_0^T \Gamma(s, T)^2 ds - \frac{1}{2} \int_0^{t_{i_j}} \Gamma(s, t_{i_j})^2 ds - \frac{1}{2} \int_{t_{i_j}}^T \sigma_s^2 ds \right) \\
&\quad \left. \exp \left(\int_0^T \Gamma(s, T) dW_s - \int_0^{t_{i_j}} \Gamma(s, t_{i_j}) dW_s + \int_{t_{i_j}}^T \rho\sigma_s dW_s + \int_{t_{i_j}}^T \sqrt{1 - \rho^2} \sigma_s dB_s \right) \right] \\
&= \frac{\prod_{j=1}^k B(0, t_{i_j})}{B(0, T)^{k-1}} \\
&\quad \exp \left(\frac{1}{2} \sum_{j=0}^k \int_{t_{i_j}}^{t_{i_{j+1}}} \left[(k-1)\Gamma(s, T)^2 - j\sigma_s^2 - \sum_{l=j+1}^k \Gamma(s, t_{i_l})^2 \right] ds \right) \\
&\quad \mathbb{E} \left[\exp \left(\sum_{j=0}^k \int_{t_{i_j}}^{t_{i_{j+1}}} \left[(k-1)\Gamma(s, T) + j\rho\sigma_s - \sum_{l=j+1}^k \Gamma(s, t_{i_l}) \right] dW_s \right. \right. \\
&\quad \left. \left. + \sum_{j=0}^k \int_{t_{i_j}}^{t_{i_{j+1}}} j\sqrt{1 - \rho^2} \sigma_s dB_s \right) \right] \\
&= \frac{\prod_{j=1}^k B(0, t_{i_j})}{B(0, T)^{k-1}}
\end{aligned}$$

$$\begin{aligned} & \exp \left(\frac{1}{2} \sum_{j=0}^k \int_{t_{i_j}}^{t_{i_{j+1}}} \left[(k-1)\Gamma(s, T)^2 - j\sigma_s^2 - \sum_{l=j+1}^k \Gamma(s, t_{i_l})^2 \right] ds \right) \\ & \exp \left(\frac{1}{2} \sum_{j=0}^k \int_{t_{i_j}}^{t_{i_{j+1}}} \left[(k-1)\Gamma(s, T) + j\rho\sigma_s - \sum_{l=j+1}^k \Gamma(s, t_{i_l}) \right]^2 ds \right. \\ & \quad \left. + \frac{1}{2} \sum_{j=0}^k \int_{t_{i_j}}^{t_{i_{j+1}}} j^2(1-\rho^2)\sigma_s^2 ds \right). \end{aligned}$$

Ainsi, le lemme est démontré. \square

Remarque

Il est possible d'avoir une formule explicite si $\sigma_t = \sigma$ et $\sigma_t^r = \sigma^r$. Pour y aboutir, on a besoin des égalités réunies dans ce lemme facile à vérifier :

Lemme 7.4.1 *Pour $k \in \{0, 1, \dots, n+1\}$, on a :*

$$\begin{aligned} \sum_{j=0}^{k-1} (k-j) \left[e^{-a(T-t_{i_{j+1}})} - e^{-a(T-t_{i_j})} \right] &= e^{-a(T-t_{i_k})} - ke^{-aT} + \sum_{j=1}^{k-1} e^{-a(T-t_{i_j})}, \\ \sum_{j=0}^{k-1} \sum_{l=j+1}^k \left[e^{-a(t_{i_l}-t_{i_{j+1}})} - e^{-a(t_{i_l}-t_{i_j})} \right] &= k - \sum_{j=1}^k e^{-at_{i_j}}, \\ \sum_{j=0}^{k-1} \sum_{l=j+1}^k \left[e^{-a(T+t_{i_l}-2t_{i_{j+1}})} - e^{-a(T+t_{i_l}-2t_{i_j})} \right] &= \sum_{j=1}^k e^{-a(T-t_{i_j})} - \sum_{j=1}^k e^{-a(T+t_{i_j})}, \\ \sum_{j=0}^{k-2} \sum_{l=j+1}^{k-1} (k-l) \left[e^{-a(t_{i_l}-t_{i_{j+1}})} - e^{-a(t_{i_l}-t_{i_j})} \right] &= \frac{k(k-1)}{2} - \sum_{j=1}^{k-1} (k-j)e^{-at_{i_j}}, \\ \sum_{j=0}^{k-2} \sum_{l=j+1}^{k-1} \sum_{m=l+1}^k \left[e^{-a(t_{i_m}-t_{i_{j+1}})} - e^{-a(t_{i_m}-t_{i_j})} \right] &= \sum_{j=1}^{k-1} \sum_{m=j+1}^k e^{-a(t_{i_m}-t_{i_j})} - \sum_{l=2}^k (l-1)e^{-at_{i_l}}, \\ \sum_{j=0}^{k-2} \sum_{l=j+1}^{k-1} \sum_{m=l+1}^k \left[e^{-a(t_{i_l}+t_{i_m}-2t_{i_{j+1}})} - e^{-a(t_{i_l}+t_{i_m}-2t_{i_j})} \right] &= \sum_{j=1}^{k-1} \sum_{m=j+1}^k e^{-a(t_{i_m}-t_{i_j})} \\ & \quad - \sum_{j=1}^{k-1} \sum_{m=j+1}^k e^{-a(t_{i_m}+t_{i_j})}. \end{aligned}$$

D'après le lemme 7.2.3, on peut écrire

$$\begin{aligned}
& c_{i_1, \dots, i_k} \\
&= \frac{\prod_{j=1}^k B(0, t_{i_j})}{B(0, T)^{k-1}} \\
&\exp \left(\frac{k-1}{2} \int_0^T \Gamma(s, T)^2 ds - \frac{1}{2} \sum_{j=0}^k \int_{t_{i_j}}^{t_{i_{j+1}}} \sigma_s^2 ds - \frac{1}{2} \sum_{j=0}^k \int_{t_{i_j}}^{t_{i_{j+1}}} \sum_{l=j+1}^k \Gamma(s, t_{i_l})^2 ds \right) \\
&\exp \left(\frac{1}{2} \sum_{j=0}^k \int_{t_{i_j}}^{t_{i_{j+1}}} \left[((k-1)\Gamma(s, T) + j\rho\sigma_s)^2 - 2((k-1)\Gamma(s, T) + j\rho\sigma_s) \sum_{l=j+1}^k \Gamma(s, t_{i_l}) \right. \right. \\
&\quad \left. \left. + \left(\sum_{l=j+1}^k \Gamma(s, t_{i_l}) \right)^2 \right] ds \right) \exp \left(\frac{1}{2} \sum_{j=0}^k \int_{t_{i_j}}^{t_{i_{j+1}}} j^2(1-\rho^2)\sigma_s^2 ds \right) \\
&= \frac{\prod_{j=1}^k B(0, t_{i_j})}{B(0, T)^{k-1}} \\
&\exp \left(\frac{k-1}{2} \int_0^T \Gamma(s, T)^2 ds - \frac{1}{2} \sum_{j=0}^k \int_{t_{i_j}}^{t_{i_{j+1}}} \sigma_s^2 ds - \frac{1}{2} \sum_{j=0}^k \int_{t_{i_j}}^{t_{i_{j+1}}} \sum_{l=j+1}^k \Gamma(s, t_{i_l})^2 ds \right) \\
&\exp \left(\frac{1}{2} \sum_{j=0}^k \int_{t_{i_j}}^{t_{i_{j+1}}} ((k-1)^2\Gamma(s, T)^2 + 2j(k-1)\rho\sigma_s\Gamma(s, T) + j^2\rho^2\sigma_s^2) \right) \\
&\exp \left(- \sum_{j=0}^k \int_{t_{i_j}}^{t_{i_{j+1}}} ((k-1)\Gamma(s, T) + j\rho\sigma_s) \sum_{l=j+1}^k \Gamma(s, t_{i_l}) ds \right) \\
&\exp \left(\frac{1}{2} \sum_{j=0}^k \int_{t_{i_j}}^{t_{i_{j+1}}} \sum_{l=j+1, m=j+1}^k \Gamma(s, t_{i_l})\Gamma(s, t_{i_m}) ds \right) \\
&\exp \left(\frac{1}{2} \sum_{j=0}^k \int_{t_{i_j}}^{t_{i_{j+1}}} j^2(1-\rho^2)\sigma_s^2 ds \right) \\
&= \frac{\prod_{j=1}^k B(0, t_{i_j})}{B(0, T)^{k-1}} \\
&\exp \left(\frac{k(k-1)}{2} \int_0^T \Gamma(s, T)^2 ds + \frac{1}{2} \sum_{j=0}^k \int_{t_{i_j}}^{t_{i_{j+1}}} (j^2 - j)\sigma_s^2 ds \right. \\
&\quad \left. + \sum_{j=0}^k \int_{t_{i_j}}^{t_{i_{j+1}}} j(k-1)\rho\sigma_s\Gamma(s, T) ds \right) \\
&\exp \left(- \sum_{j=0}^k (k-1) \int_{t_{i_j}}^{t_{i_{j+1}}} \left(\sum_{l=j+1}^k \Gamma(s, T)\Gamma(s, t_{i_l}) \right) - \rho \sum_{j=0}^k \int_{t_{i_j}}^{t_{i_{j+1}}} j\sigma_s \sum_{l=j+1}^k \Gamma(s, t_{i_l}) \right)
\end{aligned}$$

$$\begin{aligned}
& \exp \left(\sum_{j=0}^k \int_{t_{i_j}}^{t_{i_{j+1}}} \sum_{j+1 \leq l < m \leq k} \Gamma(s, t_{i_l}) \Gamma(s, t_{i_m}) ds \right) \\
&= \frac{B(0, t_{i_1}) \cdots B(0, t_{i_k})}{B(0, T)^{k-1}} \exp \left(\frac{k(k-1)}{2} \int_0^T \Gamma(s, T)^2 ds + \frac{1}{2} \sum_{j=2}^k \int_{t_{i_j}}^{t_{i_{j+1}}} (j^2 - j) \sigma_s^2 ds \right) \\
& \exp \left(-(k-1) \sum_{j=0}^k \int_{t_{i_j}}^{t_{i_{j+1}}} \sum_{l=j+1}^k \Gamma(s, T) \Gamma(s, t_{i_l}) ds + \sum_{j=0}^k \int_{t_{i_j}}^{t_{i_{j+1}}} \sum_{j+1 \leq l < m \leq k} \Gamma(s, t_{i_l}) \Gamma(s, t_{i_m}) \right) \\
& \exp \left(\rho \sum_{j=1}^k \int_{t_{i_j}}^{t_{i_{j+1}}} j(k-1) \sigma_s \Gamma(s, T) ds - \rho \sum_{j=1}^k \int_{t_{i_j}}^{t_{i_{j+1}}} j \sigma_s \sum_{l=j+1}^k \Gamma(s, t_{i_l}) ds \right).
\end{aligned}$$

Puisque $\Gamma(t, s) = \sigma^r \frac{1 - e^{-a(s-t)}}{a}$, $t \leq s$, on obtient :

$$\begin{aligned}
& c_{i_1, \dots, i_k} \\
&= \frac{B(0, t_{i_1}) \cdots B(0, t_{i_k})}{B(0, T)^{k-1}} \\
& \exp \left(\frac{k(k-1)}{2} \int_0^T \left(\frac{\sigma^r}{a} \right)^2 (1 - e^{-a(T-s)})^2 ds + \frac{1}{2} \sum_{j=2}^k (j^2 - j) \sigma^2 (t_{i_{j+1}} - t_{i_j}) \right) \\
& \exp \left(- \sum_{j=0}^k (k-1) \int_{t_{i_j}}^{t_{i_{j+1}}} \sum_{l=j+1}^k \sigma^r \frac{1 - e^{-a(T-s)}}{a} \sigma^r \frac{1 - e^{-a(t_{i_l}-s)}}{a} ds \right) \\
& \exp \left(\sum_{j=0}^k \int_{t_{i_j}}^{t_{i_{j+1}}} \sum_{j+1 \leq l < m \leq k} \sigma^r \frac{1 - e^{-a(t_{i_l}-s)}}{a} \sigma^r \frac{1 - e^{-a(t_{i_m}-s)}}{a} ds \right) \\
& \exp \left(\rho \sum_{j=1}^k j(k-1) \sigma \int_{t_{i_j}}^{t_{i_{j+1}}} \sigma^r \frac{1 - e^{-a(T-s)}}{a} ds - \rho \sum_{j=1}^k j \sigma_s \int_{t_{i_j}}^{t_{i_{j+1}}} \sum_{l=j+1}^k \sigma^r \frac{1 - e^{-a(t_{i_l}-s)}}{a} ds \right) \\
&= \frac{B(0, t_{i_1}) \cdots B(0, t_{i_k})}{B(0, T)^{k-1}} \\
& \exp \left(\frac{k(k-1)}{2} \left(\frac{\sigma^r}{a} \right)^2 \left(T - 2 \frac{1 - e^{-aT}}{a} + \frac{1 - e^{-2aT}}{2a} \right) + \frac{1}{2} \sum_{j=2}^k (j^2 - j) \sigma^2 (t_{i_{j+1}} - t_{i_j}) \right) \\
& \exp \left(- \left(\frac{\sigma^r}{a} \right)^2 \sum_{j=0}^{k-1} (k-1) \int_{t_{i_j}}^{t_{i_{j+1}}} ((k-j)(1 - e^{-a(T-s)}) - \sum_{l=j+1}^k e^{-a(t_{i_l}-s)} \right. \\
& \quad \left. + \sum_{l=j+1}^k e^{-a(T+t_{i_l}-2s)}) ds \right) \\
& \exp \left(\left(\frac{\sigma^r}{a} \right)^2 \sum_{j=0}^{k-2} \sum_{j+1 \leq l < m \leq k} \int_{t_{i_j}}^{t_{i_{j+1}}} (1 - e^{-a(t_{i_l}-s)} - e^{-a(t_{i_m}-s)} + e^{-a(t_{i_l}+t_{i_m}-2s)}) ds \right)
\end{aligned}$$

$$\begin{aligned}
& \exp \left(\frac{\rho\sigma\sigma^r}{a} \sum_{j=1}^k j(k-1) \left(t_{i_{j+1}} - t_{i_j} - \frac{e^{-a(T-t_{i_{j+1}})} - e^{-a(T-t_{i_j})}}{a} \right) \right) \\
& \exp \left(-\frac{\rho\sigma\sigma^r}{a} \sum_{j=1}^k j \left((k-j)(t_{i_{j+1}} - t_{i_j}) - \sum_{l=j+1}^k \frac{e^{-a(t_{i_l}-t_{i_{j+1}})} - e^{-a(t_{i_l}-t_{i_j})}}{a} \right) \right) \\
& = \frac{B(0, t_{i_1}) \cdots B(0, t_{i_k})}{B(0, T)^{k-1}} \\
& \exp \left(\left(\frac{\sigma^r}{a} \right)^2 \frac{k(k-1)}{2} \left(T - 2 \frac{1 - e^{-aT}}{a} + \frac{1 - e^{-2aT}}{2a} \right) + \frac{1}{2} \sum_{j=2}^k (j^2 - j) \sigma^2 (t_{i_{j+1}} - t_{i_j}) \right) \\
& \exp \left(-\left(\frac{\sigma^r}{a} \right)^2 \left\{ \sum_{j=0}^{k-1} (k-1) \left((k-j)(t_{i_{j+1}} - t_{i_j}) - (k-j) \frac{e^{-a(T-t_{i_{j+1}})} - e^{-a(T-t_{i_j})}}{a} \right) \right. \right. \\
& \quad \left. \left. - \sum_{l=j+1}^k \frac{e^{-a(t_{i_l}-t_{i_{j+1}})} - e^{-a(t_{i_l}-t_{i_j})}}{a} + \sum_{l=j+1}^k \frac{e^{-a(T+t_{i_l}-2t_{i_{j+1}})} - e^{-a(T+t_{i_l}-2t_{i_j})}}{2a} \right\} \right) \\
& \exp \left(\left(\frac{\sigma^r}{a} \right)^2 \left\{ \sum_{j=0}^{k-2} \sum_{l=j+1}^{k-1} \sum_{m=l+1}^k \left(t_{i_{j+1}} - t_{i_j} - \frac{e^{-a(t_{i_l}-t_{i_{j+1}})} - e^{-a(t_{i_l}-t_{i_j})}}{a} \right. \right. \right. \\
& \quad \left. \left. - \frac{e^{-a(t_{i_m}-t_{i_{j+1}})} - e^{-a(t_{i_m}-t_{i_j})}}{a} + \frac{e^{-a(t_{i_l}+t_{i_m}-2t_{i_{j+1}})} - e^{-a(t_{i_l}+t_{i_m}-2t_{i_j})}}{2a} \right) \right\} \right) \\
& \exp \left(\frac{\rho\sigma\sigma^r}{a} \left\{ k(k-1)(T - t_{i_k}) + \sum_{j=1}^{k-1} j(j-1)(t_{i_{j+1}} - t_{i_j}) \right. \right. \\
& \quad \left. \left. - (k-1) \frac{k - \sum_{j=1}^k e^{-a(T-t_{i_j})}}{a} + \frac{k(k-1)}{2} - \sum_{j=1}^k \sum_{l=j+1}^k e^{-a(t_{i_l}-t_{i_j})} \right\} \right).
\end{aligned}$$

Par le lemme 7.4.1, on a

$$\begin{aligned}
& c_{i_1, \dots, i_k} \\
& = \frac{B(0, t_{i_1}) \cdots B(0, t_{i_k})}{B(0, T)^{k-1}} \\
& \exp \left(\left(\frac{\sigma^r}{a} \right)^2 \frac{k(k-1)}{2} \left(T - 2 \frac{1 - e^{-aT}}{a} + \frac{1 - e^{-2aT}}{2a} \right) + \frac{1}{2} \sum_{j=2}^k (j^2 - j) \sigma^2 (t_{i_{j+1}} - t_{i_j}) \right) \\
& \exp \left(-\left(\frac{\sigma^r}{a} \right)^2 \left\{ (k-1) \sum_{j=0}^{k-1} (k-j)(t_{i_{j+1}} - t_{i_j}) - (k-1) \frac{\sum_{j=1}^{k-1} e^{-a(T-t_{i_j})} - ke^{-aT}}{a} \right. \right. \\
& \quad \left. \left. - (k-1) \frac{k - \sum_{j=1}^k e^{-at_{i_j}}}{a} + (k-1) \frac{\sum_{j=1}^k e^{-a(T-t_{i_j})} - \sum_{j=1}^k e^{-a(T+t_{i_j})}}{2a} \right\} \right) \\
& \exp \left(\left(\frac{\sigma^r}{a} \right)^2 \left\{ \sum_{j=0}^{k-2} \frac{(k-j-1)(k-j)}{2} (t_{i_{j+1}} - t_{i_j}) - \frac{k(k-1)}{2} - \sum_{j=1}^{k-1} (k-j) e^{-at_{i_j}} \right. \right.
\end{aligned}$$

$$\begin{aligned}
& \left. \begin{aligned}
& - \frac{\sum_{j=1}^{k-1} \sum_{m=j+1}^k e^{-a(t_{im}-t_{ij})} - \sum_{j=2}^k (j-1)e^{-at_{ij}}}{a} \\
& + \frac{\sum_{j=1}^{k-1} \sum_{m=j+1}^k e^{-a(t_{im}-t_{ij})} - \sum_{l=1}^{k-1} \sum_{m=l+1}^k e^{-a(t_{im}-t_{il})}}{2a}
\end{aligned} \right\} \\
& \exp \left(\frac{\rho\sigma\sigma^r}{a} \left\{ k(k-1)(T-t_{i_k}) + \sum_{j=1}^{k-1} j(j-1)(t_{i_{j+1}}-t_{i_j}) + (k-1) \frac{\sum_{j=1}^k e^{-a(T-t_{i_j})}}{a} \right. \right. \\
& \left. \left. - \frac{\sum_{j=1}^k \sum_{l=j+1}^k e^{-a(t_{il}-t_{ij})}}{a} - \frac{k(k-1)}{2a} \right\} \right).
\end{aligned}$$

En arrangeant les termes, on obtient :

$$\begin{aligned}
& c_{i_1, \dots, i_k} \\
& = \frac{B(0, t_{i_1}) \cdots B(0, t_{i_k})}{B(0, T)^{k-1}} \\
& \exp \left(\left(\frac{\sigma^r}{a} \right)^2 \left\{ \frac{k(k-1)}{2} (T-t_{i_k}) + \sum_{j=2}^{k-1} \frac{j(j-1)}{2} (t_{i_{j+1}}-t_{i_j}) - k(k-1) \frac{e^{-2aT}}{4a} - \frac{k(k-1)}{4a} \right. \right. \\
& \left. \left. + (k-1) \frac{\sum_{j=1}^k e^{-a(T-t_{i_j})}}{2a} + (k-1) \frac{\sum_{j=1}^k e^{-a(T+t_{i_j})}}{2a} \right. \right. \\
& \left. \left. - \frac{\sum_{j=1}^{k-1} \sum_{m=j+1}^k e^{-a(t_{im}-t_{ij})}}{2a} - \frac{\sum_{j=1}^{k-1} \sum_{m=j+1}^k e^{-a(t_{im}+t_{ij})}}{2a} \right\} \right) \\
& \exp \left(\frac{\rho\sigma\sigma^r}{a} \left\{ k(k-1)(T-t_{i_k}) + \sum_{j=2}^{k-1} j(j-1)(t_{i_{j+1}}-t_{i_j}) - \frac{k(k-1)}{2a} \right. \right. \\
& \left. \left. + (k-1) \frac{\sum_{j=1}^k e^{-a(T-t_{i_j})}}{a} - \frac{\sum_{j=1}^{k-1} \sum_{m=j+1}^k e^{-a(t_{im}-t_{ij})}}{a} \right\} \right) \\
& \exp \left(\frac{1}{2} \left(\sum_{j=2}^k j(j-1)(t_{i_{j+1}}-t_{i_j}) \right) \sigma^2 \right). \tag{7.4.3}
\end{aligned}$$

C'est cette dernière formule qui est utilisée pour l'implémentation numérique.

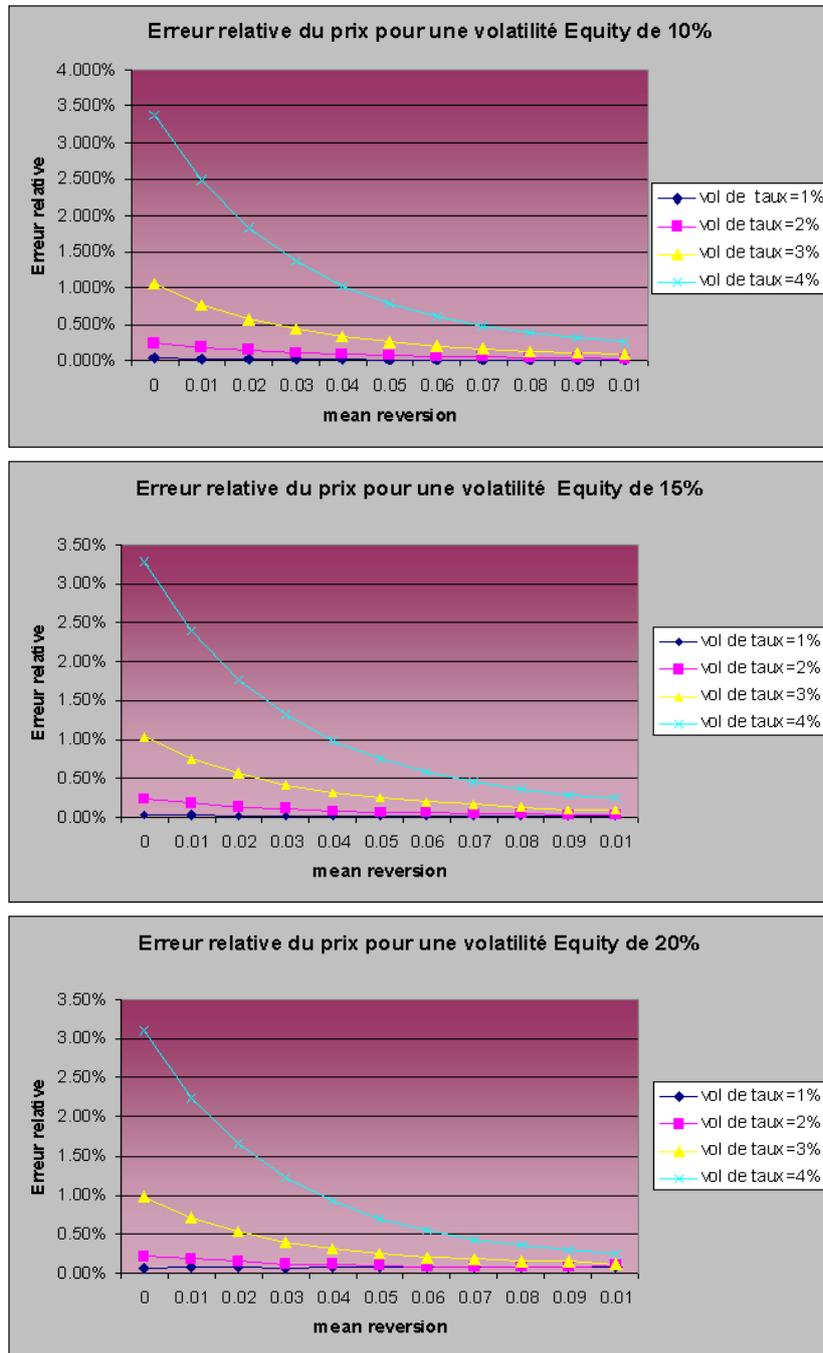


FIGURE 7.2 – Erreur relative du prix DAS5 pour $\delta=9.5$ (retrait annuel à partir de l'année 6 jusqu'à maturité (pas de retrait avant)).

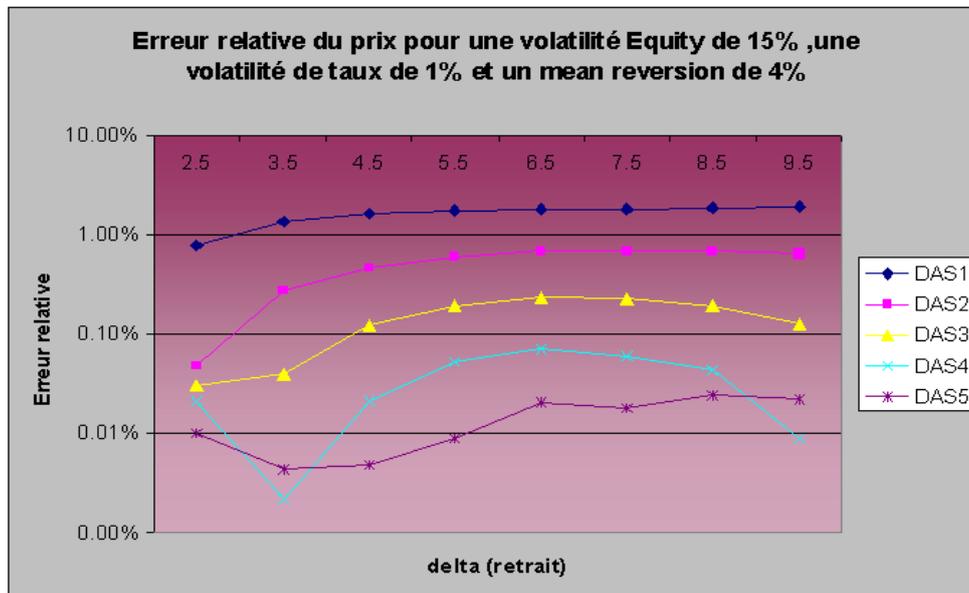


FIGURE 7.3 – Erreur relative en échelle log du prix DAS5 en fonction de δ (retrait annuel) pour $\sigma = 15\%$, $\sigma^r = 1\%$ et $a=4\%$.

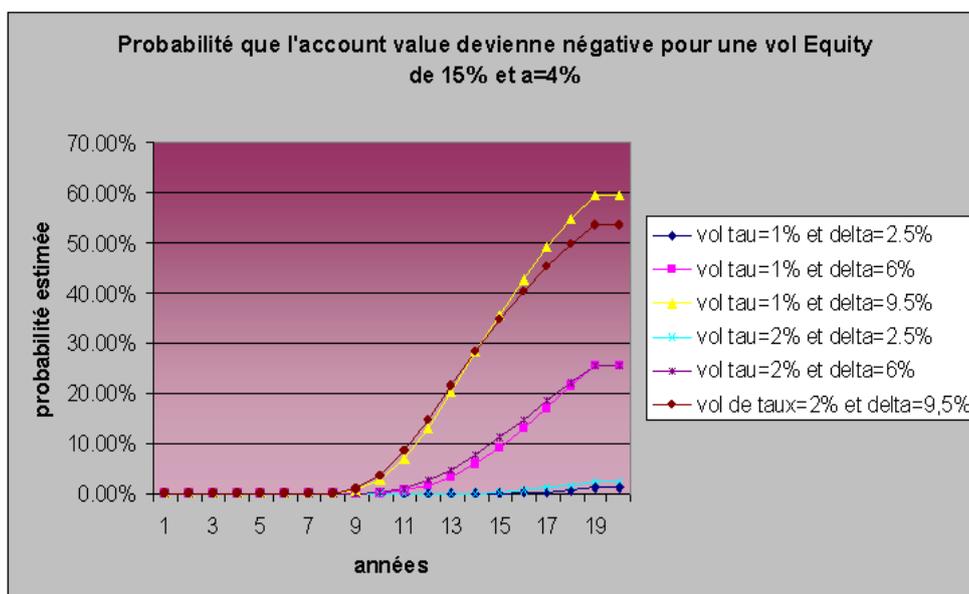


FIGURE 7.4 – Probabilité pour que l'account value devienne négative pour $\sigma = 15\%$ et $a=4\%$.

Bibliographie

- [ABF⁺01] M. Avellaneda, R. Buff, C. Friedman, N. Grandchamp, L. Kruk, and J. Newman. Weighted Monte Carlo : a new approach for calibrating asset-pricing models. *International Journal of Theoretical and Applied Finance*, 4(1) :91–119, 2001.
- [Bac03] A.R. Bacinello. Fair valuation of a guaranteed life insurance participating contract embedding a surrender option. *The Journal of Risk and Insurance*, 70(3) :461–487, 2003.
- [Bac05] A.R. Bacinello. Endogenous model of surrender conditions in equity-linked life insurance. *Insurance : Mathematics and Economics*, 37(2) :270–296, 2005.
- [BBD⁺06] C. Barrera, F. Bergeret, C. Dossal, E. Gobet, A. Meziou, R. Munos, and D. Reboul. Numerical methods for the pricing of swing options : a stochastic control approach. *Methodology and Computing in Applied Probability*, 8(4) :517–540, 2006.
- [BBM09] A.R. Bacinello, E. Biffis, and P. Millossovich. Pricing life insurance contracts with early exercise features. *Journal of Computational and Applied Mathematics*, 233(1) :27–35, 2009.
- [BBM10] A.R. Bacinello, E. Biffis, and P. Millossovich. Regression-based algorithms for life insurance contracts with surrender guarantees. *Quantitative Finance*, 10(9) :1077–1090, 2010.
- [BGM09] E. Benhamou, E. Gobet, and M. Miri. Smart expansion and fast calibration for jump diffusion. *Finance and Stochastics*, 13(4) :563–589, 2009.
- [BGM12] E. Benhamou, E. Gobet, and M. Miri. Analytical formulas for local volatility model with stochastic rates. *Quantitative Finance*, 12(2) :185–198, 2012.
- [BKR08] D. Bauer, A. Kling, and J. Russ. A universal pricing framework for guaranteed minimum benefits in variable annuities. *ASTIN Bulletin*, 38 :621–651, 2008.
- [BM07] D. Brigo and F. Mercurio. *Interest rate models, theory and practice*. Springer Finance, 2007.

- [BMOP11] A.R. Bacinello, P. Millossovich, A. Olivieri, and E. Pitacco. Variable annuities : A unifying valuation approach. *Insurance : Mathematics and Economics*, 49(3) :285–297, 2011.
- [BO93] A.R. Bacinello and F. Ortu. Pricing equity-linked life insurance with endogenous minimum guarantees. *Insurance : Mathematics and Economics*, 12 :245–257, 1993.
- [BPS01] J.P. Bouchaud, M. Potters, and D. Sestovic. Hedged Monte Carlo : Low variance derivative pricing with objective probabilities. *Physica A*, 289 :517–525, 2001.
- [BR02] T.R. Bielecki and M. Rutkowski. *Credit Risk : Modeling, Valuation and Hedging*. Springer-Verlag, 2002.
- [Bre99] H. Brezis. *Analyse fonctionnelle*. Dunod, 1999.
- [BS76] M.J. Brennan and E.S. Schwartz. The pricing of equity-linked life insurance policies with an asset value guarantee. *Journal of Financial Economics*, 3(3) :195–213, 1976.
- [BS77] P.P. Boyle and E.S. Schwartz. Equilibrium prices of guarantees under equity-linked contracts. *Journal of Risk and Insurance*, 44 :639–660, 1977.
- [BS78] D.P. Bertsekas and S.E. Shreve. *Stochastic optimal control*. Mathematics in Science and Engineering, 1978.
- [BW12] B. Bouchard and X. Warin. Monte-carlo valuation of American options : facts and new algorithms to improve existing methods. *Numerical Methods in Finance, Springer*, pages 217–258, 2012.
- [CLP02] E. Clément, D. Lamberton, and P. Protter. An analysis of a least square regression method for American option pricing. *Finance and Stochastics*, 6(4) :449–471, 2002.
- [DKZ08] M. Dai, Y.K. Kwok, and Jianping Zong. Guaranteed minimum withdrawal benefit in variable annuities. *Mathematical Finance*, 18(4) :595–611, 2008.
- [EG12] P. Eto and E. Gobet. Stochastic expansion for the pricing of call options with discrete dividends. *Applied Mathematical Finance*, 19(3) :233–264, 2012.
- [Egl05] D. Egloff. Monte Carlo algorithms for optimal stopping and statistical learning. *Annals of Applied Probability*, 15(2) :1396–1432, 2005.
- [FS86] H. Föllmer and D. Sondermann. Hedging of non-redundant contingent claims. *Contributions to Mathematical Economics. In Honor of G. Debreu*, pages 205–223, 1986.

- [GJ97] A. Grosen and P.L. Jorgensen. Valuation of early exercisable interest rate guarantees. *The Journal of Risk and Insurance*, 64(3) :481–503, 1997.
- [GJ00] A. Grosen and P.L. Jorgensen. A fair valuation of life insurance liabilities : The impact of interest rate guarantees, surrender options, and bonus policies. *Insurance : Mathematics and Economics*, 26(1) :37–57, 2000.
- [GKKW02] L. Györfi, M. Kohler, A. Krzyżak, and H. Walk. *A distribution-free theory of nonparametric regression*. Springer Series in Statistics, 2002.
- [GLW05] E. Gobet, J.P. Lemor, and X. Warin. A regression-based Monte Carlo method to solve backward stochastic differential equations. *Annals of Applied Probability*, 15(3) :2172–2202, 2005.
- [GLW06] E. Gobet, J.P. Lemor, and X. Warin. Rate of convergence of an empirical regression method for solving generalized backward stochastic differential equations. *Bernoulli*, 12(5) :889–916, 2006.
- [GVL96] G. Golub and C.F. Van Loan. *Matrix computations*. The Johns Hopkins University Press, 1996.
- [GY04] P. Glasserman and B. Yu. Number of paths versus number of basis functions in American option pricing. *The Annals of Applied Probability*, 14(4) :2090–2119, 2004.
- [HKR08] D. Holz, A. Kling, and J. Russ. GMWB for life. an analysis of lifelong withdrawal guarantees. *Working Paper, Ulm University*, 2008.
- [JRT04] P. Jaillet, E.I. Ronn, and S. Tompaidis. Valuation of commodity-based swing options. *Management Science*, 50 :909–921, 2004.
- [Kar88] I. Karatzas. On the pricing of American options. *Applied Mathematics and Optimization*, 17(1) :37–60, 1988.
- [KRR11] A. Kling, F. Ruez, and J. Russ. The impact of stochastic volatility on pricing, hedging, and hedge efficiency of withdrawal benefit guarantees in variable annuities. *ASTIN Bulletin*, 41(2) :511–545, 2011.
- [KRR12] A. Kling, F. Ruez, and J. Russ. The impact of policyholder behavior on pricing, hedging, and hedge efficiency of withdrawal benefit guarantees in variable annuities. http://www.ifa-ulm.de/downloads/Impact_Of_Stochastic_Volatility_VA_Guarantees.pdf, *Working Paper*, 2012.
- [KS88] I. Karatzas and S.E. Shreve. *Brownian motion and stochastic calculus*. Springer-Verlag, 1988.
- [KS98] I. Karatzas and S.E. Shreve. *Method of mathematical finance*. Springer-Verlag New York, 1998.

- [Lem05] J.P. Lemor. *Approximation par projections et simulations de Monte-Carlo des équations différentielles stochastiques rétrogrades*. PhD thesis, Ecole Polytechnique, 2005.
- [LS01] F. Longstaff and E.S. Schwartz. Valuing American options by simulation : A simple least square approach. *The Review of Financial Studies*, 14 :113–147, 2001.
- [Mar78] W. Margrabe. The value of an option to exchange one asset for another. 1978.
- [Mol98] T. Moller. Risk-minimizing hedging strategies for unit-linked life insurance contracts. *ASTIN Bulletin*, 28 :17–47, 1998.
- [Mol01a] T. Moller. Hedging equity-linked life insurance contracts. *North American Actuarial Journal*, 5(2) :79–95, 2001.
- [Mol01b] T. Moller. Risk-minimizing hedging strategies for insurance payment processes. *Finance and Stochastics*, 5(4) :419–446, 2001.
- [MP01] M.A. Milevsky and S.A. Posner. The titanic option : valuation of the guaranteed minimum death benefit in variable annuities and mutual funds. *Journal of Risk and Insurance*, 68(1) :91–126, 2001.
- [MS06] M.A. Milevsky and T. Salisbury. Financial valuation of guaranteed minimum withdrawal benefits. *Insurance : Mathematics and Economics*, 38(1) :21–38, 2006.
- [NS95] J.A. Nielsen and K. Sandmann. Equity-linked life insurance : A model with stochastic interest rates. *Insurance : Mathematics and Economics*, 16 :225–253, 1995.
- [PLK12] J. Peng, K.S. Leung, and Y.K. Kwok. Pricing guaranteed minimum withdrawal benefits under stochastic interest rates. *Quantitative Finance*, 12(6) :933–941, 2012.
- [Pro04] P.E. Protter. *Stochastic Integration and Differential Equations*. Springer-Verlag, second edition, 2004.
- [Sch91] M. Schweizer. Option hedging for semimartingales. *Stochastic Processes and their Applications*, 37 :339–363, 1991.
- [SX05] W. Shen and H. Xu. The valuation of unit-linked policies with or without surrender options. *Insurance : Mathematics and Economics*, 36(1) :79–92, 2005.
- [TVR01] J.N. Tsitsiklis and B. Van Roy. Regression methods for pricing complex american-style options. *IEEE Transactions on Neural Networks*, 12(4) :694–703, 2001.
- [Vil99] S. Villeneuve. Exercice regions of American options on several assets. *Finance and Stochastics*, 3 :295–322, 1999.