Le management du risque pour les compagnies d'assurance : une approche marchés financiers

Saad Mouti

École doctorale nº 386 : Sciences Mathématiques de Paris Centre

THÈSE

pour obtenir le grade de docteur délivré par l'

Université Pierre et Marie Curie

Spécialité : Mathématiques Appliquées

présentée et soutenue publiquement par

Saad MOUTI

le 6 Décembre 2017

Le management du risque pour les compagnies d'assurance : une approche marchés financiers

Composition du jury :

Directeurs :	Nicole EL KAROUI	Université Pierre et Marie Curie
	Mathieu ROSENBAUM	Ecole Polytechnique
Encadrant entreprise :	Aymeric KALIFE	AXA
Rapporteurs :	Giorgia CALLEGARO	Université de Padoue
	Jim GATHERAL	Baruch College
	Caroline HILLAIRET	ENSAE
Examinateurs :	René AID	Université Paris-Dauphine
	Romuald ELIE	Université Paris-Est Marne-la-Vallée
	Gilles PAGES	Université Pierre et Marie Curie
	Christian ROBERT	ISFA

Université Pierre et Marie Curie
Laboratoire de Probabilités et de modèles Aléatoires (LPMA)
UMR 7599, 4 Place Jussieu 75005 Paris, France

Remerciements

Je tiens en premier lieu à exprimer ma gratitude envers mes directeurs de thèse, Nicole El Karoui et Mathieu Rosenbaum, pour leur disponibilité et les précieux conseils qu'ils m'ont donnés tout au long de cette thèse. Ça a été un honneur pour moi que Nicole accepte de codiriger cette thèse. Je la remercie pour son temps, le partage de sa connaissance et de son expérience mais aussi pour sa bienveillance. Je ne remercierai jamais assez Mathieu sans qui ce projet de recherche n'aurait jamais abouti. Il m'a énormément apporté sur le plan professionnel et humain, et a su me rassurer pendant les moments de doute. Sa pédagogie et son coté humain dépassent mes attentes.

L'aventure a commencé chez AXA et je dois beaucoup à mon encadrant en entreprise Aymeric Kalife qui m'a accueilli au GRM et grâce à qui j'ai eu le financement pour la thèse. Je le remercie de m'avoir accordé sa confiance, pour les sujets intéressants qu'il m'a proposés et pour ses conseils, échanges et collaborations enrichissantes.

Je remercie Giorgia Callegaro, Caroline Hillairet, ainsi que Jim Gatheral d'avoir accepté de rapporter ma thèse. Je suis très honoré par leur lecture attentive de ce manuscrit et leur intérêt pour mon travail. Je remercie également René Aïd, Romuald Elie, Gilles Pagès et Christian Robert d'examiner ma thèse et de participer à ma soutenance.

Je suis très reconnaissant envers mes professeurs du Master à Dauphine et plus particulièrement Idriss Kharroubi, Christian Robert et Nicola Choppin qui ont contribué à la poursuite de cette thèse.

Les travaux qui forment cette thèse ont été en grande partie en collaboration. Je remercie donc Ludovic Goudenège, Giulia Livieri, Gabriela López Ruiz, Andrea Pallavicini, Xiaolu Tan et Lihang Wang pour ces collaborations instructives.

Je remercie tous mes collègues du GRM qui ont rendu mon quotidien plus agréable et avec qui j'ai partagé de très bons moments. J'ai une pensée particulière pour les anciens thésards du GRM, Fatima, Medeleine-Sophie et Nabil, mais aussi tous les collègues et personnel d'AXA avec qui j'ai pu intéragir pendant cette thèse. Je cite Claire, Grace, Jamila, Marion, Mildraid, Mouna, Rachida, Rizlaine, Ahmed, Bernard, Berthold, Cyril, Eric, Johny, Jérémie, Mohamed, Pierre, Philipe et Tom, la liste étant non-exhaustive.

Je remercie également mes frères d'armes du LPMA de la même subdivision Pamela, Omar, Thibault, Jiatu et Weibing, mes collègues de bureau Alice, Candia, Nina, Eric et Yi pour ces années enrichissantes.

Je tiens à remercier tout le personnel du secrétariat, Florence, Serena et Josette pour leur serviabilité et leur agréable compagnie dans la tour 16/26, et le personnel de l'école doctorale pour leur aide administrative.

Pour finir, je remercie mes proches, amis et familles de leur patience. Un grand merci à mes frères et mes sœurs ainsi qu'à mon épouse Fatima Ezzahra de m'avoir rassuré et soutenu pendant cette thèse. Je resterai éternellement reconnaissant envers mes parents pour leur soutien inconditionnel pendant cette thèse et bien avant.

Résumé

Cette thèse traite plusieurs aspects des risques financiers liés aux contrats d'assurance vie. Elle étudie trois sujets distincts et est composée de six chapitres qui peuvent être lus indépendamment.

Le comportement de l'assuré est un risque majeur pour les assureurs dans le cadre de produits d'assurance vie comme les annuités variables. Ainsi, nous nous penchons dans les premiers chapitres sur le comportement optimal pour deux classes de produits commercialisés. Nous traitons le cas du rachat total pour les "guaranteed minimum account benefits" (GMAB), et le retrait optimal dans le cadre des "guaranteed minimum income benefit" (GMIB). Le troisième chapitre est dédié au management et à la couverture d'une classe de produits à unité de compte également commercialisés par les assurances.

Le quatrième chapitre traite l'exécution optimal d'un large portfeuille d'options. En effet, les produits d'assurance vie sont partiellement couverts statiquement par la détention d'options vanilles. Nous considérons le cas où la taille des trades affecte le prix des options et cherchons à définir la stratégie optimale permettant de minimiser le coût de l'acquisition de ce portefeuille de couverture en prenant en compte l'impact de marché.

Enfin, le dernier thème de la thèse étudie le processus de volatilité. A cet effet, nous utilisons deux types d'estimateurs. En l'absence de données haute fréquence, les estimateurs dit de "range" permettent de revérifier que la volatilité est rugueuse. Ensuite, en utilisant les prix d'options, l'estimateur volatilité implicite court-terme et une version raffinée de cette dernière permettent encore une fois d'aboutir à la même conclusion.

Abstract

This thesis tackles several aspects of financial risks encountered in the life insurance industry and particularly in a class of products insurers offer; namely variable annuities and unit-linked products. It consists of three distinct topics and is split into six chapters that can be read independently.

In variable annuities, policyholders' behavior is a major risk for the insurer that affects the life insurance industry in almost every aspect. The first two chapters deal with policyholders' optimal behavior for two classes of these products. We address the rational lapse behavior in the guaranteed minimum account benefit, and optimal withdrawals in the guaranteed minimum income benefit. The third chapter is dedicated to a class of unit-linked products from a managing and hedging point of view.

The third chapter addresses the optimal execution of a large book of options. Typically, life insurance products are partially hedged using vanilla options. We consider the case where trades are affected by the traded quantity, and seek to find an optimal strategy based on two criteria; expected cost and mean-variance.

Finally, in the last topic we study the volatility process using two different proxies. First, range-based estimators that rely on the asset price range data allow us to double-check that volatility is a rough process in the sense that it has a scaling parameter H less than 1/2. Then, using short time-to-maturity implied volatilities as a proxy for the spot volatility, and a refined version of it, we are able to confirm that the rough aspect of volatility is universal based on different proxies.

Keywords : GMAB, Variable Annuity, rational lapse strategy, stochastic interest, PDE, ADI, high-dimensional regression, GMIB, Variable Annuities, rational behavior, optimal withdrawals, PDE, dynamic programming, CPPI, dynamic multiplier, jump processes and gap risk, vanilla and gap options, Market impact, option pricing, optimal execution, mean-variance, stochastic control, HJB equation, range-based volatility, Garman-Klass estimator, Parkinson estimator, volatility scaling, rough volatility, fractional Brownian motion, fractional Ornstein-Uhlenbeck, RFSV, volatility forecasting, implied volatility, Medvedev-Scaillet approximation.

List of papers being part of this thesis

- A. Kalife, S. Mouti and L. Wang, *Financial risk management and the rational lapse strategy in life insurance policies*, Insurance Markets and Companies, 4(2), 2013.
- A. Kalife, L. Goudenege and S. Mouti (2014), *Managing gap risks in iCPPI for life insurance companies : a risk return cost analysis*, Insurance Markets and Companies, 5(2), 2014.
- A. Kalife and S. Mouti, *On Optimal Options Book Execution Strategies with Market Impact*, Market Microstructure and Liquidity, 2(3), 2016.
- S. Mouti *Range-based proxies and rough volatility*, pre-print, 2017.
- G. Livieri, S. Mouti, A. Pallavicini and M. Rosenbaum, *Rough volatility : evidence from option prices*, submitted, 2017.

Table des matières

Table des matières ix

Introduction 1

1 Chapter 1 - Financial risk management and the rational lapse strategy in life insurance policies . . . 6
2 Chapter 2 - Optimal behavior strategy in Guaranteed Minimum Income Benefit . . . 8
3 Chapter 3 - Managing gap risks in iCPPI for life insurance companies : a risk return cost analysis . . . 10
4 Chapter 4 - Optimal execution of options book under market impact . . . 15
5 Chapter 5 - Range-based proxies and rough volatility . . . 19
6 Chapter 6 - Volatility is rough : evidence from option price data . . . 21

I Life insurance products 23

1 Financial risk management and the rational lapse strategy in life Insurance policies 25
1.1 Introduction . . . 25
1.2 Description of the contract . . . 27
1.3 Valuation of a GMAB with zero lapse . . . 28
1.4 Valuation a GMAB with rational lapse assumption . . . 29
1.5 Life insurance policy pool . . . 31
1.6 Numerical tests . . . 32
1.7 Conclusion . . . 35

2 Optimal behavior strategy in the GMIB product 43
2.1 Product description . . . 43
2.2 A brief review of the literature . . . 50
2.3 Formulation and basic notations . . . 51
2.4 Contract valuation . . . 56
2.5 Results . . . 64
2.6 Conclusion . . . 73

3 iCPPI and Gap Risk 75
3.1 Introduction . . . 75
3.2 Methodolody . . . 78
3.3 Mitigating downside risk : Preventing from breaching the floor . . . 83
3.4 Conclusion . . . 87

II Market impact and volatility **95**

4 On Optimal Options Book Execution Strategies with Market Impact **97**
4.1 Introduction . . . 97
4.2 Market impact model : A transaction costs approach . . . 100
4.3 The optimal execution problem . . . 107
4.4 Adding the agent risk aversion : A mean-variance framework . . . 110
4.5 Numerical solution and results . . . 116
4.6 Extension to a local volatility model : A numerical method for the general case . . . 123

5 Range-based proxies and rough volatility **129**
5.1 Introduction . . . 129
5.2 Overview on range-based volatility estimation . . . 131
5.3 Range-based volatility as spot volatility proxy : empirical results . . . 134
5.4 RFSV model validation using range-based proxies . . . 139
5.5 Forecasting range-based volatility using the RFSV model . . . 147
5.6 Conclusion . . . 152

6 Rough volatility : evidence from option prices **153**
6.1 Introduction . . . 153
6.2 At-the-money implied volatility with short maturity as spot volatility proxy . . . 155
6.3 A refined implied volatility based proxy for the spot volatility . . . 158
6.4 On the upward bias when estimating the Hurst parameter . . . 162
6.5 Conclusion . . . 167

Introduction

Preliminary : General review of the life insurance industry

The world's older population is growing rapidly. According to data published in 2015 by the United Nations, there was a substantial increase of 48% (from 607 to 901 million) of people aged 60 or over between 2000 and 2015. And by 2050, the population aged 60 and over might reach nearly 2.1 billion. Moreover, the "oldest-old" (aged 80 or over) population accounted for 14% of old population (aged 60 or over) in 2015, and is expected to triple 2015's value by 2050, see [133].

As a result of these demographic shifts, longer life expectancy, increasing lifestyle and health-care costs, the idea that individuals and households need to plan for their own retirement is gaining a lot of attention. On the other hand, low interest rates are putting pressure on the insurance sector, pushing providers and consumers alike to look for ways to make the most of their assets. This situation has led to a significant reduction in the demand for traditional life insurance products and an increase in the demand for annuities and other financial planning products, see [66].

Insurance companies offer a range of savings products indexed on financial assets (stocks, funds, government bonds...). These so-called unit linked products constitute a major segment of insurance companies' earnings. One reason is that they allow the transfer of the financial risk to the insured, thus reducing the required capital for the insurance company's solvency. However, in order to make these products more attractive, insurers started to offer complex guarantees that got closer and closer to banks' financial options.

Life insurance guarantees introduced numerous risks for the insurer which need to be hedged using different hedging mechanisms (static or dynamic hedging, reinsurance,...). The relatively long maturity, i.e. at least 10 years, of these products adds an additional complexity compared to classical financial products, and necessitates robust methods to estimate and hedge the resulting risks. Unsuitable risk management combined with important products sales could lead to disasters, as was seen for Equitable Life [1] in 2000.

In this thesis, we analyze some of the financial risks related to the life insurance business, and present some of the potential solutions. It is split into two major parts. The first part focuses on some of the commercialized products and the behavioral risk linked to them, as well as an alternative to products with option guarantees. The second part covers financial risk in a general perspective. In the following introduction, we present a general review of the products we are concerned about, followed by an overview of the content of each chapter.

1. Equitable Life is a life insurance company in the United Kingdom found in 1762 and closed to new business in December 2000 due to large unhedged liabilities and high guaranteed fixed returns to investors without provision for adverse market changes.

Life insurance products : Unit linked products and variable annuities

Unit linked products

Unit linked policies generally consist in insurance products which accumulate capital. They are usually taken with the purpose of accumulating a financial benefit to the policyholder at a future point in time. The policyholder can typically choose from investment funds or individual stocks. The account value is given by the number of units acquired, multiplied by the price of one unit. Unit linked products can be extended by complementary types of investment guarantees, and can also be based on asset management strategies.

For unit linked products, the amount available for annuitization after an accumulation phase depends on the development of a mutual fund. In the 1995's, equity-index annuities were introduced in the U.S. as another form of unit linked products. Their returns rate is determined according to a formula that takes into account changes in an equity index, e.g. the S&P500 or a basket of equities or mutual funds. Furthermore, these products exhibit a minimum guarantee on the premium, e.g. 90% of the premium paid and additionally 3% annual interest rate. Other variations of unit linked products include, among others, flexible premium payments during the contract term, partial withdrawals during the deferral phase, or the possibility of shortening or extending the accumulation phase.

Annuities products & variable annuities

In the last decade, equity-linked policies have become more and more popular, exposing policyholders to financial markets and providing them with different ways to consolidate investment performance over time as well as protection against mortality-related risks. The best examples of such contracts are variable annuities.

The origin of annuities dates back to the Roman empire. The term annuities comes from the Latin word "annua" which means annual income. *Annua* were ancient Roman contracts which provided an individual with a stream of payments for a specified period of time, or for life, in exchange for an upfront payment. Gnaeus Domitrius Annius Ulpianis, a roman speculator and jurist credited with creating the very first actuarial life table, is cited as one of the earliest dealers of these annuities. Annuities were also used to compensate roman soldiers for military service. Nowadays, the most famous annuities products are "variable annuities".

Variable annuities are unit linked or managed fund vehicles, which offer optional guarantee benefits for the customer. They can be purchased either by a single payment, or a series of payments. This amount constitutes the principal of the contract and, apart from some upfront costs, is entirely invested into a reference portfolio. The investment options (subaccounts) offered by the insurance companies are typically mutual funds of stocks, bonds, money market instruments, or some combination of the three.

Depending on the terms and conditions specified by the contract, the insurer promises to make periodic payments to the policyholder on predefined future dates. These payments are usually determined as percentage of the invested premium and deducted from the contract value. Policyholders can choose the date when the payout phase begins. The retirement date is often recommended to initiate annuitization. Although clients can annuitize later than this, insurers usually specify a maximum annuitization date. For example, it could be the later of the policyholder's 95th birthday or the 15th anniversary of the contract.

Upon annuitization, there are a number of possibilities on the duration of the payments :

— The lifetime of the policyholder.

— The lifetime of the named beneficiary.

— A specified period such as 20 years.

— The longer of the policyholder's lifetime and a certain period.

Another feature that contributed to make these products attractive is the presence of tax incentives introduced by governments, particularly in the U.S., to support the development of individual pension solutions and contain public expenditure. But the main feature of variable annuities remains the possibility to benefit from various guarantees against investment and mortality/longevity risks. These guarantees are usually referred to as GMxB, where x stands for the class of benefits involved. The most famous GMxB riders on the market are :

— **Guaranteed Minimum Accumulation Benefits (GMAB)** : Also known as maturity guarantees, see [39], GMABs are one of the earliest products in the family of GMxB riders on variable annuities. The original form of this put-like rider is purchased for a fixed term (e.g. 10 years). The GMAB contract gives policyholders the ability to protect their retirement investments against downside market risk. They allow the policyholder to receive the greater of the account value and the benefit base at maturity, and may include some specific features.

— **Guaranteed Minimum Death Benefits (GMDB)** : Introduced in the market in the 1990's, GMDBs are guarantees in case of policyholder's death. Upon his death during the term of the contract, a specific monetary amount is passed on to a person of the policyholder's choosing, i.e. usually spouse or children. There are several variations for the death benefit. It may simply be the original premium when combined with other riders for example, or may accumulate deposits at a fixed rate.

— **Guaranteed Minimum Income Benefits (GMIB)** : GMIB riders, also launched in the 1990's, provide policyholders the right to convert the benefit base at the end of deferral period into annuities for life with a constant rate fixed at inception. Generally speaking, the value of the benefit base is not less than the initial account value paid by policyholders. Due to enduring competitions, most insurers currently add some "features" for these guarantees. For example, the benefit base can be reset to the high-water mark of the account value on anniversary dates (step-up or ratchet) when the market has performed well, or can roll up with a fixed percentage (known as roll-up rate, e.g. 2%), regardless of market conditions.

— **Guaranteed Minimum Withdrawal Benefits (GMWB) and Guaranteed Lifelong Withdrawal Benefits (GLWB)** : GMWB are a relatively recent innovation in the life insurance market. As their predecessors GMIBs, GMWB riders also promise a minimal annuity level from an initial investment capital, regardless of the performance of the underlying account value. The most important difference between GMIB and GMWB is the policyholder's surrender right after the end of the deferred period. While the policyholder has to give up all the asset invested once the contract enters into payment period for GMIB contract, in case of GMWB product, the insurer deducts the annuity and some charge fees. GMWB contract sets a limited payment period at inception. Thus, the insurer will stop paying annuities and return the remaining account value (if not exhausted) to the policyholder. GLWB contracts, also known as GMWB for life (GMWB-L), on the other hand, pay annuities until the policyholder's death.

— **Combo variable annuities :** In order to attract more investors, insurance companies propose combined variable annuities that offer more than one guarantee in a single contract. A typical example is the GMIB-DB, which combines an income benefit if the policyholder is alive at the contract maturity and a death benefit if he dies before.

Compared to fixed annuities, variable annuities are designed to protect against the effect of inflation on fixed income over the long run. Based on numbers from the Life Insurance and Market Research Association (LIMRA) [2], Figure 1 illustrates individual sales in variable annuities from the last decade in the U.S. In particular, sales were particularly high in 2006 and 2007 due to attractive guaranteed lifetime payments like GLWB offerings. However, the financial crisis brought about a strategic reset for many insurance carriers as interest rates were lowered and equity volatility rose. We also see from Fi-

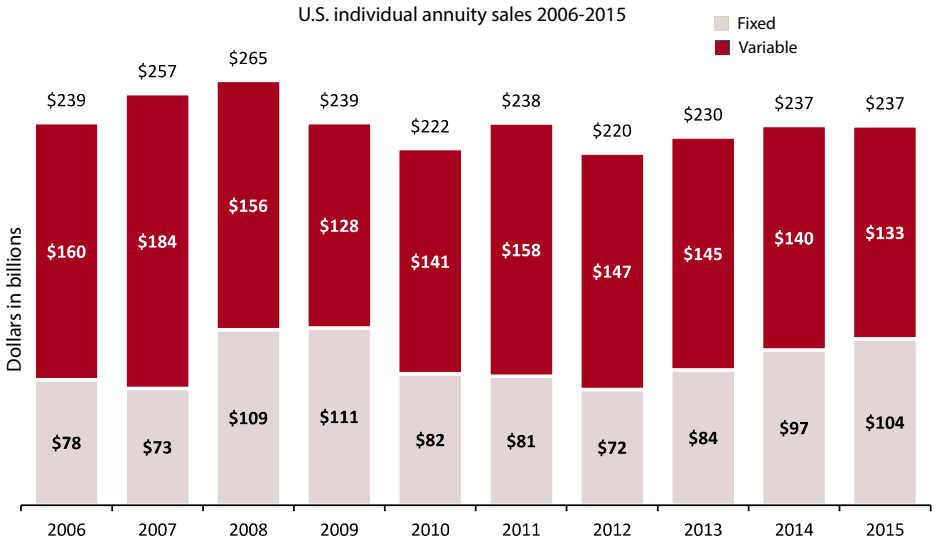


FIGURE 1 – Annuity Sales Estimates in the U.S. in 2006-2015. Source : *LIMRA Secure Retirement Institute U.S. individual annuities sales survey*

gure 1 a noticeable superiority of variable annuities in sales with respect to fixed annuities for almost a decade. Nevertheless, there is also an evident decline on variable annuities sales in recent years which can be explained by market volatility, growing popularity of indexed annuities, and a looming labor department, see [103]. Table 2 shows the top 20 sales leaders for total, variable and fixed annuities, as well as the top manufacturers of indexed annuities in 2015 published by LIMRA Secure Retirement on March 2016. AXA is ranked 8th in the total annuities sales and 5th in the variable annuities sales.

Variable annuities business seems to struggle to restart after 2008 crisis. In fact, years 2009 and beyond have been marked by the use of more dynamic asset allocation. Dynamic asset allocation comprises the use of "portfolio insurance" strategies and risk-control strategies. They allow insurers to tailor products that necessitate less capital requirement, and attract new costumers. In the last section, we will see one of these strategies and the product based upon it : the individualized constant proportion portfolio insurance product.

Individualized Constant Proportion Portfolio Insurance products

Constant proportion portfolio insurance is a dynamic asset allocation strategy between two pools of assets : a risky basket that is intended to provide the returns, and a safe basket that provides some level of a predefined capital protection. The percentage allocated to each depends on the "cushion" value, defined as (current portfolio value – floor value), and a multiplier coefficient, where a higher number denotes a more aggressive strategy.

2. LIMRA is an organization that conducts research on distribution systems for the life and health insurance products on behalf of its member companies.

Rank	Company name	Total	Company name	Variable	Company name	Fixed
1	Jackson National Life	24,491,828	Jackson National Life	23,109,447	Allianz Life of North America	8,773,123
2	AIG Companies	19,999,606	TIAA	12,752,518	New York Life	8,644,037
3	Lincoln Financial Group	14,638,405	Lincoln Financial Group	11,507,596	AIG Companies	8,508,110
4	TIAA	12,752,518	AIG Companies	11,491,496	American Equity Investment Life	7,083,967
5	New York Life	12,015,254	AXA US	9,848,026	Forethought Annuity	5,258,481
6	Allianz Life of North America	10,783,660	Prudential Annuities	8,722,770	Symetra Financial	4,085,009
7	MetLife	10,149,277	Transamerica	7,786,784	Great American	4,061,020
8	AXA US	9,875,961	MetLife	7,046,118	Nationwide	3,281,000
9	Prudential Annuities	9,539,028	Nationwide	5,374,000	Lincoln Financial Group	3,130,809
10	Nationwide	8,655,000	RiverSource Life Insurance		MetLife	3,103,160
11	Transamerica	7,882,188	Pacific Life	3,655,793	Midland National	3,056,870
12	American Equity Investment Life	7,083,967	New York Life	3,371,217	Pacific Life	2,931,193
13	Pacific Life	6,586,985	Thrivent Financial for Lutherans	3,312,568	Athene Annuity & Life	2,477,932
14	Forethought Annuity	6,452,977	Allianz Life of North America	2,010,537	Principal Financial Group	2,272,371
15	RiverSource Life Insurance	5,522,156	Fidelity Investments Life	2,007,503	Fidelity & Guaranty Life	2,059,554
16	Symetra Financial	4,113,791	Ohio National Life	1,952,352	Security Benefit Life	1,963,292
17	Great American	4,094,123	Northwestern Mutual Life	1,475,214	North American Company for Life and Health	1,951,044
18	Thrivent Financial for Lutherans	3,862,528	Forethought Annuity	1,194,495	EquiTrust Life	1,945,253
19	Midland National	3,436,721	Protective Life	1,155,228	Voya Financial	1,802,508
20	Principal Financial Group	3,232,160	Massachusetts Mutual Life	1,064,262	Western Southern Group	1,662,688
	Top 20	**$185,168,133**		**$124,066,116**		**$78,051,421**
	Total industry	**$236,677,000**		**$133,000,000**		**$103,677,000**
	Top 20 share	**78%**		**93%**		**75%**

FIGURE 2 – 2015 year-end U.S. Individual Annuity Sales given in thousands of dollars. Source : *LIMRA Secure Retirement Institute U.S. Individual Annuities Sales Survey*

In rising markets, the asset allocation mechanism (which is typically algorithmic and fully deterministic) is designed to either i) maintain maximum permitted exposure to the risky basket or ii) increase its exposure whenever the maximum permitted allocation has not been reached. In falling markets, the asset allocation mechanism will allocate more to the safe basket and in extreme markets will "monetize", i.e. allocate 100% of the portfolio value to the safe basket to ensure the guarantee at maturity.

In managing the portfolio assets in this manner, the CPPI asset allocation mechanism aims (but does not guarantee) to provide returns via the risky assets subject to meeting the predefined capital protection constant. Such mechanism, however, is also subject to major risks, such as high transaction costs, market liquidity risk, discontinuous price process, or unexpected changes in the volatility of the underlying stocks, which might imply a failure of the strategy, see [148]. This also includes the risk of monetization, which implies that a reallocation into the risky asset and a further participation in market upturns are no longer possible.

Individualized CPPI (iCPPI) employs the exact same mechanism as CPPI to the individual policyholder level rather than a pooled set of policies invested in the same CPPI fund, thereby allowing a life insurance company to provide protected / guaranteed solutions fully tailored to the individual policyholder. Furthermore, the market risk associated with providing such guarantees can be precisely and fully hedged out to a third party (typically an investment bank with iCPPI capabilities), leaving the life company with the residual actuarial risk that it is best positioned to manage.

Part I : Policyholder behavior in variable annuities and iCPPI gap risk

In variable annuities, policyholders' behavior is a major risk for the insurer, and a complex issue that affects life insurance industry in almost every aspect ; product design, pricing, marketing and distribution, financial reporting and risk management. Insurers' concern about policyholders behaviors risk is not new, and is still in the early stages of understanding and modeling. The first two chapters of this first part deal with this risk. We address the rational lapse behavior in GMABs, and optimal withdrawals in GMIBs. The third chapter is dedicated to the iCPPI product from a managing and hedging point of view.

1 Chapter 1 - Financial risk management and the rational lapse strategy in life insurance policies

In Chapter 1, we address the problem of pricing a GMAB contract under rational lapse assumption. A rational lapse assumes the policyholder behaves rationally through maximizing a certain criterion. In our case, we take the expected value.

We consider a policyholder possessing a GMAB contract defined in the preliminary section. The payoff of the guarantee is expressed as follows :

$$\mathrm{H}(\mathrm{T},\mathrm{A}(\mathrm{T}),\mathrm{G}) = \max(\mathrm{A}(\mathrm{T}),\mathrm{G}) = \mathrm{A}(\mathrm{T}) + (\mathrm{G}-\mathrm{A}(\mathrm{T}))^{+},$$

where A(T) is the account value at maturity, T denotes the maturity of the contract, and G the guaranteed minimum return, also known as the benefit base.

We assume that the account value A is invested in a single underlying asset, denoted by $\mathrm{S} = (\mathrm{S}(t))_{t\geq 0}$, following the Black-Scholes framework. The short term interest rate $r = r(t)_{t\geq 0}$ is driven by the one factor Hull and White model :

$$\begin{cases} d\mathrm{S}(t) = r(t)\mathrm{S}(t)dt + \sigma\mathrm{S}(t)d\mathrm{W}(t) \\ dr(t) = a(\theta(t) - r(t))dt + \sigma_r d\mathrm{Z}(t)\ ,\ \mathrm{Z} := (1-\rho^2)^{\frac{1}{2}}\mathrm{W}^{\perp} + \rho\mathrm{W}, \end{cases}$$

where a and σ_r are positive constants, θ is a deterministic Lebesgue-integrable function, σ is the instantaneous volatility of the asset return and Z and W are standard Brownian motions with correlation ρ.

We assume that annual fees α are deducted continuously from the policyholder's account value A. It is given by $\mathrm{A}(t) = e^{-\alpha t}\mathrm{S}(t)$ and can be expressed through the following stochastic differential equation :

$$d\mathrm{A}(t) = (r(t)-\alpha)\mathrm{A}(t)dt + \sigma\mathrm{A}(t)d\mathrm{W}(t),$$

where α is a constant corresponding to the total fees deducted from the account value.

If we assume that the policyholder does not lapse the contract prior to its maturity T, the guarantee becomes a European-style option. In this case, we can use the formula given in [102] to express the forward liability $\tilde{v}^{\mathrm{E}}$ of GMAB contract as :

$$\tilde{v}^{\mathrm{E}}(t,\mathrm{F}^{\mathrm{T}}(t)) = e^{-\alpha(\mathrm{T}-t)}\mathrm{F}^{\mathrm{T}}(t) + \mathrm{GN}(-d_2) - e^{-\alpha(\mathrm{T}-t)}\mathrm{F}^{\mathrm{T}}(t)\mathrm{N}(-d_1),\ \ d_{1,2} = \frac{\log(\mathrm{F}^{\mathrm{T}}(t)/\mathrm{G}) - \alpha(\mathrm{T}-t)}{\Gamma} \pm \frac{\Gamma}{2}.$$

where $F(t)^T$ is the forward value of $A(t)$ at time T observed at time t, $\Gamma = \sqrt{\int_0^{T-t} \omega_s^2 ds}$, $\omega_s^2 = \sigma^2 + \nu^2 B_r^2(s) + 2\rho\sigma\nu B_r(s)$, $B_r(s) = \frac{1-e^{-as}}{a}$ and $N(.)$ is the cumulative distribution function of the standard normal density.

Unlike European-style options, GMAB contracts allow the policyholder to get his account value before the maturity. As a consequence, a zero lapse assumption is not consistent with market practice. We observe that the lapse rate changes significantly in different market conditions (equity and interest rate level). Such variability has a notable impact on the liability value and the insurer's hedging strategy. This is known as the rational lapse strategy, and is similar to the optimal early-exercise strategy of classical Bermudian options, see [36].

Question. *Assuming the policyholder can lapse the contract at the end of each policy year, how can we define a rational strategy?*

Let $0 = t_0 < .. < t_n < ... < t_N = T$, $n = 1, 2, ..., N$ be the policy years. Lapses can take place at each anniversary date but not between two successive ones. We denote by t_n^- the time right before an anniversary date t_n, i.e. before the policyholder's decision to stay in the contract or lapse.

Due to the similarity to the optimal early-exercise strategy of classical Bermudian options, the value of the Bermudian-style liability $\tilde{v}^B$ under the forward measure $\mathbb{Q}^T$ satisfies the following equation :

$$\tilde{v}^B(t_n^-, F^T(t_n^-)) := \max(F^T(t_n), \mathbb{E}^{\mathbb{Q}^T}[\tilde{v}^B(t_{n+1}^-, F^T(t_{n+1}^-)) \mid \mathscr{F}_{t_n}],$$

where the terminal condition is

$$\tilde{v}^B(T, F^T(T)) = F^T(T) + (G - F^T(T))^+.$$

Result 1. *Assuming the benefit base* G *is fixed at inception, $\tilde{v}^B(t, f)$ is convex and nondecreasing w.r.t the forward price f, thus, for each t_n $n = 1, ..., N$ there exists a real number $f^*(t_n^-)$ such that*

$$\begin{array}{lll} 0 \leqslant f < f^*(t_n^-) & \Rightarrow & \tilde{v}^B(t_n^-, f) > f \quad (\textit{No Lapse}), \\ f \geqslant f^*(t_n^-) & \Rightarrow & \tilde{v}^B(t_n^-, f) = f \quad (\textit{Lapse}). \end{array}$$

$f^*(t_n^-)$ is referred to as the "critical forward account value" since the policy should be lapsed as soon as the forward account value increases to this level at time t_n.The existence of such critical value results in an increase of the value of the liability compared to the European-style one. As a consequence, the rational lapse case is the worst case scenario the insurer can face, therefore, it is crucial to take it into account.

To evaluate the liability $\tilde{v}$ and the critical forward value f^* of a single GMAB rider, we use two numerical methods : PDE schemes and Monte Carlo simulations. To our knowledge, this study is the first to apply these methods for a GMAB contract. Numerical tests, detailed in the chapter, show not only the consistency between the PDE and Monte Carlo methods, but also the precision of both methods. We also find that PDE method is faster and more precise than Monte Carlo. Moreover, it can calculate the price and other important Greeks for different f and t at the same time. On the other hand, Monte Carlo method is more flexible, easier to implement and can be extended to other high-dimensional problems.

Insurers are aware that they can neither rely on zero lapse strategy, nor consider a full rational behavior. In fact, they deal with an inhomogeneous pool of policies that can be terminated for various reasons which are not necessarily rational, and others that remain until their expiration. In this case, the lapse strategy can be represented by the frequency of the policies that are early terminated at a given anniversary date t_n. The goal is to find a way to estimate a "reasonable" lapse rate.

Question. *How can we use the rational strategy results to estimate a "reasonable" lapse frequency?*

Assume we want to predict policyholders lapses at a future anniversary date t_n. We can either use past observations and a regression model to predict future lapses, or use the results based on rational lapse strategy given present market observations. The first approach can underestimate lapse risk, while the second one focuses on market information and might be too extreme.

We denote by $p(t_n)$ the proportion of lapses at t_n. The rational lapse strategy allows us to express the proportion $p(t_n)$ as a deterministic function h of the forward account value $\mathrm{F}^{\mathrm{T}}(t_n)$, that is $h(\mathrm{F}^{\mathrm{T}}(t_n)) = \mathbb{1}_{\{\mathrm{F}^{\mathrm{T}}(t_n) \geq f^*(t_n^-)\}}$. This implies that all policyholders lapse the contract once $\mathrm{F}^{\mathrm{T}}(t_n)$ touches the critical boundary $f^*(t_n^-)$, and hold the policy otherwise.

Result 2. *Inspired by the mortgage prepayment models and evaluation approaches of surrender options for other life insurance products, we assume the lapse function is nonincreasing piece-wise linear of the value* $\mathrm{F}^{\mathrm{T}}(t_i)$. *When* $\mathrm{F}^{\mathrm{T}}(t_i) < f^*(t_i^-)$, *the lapse rate is not zero due the policyholder circumstances, and when* $\mathrm{F}^{\mathrm{T}}(t_i) \geq f^*(t_i^-)$, *some rational lapses never occur. Therefore we present a reasonable lapse frequency given by Figure 3.*

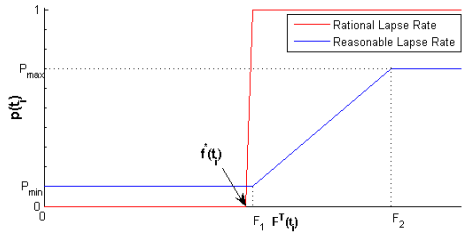


FIGURE 3 – Comparing the rational lapse function with the reasonable lapse function.

The four parameters F_1, F_2, $\mathrm{P}_{\min}$ *and* $\mathrm{P}_{\max}$ *are chosen by the insurer to match some empirical tests.*

By adopting this framework, insurers are able to incorporate policyholders rationality into their lapse assumptions, instead of limiting their modeling to empirical estimations, which happen to be quite reckless.

2 Chapter 2 - Optimal behavior strategy in Guaranteed Minimum Income Benefit

Rational behavior for GMAB products is limited to lapsing the contract. Other products are subject to optimal behaviors that include withdrawing money without necessarily terminating the contract, i.e. partial withdrawals. These so-called optimal withdrawals are for example experienced in GMWBs and GMIBs, i.e. see the preliminary section for details on these products. Many authors address this problem in the case of GMWBs, see for example [61, 110, 131, 143, 149]. In this chapter, we are interested in GMIBs which has got less attention because they were thought to be "safer". However, due to increasing guarantees and features, GMIBs became riskier and more exposed to optimal behaviors.

Consider an x-year old policyholder possessing a GMIB rider with an income benefit in case the insured is alive and the possibility of a death benefit in case he dies. Again, the account value A of the GMIB is invested in a single underlying asset, denoted by S_t following the Black-Scholes framework. Due to the complexity of the contract, we consider the simple case of constant interest rate.

Let $0 = t_0 < ... < t_n < ... < t_N = T$, n=1,2,...,N be the policy years, where t_0 is the contract inception and $t_N = T$ its maturity. All events related to the contract are considered to take place at these dates and not between two consecutive ones, i.e. evolution of the benefit base, withdrawals, lapse, payments, etc...

The value of the GMIB contract at time t is determined by three main state variables :

— The account value A_t which represents the wealth of the policyholder and evolves according to S_t performance.

— The benefit base G which is updated at anniversary dates by following a roll-up, i.e. $G_{t_n} = (1+\eta)G_{t_{n-1}}$, a ratchet, i.e. $G_{t_n} = \max(G_{t_{n-1}}, A_{t_n})$, or by combining the two, i.e. $G_{t_n} = \max((1+\eta)G_{t_{n-1}}, A_{t_n})$. The roll-up rate η is a constant fixed at inception by the insurer.

— A discrete state variable $I_n = \{0,1\}$ which informs if the policyholder died between $(t_{n-1}, t_n]$ or is still alive at time t_n. Therefore, we denote the death probability during $(t_{n-1}, t_n]$ as $q_n = \mathbb{P}(I_n = 0 \mid I_{n-1} = 1)$, and the probability of being alive at time t_n as p_n. Note that q_n and p_n depend on the age of the contract holder at time t_n and thus on the age x at $t_0 = 0$.

At time 0, A_0 and G_0 are set equal to the upfront premium. We denote by γ_n withdrawals at time $t_{n=1,...,N-1}$, and the values of the benefit base (resp. account value) just before and after events take place at t_n by $G_{t_n^-}$ and $G_{t_n^+}$ (resp. $A_{t_n^-}$ and $A_{t_n^+}$).

Let t_n be an arbitrary policy year. We consider the following modeling notations and assumptions :

— Fees deducted from both the account value and benefit base between t and $t+dt$ for $t \in (t_{n-1}, t_n]$ are expressed as :

$$\alpha_{tot}(A_t, G_t)\,dt = (\alpha^A A_t + \alpha^G G_{t_{n-1}^+})\,dt,$$

where α^A and α^G are constants defined by the insurer at inception.

— At an arbitrary anniversary date t_n, the contract guaranteed withdrawal is a proportion of the benefit base at time t_n^-, i.e. $\gamma_n^{gua} = \eta G_{t_n^-}$, where η is the guaranteed rate (which is the same as the roll-up rate).

— Upon withdrawing an amount γ_n, the account value and benefit base are reduced. Their evolution depends on both their values before the withdrawal and the withdrawn amount. We express those jump conditions as :

$$A_{t_n^+} := h^A(A_{t_n^-}, G_{t_n^-}, \gamma_n),$$
$$G_{t_n^+} := h^G(A_{t_n^-}, G_{t_n^-}, \gamma_n),$$

where the withdrawal amount γ_n belongs to an admissible space $\mathscr{A}_n(A_{t_n^-}, G_{t_n^-})$, and $h^A(.)$ (resp. $h^G(.)$) is some function that determines the change in the account value (resp. benefit base) subject to withdrawing γ_n. The change in A and G upon withdrawing γ_n may include penalties for excess withdrawals, i.e. withdrawals that exceed the guaranteed amount.

— The income benefit can be activated at any time between a starting date fixed by the insurer t_I, and the maturity T, i.e. typically the policyholder is given the choice to start his income period between the 10th anniversary of the contract and the 85th or 95th policyholder's birthday.

— If the policyholder dies during $(t_{n-1}, t_n]$, the contract allows for a payout of the death benefit $D(t_n, A_{t_n^-}, G_{t_n^-})$ to the beneficiary at t_n.

— The cash flow received by the policyholder at t_n is a function of the current account value, benefit base, and withdrawn amount $f_n(A_{t_n^-}, G_{t_n^-}, \gamma_n)$. It is usually equal to the withdrawn amount.

The specifications vary across different proposed products and different insurers and are quite painful to extract from the long product specifications document. Moreover, the academic literature usually presents different specifications for GMIBs than those commercialized.

We are interested in the worst case scenario from an insurer point of view in terms of policyholders behavior. This corresponds to the strategy that maximizes the expected value of their discounted future cash flows. Therefore, we will consider this pricing strategy to define a fair price and extract the optimal withdrawals at each anniversary date for a given state, i.e. level of the account value and benefit base. The following assumptions are considered :

— Financial risk can be eliminated by continuous hedging.

— Mortality risk is fully diversified via selling the contract to many people of the same age.

— Financial risk and mortality risk are independent.

By exploiting the Markovian property of the state variables, and taking the expectation w.r.t mortality, we can calculate the price under the optimal strategy. This is justified by the fact that the financial risk and mortality are independent[3], and that the policyholder's decision does not affect mortality.

Result 3. *The value function Φ which solves the optimal withdrawals problem is given by the following explicit recursion*

$$\Phi(t_n^+, \mathrm{A}, \mathrm{G}) = \mathbb{E}_{t_n^+}\left[\mathrm{B}_{n,n+1}\Phi(t_{n+1}^-, \mathrm{A}_{t_{n+1}^-}, \mathrm{G}_{t_{n+1}^-}) \mid \mathrm{A}, \mathrm{G}\right],$$
$$\Phi(t_n^-, \mathrm{A}, \mathrm{G}) = \max(\mathrm{P}(t_n^-, \mathrm{A}, \mathrm{G}), \max_{\gamma_n \in \mathscr{A}_n}\left(\Phi(t_n^+, h^{\mathrm{A}}(\mathrm{A}, \mathrm{G}, \gamma_n), h^{\mathrm{G}}(\mathrm{A}, \mathrm{G}, \gamma_n)) + p_n\gamma_n + p_{n-1}q_n\mathrm{D}(t_n^+, \mathrm{A}, \mathrm{G})\right),$$

where $\mathrm{B}_{n,n+1}$ *is the actualization factor between* t_n *and* t_{n+1} *and* $\mathscr{A}_n$ *is the set of admissible strategies.*

Remark 1. *The first equation translates the transition between t_{n+1}^- and t_n^+ backwards, while the second one provides the jump condition upon choosing between starting the income benefit, or staying in the contract. In the second case, the withdrawal is weighted by the probability that the insurer is alive at time t_n, i.e. p_n, and the death payout by the probability that he has lived up to t_{n-1} and dies in the interval $(t_{n-1}, t_n]$, i.e. $p_{n-1}q_n$. A linear search of the optimum is performed over the set of withdrawal strategies, which allows to find the optimal strategy $\boldsymbol{\gamma}_n^*$ at each time t_n for each account value* A *and benefit base* G.

The solution of the optimal withdrawals problem allows the insurer to mitigate policyholders' behavior risk related to GMIB contracts. Our analysis points out the following observations :

— The optimal withdrawal strategy of the policyholder is limited to 4 possibilities ; zero withdrawal, guaranteed withdrawal, income benefit election and lapse. Intermediate withdrawals do not seem to be among those decisions.

— Given the fees levels, the contract seems to be underpriced under the optimal withdrawal strategy for most cases. Either increasing the fees or adjusting the roll-up rate can be a solution to overcome this issue.

3 Chapter 3 - Managing gap risks in iCPPI for life insurance companies : a risk return cost analysis

In order to diversify their products, improve their attractiveness, and reduce their capital requirement, insurance companies offer a range of savings other than variable annuities. The individualized

3. It is a common hypothesis in life insurance industry to consider that financial and demographic risks are independent to price variable annuities.

constant proportion portfolio insurance (iCPPI) is one of them.

As mentioned in the preliminary section, an iCPPI product is based on the CPPI mechanism. It consists in decomposing the portfolio value V_t, at each time t, into a sum of a risky and non-risky exposure. Given a multiplier m, the amount invested in the risky asset is leveraged, such that, at each time t, the risky exposure e_t is equal to $m(V_t - F_t)$, where the "floor" F_t is the actualized guarantee G at time t, i.e. $F_t = Ge^{-r(T-t)}$. When the portfolio value "breaches" the floor, i.e. $V_t \leq F_t$, all its assets are switched to the non-risky part. Ideally, by doing so, the insurer is guaranteed to recover G at maturity.

The portfolio rebalancing occurs at discrete times, i.e. on a daily or weekly basis, and the value of the fund can be defined using a recursive formula :

$$V_{t_{k+1}} - F_{t_{k+1}} = \begin{cases} (V_{t_k} - F_{t_k})\left(\frac{S_{t_{k+1}}}{S_{t_k}} - (m-1)e^{r\frac{T}{N}}\right) & \text{if} \quad V_{t_k} - F_{t_k} > 0 \\ (V_{t_k} - F_{t_k})e^{r\frac{T}{N}} & \text{if} \quad V_{t_k} - F_{t_k} \leq 0. \end{cases}$$

We define the probability of breaching the floor as the probability that the portfolio final value falls below the guarantee, i.e. $P^{BF} := \mathbb{P}(V_T \leq G)$. This probability is null for continuous rebalancing but is strictly positive in practice. Moreover, it depends on the multiplier m, the change in the asset price between two rebalancing dates, the rebalancing frequency, and the interest rate.

If the underlying asset has jumps, we can no longer control the probability of breaching the floor, and the insurer can be subject to "gap risk", which occurs when the underlying asset has a negative jump. We study two solutions :

— Adjusting the multiplier m to reduce the probability of breaching the floor.

— Hedging strategies in case the underlying asset has jumps.

Question. *How can the insurer reduce the probability* P^{BF} *of breaching the floor in case of volatile market?*

This issue suggests choosing a time-varying multiplier that adapts to market conditions, namely the volatility, drift and interest rate level. A few studies were held in this sense to define a dynamic multiplier. Among them, the long-term risk sensitive portfolio optimization, see [93], gives a multiplier under which the strategy is optimal. This multiplier is equal to the excess return of the strategy divided by the variance. Following Merton's optimum consumption and portfolio rules, see [125], one can define a multiplier based on the the optimal certainty equivalent returns (CERs) approach using HARA utilities. These multipliers are, however, very small compared to the ones used in practice, reducing the upside potential of the risky asset. Our approach suggest choosing a value-at-risk based multiplier in which the investor defines his risk tolerance.

Result 4. *Based on the "value-at-risk based portfolio insurance" (VBPI) introduced in [106], our dynamic multiplier is expressed by*

$$m_t = \frac{1}{1 - \exp\left(\left(\mu - r - \frac{1}{2}\sigma^2\right)(T-t) - z_p\sigma\sqrt{T-t}\right)},$$

where $z_p = \Phi^{-1}(p)$ *is the quantile function of the standard normal distribution.*

Such multiplier allows the manager to adjust the drift and volatility in this expression to the market risk and return. As shown through backtestings, dynamic-CPPI performs better that a classical CPPI with fixed multiplier.

However, the dynamic multiplier does not totally eliminate the risk of sudden negative jumps observed in the market. To incorporate jumps in the asset price dynamic, we use Lévy processes. For this analysis we consider the Kou model introduced in [112].

The dynamic of the asset price S_t is given by :

$$\frac{dS_t}{S_t} = \mu dt + \sigma dW_t + d\Big(\sum_{i=1}^{N_t}(e^{Y_i}-1)\Big),$$

where μ and σ are constants, W_t is a standard Brownian motion, N_t is a Poisson process with rate λ, and Y_i a sequence of independent identically distributed (i.d.d.) non-negative random variables with a double exponential distribution with density :

$$f_Y(y) = q\eta^+ e^{-\eta^+ y}\mathbb{1}_{y\geq 0} + p\eta^- e^{\eta^- y}\mathbb{1}_{y<0},$$

where $\eta^+ > 1$, $\eta^- > 0$ and p, q, $p+q=1$, represent the probabilities of upward and downward jumps.

Question. *How can the manager deal with rapid downside movements ?*

Result 5. *Rapid downside market movements, known as "gaps", can be hedged using :*

— *Semi-static hedging with vanilla put options.*

— *Static hedging with gap options.*

These solutions were analyzed in [57]. The first one consists in buying vanilla puts at each rebalancing date t_k, maturing the following one t_{k+1} with strike $(1-1/m)e^{rT/N}S_{t_k}$. The insurer can deduct the hedging costs from the portfolio final value and, in this case, they represent the sum of all the put options required for the whole period :

$$\bar{c} = \sum_{k=0}^{n-1} m\frac{C_{t_k}}{S_{t_k}}\mathbb{E}^{\mathbb{Q}}\Big[\Big((1-1/m)e^{rT/n}S_{t_k} - S_{t_{k+1}}\Big)\Big],$$

where the notation C_{t_k} is used for the cushion $V_{t_k} - F_{t_k}$ at time t_k.

In practice, the portfolio is self-financed and the price of the put used for the hedge is deducted directly from the portfolio's value. This is expressed through an adjusted cushion $\tilde{C}_t$ given by the following recursive formula :

$$\tilde{C}_{t_{k+1}} = e^{-rT/N}\tilde{C}_{t_k}\frac{\Big(m\frac{S_{t_{k+1}}}{S_{t_k}} + (1-m)e^{rT/N}\Big)^+}{\mathbb{E}^{\mathbb{Q}}\Big[\Big(m\frac{S_{t_{k+1}}}{S_{t_k}} + (1-m)e^{rT/N}\Big)^+ \mid \mathcal{F}_t\Big]}.$$

In practice, using options can be hardly applicable for the following reasons :

— The maturity is of 1 to 5 days and options reaching maturity are very volatile or illiquid.

— The insurer usually uses "exotic" underlying assets like funds or funds of funds, or detains a mixed portfolio of different funds, indexes, bonds etc...

— The cost of the strategy can be very high due to the number of options needed for the whole period.

Gap options are derivatives which allow for a protection against sudden significant downside moves of an underlying asset. If a gap event occurs between two consecutive dates, i.e. the asset performance is below a certain level called trigger, the option is exercised and the buyer receives the difference between the performance of the underlying asset, and a fixed threshold. These options are particularly useful in the case of iCPPI products. Indeed, through dynamic allocation strategy, the insurer's protection only fails if $m\frac{S_{t_{k+1}}}{S_{t_k}} + (1-m)e^{r\mathrm{T/N}} < 0$. By choosing a put gap option, i.e. with payoff $f(x) = (\mathrm{K}-x)^+$, and a strike $\mathrm{K} = (1-1/m)e^{r\mathrm{T/N}}$, the insurer is protected against gap events and ensured to recover the guarantee at maturity.

Our numerical analysis is based on the Kou model under some mild assumptions. The gap option price has the following reduced closed form :

$$\mathrm{G}_h \approx \frac{\lambda p \eta^-}{1+\eta^-} \mathrm{K}^{1+1/\eta^-} \frac{1-e^{-\mathrm{T}(r+\lambda p e^{\beta/\eta^-})}}{r+\lambda p e^{\beta/\eta^-}},$$

where p is the probability that a jump is negative, η^- its intensity, λ the Poisson process rate, $h = \mathrm{T/N}$ is the time step, and β the log return level which triggers the gap option.

The comparison between hedging using vanilla and gap options reveals that, even though both are equivalent, the second choice is less costly. Indeed, gap options are priced over-the-counter based on very rare events. Furthermore, while put options need to be bought at each rebalancing date with the following one as their maturity, and use a floating strike, gap options need to be acquired at time 0 for the whole period with a fixed strike.

Part II : Focus on market impact and volatility

Insurance products like variable annuities are characterized by three main features : long-term duration, large volumes, and significant market risk exposure. Large volumes means a greater risk of impacting the market when executing a large order, for hedging purposes for example. Market risk exposure and long-term duration induce an important volatility risk, which, in light of all the recent developments in volatility modeling, suggests a need to revisit some of the fundamentals of volatility and its properties.

In this second part of the thesis, we address these two problems ; market impact and volatility. Before exposing the analyses and results for each of these subjects, we give a brief review and motivate the choice for these problematics. In Part II-A, which addresses the optimal execution of a large book of options, we give a review of the current practices in hedging variable annuities and long-term equity-linked products.

Part II- A : Hedging life insurance products and the market impact dilemma

Guarantees in variable annuities are similar to options. The underlying asset price is the fund value, and the insurer plays the role of the option writer. However, unlike options where the premium is paid upfront, the costs of variable annuities are paid periodically as a percentage of the account value throughout the life of the contract. The fees collected should then be partially used to hedge the provided guarantees. Hedging variable annuities can be done via dynamic hedging, static hedging or semi-static hedging.

— **Dynamic hedging :** It is the most common approach to hedge financial risks in variable annuities embedded guarantees, see [109] and [110]. The principal of dynamically hedging a guarantee that depends on a tradable asset S_t, is to hold, at any time t, Δ_t shares of the underlying asset. In a complete market with continuous rebalancing and no transaction costs, the guarantee can be perfectly hedged in a self-financing manner.

— **Static hedging :** Investigated by [96] and [123], it suggests replicating the embedded guarantees with a static position in put options. The agent takes position, at inception, in a portfolio of financial instruments available in the market, so that the variable annuities cash flows can be replicated and match the cash flows of the hedge portfolio. Once the position set, static hedging assumes no intermediary costs between the contract's inception and maturity. Consequently, static hedging is considered to be highly robust, model independent, and does not involve any rebalancing throughout the life of the product. Unfortunately, this approach is far from being perfect. In fact, long-term options with maturities of at least 10 years are not available, illiquid, or subject to counterparty risk. Moreover, most guarantees are path-dependent and vanilla options are not suitable to hedge them. Last but not least, variable annuities fees are collected periodically, which makes it hard to match future fees and the amount borrowed at the inception to purchase the hedge.

— **Semi-static hedging :** Similar to static hedging, semi-static hedging exploits available financial instruments to hedge the guarantee. Instead of fixing the hedging portfolio from the contract inception until its maturity, here the insurer constructs a hedging portfolio at each rebalancing date by following an optimal hedging strategy for some optimality criterion. Several authors have studied this approach, see [25, 53, 54, 111]. In particular, static-hedging is investigated for a GMDB contract with ratchet features in [53, 54]. Under certain assumptions, the GMDB rider with a ratchet feature can be seen as a lookback option. They show that using vanilla options to hedge the guarantee can be significantly more effective than delta-hedging. [111] propose to hedge a path-dependent option by taking a position in an optimally chosen European option. They present results of a simulation study of a version of the GMWB product and use local risk minimization as the optimality criterion for each hedging date.

A review of derivatives holdings in the U.S. insurance industry

In its review of derivatives holdings and exposure trends in the U.S. insurance industry, the NAIC's Capital Markets Bureau [4] reports a $2 trillion total notional value of derivatives over the year-end 2014. An overwhelming 94% of the total notional value is used for hedging purposes. Out of the 94%, 49% is related to interest rate hedges, while 25% is aimed at hedging equity risk. Swaps accounted for the largest share (50%) of total notional value, followed by options (44%), futures (3%) and forwards (3%). After several consecutive years of increase, U.S. insurer's derivatives leveled off in 2015 in terms of notional value.

As given in Figure 4, derivatives exposure in BACV as of Dec. 31, 2015, totaled $55 billion, accounting 1% of total cash and invested assets, and representing a decrease of 4% from year-end 2014. In the life industry segment, only 18% of insurance companies had derivatives exposure. However, those involved with derivatives are larger, accounting for $3.28 trillion, or 87% of the segment total. Derivatives positions can be quite large; the average position size was $26.4 million. The largest single position

4. The NAIC's Capital Markets Bureau monitors developments in the capital markets globally and analyzes their potential impact on the investment portfolios of U.S. insurance companies. They published several reports concerning derivatives. These reports provide insight into exposure trends, credit default swaps, hedging, reporting requirements, and market developments. In particular, they review U.S insurer's derivatives holdings and exposure trends. A list of archived Capital Markets Bureau Special Reports is available via :http://www.naic.org/capital_markets_archive_index.htm

open at Dec. 31, 2015, in terms of notional value, was $10 billion corridor option [5] that expires in 2021, which was purchased as an interest rate hedge for a company's fixed income portfolio.

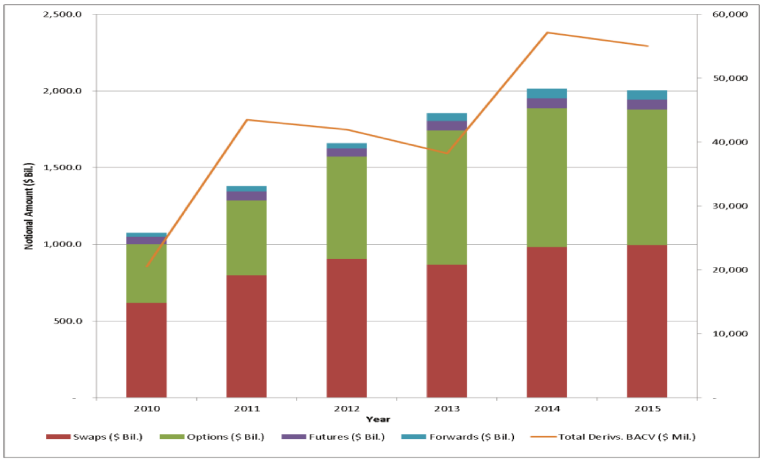


FIGURE 4 – U.S. insurance industry derivatives exposure as of Dec. 31, 2015. *Source : NAIC*

4 Chapter 4 - Optimal execution of options book under market impact

Market impact is defined as the effect that a market participant has when he buys or sells a given volume of an asset. In that respect, it is important that insurance companies, who trade large quantities of options for hedging purposes, take into account this effect when trading options by splitting trades intelligently to minimize its cost.

Inspired by the large literature on equities, and particularly the work of Almgren and Chriss [7], we tackle in this chapter the optimal execution of a book of options under market impact. To do so, we consider a universe with two agents : the option market maker who sells the option, and the end-user who buys the option to hedge an external risk.

Following the steps of usual optimal execution problems with market impact, our approach is three steps; first we need to define the option price under market impact constraints. Second, we choose the risk criterion and derive the optimal execution problem. Finally, depending on the complexity of the formulation, we use an appropriate method to solve the optimization problem.

Question. *How can we define an option price that incorporates the execution quantity?*

Our approach is inspired by the feedback effect and option replication with transaction costs see [78] and [115]. The market maker who dynamically hedges the option to the first order, i.e. delta-hedging, moves the underlying asset price accordingly. Our assumption is based on a linear impact function, i.e. buying x shares drives up the asset price by a cost proportional to x, and selling x shares drives it down by the same proportion cost.

Following Leland's transaction costs framework, see [115], we can derive a price for the option under market impact constraints. For simplicity, we assume that interest rates are null. Furthermore, we assume that the asset price is modeled by a geometric Brownian motion in the absence of market impact.

5. A corridor option is a derivative whose payoff at maturity depends on the amount of time a specified spot rate remains within a specified range during the option's life.

Result 6. *The option execution price under market impact constraints can be expressed through a Black-Scholes like PDE with an enlarged volatility* $\tilde{\sigma}$ *such that :*

$$\tilde{\sigma}^2 = \sigma^2 + f(t, \dot{x}_t, x_t, \sigma),$$

where f *is the market impact function which depends on time* t*, volatility* σ*, inventory* x_t *and trading speed* $\dot{x}_t$*.*

The previous formulation can be generalized to local volatility by following Lepinette's arguments, see [117]. We can also use a simple Taylor approximation and Black-Scholes closed formula to rewrite the option execution price as the sum of the option price without market impact and a linear impact term :

$$\tilde{P}(t, S_t, \dot{x}_t, x_t) = P(t, S_t) + \frac{1}{2}\left\{\tilde{\eta}\dot{x}_t + \tilde{\gamma}(x_t - x_0)\right\}\sigma S_t^2\left(\hat{T} - t\right)^{3/2}\Gamma(t, S_t),$$

where $\tilde{\eta}$ and $\tilde{\gamma}$ are constant parameters associated with the market impact temporary and permanent components [6], $\hat{T}$ is the option maturity and Γ its second derivative w.r.t asset price.

Note that the option gamma is known to explode close to maturity. However, our framework assumes the trades are performed far from the option expiry, i.e. $T \ll \hat{T}$, which removes any undesirable irregularities in the option price or the impact function.

Let us consider a trade execution strategy in which an initial long position of X options, with fixed strike K and maturity $\hat{T}$, is liquidated by a fixed time horizon $[0, T]$, where $T \ll \hat{T}$ is the end time. The asset position x_t is nonincreasing with $x_0 = X < 0$ and $x_{T^+} = 0$ for a pure buy strategy. The optimal strategy is based on the mean-variance criterion $\mathbb{E}[\mathscr{C}(x)] + \lambda\mathbb{V}ar[\mathscr{C}(x)]$. It is reduced to the simple expected cost for $\lambda = 0$, which has an explicit solution under our hypotheses.

Result 7. *Assuming only temporary impact, i.e.* $\tilde{\gamma} = 0$*, and under the Black-Scholes framework, the optimal strategy* x^* *resulting in minimizing the expected cost is given by :*

$$\dot{x}^*(t) = \frac{K_1}{(\hat{T} - t)^{3/2}}$$
$$x^*(t) = \frac{K_1}{(\hat{T} - t)^{1/2}} + K_2$$

where $K_1 = \frac{X}{2\left(\hat{T}^{-\frac{1}{2}} - (\hat{T} - T)^{-\frac{1}{2}}\right)}$ *and* $K_2 = -2K_1(\hat{T} - T)^{-1/2}$.

We recall that the expected cost optimal strategy for the equity case is characterized by having a constant trading rate $x_t^* = -\frac{X}{T}$, as shown in [27] in a discrete setting. In our case, the trading speed is an increasing convex function of time.

For $\lambda > 0$, the mean-variance minimization problem necessitates to establish a proper stochastic dynamic programming framework. We parameterize the trading strategies x by their speed of trading $\alpha_t = -\dot{x}_t$ and introduce $\mathscr{A}(T, X)$ the set of admissible strategies such that the parametrized strategy x^α satisfies necessary conditions.

We restrict our framework to Markovian controls and thus, solving the optimal stochastic control problem at time 0 is brought to a more general case where the agent starts buying at any arbitrary time

6. The temporary impact component reflects the instantaneous effect of the trade, while permanent component transmits the information of buy/sell impact to the market on the long run.

$t \in [0,\mathrm{T}]$ with an initial quantity x without loosing the optimality. We then define the value function $\mathrm{U}(t,\mathrm{S},x)$ for the mean-variance framework as :

$$\mathrm{U}(t,\mathrm{S},x) = \inf_{\alpha\in\mathscr{A}(\mathrm{T},\mathrm{X})} \mathbb{E}_t\Big[\int_t^{\mathrm{T}} \Big\{\alpha_u^2 \mathrm{S}_u^2 \big(\hat{\mathrm{T}}-u\big)^{3/2} \Gamma(u,\mathrm{S}_u) + \lambda\sigma^2 (x_u^\alpha)^2 \mathrm{S}_u^2 \Delta^2(u,\mathrm{S}_u)\Big\} du\Big].$$

Note that the value function $\mathrm{U}(t,\mathrm{S},x)$ satisfies the so-called finite-fuel condition. A state with a non-zero option position with no time left for its liquidation, means that the liquidation task has not been performed, and thus should receive an infinite penalty. We replace this constraint by a finite terminal condition with large penalty and denote by U_ε the value function in this case. By applying the classical framework to derive the Hamilton-Jacobi-Bellman equation, we find :

Result 8. *Let* U_ε^* *be a regular function which solves the PDE :*

$$\begin{cases} \partial_t \mathrm{U}_\varepsilon^* + \frac{1}{2}\sigma^2\mathrm{S}^2\partial_{\mathrm{SS}}\mathrm{U}_\varepsilon^* + \lambda x^2\sigma^2\mathrm{S}^2\Delta^2(t,\mathrm{S}) - \frac{\left(\partial_x \mathrm{U}_\varepsilon^*\right)^2}{4(\hat{\mathrm{T}}-t)^{3/2}\Gamma(t,\mathrm{S})} = 0 \\ \mathrm{U}_\varepsilon^*(\mathrm{T},\mathrm{S}_\mathrm{T},x_\mathrm{T}) = \frac{1}{\varepsilon}\psi(x_\mathrm{T}^\alpha). \end{cases}$$

Then U_ε^* *is the unique solution to the optimal execution problem. Moreover, the optimal execution rate* $\alpha_t^* = -\dot{x}_t^*$ *is given by :*

$$\alpha_t^* = \frac{\partial_x \mathrm{U}_\varepsilon^*(t,\mathrm{S}_t,x_t^*)}{4(\hat{\mathrm{T}}-t)^{3/2}\mathrm{S}_t^2\Gamma(t,\mathrm{S}_t)}.$$

The PDE satisfied by U_ε^* is quadratic in its first derivative w.r.t x, and without knowing the dependence of U_ε^* to x it is difficult to solve it accurately. Fortunately, we are able to reduce the problem's dimension by either reparameterizing the state variable, or using an ansatz to separate the spacial variables. We write $\mathrm{U}_\varepsilon(t,s,x) := x^2 u_\varepsilon(t,s)$ where u_ε is the reduced problem, and use the rate of trading $\kappa_t = -x_t\frac{dx_t}{dt}$ as a control variable. It follows the result :

Result 9. *Let* u_ε^* *be a regular function verifying the following PDE*

$$\begin{cases} \partial_t u_\varepsilon^* + \frac{1}{2}\sigma^2\mathrm{S}^2\partial_{\mathrm{SS}}u_\varepsilon^* + \lambda\sigma^2\mathrm{S}^2\Delta^2(t,\mathrm{S}) - \frac{1}{(\hat{\mathrm{T}}-t)^{3/2}\mathrm{S}^2\Gamma(t,\mathrm{S})}(u_\varepsilon^*)^2 = 0 \\ u_\varepsilon^*(\mathrm{T},s) = \frac{1}{\varepsilon}. \end{cases} \tag{1}$$

Then u_ε^* *is the unique solution to the reduced optimization problem (4.29). The optimal trading rate* κ_t^* *is defined by :*

$$\kappa^*(t,\mathrm{S}) = \frac{u_\varepsilon(t,\mathrm{S})}{(\hat{\mathrm{T}}-t)^{3/2}\mathrm{S}^2\Gamma(t,\mathrm{S})}.$$

In this case u_ε solves a nonlinear PDE with quadratic term of order 0. This is much simpler than the previous PDE. Adding appropriate boundary conditions, the problem can be solved numerically by using finite differences schemes. In Figure 5, we present the trading rate $\kappa_t = -\frac{\dot{x}_t}{x_t}$ surface as a function of time and asset price for $\lambda = 100$. We can see that it depends on the asset level and that it increases as function of time.

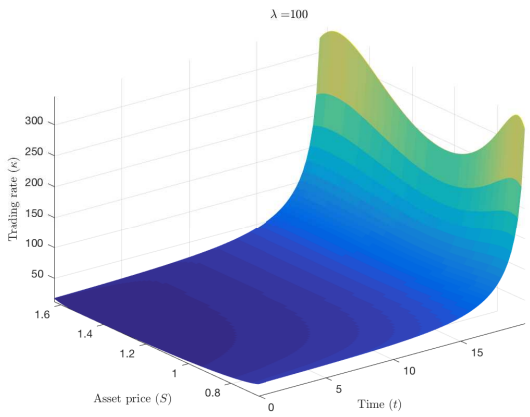


FIGURE 5 – The rate of trading κ as a function of the underlying price S and time t for λ = 100. The strike of the option $K = S_0$ is fixed at time 0.

Part II-B : Volatility models

In the insurance industry, variable annuities and other savings and investment products are subject to market risk like any other financial product. Therefore, it is important to be able to make quick adjustments and build hedging strategies in response to market changes.

Volatility is one of the main drivers of market prices and a key parameter to assess risk, and a large literature tackles its modeling. Since the seminal work of Black and Scholes [30], the most classical way to model the behavior of an asset price S_t is to use continuous semi-martingale dynamic of the form

$$d\log S_t = \mu_t dt + \sigma_t dW_t,$$

where μ_t a drift process and W_t a Brownian motion. The coefficient σ_t is the so-called volatility process.

Following the pioneering approach of [30], practitioners have first considered the case where the process σ_t is constant or deterministic, that is the Black and Scholes model. However, in the late eighties, it became clear that such specification for the volatility is inadequate. In particular, the Black and Scholes model is inconsistent with observed prices for liquid European options. Indeed the implied volatility, that is the volatility parameter that should be plugged into the Black-Scholes formula to retrieve a market option price, depends in practice on the strike and maturity of the considered option, whereas it is constant in the Black-Scholes framework.

Hence more sophisticated models have been introduced. A first possible extension, proposed by Dupire [69] and Derman and Kani [63], is to take σ_t as a deterministic function of time and asset price. Such models, called local volatility models, enable us to perfectly reproduce a given implied volatility surface. However, its dynamic is usually quite unrealistic under local volatility. Another approach is to consider the volatility σ_t itself as an Ito process driven by an additional Brownian motion, typically correlated to W. Doing so one obtains less accurate static fits for the implied volatility surface but more suitable dynamics. Among the most famous of these stochastic volatility models are the Hull and White model [101], the Heston model [97] and the SABR model [95]. More recent market practice is to use the so-called local-stochastic volatility models which both fit the market exactly and generate

reasonable dynamics.

In all the Brownian volatility models mentioned above, the smoothness of the sample path of the volatility is the same as that of a Brownian motion, namely $1/2-\varepsilon$ Hölder continuous, for any $\varepsilon > 0$. However, it is shown in [88] that in practice, spot volatility is much rougher than this. This result in [88] is based on a statistical analysis of historical data using sophisticated high frequency estimation methods. More precisely, it is established in [88] that the dynamic of the log-volatility process is very close to that of a fractional Brownian motion with Hurst parameter smaller than $1/2$. Recall that a fractional Brownian motion W^H with Hurst parameter $H \in (0,1)$ is a centered Gaussian process with stationary increments such that

$$\mathrm{Cov}[W_t^H, W_s^H] = \frac{1}{2}\left(|t|^{2H} + |s|^{2H} - |t-s|^{2H}\right).$$

The Hölder regularity of W^H is $H-\varepsilon$ for any $\varepsilon > 0$ and for $H = 1/2$ we retrieve the classical Brownian motion. Therefore, models where the volatility is driven by a fractional Brownian motion with $H < 1/2$ are called rough volatility models. Beyond fitting almost perfectly historical volatility time series, rough volatility models enable us to reproduce important stylized facts of liquid option prices that local/stochastic volatility models typically fail to generate. In particular, the exploding term structure when maturity goes to zero of the at-the-money skew (the derivative of the implied volatility with respect to strike) is readily obtained, see [20, 79]. Other developments about rough volatility models can be found in [22, 23, 70, 71, 74, 80, 94, 105, 134].

In this part of the thesis we revisit the finding in [88] using range-based and option-based data. Indeed in [88], the authors work with realized volatility based on high frequency data to estimate spot volatility. Access to high frequency data is sometimes costly and/or unavailable for certain assets. Therefore, other proxies are used to estimate daily volatility. Here we use two spot volatility proxies.

In Chapter 5, we use range-based estimators are based on open, high, low and close daily prices. We particularly focus on Garman-Klass and Parkinson estimators, see [86, 139]. In Chapter 6 we use a spot volatility proxy which is not based on historical price data, but on implied volatility. More precisely, we approximate the spot volatility by the implied volatility of an at-the-money liquid option with short maturity (or a refined version of it). This idea can be justified by the fact that in most models, the at-the-money implied volatility tends to the spot volatility as maturity goes to zero, see for example [132].

Our main results are a confirmation of that in [88] : when using alternate spot volatility measurement methods, we can still conclude that volatility is rough.

5 Chapter 5 - Range-based proxies and rough volatility

Question. *What is the regularity of the volatility based on range-based data ?*

We look at the measure $m(q,\Delta)$ defined by :

$$m(q,\Delta) = \mathbb{E}[|\log(\sigma_{t+\Delta}) - \log(\sigma_t)|^q]$$

By plotting $\log(m(q,\Delta))$ against $\log(\Delta)$ for different values of q, from 1 day to a few months, we try to find a form for the scaling function ζ such that :

$$m(q,\Delta) = K_q \Delta^{\zeta_q}.$$

Result 10. *The increments of the log-volatility behave as that of a fractional Brownian motion with small Hurst exponent* H *for a very large range of time scales :*

$$\zeta_q \approx \mathrm{H} q,$$

where $\mathrm{H} < 0.1$ *for all assets*[7] *we performed the study on.*

We also find that log-volatility increments over a lag Δ have a similar distribution than a centered normal distribution with variance Δ^{H}. As a result, log-volatility increments can be modeled by the increments of a fBm :

$$\sigma_t = \sigma_0 e^{\nu \mathrm{W}_t^{\mathrm{H}}}.$$

This model is unfortunately not stationary. Stationarity being a desirable property in time series, volatility can be modeled through a fractional Ornstein-Uhlenbeck volatility model with Hurst exponent $\mathrm{H} < 1/2$, and a mean-reverting parameter α such that $1/\alpha$ (the mean-reversion time scale) is very large compared to the time scales of interest. This so-called fractional rough volatility model (RFSV), see [88], is fundamentally different than Compte and Renault fractional volatility model (FSV), see [56]. Indeed, FSV considers $\mathrm{H} > 1/2$ and $\alpha \gg 1$.

More precisely, we have :

$$d\log(\sigma_t) = -\alpha(\log(\sigma_t) - m) + \nu d\mathrm{W}_t^{\mathrm{H}},$$

with $\mathrm{H} < 1/2$ and $\alpha \ll 1$.

Question. *Can we validate the RFSV model? Doesn't the FSV model exhibit the same behavior?*

Result 11. *Empirical results and simulations allow to exclude the FSV model as a data consistent model. The RFSV model, on the other hand did not fail the tests we performed.*

To perform our tests, we simulate the FSV and FRSV models with known parameters. We then estimate realized and range-based volatility proxies. By looking at the behavior of $\log(m(q,\Delta))$ versus $\log\Delta$, we find that FSV model can be misleading. In fact, as illustrated by the sketch in Figure 6, $\log(m(q,\Delta))$ exhibits three types of behaviors :

— For small Δ, the estimated smoothness parameter H is affected by noise and is thus close to 0.

— For intermediate Δ, one can approach the true value of H but it can be relatively difficult to find it with good accuracy.

— For large Δ, the stationarity (through the mean-reversion) governs the value of the estimated H, which is again very small compared to the real one.

For RFSV model, the estimated smoothness for the RFSV is consistent whatever the lag Δ. Hence, the FSV model does not seem to be consistent with observed data, meanwhile RFSV simulated volatility is much closer to observations. We conclude the analysis by studying the prediction power of RFSV. Using result by Nuzman and Poor, see [135], we can use the conditional expectation of the fractional Brownian. As a result, the conditional expectation of log-variance and variance w.r.t past information $\mathscr{F}_t$ satisfies

$$\mathbb{E}[\log(\sigma^2_{t+\Delta}) \mid \mathscr{F}_t] = \frac{\cos(\mathrm{H}\pi)}{\pi} \Delta^{\mathrm{H}+1/2} \int_{-\infty}^{t} \frac{\log(\sigma_s^2)}{(t-s+\Delta)(t-s)^{\mathrm{H}+1/2}},$$

7. We estimated the scaling of the volatility for liquid and less liquid assets based on the Garman-Klass and Parkinson proxies

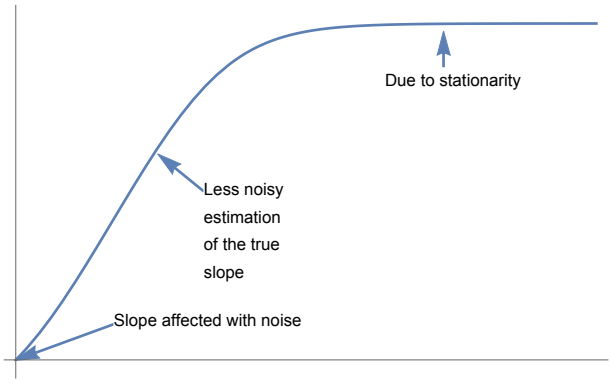


FIGURE 6 – Observed behavior for $\log(m(q,\Delta))$ as a function of $\log(\Delta)$ using the realized-volatility proxy

and

$$\mathbb{E}[\sigma^2_{t+\Delta} \mid \mathscr{F}_t] = \exp\left(\mathbb{E}[\log(\sigma^2_{t+\Delta}) \mid \mathscr{F}_t] + 2c\nu^2\Delta^{2H}\right),$$

where $c = \frac{Gamma(3/2-H)}{\Gamma(H+1/2)\Gamma(2-2H)}$.

Result 12. *RFSV outperforms other times series models such as AR, HAR and GARCH in predicting future log-volatility and variance.*

6 Chapter 6 - Volatility is rough : evidence from option price data

Instead of focusing on volatility proxies recovered from the underlying historical price, we wish to use option price based data. Our goal is again to study the instantaneous volatility smoothness. To do so, we exploit asymptotic properties of the implied volatility, and particularly, the fact that at-the-money implied volatility converges to spot volatility as time-to-maturity goes to zero.

Result 13. *Using the same procedure as in Garman-Klass and realized volatility, we find that the scaling of at-the-money short-term implied volatility is fractional, i.e. :*

$$\zeta_q = qH,$$

with $H \backsimeq 0.32$*, meaning volatility is rough.*

Question. *Why is the value large compared to realized and Garman-Klass volatilities? Can we improve the result?*

One reason for this relatively high value, is that our options have a significant remaining time to maturity of one month. This induces a smoothing phenomenon in the estimation of the Hurst parameter. This effect is of the same nature as that described and explained in [88], caused by the discrepancy between spot and integrated volatility over a short time interval.

To improve the result, one solution would be to use asymptotic methods to estimate implied volatility as time-to-maturity tends to zero, i.e. $\tau \to 0$, and raw option data for the estimation. We choose the methodology by Medveded and Scaillet who give an asymptotic formula for implied volatility, see Chapter 6 and [124].

Result 14. *The Medvedev-Scaillet estimator of spot volatility is rough, with* $H \backsimeq 0.3$.

The value of the smoothness parameter is still high compared to H found for realized-volatility and Garman-Klass, even though we used an asymptotic approach to approximate close-to-maturity implied volatility. We believe this is still due to the remaining time-to-maturity as the option data we used for the Medvedev-Scaillet proxy is still of at least 15 days time-to-maturity.

To understand this phenomenon, we perform a simulation study where spot volatility is rough (H = 0.04). Under mild assumptions, we find that 1 day ATM-IV has H = 0.06, while 20 days ATM-IV has H around 0.27.

Result 15. *Implied volatility regularity increases as time-to-maturity increases as given by Figure 7.*

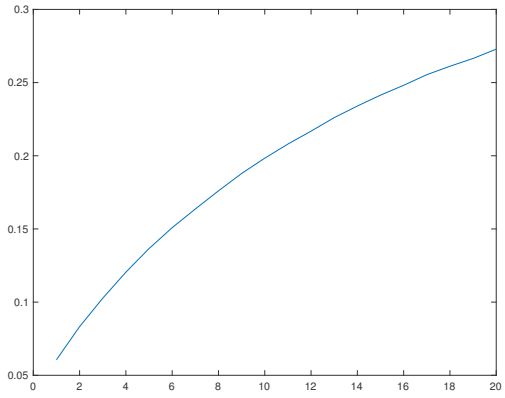


FIGURE 7 – Estimated values of the Hurst parameter using implied volatilities as a function of time to maturity.

To provide a more quantitative understanding of such upward bias, we consider the crude approximation of the Black-Scholes implied volatility :

$$\hat{v}^{\tau}(t) := (\sigma^{\tau}_{imp}(t))^2 = \frac{1}{\tau}\int_t^{t+\tau} \mathbb{E}[\sigma_t^2 \mid \mathscr{F}_t]\, du,$$

using the simplified rough volatility model :

$$\sigma_u^2 = \sigma_0^2 + \nu W_u^H.$$

Result 16. *The scaling based on the moment of order 2 is given by equation :*

$$\mathbb{E}[(\hat{v}^{\tau}(\Delta) - \hat{v}^{\tau}(0))^2] \propto f(\tau/\Delta)\Delta^{2H},$$

where $f(\tau/\Delta) \underset{\tau/\Delta \to 0}{\longrightarrow} 1.$

This means that the same scaling relationship, as that associated to the spot volatility, is approximately satisfied when considering implied volatilities with small enough times to maturity. Otherwise, one should add a multiplicative factor.

Première partie

Life insurance products

Chapitre 1

Financial risk management and the rational lapse strategy in life Insurance policies

***Abstract*— Over the past decade, Variable Annuities have experienced tremendous growth accounting for half of the life insurance industry, as unit-linked products offering both participation in equity market and guarantees at key life moments (retirement, death).

The recent Quantitative Impact Study (QIS 5) of the Solvency II framework showed that lapse risk is the most important risk among life underwriting risks for Variable Annuities, as illustrated by solvency issues experienced by the policyholder run in the late 1980's. Thus research on lapse rates is crucial to a proper calibration of regulatory standard models and internal risk models.
Usually the lapse behavior has been modeled by historical or backward looking statistical regressions which have empirically underestimated the risk due to the scarcity of extreme scenario samples and the inability to dynamically extrapolate the observed behavior to various market conditions. In contrast, a "rational" lapse strategy valuation is a prudent forward looking approach where policyholders lapse in a way that maximizes the net present value of the future cash-flows, depending on key drivers. Empirically consistent with herd behavior as experienced in the last financial crisis, this approach is illustrated on a GMAB VA product using two alternatives numerical schemes (PDE and Monte Carlo).

However, as policyholders cannot be expected to lapse all at the same time, this rational lapse framework is slightly amended by introducing a proportion of lapses among the contract still active, which translates into the notion of "reasonable" lapse more consistent with empirics.

Keywords : GMAB; Variable Annuity; rational lapse strategy; stochastic interest; PDE; ADI; high-dimensional regression.

1.1 Introduction

The VA product, a popular retirement savings vehicle in the US, is starting to emerge as a viable option in other markets, including Europe and Asia. The GMAB riders written on VAs (also known as Maturity Guarantees, see [40]) provide policyholders a guaranteed amount at a fixed expiration date, so this kind of products have some similar properties as long-term vanilla puts. One important attractiveness of GMAB products is that this guarantee gives policyholders the ability to protect their

retirement investments against downside market risk by allowing the policyholder to receive the greater of the account value and the benefit base at the maturity. The benefit base can either step up to the high-water mark of the account value at the end of each policy year (annual ratchet), or can roll up with a fixed percentage (the roll-up rate, e.g. 2%), regardless of the market conditions. Thanks to these new product characteristics, the guarantee not only protects policyholders against investment losses, but also allows customers taking advantage of upside gain from the market. In exchange for this benefit, the policyholder pays a charge fee each year.

The recent Quantitative Impact Study (QIS 5) of the Solvency II framework showed that lapse risk is the most important risk among life underwriting risks for Variable Annuities, as illustrated by solvency issues experienced by the policyholder run in the late 1980's. Thus research on lapse rates is crucial to a proper calibration of regulatory standard models and internal risk models.

The dynamic behavior is essentially a selection process of the policyholders against the VA writer, where an increase leaves fewer insured to ultimately make a claim on the guarantees but reduces the fees the insurer can collect. The general pattern is that more policies will lapse when the capital market is up, and fewer policies will lapse when the capital market is down.

— In an up market, the value of the minimum guarantee diminishes as the account value is likely to exceed the minimum guarantee values. As such, surrendering the policy does not create much loss to the policyholder.

— On the other hand, a down market can result in the surrender value being less than the guarantee value causing the policy to be in-the-money. If the policyholder surrounds at this time then he or she can only get the reduced surrender value, forfeiting the added value from the guarantee rider. The result is that there is strong incentive for the policyholder to keep the in-the-money VA contract in force.

As the lapse assumption may impact significantly the profitability of GMAB riders, a rigorous modeling framework of the lapse rate is necessary for both pricing and hedging purpose. During the last decade, the literature on pricing and risk management of these guarantees has been evolving.

— Tradionnally the lapse behavior has been modeled by historical or backward looking statistical regressions which have empirically underestimated the risk due to the scarcity of extreme scenario samples for these new products and the inability to dynamically extrapolate the observed behavior to various market conditions.

— In contrast, a "rational" lapse strategy valuation is a prudent forward looking approach where policyholders lapse in a way that maximizes the net present value of the future cash-flows, depending on key drivers. This reflects a potential extreme policyholder behavior, as experienced in the last market crash, with an initial immediate and sustained fall in lapses right after the crash, before an abrupt recovery consistent with the interest rates. In contrast, dynamic lapses modeling are usually unable to provide such empirical dynamics.

This approach is illustrated on a GMAB VA product using two alternatives numerical schemes (PDE and Monte Carlo), as the valuation of a Bermudan-style contingent claim for the insurer, where the contingency is closely related to equity market conditions and the interest rate level, see [35].

The price evaluated by this approach can be interpreted as the fair value of the policy if all policyholders use the same rational lapse strategy, which is similar with the optimal early-exercise strategy of Bermudan options. However, as policyholders cannot be expected to lapse all at the same time,

this rational lapse framework is slightly amended by introducing a proportion of lapses among the contract still active, which translates into the notion of "reasonable" lapse more consistent with empirics. Note that this is only an interpretation, and that the critical aim is to make sure the lapse risk can be hedged no matter which strategy the holders use.

The remainder of this paper is organized as follows. Firstly, in Section 2 the GMAB policy is explained in full details. Section 3 introduces the modeling framework to evaluate the liability of GMAB policies in no-lapse assumption. The rational lapse strategy and critical lapse boundaries are studied in Section 4. Section 5 addresses the pooling of lapse risks. In Section 6 we implement two numerical methods, the PDE approach and the high-dimensional regression method (Monte Carlo, see [37]) to calculate the no-arbitrage price of GMABs in the Hull-White interest rate model. Numerical results of these two methods are shown in Section 6 and conclusions in Section 7.

1.2 Description of the contract

In practice, most GMAB policies are purchased in a lump sum. We assume that a single premium is paid at inception of the contract and denoted by A(0) = 100$ the initial account value at time zero after the upfront fees have been paid. The account value is invested in a portfolio consisting mainly of equities and bonds. At the end of each policy year t_i, the insurer deduct a charge fee $\bar{\alpha}A(t_i)$ on the account value, where $\bar{\alpha} = 2\%$ is the annual charge rate. The life time of the policy is T = 10 years if there is neither early termination nor rollover.

For a contract that is held until the maturity, there is a guaranteed minimum return paid to the policyholder. We represent this guarantee to the policyholder as G, which is called the benefit base for insurers. In other words, at the maturity, the policyholder has the right to receive a cash payment equal to either A(T) or to the benefit base G. Consequently, at maturity, the value of the policy is $\max(A(T), G)$. This payoff can be decomposed to the sum of the account value A(T) and a vanilla put $(G - A(T))^+$ (the strike level is G). The benefit base G is fixed at inception, which is equal to $A(0)(1+\bar{r})^T$, where $\bar{r}$ is the roll-up rate. In most cases, $\bar{r}$ is approximately equal to the yield of zero-coupon bonds maturing at T. We assume that one GMAB policy is purchased in 2000 and hold until 2010, and the account value

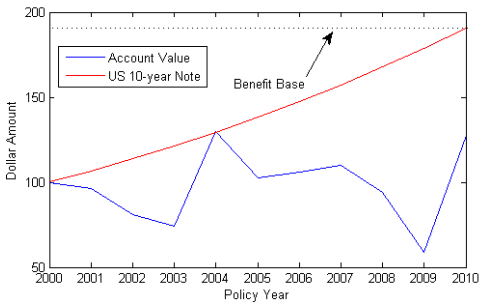


FIGURE 1.1 – The illustration of the account value over time compared with the US 10-year note.

is invested in S&P 500 at inception. The roll-up rate is set at 6.67%, which is equal to the yield of US 10-year notes in January 2000. All other parameters are the same as those mentioned above. Figure 1.1 plots over time the account value $A(t)$. By comparing the net return of 10-Year bonds with that of the roll-up GMAB rider, we can see that the roll-up benefit base can not only protect policyholders from catastrophes in stock market, but also from risks of the persistent decrease of the interest rate [1].

1. In fact, the US 10-year bond yield was 6.67% in January 2000, while it was 3.61% ten years later.

1.3 Valuation of a GMAB with zero lapse

Firstly we establish the general modeling framework to evaluate the liability of GMAB with zero lapse (European GMAB). From now on, we let $(\Omega, \mathscr{F}, \mathbb{F} = (\mathscr{F}_t)_{t\leqslant \mathrm{T}}, \mathbb{Q})$ denote a complete filtered probability space supporting two independent standard one dimensional Brownian motions W and $\mathrm{W}^{\perp}$. Here $\mathrm{T} > 0$ is a fixed time horizon. We assume that the filtration $\mathbb{F}$ is the completion of the rough filtration generated by $(\mathrm{W}, \mathrm{W}^{\perp})$, so that any martingale $(\mathbb{Q}, \mathbb{F})$-martingale can be represented as a stochastic integral with respect to $(\mathrm{W}, \mathrm{W}^{\perp})$.

During the last decade the literature on pricing variable annuities has evolved, but many evaluation approaches proposed (e.g. [18, 129]) are still based on the assumption of deterministic interest rates. Such an assumption is harmless in most situations since the interest-rates variability is usually negligible when compared to the variability observed in equity markets. While pricing a long-maturity securities such as VA guarantees, however, the volatile feature of interest rates can have stronger impacts on the liability of GMAB. In such case it is therefore advisable to use stochastic interest rate models.

In this paper, we assume that the short term interest rate $r = (r(t))_{t\geqslant 0}$ is driven by the one factor Hull and White model, and the underlying asset $\mathrm{S} = (\mathrm{S}(t))_{t\geqslant 0}$ in which the account value is invested follows a Black and Scholes type dynamics, namely :

$$\begin{cases} d\mathrm{S}(t) = r(t)\mathrm{S}(t)dt + \sigma \mathrm{S}(t)d\mathrm{W}(t) \\ dr(t) = a(\theta(t) - r(t))dt + \sigma_r d\mathrm{Z}(t)\ ,\ \mathrm{Z} := (1-\rho^2)^{\frac{1}{2}}\mathrm{W}^{\perp} + \rho \mathrm{W} \end{cases} \tag{1.1}$$

Here, a and σ_r are positive constants, θ is a deterministic Lebesgue-integrable function, σ is the instantaneous volatility of the asset return, and ρ is the correlation [2] between the account value and the interest rate. Note that the above financial market is complete whenever S and a zero-coupon bond with maturity T can be freely traded, and that $\mathbb{Q}$ is the only martingale (risk neutral) measure.

For the account value, a charge fee is deducted at a rate α continuously, where $\alpha = -\log(1-\overline{\alpha})$. This means that $\mathrm{A}(t)$ evolves according to

$$d\mathrm{A}(t) = (r(t) - \alpha)\mathrm{A}(t)dt + \sigma \mathrm{A}(t)d\mathrm{W}(t) \tag{1.2}$$

Since (r, A) is a Markov process, the European-style liability V^{E} of a single GMAB rider can be identified to a deterministic liability function v^{E} by :

$$\mathrm{V}^{\mathrm{E}}(t) : v^{\mathrm{E}}(t, r(t), \mathrm{A}(t)) = \mathbf{E}^{\mathbb{Q}}[\mathrm{D}_t^{\mathrm{T}} \max(\mathrm{A}(\mathrm{T}), \mathrm{G})|\mathscr{F}_t] = \mathbf{E}^{\mathbb{Q}}[\mathrm{D}_t^{\mathrm{T}}(\mathrm{A}(\mathrm{T}) + (\mathrm{G} - \mathrm{A}(\mathrm{T}))^+)|\mathscr{F}_t] \tag{1.3}$$

where $\mathrm{D}_{t_1}^{t_2}$ represents the stochastic discount factor between t_1 and t_2

$$\mathrm{D}_{t_1}^{t_2} := \exp(-\int_{t_1}^{t_2} r(s)ds)\ .$$

Equation (1.3) shows that the European-style liability of GMAB riders can be considered as the sum of a forward contract of the account value ending at T and a vanilla put with the maturity T and the strike level G. In the Hull-White interest rate model, this liability value can be easily calculated analytically.

However, it does not always exist some closed formula of the liability value, especially when the early-lapse premiums are taken into account. Thus in practice, we need to use some numerical methods,

2. For VAs, the correlation is often negative, as most portfolio contains fixed income assets, such as bonds.

such as PDE or Monte-Carlo based algorithms to evaluate the policies. For GMAB riders, it is sometimes more convenient to price the liability under the so-called forward measure rather than the risk-neutral measure $\mathbb{Q}$. Because in $\mathbb{Q}^T$, we can reduce the number of dimensions of the liability evaluation problem (1.3) from three to two and the pricing process can be significantly accelerated.

To facilitate the following study, we evaluate the GMAB riders in the forward measure $\mathbb{Q}^T$. Firstly, we introduce the forward value of $A(t)$ at T observed at date t, denote by $F^T(t) = A(t)/Z_t^T$, where Z_t^T is the price of a zero-coup bond maturing at T. Applying Ito's lemma to $F^T(t)$, we get the dynamics of the forward account value :

$$\frac{dF^T(t)}{F^T(t)} = (\nu^2 B_r^2(u) + \rho\nu\sigma B_r(u) - \alpha)dt + \sigma dW(t) + \nu B_r(u)dZ(t),$$

where $u = T - t$ is the time to maturity and the function $B_r(u) = (1 - e^{-au})/a$. By doing the following transformations of the Brownian motions from $\mathbb{Q}$ to $\mathbb{Q}^T$:

$$\begin{aligned} dZ(t) &\rightarrow dZ^T(t) - \nu B_r(u)dt \\ dW(t) &\rightarrow dW^T(t) - \rho\nu B_r(u)dt \end{aligned} \tag{1.4}$$

we have that, under $\mathbb{Q}^T$, the dynamics of $F^T(t)$ can be written as :

$$\begin{aligned} dF^T(t) &= -\alpha F^T(t)dt + \sigma F^T(t)dW^T(t) + \nu B_r(u)F^T(t)dZ^T(t) \\ &= -\alpha F^T(t)dt + \omega_u F^T(t)d\tilde{W}^T(t) \end{aligned} \tag{1.5}$$

where $\omega_u^2 = \sigma^2 + \nu^2 Br^2(u) + 2\rho\sigma\nu Br(u)$ and $\tilde{W}^T$ is a Brownian motion in $\mathbb{Q}^T$. The results above allow us to simplify the pricing problem of GMAB riders. Instead of computing the liability under risk-neutral measure $\mathbb{Q}$ (as in (1.3)), we evaluate the forward liability $\tilde{v}^E(t,f)$ in $\mathbb{Q}^T$:

$$\tilde{v}^E(t, F^T(t)) := \frac{v^E(t, r(t), A(t))}{Z_t^T} = \mathbf{E}^{\mathbb{Q}^T}[F^T(T) + (G - F^T(T))^+ | \mathscr{F}_t] \tag{1.6}$$

Equation (1.5) and (1.6) show that the European-style forward liability $\tilde{v}^E$ can be evaluated by the following analytical formula :

$$\tilde{v}^E(t, F^T(t)) = e^{-\alpha u}F^T(t) + GN(-d_2) - e^{-\alpha u}F^T(t)N(-d_1), \quad d_{1,2} = \frac{\log(F^T(t)/G) - \alpha u}{\Gamma} \pm \frac{\Gamma}{2} \tag{1.7}$$

where $\Gamma = \sqrt{\int_0^{T-t} \omega_s^2 ds}$ and $N(\cdot)$ is the cumulative distribution function of the standard normal distribution.

1.4 Valuation a GMAB with rational lapse assumption

In the previous section, we have formulated the pricing issue of GMAB riders under the no-lapse assumption. If policyholders are not allowed to lapse contracts before maturity, the liability of GMAB riders can be calculated analytically by (1.7). However, in practice we can not assume the lapse rate to be zero or some other constant, as we observe that the lapse rate does change significantly in different market conditions (equity market and interest rate level) and this fluctuation of lapse rate has notable impacts on the liability value and insurer's hedging strategy.

As explained in the introduction, we consider the pricing problem of liabilities with lapse options as the valuation of a Bermudan-style contingent claim, see [42]. Because the rational lapse strategy

discussed here are similar with the optimal early-exercise strategy of classic Bermudan options. We assume that the policyholder can lapse the contract at the end of each policy year, noted as t_i, $i = 1,2,...,N$, where $t_N = T$. According to the definition, we have

$$\begin{array}{lcl} v^B(T, r(T), A(T)) & = & \max(A(T), G) \\ v^B(t_i-, r(t_i-), A(t_i-)) & = & \text{ess sup}_{\tau_i \in \mathscr{T}_i} \mathbf{E}^{\mathbb{Q}}[D_i^{\tau_i} A(\tau_i) + \mathbf{1}_{\{\tau_i = T\}} D_i^T \max(A(T), G) | \mathscr{F}_{t_i}] \end{array} \tag{1.8}$$

where $\mathscr{T}_i$ is the set of all stopping times taking values in $\{t_i, t_{i+1}, \cdots, T\}$ and τ_i denotes the stopping time of the rational lapse strategy since time t_i. According to the assumption, the policyholder is not authorized to lapse the contract between two purchase anniversaries t_i and t_{i+1}, so the process V^B of the liability should evolve in the same way as V^E for $t_i \leqslant t < t_{i+1}$. Applying the fact that $r(t)$ and $A(t)$ are all Markov processes, we have,

$$\forall i < N, \quad v^B(t_i, r(t_i), A(t_i)) = \mathbf{E}^{\mathbb{Q}}[D_i^{i+1} v^B(t_{i+1}-, r(t_{i+1}-), A(t_{i+1}-)) | \mathscr{F}_{t_i}] \tag{1.9}$$

Similarly with Bermudan-style options, at discrete time points t_i, the policyholder is supposed to compare the account value with the value of the liability to decide whether lapse or not. If the account value is bigger than the liability, policyholders surrender the contract and get back $A(t_i)$. Otherwise, they continue to hold the policy. That is to say, at time t_i, the Bermudan-style liability should evolve as following

$$v^B(t_i-, r(t_i-), A(t_i-)) = \max(A(t_i), \ v^B(t_i, r(t_i), A(t_i))) \tag{1.10}$$

Equation (1.10) reflects the fact that the liability before the annuity payment is equal to the greater of the current account value $A(t_i^-)$ and the value of continuation. To simplify the pricing process of the Bermudan-style liability, we can calculate the expectations under the forward measure $\mathbb{Q}^T$ instead of the risk neutral measure $\mathbb{Q}$. That is to say, we write (1.10) as

$$\begin{array}{rcl} \tilde{v}^B(t_i-, F^T(t_i-)) & := & \dfrac{v^B(t_i-, r(t_i-), A(t_i-))}{Z_i^T} = \max(F^T(t_i), \ \tilde{v}^B(t_i, F^T(t_i))) \\ & = & \max(F^T(t_i), \ \mathbf{E}^{\mathbb{Q}^T}[\tilde{v}^B(t_{i+1}-, F^T(t_{i+1}-)) | \mathscr{F}_{t_i}]) \end{array} \tag{1.11}$$

where the boundary condition at maturity is

$$\tilde{v}^B(T, F^T(T)) = F^T(T) + (G - F^T(T))^+ \tag{1.12}$$

Another important issue related with the evaluation problem of the Bermudan-style liability is the determination of the rational lapse strategy to be followed. As the benefit base G is fixed at inception, according to (1.11) and (1.12), we have that $\tilde{v}^B(t, f)$ is a convex, nondecreasing function of f. In addition, it is also a positive function on $(t, f) \in [0, T) \times [0, \infty)$, for $\tilde{v}^B(t, f) > \tilde{v}^E(t, f) > 0$. Finally, for the charge fee $\alpha > 0$, (1.11) and (1.12) also imply that $\lim_{f \to \infty}(f - \tilde{v}^B(t_i, f)) > 0$. It follows from the previous arguments that, for each $t \in \{t_0, \cdots, t_i, \cdots, t_{n+1}\}$, there exists a real number $f^*(t_i-)$,

$$\begin{array}{lcll} 0 \leqslant f < f^*(t_i-) & \Rightarrow & \tilde{v}^B(t_i-, f) > f & \text{(Not Lapse)} \\ f \geqslant f^*(t_i-) & \Rightarrow & \tilde{v}^B(t_i-, f) = f & \text{(Lapse)} \end{array} \tag{1.13}$$

In this paper, $f^*(t_i-)$ is referred to as the "critical forward account value" since the policy should be lapsed as soon as the forward account value increases to this level at time t_i. As it is shown by (1.13), thanks to the change of measure, the critical boundary here depends only on the forward account value, rather than on both the interest rate and the account value level. However, it is not always possible to do this kind of simplifications when we evaluate Bermudan-style options, because sometimes the intrinsic payoff (such as f for $\tilde{v}^B(t, f)$) is not a linear function of the underlying (e.g. American vanilla options).

The objective now is to evaluate the liability $\tilde{v}^{\mathrm{B}}$ and the critical boundary f^* of a single GMAB rider. Although many analytical approximations exist in academy literatures, see [129], most of them are not sufficiently precise due to the long maturity property. In the present paper, we propose two numerical methods : PDE and Monte Carlo schemes (see Appendix C for the description of the numerical schemes and numerical tests for the results), to calculate both the Bermudan-style liability v^{B} and the critical lapse surface. As we have mentioned at the beginning, the PDE method is precise for low-dimensional problems (< 3), while the Monte Carlo is more efficient when there are more than three dimensions in the pricing problem (e.g. multi-asset account value or stochastic volatility models).

1.5 Life insurance policy pool

The analysis above is focused on the rational lapse strategy and the no-arbitrage value of a single GMAB policy. The liability v^{B}, which takes into account the lapse risks, allows the insurer to hedge the uncertain customer behavior no matter what the lapse strategy of the policyholder is. But in practice, the insurers often need to estimate the lapse risks of a pool of life insurance policies, and in this case, the lapse strategy can be represented by the frequency $p(t_i)$ of the policies that are early terminated at time t_i, see [3]. Actually, the common sense and experience tells us that not all policyholders will lapse the contract at the same time, so we need to slightly change the function $p(t_i)$ to estimate the real lapse rate of a policy pool.

For a pool of GMAB policies, we denote by $p(t_i)$ the proportion of lapses at date t_i among the contracts still active in the pool. According to the rational lapse strategy, we can express $p(t_i)$ as a deterministic function h of the forward account value $\mathrm{F}^{\mathrm{T}}(t_i)$, that is $h(\mathrm{F}^{\mathrm{T}}(t_i)) = \mathbf{1}_{\{\mathrm{F}^{\mathrm{T}}(t_i) \geqslant f^*(t_i-)\}}$. This lapse function implies that once $\mathrm{F}^{\mathrm{T}}(t_i)$ touches $f^*(t_i-)$, all policyholders lapse the contract and otherwise everybody hold the policy.

Inspired by the mortgage prepayment models, see [137], and evaluation approaches of surrender options for other life insurance products, see [3], we assume that h is a nondecreasing piece-wise linear function of the variable $\mathrm{F}^{\mathrm{T}}(t_i)$. When $\mathrm{F}^{\mathrm{T}}(t_i) < f^*(t_i-)$, the lapse rate is not zero due to policyholders' personal circumstances (including liquidity and death), which is independent of financial considerations. These "irrational" lapses are analogous to noneconomic prepayment on low-rate mortgages. While when $\mathrm{F}^{\mathrm{T}}(t_i) \geqslant f^*(t_i-)$, some rational lapses never occur, and a reasonable specification of $p(t_i)$ may be illustrated as that in Figure 1.2 :

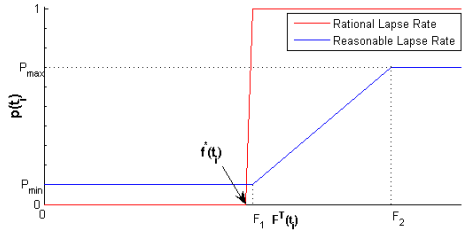


FIGURE 1.2 – Comparing the rational lapse function with the reasonable lapse function.

The four parameters F_1, F_2, $\mathrm{P}_{\min}$ and $\mathrm{P}_{\max}$ are determined by insurers according to some empirical tests. To be consistent with the rational lapse assumption, F_1 should be very close to $f^*(t_i)$ and $\mathrm{P}_{\max}$ should be set high enough (normally $\geqslant 50\%$). Under the reasonable lapse assumption, once the critical lapse level and the reasonable lapse function are determined, the GMAB liability can be simply

evaluated as an European-style option, see [3]. However, it is worthy to mention that, unlike the rational lapse approach, the reasonable lapse assumption makes the insurer partially exposed to the risk of lapses in the future.

1.6 Numerical tests

In this section, we use two numerical methods (PDE and Monte Carlo) introduced above to evaluate the Bermudan-style liability of one standard GMAB policy. Our final results show not only the consistency between these two methods, but also the efficiency and precision of both methods. In addition, the option value of GMAB (defined later), the forward delta and the "critical boundary" found by these two methods are also compared.

For the tested policy, The account value is supposed to evolve according to Hull-White model, where the principle model inputs are listed in Table 1.1. The initial equilibrium short-rate curve $\theta(t)$ is supposed to be flat (θ constant) and the short rate at inception is denoted by r_0. All other parameters of product properties will be clarified later.

TABLEAU 1.1 – Hull-White Model Inputs

σ	r_0	θ	a	σ_r	ρ
0.2	0.02	0.02	0.03	0.01	0

For simplicity, we also assume that the policyholder is alive at the maturity of GMAB policies. Although we are focused on the liability of a single policy in the following numerical tests, the methodology we propose here can be easily extended to evaluate a GMAB policy pool by adding up policies of different maturities with a proper weight indicated by mortality rate assumptions.

1.6.1 Bermudan-style GMAB Liability

Firstly, we calculate the liability of a standard GMAB policy. The principle product parameters are listed in Table 1.2, where the charge fees rate is $\alpha = 2\%$, the maturity is 10 years and the benefit base, fixed at inception, is 100\$ for one policy. For simplicity, we assume that the policyholders are allowed to lapse the contract only at one specific date of each policy year.

Figure 1.3 shows the forward Bermudan GMAB liability $\tilde{v}^{\mathrm{B}}(t,f)$ computed by the PDE scheme for different forward account values through time. In addition, the intrinsic value of GMAB policies, which is equal to the instantaneous forward account value, is also recorded in Figure 1.3. It is worthy to mention that at the dates when lapses are allowed, once we have $\tilde{v}^{\mathrm{B}}(t,f) = f$ for f big enough, policyholders should lapse the contract immediately. This phenomena is consistent with our intuitive, as the higher the account value is, the less the GMAB guarantees worth and the more likely that policyholders lapse the contract and get back the intrinsic value immediately.

TABLEAU 1.2 – Product Parameters of the GMAB Policy

α	G	T	Lapse Date Frequency
2%	100\$	10 years	1/year

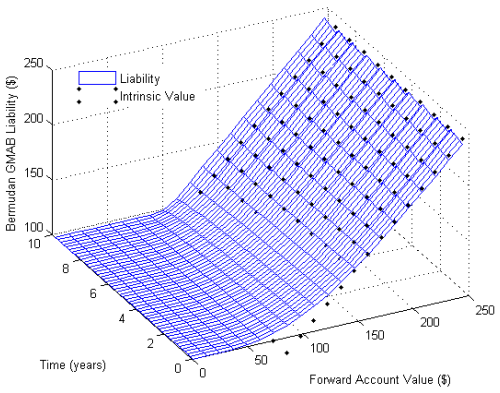


FIGURE 1.3 – The forward Bermudan liability ($\tilde{v}^{\mathrm{B}}$) of a standard GMAB policy for different forward account values and time points, compared with the intrinsic value at all exercisable dates.

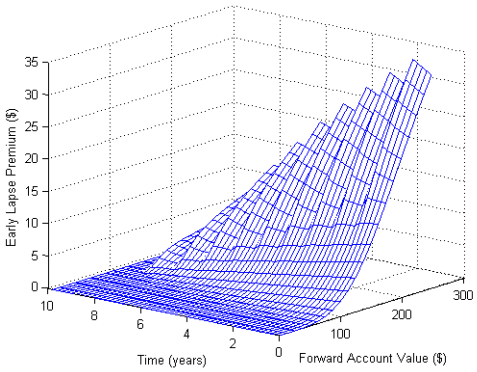


FIGURE 1.4 – The forward early lapse premium ($\tilde{v}^{\mathrm{B}} - \tilde{v}^{\mathrm{E}}$) of a standard GMAB policy for different forward account values and time points.

To be further protected from potential lapse waves or other financial risks, the insurers can charge the policyholders an up-front fee, which is equal to $Z_0^{\mathrm{T}}(\tilde{v}^{\mathrm{B}}(0,f_0)-f_0)$ (the difference between the liability and the asset), to make sure the balance sheet is in equilibrium at inception, see [35].

Figure 1.4 shows the early lapse premium (the difference between $\tilde{v}^{\mathrm{B}}$ and $\tilde{v}^{\mathrm{E}}$) calculated by the PDE scheme. In this figure, we observe that the Bermudan liability is almost equal to the European-style liability when the account value falls to very low levels. Because in this case, the probability that policyholders lapse the contract before the maturity is extremely small. While when the forward account value increases to very high levels, the early lapse premium grows almost linearly with f. This is due to the fact that when $f \gg f^*(t)$, $\tilde{v}^{\mathrm{B}}(t,f) = f$ and $\tilde{v}^{\mathrm{E}}(t,f) \approx e^{-\alpha(\mathrm{T}-t)} f$. Finally we observe that, like other Bermudan contingent claims, the early lapse premium of GMAB policies reduces gradually to 0 at the expiration date.

1.6.2 Option value of the GMAB policy

Firstly, we define the forward option value, denoted as $\tilde{w}^{\mathrm{B}}(t,f)$, of the Bermudan-style GMAB policy,

$$\tilde{w}^{\mathrm{B}}(t,f) := \tilde{v}^{\mathrm{B}}(t,f) - f \tag{1.14}$$

The notation $\tilde{w}^{\mathrm{B}}$ implies that the option value has many similar properties as a vanilla put[3], see Appendix B. In fact, $\tilde{w}^{\mathrm{B}}(t,f)$ can be simply interpreted as the difference between the liability $\tilde{v}^{\mathrm{B}}(t,f)$ and the asset f of GMAB policy issuers. That is to say, $\tilde{w}^{\mathrm{B}}(t,f)$ is the option that the insurers should replicate in practice.

In Figure 1.5, we provide the numerical results obtained by the PDE scheme for different forward account values and different dates from the inception to the expiration of the policy. In this figure, we observe that the forward option value $\tilde{w}^{\mathrm{B}}(t,f)$ evolves similarly as a vanilla put, see Appendix B. In addition, when the forward account value is significantly higher than the critical lapse boundary, the option value becomes negative between two discrete exercisable dates. This is due to the charge fees that policyholders are obliged to pay to insurers.

3. For European GMAB policies, we know that the option value is in fact a vanilla put (see (1.7)).

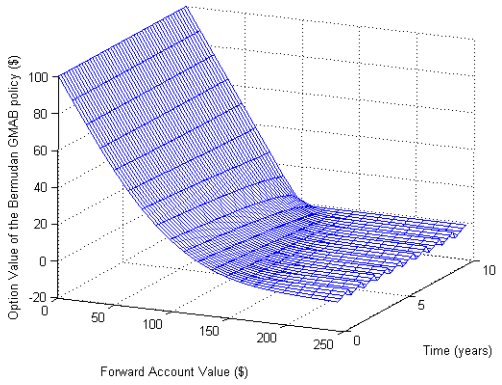


FIGURE 1.5 – The forward option value ($\tilde{w}$) of a standard GMAB policy for different forward account values and dates from the inception to the expiration of the policy.

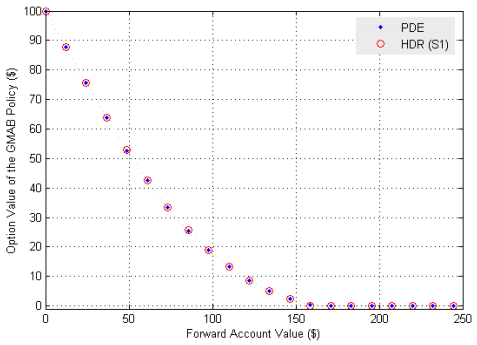


FIGURE 1.6 – Comparing the numerical results of the option value $\tilde{w}(0,f)$ computed by two different methods : PDE and Monte Carlo.

Figure 1.6 compares the numerical results of $\tilde{w}^{\mathrm{B}}(0,f)$ computed by two methods : PDE and Monte Carlo. For the Monte Carlo method used here, we simulate 10,000 scenarios with the step length of 0.1 year. To evaluate $\tilde{w}^{\mathrm{B}}(t,f)$ in this example, the average computing time of Monte Carlo method is about 1 to 2 seconds. Figure 1.6 shows that the prices calculated by Monte Carlo-S1 method is consistent with the results of PDE method.

We also compare the numerical results of the forward delta of the option value computed by PDE and Monte Carlo-S1 in Figure 1.8. We observe that the forward delta jumps up to 0 very quickly when f approaches to the critical boundary. This is easy to understand, as in this case all policyholders are supposed to lapse the contract and the insurers have no more need to hedge their liabilities. Fortunately, this difficulty of hedging lapse risks near the critical boundary can be partly overcome by diversifying the portfolio of GMAB policies (e.g. different maturities and benefit base levels).

Figure 1.9 compares the critical lapse boundary for different discrete dates computed by the two methods introduced above. We observe that the numerical results of Monte Carlo method becomes more and more instable from the expiration to the inception of the policy. This is because of the accumulation of pricing errors caused by linear regressions backward through time.

In summary, we find that in our specific example here, the PDE method is faster and more precise, especially for f nearing the critical boundary, than Monte-Carlo based methods. In addition, this method can calculate the price and other important Greeks for different f and t at the same time. However, compared with the PDE method, the Monte Carlo method is much more flexible and easier to be implemented. Moreover, unlike PDE based methods, the Monte Carlo method can be extended to other high-dimensional problems, such as path-dependent payoffs, stochastic volatility models or basket account values, see [35, 37] .

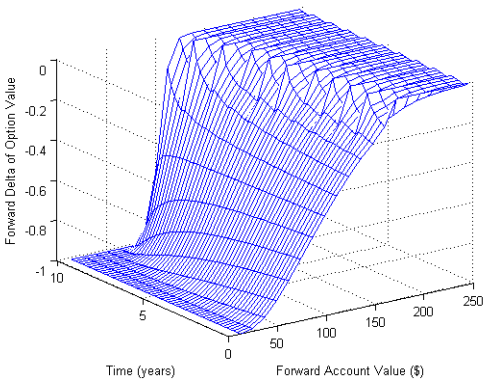


FIGURE 1.7 – The forward delta of $\tilde{w}(t,f)$ of a standard GMAB policy for different forward account values and dates from the inception to the expiration of the policy.

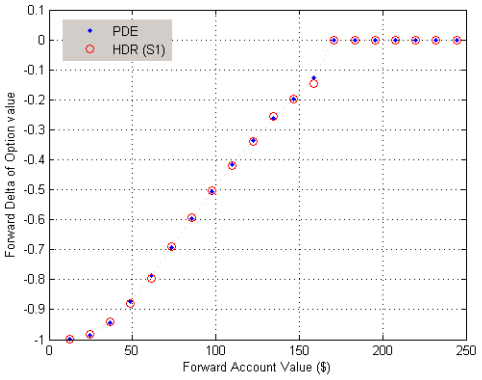


FIGURE 1.8 – Comparing the numerical results of the forward detla of the option value computed by two different methods : PDE and Monte Carlo-S1.

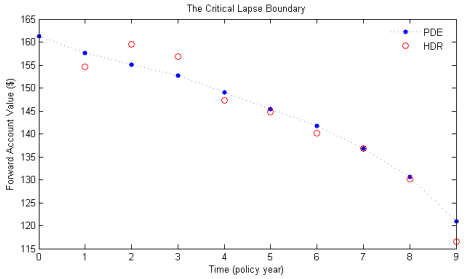


FIGURE 1.9 – Comparing the critical lapse boundary computed by two methods : PDE and Monte Carlo.

1.7 Conclusion

In this paper, we introduce a framework to evaluate the liability of GMAB polices under rational lapse assumption. We study in full details not only the financial sensitivities, but also the rational lapse strategy of GMAB products in the stochastic interest rate model. Two numerical methods, the PDE and Monte Carlo, are implemented to price the policy and also to determine the critical lapse boundary. Moreover, we find a semi-analytical formula to approximate the lapse premium of the GMAB. Inspired by the rational lapse assumption, we finally introduce the reasonable lapse assumption to help insurers to measure the lapse risks of a policy pool.

Appendix A : American-style GMAB

To the best of our knowledge, there is no closed formula to evaluate the Bermudan-style liability $\tilde{v}^{\mathrm{B}}(t,f)$. However, we can use some semi-closed formulas to approximate the liability if we assume that the policyholders can lapse the contract at any time. In this case, we denote the GMAB liability as $\tilde{v}^{\mathrm{A}}(t,f)$, which is in fact an American-style contingent claim.

In the past twenty years, many analytical approaches for evaluating American-style options in the Black-Scholes model are published, such as [15, 28, 46, 107], etc. However, most of them are not flexible for different payoff functions. In this paper, we find that the BAW and JZ approaches, see [15] and [107], can be extended to estimate $\tilde{v}^{\mathrm{A}}(t,f)$ in the one factor Hull-White model. Numerical tests show that these two approximations are efficient and precise for GMAB policies with 20 years maturity.

Firstly, we show how the BAW method can be applied directly to estimate the American–style liability $\tilde{v}^{\mathrm{A}}(t,f)$. It is obvious that $\tilde{v}^{\mathrm{A}}(t,f)$ is the solution of (1.29),

$$\frac{\partial \tilde{v}^{\mathrm{A}}}{\partial t} - cf\frac{\partial \tilde{v}^{\mathrm{A}}}{\partial f} + \frac{w^2_{\mathrm{T}-t} f^2}{2}\frac{\partial^2 \tilde{v}^{\mathrm{A}}}{\partial f^2} = 0 \tag{1.15}$$

and is subject to the boundary conditions $\tilde{v}^{\mathrm{A}}(t-,f) = \max(f, \tilde{v}^{\mathrm{A}}(t,f))$ for $0 \leqslant t \leqslant \mathrm{T}$. The key insight of BAW approximation, see [15], is that if both American options and European options are solutions of (1.29), then the early exercise premium $\psi(t,f)$ of GMAB policies, which is equal to $\tilde{v}^{\mathrm{A}}(t,f) - \tilde{v}^{\mathrm{E}}(t,f)$, is also a solution of (1.29). Defining $\tau = \mathrm{T} - t$ and changing the variable of ψ from t to τ, we have that $\psi(\tau,f)$ is the solution of the following equation,

$$-\frac{\partial \psi}{\partial \tau} - cf\frac{\partial \psi}{\partial f} + \frac{1}{2}\omega^2_\tau f^2 \frac{\partial^2 \psi}{\partial f^2} = 0 \tag{1.16}$$

In practice, it is very difficult to find the solution of (1.16) analytically. So the authors of [15] developed an approximation method to estimate $\psi(\tau,f)$.

According to the BAW method, the early exercise premium can be approximated by the function $\overline{\psi}(\tau,f) = h(\tau)u(h,f)$, where $h(\tau) = 1 - e^{-g\tau}$ (numerical tests show that $g = -\log(\mathrm{Z}^{\mathrm{T}}_t)/\tau$ could be a good choice) and $u(h,f)$ is a function to determine. Replacing $\psi(\tau,f)$ by $\overline{\psi}(\tau,f)$ in (1.16) and neglecting the term $\partial u/\partial h$, we have :

$$-\frac{g e^{-g\tau}}{1-e^{-g\tau}}u - cf\frac{\partial u}{\partial f} + \frac{1}{2}\omega^2_\tau f^2 \frac{\partial^2 u}{\partial f^2} = 0 \tag{1.17}$$

The general solution $u(h,f)$ of (1.17) is :

$$u(h,f) = \mathrm{A}_1 f^{\lambda_1} + \mathrm{A}_2 f^{\lambda_2}, \text{ where } \lambda_{1,2} = \frac{\omega^2_\tau + 2c \pm \sqrt{(\omega^2_\tau + 2c)^2 + 8g e^{-g\tau}\omega^2_\tau / h(\tau)}}{2\omega^2_\tau}$$

As $\lambda_2 < 0$ while the early exercise premium is worthless when the asset price drops to zero, the coefficient A_2 must be zero. Thus when $f < f^*(t)$ at time t, the American-style liability can be approximated by :

$$\tilde{v}^{\mathrm{A}}(t,f) \approx \tilde{v}^{\mathrm{E}}(t,f) + h(\tau)\mathrm{A}_1 f^{\lambda_1} \tag{1.18}$$

It remains the coefficient A_1 and the critical forward account value $f^*(t)$ to find. In fact, (1.13) implies that at $f^*(t)$, $\tilde{v}^{\mathrm{A}}(t,f^*)$ is equal to the forward account value, that is

$$f^*(t) = \tilde{v}^{\mathrm{E}}(t,f^*(t)) + h(\tau)\mathrm{A}_1 f^*(t)^{\lambda_1} \tag{1.19}$$

and the slope of the exercisable value, which is the forward account value, is set equal to the slope of $\tilde{v}^{\mathrm{A}}(t,f)$ at $f^*(t)$, that is,

$$1 = \frac{\partial \tilde{v}^{\mathrm{E}}(t,f)}{\partial f}|_{f=f^*(t)} + h(\tau)\lambda_1 \mathrm{A}_1 f^*(t)^{\lambda_1 - 1} \tag{1.20}$$

Solving (1.19) and (1.20) by the algorithm of Newton-Raphson, see [15], we can find both $f^*(t)$ and A_1 at time t.

The numerical tests show that the pricing error of BAW method is tiny if the forward account value f is not too small. However, when the GMAB policies are deep in the money, the BAW approximation becomes less precise. To improve the precision in this case, we extend the method developed by Ju and Zhong (JZ method, see [107]) to our specific evaluation problem here. In fact, the authors of

[107] proposed to add a perturbation term to the function $\overline{\psi}(\tau, f)$ to improve the precision of the early exercise premium. This corrected function, denoted by $\overline{\psi}_j$, is defined as :

$$\overline{\psi_j}(\tau, f) := (1+\epsilon(h,f))\overline{\psi}(\tau,f) = (1+\epsilon(h,f))h(\tau)u(h,f) \tag{1.21}$$

where $\epsilon(h,f)$ is a function to determine. Replacing ψ by $\overline{\psi_j}$ in (1.16) and applying (1.17), we obtain an equation for $\epsilon(h,f)$ at time 0,

$$-\frac{\partial h}{\partial \tau}\frac{\partial u}{\partial h}(1+\epsilon) - u\frac{\partial h}{\partial \tau}\frac{\partial \epsilon}{\partial h} + (\omega_\tau^2 f^2 \frac{\partial u}{\partial f} - cfu)\frac{\partial \epsilon}{\partial f} + \frac{1}{2}\omega_\tau^2 f^2 u \frac{\partial^2 \epsilon}{\partial f^2} = 0 \tag{1.22}$$

After a series of approximations, see [107], we get the corrected approximation to the American-style liability $\tilde{v}^{\mathrm{A}}(t,f)$:

$$\tilde{v}^{\mathrm{A}}(t,f) \approx \tilde{v}^{\mathrm{E}}(t,f) + \frac{d}{1-bx^2-\alpha x}\left(\frac{f}{f^*(t)}\right)^{\lambda_1} \tag{1.23}$$

where $x = \log(f/f^*(t))$ and a, b, c and d are four parameters to be determined by (1.22). Figure 1.10 compares the American liability v^{A} computed by (1.18) and (1.23) with the numerical results of PDE

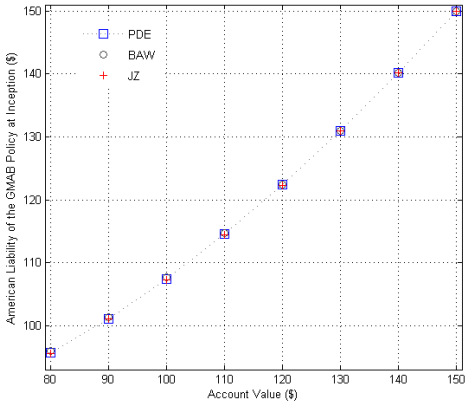


FIGURE 1.10 – Comparing the liability calculated by two approximation methods with the numerical results of PDE.

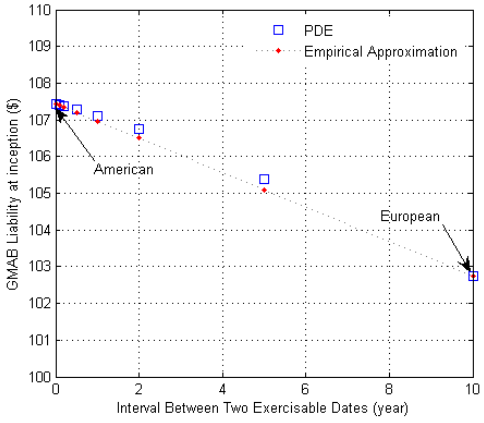


FIGURE 1.11 – Comparing the empirical approximation methods with the numerical results of PDE.

scheme [4], which is considered as the benchmarks here. We observe that the approximation methods are precise for a wide range of initial account values.

In addition, we find an empirical relationship between the Bermudan GMAB liabilities and the American ones, which can be simply written as

$$v^{\mathrm{B}}(0,f) \approx v^{\mathrm{A}}(0,f) - (v^{\mathrm{A}}(0,f) - v^{\mathrm{E}}(0,f))\frac{\Delta t}{\mathrm{T}} \tag{1.24}$$

where T is the maturity and Δt is the interval between two exercisable dates of Bermudan GMAB liabilities. Figure 1.11 verifies the empirical approximation (1.24) by the numerical results of PDE scheme. In fact, our numerical tests show that (1.24) is also applicable for other long-term Bermudan contingent claims (e.g. vanilla puts with maturities longer than 5 years).

Appendix B : Option value of GMAB policies

4. In this numerical test, the discrete time step of the PDE scheme is 0.01 year.

The forward option value $\tilde{w}^{\mathrm{B}}(t,f)$ of GMAB policies, defined by (1.14), is what the insurers should replicate in practice once they write GMAB contracts. According to the definition, we can decompose $\tilde{w}^{\mathrm{B}}(t,f)$ into two parts :

$$\tilde{w}^{\mathrm{B}}(t,f) = \tilde{v}^{\mathrm{B}}(t,f) - f = [\tilde{v}^{\mathrm{B}}(t,f) - e^{-\alpha\tau} f] - (1-e^{-\alpha\tau}) f \tag{1.25}$$

For simplicity, we define $\tilde{u}^{\mathrm{B}}(t,f) = \tilde{v}^{\mathrm{B}}(t,f) - e^{-\alpha\tau} f$. Applying (1.11), it is easy to verify that $\tilde{u}^{\mathrm{B}}(t_i-, \mathrm{F}^{\mathrm{T}}(t_i-))$ evolves as

$$\begin{aligned} \tilde{u}^{\mathrm{B}}(t_i-, \mathrm{F}^{\mathrm{T}}(t_i-)) &= \max((1-e^{-\alpha(\mathrm{T}-t_i)})\mathrm{F}^{\mathrm{T}}(t_i),\ \tilde{u}^{\mathrm{B}}(t_i, \mathrm{F}^{\mathrm{T}}(t_i))) \\ &= \max((1-e^{-\alpha(\mathrm{T}-t_i)})\mathrm{F}^{\mathrm{T}}(t_i),\ \mathbf{E}^{\mathbb{Q}^{\mathrm{T}}}[\tilde{u}^{\mathrm{B}}(t_{i+1}-, \mathrm{F}^{\mathrm{T}}(t_{i+1}-))|\mathscr{F}_{t_i}]) \end{aligned} \tag{1.26}$$

and at the maturity, we have $\tilde{u}^{\mathrm{B}}(\mathrm{T}, \mathrm{F}^{\mathrm{T}}(\mathrm{T})) = (\mathrm{G} - \mathrm{F}^{\mathrm{T}}(\mathrm{T}))^+$. Therefore, we can interpret $\tilde{u}^{\mathrm{B}}(t,f)$ as a Bermudan put option with the exercisable value $(1-e^{-\alpha\tau})f$ at dates t_i. Some insurers call $\tilde{u}^{\mathrm{B}}(t,f)$ as the forward value of claims, and $(1-e^{-\alpha\tau})f$ as the forward value of charges, for this term is in fact the expectation of forward charge fees insurers can receive if the policyholder holds the contract to the maturity. According to (1.25), the forward option value $\tilde{w}^{\mathrm{B}}(t,f)$ is equal to the difference between the forward value of claims and the forward value of charges.

Appendix C : Numerical schemes

It follows from the definition of the forward Bermudan-style liability $\tilde{\mathrm{V}}^{\mathrm{B}}(t)$ with an optimal stopping time $\tau \in \{t_1, t_2, \cdots, \mathrm{T}\}$, that the process of the forward liability $\tilde{\mathrm{V}}^{\mathrm{B}}(t)$ satisfies the backward programming equation, for $0 < t_i < \mathrm{T}$

$$\tilde{\mathrm{V}}^{\mathrm{B}}(t_i-) = \max\{\mathrm{F}^{\mathrm{T}}(t_i),\ \mathbf{E}^{\mathbb{Q}^{\mathrm{T}}}[\tilde{\mathrm{V}}^{\mathrm{B}}(t_{i+1}-)|\mathscr{F}_{t_i-}]\} \tag{1.27}$$

and at the maturity, we have $\tilde{\mathrm{V}}^{\mathrm{B}}(\mathrm{T}) = \max(\mathrm{G}, \mathrm{F}^{\mathrm{T}}(\mathrm{T}))$.

Thanks to the martingale property of $\tilde{\mathrm{V}}^{\mathrm{B}}$ on the interval $[t, \hat{\tau}_i)$, we have for $0 < t_i < \mathrm{T}$,

$$\tilde{\mathrm{V}}^{\mathrm{B}}(t_i-) = \mathbf{E}^{\mathbb{Q}^{\mathrm{T}}}[\mathrm{F}^{\mathrm{T}}(\hat{\tau}_i) + \mathbf{1}_{\{\hat{\tau}_i=\mathrm{T}\}} \max(\mathrm{G}, \mathrm{F}^{\mathrm{T}}(\mathrm{T}))|\mathscr{F}_{t_i}] \tag{1.28}$$

where the optimal stopping time $\hat{\tau}_i$ is defined as : $\hat{\tau}_i := \inf\{t_j \geqslant t_i : \tilde{\mathrm{V}}^{\mathrm{B}}(t_j-) = \mathrm{F}^{\mathrm{T}}(t_j-)\}$.

To the best of our knowledge, it is difficult to find precise analytical formulas to evaluate the Bermudan-style contingent claims in practice. In this paper, we extend the traditional semi-analytical methods, see [15, 107], to estimate the spot Bermudian-style liability $\tilde{v}^{\mathrm{B}}(t,f)$ of GMAB polices, see Appendix A. However, the approximation method introduced here is not as flexible as numerical approaches, especially for high dimensional problems. Thus in most cases, we need to use numerical methods, such as PDE and Monte Carlo, to calculate the Bermudan liability $\tilde{v}^{\mathrm{B}}(t,f)$.

1. PDE scheme

In this paper, we transform the evaluation problem (1.11) of Bermudan-style liability $\tilde{v}^{\mathrm{B}}(t,f)$ into a free-boundary partial differential equation, for which $\tilde{v}^{\mathrm{B}}(t,f)$ is the solution. For GMAB policies, the Bermudan-style liability $\tilde{v}^{\mathrm{B}}(t,f)$ is represented as a function of two variables : the time t and the forward account value f. Applying Itô's lemma and the martingale representation theorem together, we

know that the liability $\tilde{v}^{\mathrm{B}}(t,f)$ is the solution of a one dimensional PDE. By adding the free-boundary constraint implied by equation (1.11) to this PDE, we have

$$\frac{\partial \tilde{v}^{\mathrm{B}}}{\partial t} - cf\frac{\partial \tilde{v}^{\mathrm{B}}}{\partial f} + \frac{w_{\mathrm{T}-t}^2 f^2}{2}\frac{\partial^2 \tilde{v}^{\mathrm{B}}}{\partial f^2} = 0 \tag{1.29}$$

on $\{(t,f) : t_{i-1} \leqslant t < t_i, f > 0\}$, subject to the boundary conditions at time points $0 < t_i < t_{n+1}$

$$\tilde{v}^{\mathrm{B}}(t_i-, f) = \max(f, \tilde{v}^{\mathrm{B}}(t_i, f)) \tag{1.30}$$

and at the maturity T, we have

$$\tilde{v}^{\mathrm{B}}(\mathrm{T}, f) = \max(f, \mathrm{G}) \tag{1.31}$$

On each of the intervals $[t_{i-1}, t_i)$, the PDE (1.29) can be calculated numerically by using the Crank-Nicolson method, see [59], for $f \in [0, \mathrm{F})$, where F is the upside boundary of the numerical solution. While at discrete time points t_i, the critical lapse surface $f^*(t_i)$ can be easily found by the free-boundary constraint indicated in (1.30). On the boundary, we impose the zero-convexity conditions [5] :

$$\tilde{v}^{\mathrm{B}}(t,f)\,|_{f=0} = \mathrm{G}; \qquad \frac{\partial^2 \tilde{v}^{\mathrm{B}}}{\partial f^2}|_{f=\mathrm{F}} = 0$$

In fact, according to [14], the precision of the final solution is not very sensible to the error on boundaries if the solution domain of parabolic equation is large enough. So in most cases, the practitioner can choose other boundary conditions instead of those we propose here [6].

2. Monte-Carlo scheme

The liability $\tilde{\mathrm{V}}^{\mathrm{B}}(0)$ is estimated as the conditional expected value of the forward liability based on Monte-Carlo simulation.

The forward liability satisfies two conditions :

— the backward programming equation :

$$\begin{cases} \tilde{\mathrm{V}}^{\mathrm{B}}(t_i-) = \max\{\mathrm{F}^{\mathrm{T}}(t_i),\ \mathbf{E}^{\mathbb{Q}^{\mathrm{T}}}[\tilde{\mathrm{V}}^{\mathrm{B}}(t_{i+1}-)|\mathscr{F}_{t_i-}]\} \\ \tilde{\mathrm{V}}^{\mathrm{B}}(\mathrm{T}) = \max(\mathrm{G}, \mathrm{F}^{\mathrm{T}}(\mathrm{T})) \end{cases} \tag{1.32}$$

— the martingale property of $\tilde{\mathrm{V}}^{\mathrm{B}}$ on each $[t, \tilde{\tau}_i)$ traduced by :

$$\begin{cases} \tilde{\mathrm{V}}^{\mathrm{B}}(t_i-) = \mathbf{E}^{\mathbb{Q}^{\mathrm{T}}}[\mathrm{F}^{\mathrm{T}}(\hat{\tau}_i) + \mathbf{1}_{\{\hat{\tau}_i=\mathrm{T}\}}\max(\mathrm{G}, \mathrm{F}^{\mathrm{T}}(\mathrm{T}))|\mathscr{F}_{t_i}] \\ \hat{\tau}_i := \inf\{t_j \geqslant t_i : \tilde{\mathrm{V}}^{\mathrm{B}}(t_j-) = \mathrm{F}^{\mathrm{T}}(t_j-)\} \end{cases} \tag{1.33}$$

As pointed out in [37], these equations ((1.27) and (1.28)) lead to two algorithms, referred to as S1 and S2 hereafter.

The first algorithm S1 computes the optimal stopping time to lapse in three steps :

1. Simulate N discrete scenarios of the forward account value, denoted as $\mathrm{F}^{\mathrm{T}(k)}$ ($0 \leqslant i \leqslant n+1$ and $0 < k \leqslant \mathrm{N}$), according to (1.5).

5. This assumption is based on the fact that the gamma of the liability is small on the boundary.

6. In the specific case here, the first or second order derivative boundary condition is preferred to the Dirichlet condition. As the latter could lead to significant errors on the boundary.

2. Set the forward Bermudan-style liability at maturity for each scenario : $\tilde{V}^{B(k)}_{[1]}(T) = F^{T}(T)$.
3. Apply (1.27) from t_n to t_0. For $i = n$ to 0 :

$$\text{if } F^{T(k)}(t_i-) < B: \quad \tilde{V}^{B(k)}_{[1]}(t_i-) = \tilde{V}^{B(k)}_{[1]}(t_{i+1}-)$$
$$\text{if } F^{T(k)}(t_i-) \geqslant B: \quad \tilde{V}^{B(k)}_{[1]}(t_i-) = \max\{F^{T(k)}(t_i-),\ \tilde{\mathbf{E}}^{\mathbb{Q}^T}[\tilde{V}^{B}_{[1]}(t_{i+1}-)|F^{T(k)}(t_i-)]\}$$

From step 3 of scheme S1, we can identify the estimated rational lapse time $\tilde{\tau}^{(k)}_0$ as the first time for the k-th scenario where the liability equals the account value. Once $\tilde{\tau}^{(k)}_0$ is recorded for each path, we can estimate the Bermudan-style liability by scheme S2 where we regress the cash flows on a set of basis functions .

This latest computes the corresponding liability following four steps :

1. Simulation : Use the same N simulated scenarios as in S1.
2. Initialization : Set the rational lapse time $\tilde{\tau}^{(k)}_0 = t_{n+1}$, for $0 < k \leqslant N$.
3. Backward induction : For $i = n$ to 0, $\tilde{\tau}^{(k)}_i = i\mathbf{1}_{\{(k)\in\mathfrak{L}_i\}} + \tilde{\tau}^{(k)}_{i+1}\mathbf{1}_{\{(k)\in\mathfrak{L}^c_i\}}$. (where $\mathfrak{L}_i := \{(k) : V^{B(k)}_{[1]}(t_i-) = F^{T(k)}(t_i-)\}$ and $\mathfrak{L}^c_i\{(k) : V^{B(k)}_{[1]}(t_i-) > F^{T(k)}(t_i-)\}$ its complement)
4. Price estimator at 0 : $\tilde{V}^{B}_{[2]}(0) := \frac{1}{N}\sum_{k=1}^{N}[F^{T(k)}(\tilde{\tau}^{(k)}_0) + \mathbf{1}_{\{\tilde{\tau}^{(k)}_0 = T\}}\max(G, F^{T(k)}(T))]$.

In [37], the authors find the following relation with the two estimators $\tilde{V}^{B}_{[2]}$ and $\tilde{V}^{B}_{[1]}$ computed above :

$$\mathbf{E}[\tilde{V}^{B}_{[2]}(0)] \leqslant \tilde{V}^{B}(0) \leqslant \mathbf{E}[\tilde{V}^{B}_{[1]}(0)] \tag{1.34}$$

In the numerical tests in the section below, we calculate both $\tilde{V}^{B}_{[1]}(0)$ and $\tilde{V}^{B}_{[2]}(0)$ to construct confidence intervals $[\tilde{V}^{B}_{[2]}(0), \tilde{V}^{B}_{[1]}(0)]$ for the true value $\tilde{V}^{B}(0)$.

Appendix D : Linear regression vs global polynomial regression

We now introduce the scheme used to calculate the conditional expected value of continuation for scenarios such that $F^{T(k)}(t_i-) > B$. Here we use the local linear regression approach proposed in [37] to calculate this value, as opposed to the global polynomial regression method developed in [120]. The reason for this is that the latter can lead to some instability in the regression process for high dimensional and long maturity problems, see [37].

For our specific problem, we have only one dimension : the forward account value F^T. The idea is to use, at each time step t_i, a set of functions ψ_d having local hypercube supports D_l, where the space is cut into I regions, $l = 1$ to I and $\{D_d\}$ is a partition of $[\min_{\{k=1,N\}} F^{T(k)}(t_i), \max_{\{k=1,N\}} F^{T(k)}(t_i)]$. The index $(\cdot)^{(k)}$ denotes the k-th simulated scenario. On each support D_l, we define a linear function Ψ_l with 2 degrees of freedom, which are represented by a constant and F^T. Our goal now is to regress the future cash flow of liability on the function Ψ_l to estimate the relevant conditional expectation. The two regression basis of Ψ_l, noted as (ψ^0_l, ψ^1_l) [7].

For simplicity, we define the function $G^N(t_i, F^T(t_i))$ as the conditional expectation at time t_i, we have :

$$G^{N(k)}_i :\ G^N(t_i-, F^{T(k)}(t_i-)) = \mathbf{E}^{\mathbb{Q}^T}[\tilde{V}^{B}_{[1]}(t_{i+1}-, F^T(t_{i+1}-))|F^{T(k)}(t_i-)]$$

7. The two regression basis correspond to the constant and the forward account value.

where $F_i^{N(k)}$ is the conditional expectation associated with the k-th path at time t_i. In the context of S1, the numerical procedure to calculate $G_i^{N(k)}$ reads as follows :

Scheme S_c : estimator of $G_i^{N(k)} (0 \leqslant i \leqslant n)$ with regression :

1. At time t_i, realize a quick-sort of $F^{T(k)}(t_i)$ for N scenarios and identify the support D_l of the functions Ψ_l so that each support contains approximately the same number of scenarios.
2. For each scenario $0 < k \leqslant N$, set the three regression basis of Ψ_l :(ψ_l^0, ψ_l^1), where $\psi_l^0(\cdot) = 1$, $\psi_l^1(F^{T(k)}(t_i-)) = F^{T(k)}(t_i-)$.
3. On each support D_l, regress $\{V_{[1]}^{B(k)}(t_{i+1}-)\}_{k \leqslant N}$ on Ψ_l. In other words, for $\forall l$, we calculate the coefficients (α_l^0, α_l^1) that minimize $\sum_{k=1}^{N} |V_{[1]}^{B(k)}(t_{i+1}-) - \sum_{m=0}^{1} \alpha_l^m \psi_l^m(\cdot^{(k)})|^2$, and set $G_i^{N(k)} = \sum_{m=0}^{1} \alpha_l^m \psi_l^m(F^{T(k)}(t_i^-))$.

Chapitre 2

Optimal behavior strategy in the GMIB product

***Abstract*— This chapter falls within the scope of quantitative studies done at AXA Group Risk Management. It focuses on a variation of the guaranteed minimum income benefit with and without a death benefit. This variable annuity rider offers the policyholder the possibility to convert the benefit base into annuities for life. The income rate is fixed by the insurer at inception, depending on the insured age at election. From the insurer's point of view, this product embeds the policyholder "optimal behavior". In this chapter we study such behavior from the policyholder's perspective who maximizes the expectation of his future cash flows. We develop this analysis in a dynamic programming framework. Using convenient scaling properties of the contract value, we reduce the dimension of the problem. We policyholder's decision as a function of the contract moneyness. Furthermore, we analyze the sensitivity of the product to different drivers like the volatility, interest rate and roll-up rate. In particular, we find that the contract is usually underpriced under optimal behavior.**

Keywords : GMIB; Variable Annuities; rational behavior; optimal withdrawals; PDE; dynamic programming.

2.1 Product description

Guaranteed Minimum Income benefit (GMIB) product appeared in the market in 1996. This guarantee enables policyholders to make annual partial withdrawals (typically 4% to 7%) of their guaranteed protection amount and ensures an analogous percentage of the GMIB benefit base for their entire lifetime, no matter how the investments in the sub-accounts perform. It combines longevity protection with withdrawal flexibility, hence it is seen as a "second-generation" guarantee. The guarantee can concern one or two lives (typically spouses). Each annual withdrawal does not exceed some maximum value, but it is evident that the total amount of withdrawals is limited only to the exhaustion of the client's account value. Annual withdrawals of about 5% of the (single initial) premium are commonly guaranteed for insured aged 60+. In case of death, any remaining fund value is paid to the insured dependents. To satisfy the new needs of an ageing population, insurance companies have started offering a lifetime benefit feature with GMIB.

While the GMIB product was defined in the introduction, its form Retirement Cornerstone® commercialized by AXA will be explained and illustrated in this section.

2.1.1 Retirement Cornerstone

Retirement Cornerstone® is a long-term GMIB-type product designed for retirement purposes. This deferred variable annuity is commercialized by AXA Equitable life insurance company in the U.S. since 2011. It offers tax-deferred growth potential and living and death benefits as optional features.

Benefit Base and partial withdrawals

To describe the benefit features of this GMIB clearly, some key terms must be addressed : GMIB benefit base and guaranteed withdrawal amount (GWA).

— **GMIB benefit base :** The GMIB benefit base is an amount used to determine the guaranteed annual withdrawal amount and lifetime payments. The GMIB benefit base is created and increased by allocations and transfers to the account value, as well as annual withdrawn amounts. This percentage is know as roll-up rate and deferral roll-up rate. It must be noticed that the GMIB benefit base is a "fictive" amount, i.e., it can not be considered as an account nor cash value, only as a reference in the computation of lifetime payments and guaranteed withdrawal amounts.

— **(Annual) guaranteed withdrawal amount (GWA) :** The "annual guaranteed withdrawal amount" is the withdrawal amount suggested by the insurer. It is equal to the annual roll-up rate in effect on the first day of the contract year, multiplied by the current benefit base. It is also the maximum amount upon which the benefit base is reduced without penalty, in contrast to excess withdrawals.

In Table 2.1, we illustrate a concrete example of the calculations of a GMIB benefit. BB designates the benefit base and AV refers to the account value. The contract initial premium is $100,000. The policyholder does not make any withdrawal till the 6th contract year and once he/she does, all withdrawals stay within the boundaries of the GWA. Therefore, the GMIB benefit base does not diminish. The effect of excess withdrawals will be discussed later. Till the 5th contract anniversary, the deferral roll-up rate is used to calculate the amount that is credited to the benefit base each year since no withdrawal has been made. For example, the GMIB benefit base in the 5th year is computed by taking the value of the previous year and adding the corresponding 5%, i.e. $120,510+5.4%×$120,510 =$120,510+6,507.54=127,017.54.

Year	Deferral/ Roll-up rate	GMIB BB	WA	Percentage of GMIB BB	GMIB BB after withdrawal
0	-	$100,000	$0	0%	$100,000
1	4.8%	$104,800	$0	0%	$104,800
2	4.3%	$109,830	$0	0%	$109,830.40
3	5.2%	$114,553	$0	0%	$114,553.11
4	5.4%	$120,510	$0	0%	$120,509.87
5	5.0%	$127,017	$0	0%	$127,017.40
6	4.7%	$133,368	$6,268.31	4.7%	$127,099.96
7	5.2%	$133,368	$6,935.15	5.2%	$126,433.12
8	5.4%	$133,368	$7,201.89	5.4%	$126,166.39
9	6,0%	$133,368	$5,334.73	4,0%	$128,033.54
10	7.3%	$136,036	$5,441.43	4,0%	$130,594.21

TABLEAU 2.1 – GMIB benefit base evolution for an allocation of $100,000 with partial withdrawals that does not affect the value of the guaranteed account.

Once the client proceeded to his/her first withdrawal, the roll-up rate determines the evolution of the GMIB benefit base and the annual GWA. If the maximum guaranteed quantity is withdrawn, the benefit base remains unchanged, as shown in years 6 to 9, i.e. the same quantity withdrawn from the benefit base is added by the roll-up amount. In case of withdrawing less than the annual GWA, a greater value for the benefit base is obtained. In the 10th year, the policyholder only withdraws 4 % of the GMIB guaranteed value. Consequently, the benefit base becomes \$133,368-\$5,334.73+\$133,368×6.0% =\$128,033.54+\$8002.08=\$136,035.62.

The step-up option enables the policyholder to reset the guaranteed withdrawal balance to the current higher account value when investment performance is strong. By choosing to reset the benefit base, the policyholder is able to increase the total benefit amount and the annual guaranteed withdrawal amount. The option may reduce the inflation effect on incomes when the account value goes up and the step-up option is available. Accordingly, the period over which lifetime payments can begin is extended of 10 years. Nevertheless, at policyholder's 95th birthday, the lifetime payments are set to automatically begin no matter how many times the reset option has been chosen. In Figure 2.1 we illustrate the evolution of the protected account value, and the different options for the GMIB benefit base. The contract is issued at the policyholder's 50th birthday. Partial withdrawals from this annuity

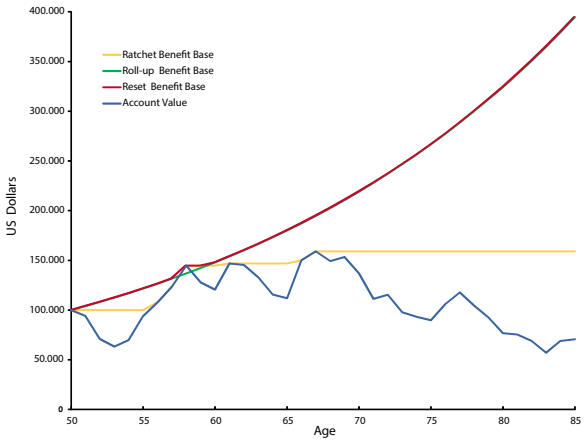


FIGURE 2.1 – Behavior of the account value and the different options for the GMIB benefit base when no withdrawals have been performed.

contract are taxable as ordinary income and, if made prior to age 59$^1/_2$ may be subject to an additional 10% federal tax and withdrawal charges. All amounts invested in an annuity's portfolios are subject to fluctuation in value and market risk, including loss of principal. The account value may be reduced due to fees and charges such as operations and sales charges, administrative fees, and optional benefits additional charges.

Fees and charges

The fee structure has an impact on the GMIB price. The GMIB charge is deducted from the contract value periodically. It is usually presented as a percentage of the current account value, although it can also be a percentage of the initial premium, a percentage of the remaining guaranteed benefit amount, or the greater of these two. The annual charge ranges for Retirement Cornerstone® is comprised between 20 and 75 basis points depending on the nature of the benefit.

Lifetime payments

Lifetime GMIBs provide guaranteed annual income until death. Policyholders are also able to access potentially increased account values, and control the asset allocation in ways that the traditional variable annuitization normally does not allow. Lifetime GMIBs usually have two options : single life or joint spousal life. For the single life option, the benefit payments end at the death of the person covered. For the joint spousal life option, the benefit payments end when the remaining spouse dies. The fee rate for the single life option ranges from 25 to 55 basis points, while the spousal life option tends to be 10-20 basis points higher.

For the single life option, a spouse continuation option is available upon the first death with the same charge, but the account value and the benefit amount may be adjusted. For the joint spousal life, there will be no recalculation of the benefit amount when the first death occurs. The annual benefit payment amount is a percentage of the initial guaranteed benefit amount. The older the policyholder is, the larger the value of the lifetime payments will be. For example, the guaranteed factor to compute the lifetime payment for the legacy product Accumulator 7 is 5.3% if the annuity starts at age 73 and 7.1% if the attained age is 83. The factor used to compute the lifetime payments is given by the insurer and depends on the policyholder's age at inception.

Annual lifetime payments in GMIB products begin as follows :

(i) The next contract year following the date the account value falls to zero.

(ii) The contract date anniversary following the policyholder's 95th birthday.

(iii) The policyholder's election to exercise the GMIB.

Similarly to GLWB product, GMIB is subject to a waiting period which begins on the date when the account value is first fund, and it ranges from 10-15 years depending on the policyholder's age.

If an excess withdrawal, i.e. withdrawal superior to the guaranteed withdrawal amount, reduces the account value to zero, the GMIB will be terminated. Even if an excess withdrawal does not cause the contract to terminate, it can greatly reduce the GMIB benefit base and the value of the benefit since it is done in pro rata basis.

In Table 2.2 cash flows for a maximum annual guaranteed withdrawal strategy are shown. The current and guaranteed factors [1] are illustrated for the case of a 60 year old male that acquires the contract in 2016. These factors take into account the age of the policyholder when the contract is issued, his/her gender and probability of survival. In the particular case of the current annuity factor, the market interest rates play a key role. A constant net return of 3 % and fees of 4.5% has been considered to facilitate comprehension. The factors need to be recalculated when they are faced to changing market interest rates.

In the case illustrated in Table 2.2, the policyholder takes the annual guaranteed withdrawal amount till his account value turns to zero at age 76. At that moment, lifetime payments begin and the owner of the policy faces two options : to annuitize the GMIB benefit base or the account value. Annual payments will be based on the guaranteed or current factor depending on the policyholder's choice. This is possible till the client's 85th anniversary, otherwise he will lose the possibility to transform the contracts benefit base into annual income payments. In the described scenario and taking the discount factors into account, the policyholder will obtain $86,981.33, meaning he will not recover the

1. The current (resp. guaranteed) factor is the annual income rate when the income benefit is based on the account value (resp. benefit base).

Contract Year	AV	Roll-up rate	WA	GMIB BB	GMIB factor	Current factor	Annuitization BB	Annuitization AV	Lifetime payments
0	$100,000		$0	$100,000	4.2%	5.4%	$4,204.79	$5,367.16	-
1	$99,500	6%	$6,000	$100,000	4.3%	5.5%	$4,269.15	$5,444.66	-
2	$93,032.50	6%	$6,000	$100,000	4.3%	5.6%	$4,336.57	$5,193.53	-
3	$86,597.34	6%	$6,000	$100,000	4.4%	5.7%	$4,407.24	$4,935.07	-
4	$80,194.35	6%	$6,000	$100,000	4.5%	5.8%	$4,481.31	$4,668.45	-
5	$73,823.38	6%	$6,000	$100,000	4.6%	6.0%	$4,558.97	$4,392.83	-
6	$67,484.26	6%	$6,000	$100,000	4.6%	6.1%	$4,640.42	$4,107.29	-
7	$61,176.84	6%	$6,000	$100,000	4.7%	6.2%	$4,725.87	$3,810.86	-
8	$54,900.96	6%	$6,000	$100,000	4.8%	6.4%	$4,815.53	$3,502.52	-
9	$48,656.45	6%	$6,000	$100,000	4.9%	6.5%	$4,909.63	$3,181.14	-
10	$42,443.17	6%	$6,000	$100,000	5.0%	6.7%	$5,008.38	$2,845.48	-
11	$36,260.95	6%	$6,000	$100,000	5.1%	6.9%	$5,111.95	$2,494.23	-
12	$30,109.65	6%	$6,000	$100,000	5.2%	7.1%	$5,220.49	$2,126	-
13	$23,989.10	6%	$6,000	$100,000	5.3%	7.3%	$5,334.10	$1,739.42	-
14	$17,899.16	6%	$6,000	$100,000	5.5%	7.4%	$5,452.86	$1,333.15	-
15	$11,839.66	6%	$6,000	$100,000	5.6%	7.7%	$5,576.81	$905.97	-
16	$5,810.46	6%	$5,810.46	$100,189.54	5.7%	7.9%	$5,716.80	$456.79	-
17	$0	6%	$0	$0	5.8%	8.1%	$0	$0,00	$5,716.80
18	$0	6%	$0	$0	6.0%	8.3%	$0	$0,00	$5,716.80
19	$0	6%	$0	$0	6.1%	8.5%	$0	$0,00	$5,716.80
20	$0	6%	$0	$0	6.3%	8.7%	$0	$0	$5,716.80

TABLEAU 2.2 – Protected account value and GMIB behavior given a static withdrawal strategy and lifetime payments.

initial premium invested on the contract.

Death benefit

The Retirement Cornerstone® also offers the possibility of combining the GMIB with Guaranteed Minimum Death Benefit (GMDB) when the contract is purchased. This is not unusual since variable annuities typically provide a guarantee if the policyholder dies before receiving any income.

The death benefit often equals the greater of the account value and total premiums paid less any withdrawals. For example, a person had paid premiums totaling $100,000, and had made withdrawals equaling $15,000. The account value stands at $80,000 because of these withdrawals and investment losses. If he were to die, his beneficiary would receive the aforementioned quantity. Within Retirement Cornerstone® product this type of death benefit is known as **"Return of Principal"** and is considered without any extra charges.

Retirement Cornerstone® also offers optional death benefits in the form of roll-up or annual ratchet and reset with extra charges. These options are :

— **Highest anniversary value death benefit**- The "Highest anniversary value death benefit" is an optional guaranteed minimum death benefit in connection with the account value. The death benefit is calculated using the highest value of the account on the contract date anniversary.

— **Roll-up to age 85 benefit base**-The "Roll-up to age 85 benefit base" is equal to the GMIB benefit base, i.e., it is reduced dollar-for-dollar in the case of partial withdrawals being done within the limits of the guaranteed withdrawal amount and pro-rata, when an excess withdrawal has been made by the policyholder. It is favored by the roll-up amount till the policyholder's 85th birthday. This option is tied only to "Greater of" death benefit, i.e. it can not be chosen individually.

— **"Greater of" death benefit**- The "greater of" death benefit is an optional guaranteed minimum

death benefit in connection with the protected benefit account value only. The death benefit is calculated using the greater of two benefit bases- the greater of the roll-up to age 85 benefit base and the highest anniversary value benefit base. There is an additional charge for the "greater of" death benefit under the contract.

Once the lifetime payments corresponding to the GMIB start, the policyholder loses the possibility of keeping the GMDB. This right is lost at policyholder's 95th anniversary since the lifetime payments start automatically. The return of principal, highest anniversary value, and "greater of" guaranteed minimum death benefits will terminate without value if the account value falls to zero as a result of withdrawals or payment of any applicable charges. This will happen whether the policyholder elects the GMIB or receive lifetime GMIB payments or not.

Policyholders can elect the optional death benefit guarantees between age 20 and 68, implying that this product targets a "younger" sector of the population compared to the "'return of principal' which can be chosen till age 80.

Some numerical examples will be presented to illustrate the evolution of the GMIB and GMDB under partial withdrawals. A premium of $100,000-dollar is considered for a policyholder aged 60, with no additional contributions, and no transfers. Throughout these examples, no charges are deducted from the account value and there is a fixed roll-up rate of 4 %. The assumed returns do not follow any market trends and were chosen to serve the purposes of illustrating two types of scenarios.

We define the following notation :

GMIB BB	:	Guaranteed minimum income benefit base
RP BB	:	Return of principal benefit base
RU BB	:	Roll-up to age 85 benefit base
HA BB	:	Highest anniversary value benefit base
GO BB	:	"Greater of" benefit base

					Guaranteed minimun death benefit				
Year	Net Return	AV	WA	Roll-up rate	GMIB BB	RP BB	HAV BB	RU BB	GO BB
0		$100,000			$100,000	$100,000	$100,000	$ 100,000	$100,000
1	3%	$103,000	$0	4%	$104,000	$100,000	$103,000	$ 104,000	$104,000
2	4%	$107,120	$0	4%	$108,160	$100,000	$107,120	$ 108,160	$108,160
3	6%	$113,547.20	$0	4%	$113,547.20	$100,000	$113,547.20	$ 113,547.20	$113,547.20
4	6%	$120,360.03	$0	4%	$120,360.03	$100,000	$120,360.03	$ 120,360.03	$120,360.03
5	7%	$128,785.23	$0	4%	$128,785.23	$100,000	$128,785.23	$ 128,785.23	$128,785.23
Alternative 1 : annual guaranteed withdrawal amount (dollar-for-dollar)									
6	-5%	$122,345.97	$5,151.41	4%	$128,785.23	$95,789.47	$123,633.82	$ 128,785.23	$128,785.23
7	1%	$118,418.02	$5,151.41	4%	$128,785.23	$91,622.45	$118,482.42	$ 128,785.23	$128,785.23
8	-2%	$116,049.66	$5,151.41	4%	$128,785.23	$87,555.36	$113,331.01	$ 128,785.23	$128,785.23
9	2%	$118,370.66	$5,151.41	4%	$128,785.23	$83,745.01	$108,179.60	$ 128,785.23	$128,785.23
10	2%	$120,738.07	$5,151.41	4%	$128,785.23	$80,171?94	$103,028.19	$ 128,785.23	$128,785.23
Alternative 2 : excess withdrawal (pro-rata)									
6	-5%	$122,345.97	$7,000	4%	$126,839.35	$94,278.52	$121,765.78	$ 126,839.35	$126,839.35
7	3%	$119,016.35	$7,000	4%	$124,786.30	$88,733.49	$114,803.39	$ 126,839.35	$126,839.35
8	-2%	$109.636,02	$4,991.45	4%	$124,786.30	$83,742.03	$109,811.94	$ 126,839.35	$126,839.35
9	2%	$106,837.29	$4,991.45	4%	$124,786.30	$78,750.58	$104,820.49	$ 126,839.35	$126,839.35
10	2%	$103,982.59	$4,991.45	4%	$124,786.30	$73,759.13	$ 99,829.04	$ 126,839.35	$126,839.35

TABLEAU 2.3 – GMIB and GMDB behavior given policyholder's withdrawals when the account value is less than the GMIB benefit base at the time of the first withdrawal.

Table 2.3 shows that the account value is reduced dollar-for-dollar by the withdrawal amount before considering market behavior no matter the size of withdrawals. In alternative 1, when the owner withdraws the annual guaranteed withdrawal amount [4% (roll-up rate) × $128,785 (the roll-up benefit bases as of the 6th contract anniversary)] the GMIB and roll-up to age 85 benefit bases neither decrease nor increase. The return of principal benefit base is reduced prorata as follows : since the withdrawal amount of $5,151 equals 4.21% of the account value ($5,151 = 4.21% × $122,346), the return of principle (RP) benefit base is also reduced by 4.21% ; while the highest anniversary value (HAV) benefit base is reduced dollar-for-dollar, i.e., $128,785 (HA BB as of the last contract date anniversary) - $5,151 = $123,634 for the 6th contract year.

In the case of an excess withdrawal, as it is the case of contract years 6 and 7 of the second scenario, the Return of principal is reduced in the same way : since the withdrawal amount of $7,000 equals 5.721% of the account value in 6th year ($7,000 divided by $122,346 = 5,721%), the RP benefit base is reduced by 5.721%. The pro-rata reduction of the roll-up benefit bases is as follows : $7,000 (the amount of the withdrawal, including any applicable withdrawal charge) - $5,151 (GWA) = $1,849 (”excess”) which represent 1.511% of the account value, there is a decrease of 1.511% in the Roll-up benefit bases. The highest anniversary value benefit base is reduced dollar-for-dollar and pro-rata, as follows : $128,785 (HAV BB as of the last contract date anniversary) - $5,151 (GWA) = $1,868 [($128,785 - $5,151) × 1.511%] = $121,766.

In Table 2.4, the account value is reduced dollar-for-dollar by the withdrawn amount no matter

					Guaranteed minimun death benefit				
Year	Net return	AV	WA	Roll-up rate	GMIB BB	RP BB	HAV BB	RU BB	GO BB
0		$100,000			$ 100,000	$100,000	$ 100,000	$100,000	$ 100,000
1	3%	$103,000	$0	4%	$104,000	$100,000	$ 103,000	$104,000	$ 104,000
2	4%	$107,120	$0	4%	$108,160	$100,000	$ 107,120	$108,160	$ 108,160
3	6%	$113,547.20	$0	4%	$113,547.20	$100,000	$ 113,547.20	$113,547.20	$ 113,547.20
4	6%	$120,360.03	$0	4%	$120,360.03	$100,000	$ 120,360.03	$120,360.03	$ 120,360.03
5	7%	$128,785.23	$0	4%	$128.785.23	$100,000	$ 128,785.23	$128,785.23	$ 128,785.23
Alternative 1 : annual withdrawal amount									
6	5%	$135,224.50	$ 5,151.41	4,00%	$130,073.09	$96,190.48	$ 130,073.09	$130,073.09	$ 130,073.09
7	3%	$133,975.28	$ 5,202.92	4%	$130,073.09	$92,454.92	$ 128,772.36	$130,073.09	$ 130,073.09
8	-2%	$126,196.91	$ 5,202.92	4%	$130.073,09	$88,643.14	$ 123,569.43	$130,073.09	$ 130,073.09
9	2%	$123,413.86	$ 5,202.92	4%	$130,073.09	$84,906.09	$ 118,366.51	$130,073.09	$ 130,073.09
10	2%	$120,575.16	$ 5,202.92	4%	$130,073.09	$81,242.32	$ 115,372.24	$130,073.09	$ 130,073.09
Alternative 2 : excess withdrawal									
6	5%	$135,224.50	$ 7,000	4%	$ 128,224.50	$94,823.42	$ 128,224.50	$128,224.50	$ 128,224.50
7	3%	$132,071.23	$ 7,000	4%	$ 125,071.23	$89,797.62	$ 125,071.23	$125,071.23	$ 125,071.23
8	-2%	$122,569.81	$ 5,002.85	4%	$ 117,566.96	$86,132.41	$ 120,068.38	$117,566.96	$ 120,068.38
9	2%	$119,918.30	$ 4,802.74	4,00%	$ 115,115.56	$82,682.80	$ 115,265.65	$115,115.56	$ 115,265.65
10	2%	$117,417.87	$ 4,610.63	4%	$ 112,807.25	$79,436.11	$ 112,807.25	$112,807.25	$ 112,807.25

TABLEAU 2.4 – GMIB and GMDB behavior given policyholder’s withdrawals when the account value is greater than the GMIB benefit base at the time of the first withdrawal.

the size of withdrawal : in year 7 the AV is computed as $135,224.50 [AV as of the last contract date anniversary]-$7,000 (the amount of the withdrawal, including any applicable withdrawal charge)×(1+0.03 [assumed net return for the 7th contract anniversary])=$128,224.50×1.03 = $132,071.23. When the owner limits himself to making only annual guaranteed withdrawals [4% (roll-up rate) x $128,785 (the roll-up benefit bases as of the 6th contract anniversary)] the GMIB and roll-up to age 85 benefit are reduced dollar-for-dollar but since AV after withdrawal is greater than the aforementioned quantity, they are automatically set to $130,073.

As a result of the GMIB benefit base increase in contract year 6, the annual withdrawal amount in contract year 7 is $5,203 [4% (roll-up rate) x $130,073 (the roll-up benefit bases as of the sixth contract anniversary)]. The return of principal benefit base is reduced pro-rata as shown in the previous examples and the highest anniversary value benefit base is reduced dollar-for-dollar as follows : $128,785 (highest anniversary value benefit base as of the 5th contract date anniversary)-$5,151 = $123,634. The highest anniversary value benefit base is reset to the protected benefit account value after withdrawal ($130,073).

In the case of an excess withdrawal and similar to the previous example, the roll-up bases will be reduced in the same percentage as the excess. Taking the 6th contract anniversary as an example, it is reduced by 5.177% ($7,000 divided by $135,224) which gives $121,944. The roll-up to age 85 benefit base and GMIB benefit base are then set to the protected account value after withdrawal $128,224 since this value is clearly superior that of the benefit bases after the pro-rata reduction.
The RP benefit base continues to be reduced in pro-rata basis. The highest anniversary value benefit base is reduced dollar-for-dollar and pro-rata, as follows : $128,785 (highest anniversary value benefit base as of the 5th contract anniversary)-$5,151 (annual withdrawal amount) -$1,690 [($128,785 - $5,151)× 1.367%]= $121,944. Here 1.367% represents the percentage of the excess in withdrawal with respect to the GWA ($1,849 divided by $135,224 = 1.367%). The highest anniversary value benefit base is also reset to the protected account value after withdrawals ($128,224).

In the following section, we will set up the mathematical formulation of the product for the purpose of studying its valuation. For the sake of simplicity, we will limit our study to a single benefit base.

2.2 A brief review of the literature

There is a large literature on pricing and hedging variable annuities guarantees. Most of it addresses on individual variable annuities contracts. Milevsky and Posner, see [127], price a GMDB contract using the usual risk-neutral valuation theory. Gerber and Shiu, see [91], exploit the closed-form solution of European look-back options to price complex guarantees embedded in some equity-linked annuities. Milevsky and Salisbury, see [128], study the impact of policyholder behavior on the cost and value of the GMWB rider and argue that the current pricing is not sustainable. An analysis of the design of general equity-indexed annuities from the investor's perspective and a generalization of the conventional design are proposed in [38].

The optimal behavior approach in a GMWB valuation was formalized by Dai and co-authors in [61]. They develop a singular stochastic control problem in a continuous framework, and also construct discrete pricing formulation that models withdrawals on discrete dates. In [19], the authors develop an extensive and comprehensive framework, to price any of the common guarantees available with VAs, using Monte Carlo simulations in deterministic withdrawals scenarios. On the other hand, in [50], the authors explore the effect of various modeling assumptions on the optimal withdrawal strategy of the policyholder, and examine the impact on the guarantee value under sub-optimal withdrawal behavior. Shah and Bertsimas, see [149], analyze the GLWB option in a time continuous framework considering simplified assumptions on population mortality, and adopting different asset pricing models. In [11], a number of guarantees under a more general financial model with stochastic interest rates, volatility, and mortality are considered. A utility-based approach, see [82], is used to study the valuation of the GMDB rider.

Holz and co-authors, see [98], price GLWB contracts for different product designs and model para-

meters under the geometric Brownian Motion dynamic. They consider various policyholders behaviors assumptions including deterministic, probabilistic and stochastic models. The GMIB is studied in [62] under a local volatility framework. The authors argue that an appropriate volatility modeling is important to the long-dated guarantees. Finally, Dai and Yang, see [156], develop a tree model to price the GMWB rider embedded in deferred life annuity contracts. Other papers investigate the impact of volatility risk, or assess the mortality risk in GLWB, or analyze equity and systematic mortality risks, see for example [81, 92, 143]. Recently, the work by Shevchenko and co-authors, see [150], provides a useful general framework to price different living and death guarantees. They use a direct integration method to solve the problem and compare it to PDE-based methods. In the following we will focus on the PDE method for the GMIB product. Our goal is to be able to analyse the impact of different market drivers and product design on policyholders' behaviors and the value of the contract.

2.3 Formulation and basic notations

In summary, GMIB contracts promise a policyholder an income stream at maturity for the rest of his life. Before the contract maturity, the insured is allowed to withdraw a certain amount on a yearly basis, called a withdrawal. If the GMIB contract contains a death benefit (GMIB-DB), then a certain amount is paid to the beneficiaries in case the policyholder dies during the term of the contract.

To formulate our problem, we consider an x-year old policyholder possessing a GMIB contract. At inception, an initial endowment is invested in a risky asset S_t. The specifications of the contract include a set of dates $0 = t_0 < t_1 < ... < t_n < ... < t_N = T$, where $t_0 = 0$ is the contact inception and $t_N = T$ its maturity. These so-called contract anniversaries are the dates in which events can take place, i.e. bonuses, withdrawals, payments, etc...

2.3.1 The contract assumptions

The financial market

Variable annuities pricing is based on the common pricing literature which assumes the existence of a risk neutral measure $\mathbb{Q}$ under which future cash flows can be valued as their expected discounted values. The existence of such measure implies an arbitrage free financial market. Moreover, the derivative's payoff can be replicated by a self-financing strategy, which allows the insurer to hedge the liabilities.

We assume that the risky asset S_t, which serves as an underlying mutual fund for the variable annuity, follows a Geometric Brownian motion with constant coefficients under $\mathbb{Q}$:

$$dS_t = rS_t dt + \sigma S_t dW_t,$$

where σ is the volatility or the risky asset, r the risk-free rate and W a standard Brownian motion under $\mathbb{Q}$.

The money market evolves with risk-free interest rate, and the numeraire process B_t is given by :

$$dB_t = rB_t dt.$$

Under the risk neutral probability measure, the discounted asset process $B_t^{-1}S_t$ is a martingale.

The mortality assumption

It is common practice among insurers to use deterministic mortality rate to evaluate and replicate their policy pool. We also use this assumption in this chapter by considering future mortality rates as a deterministic curve. Moreover, we make the common assumption that financial markets and biometric events are independent. Let us introduce the mortality notations as :

— x_0 : the policyholder's age at the contract inception.

— q_n : the probability that the policyholder, aged x_0 at inception, dies between time t_{n-1} and t_n.

— p_n : the probability that the policyholder, aged x_0 at inception, is alive at time t_n.

— ω : the limiting age beyond which survival is impossible.

According to the definition, we have $p_n = (1 - q_n)p_{n-1}$, where $n \in \{1, 2, ..., N\}$. From the insurer's perspective, the percentage of active contracts in a large policy pool of policyholders aged $x = x_0 + t_n$ at a given time t_n is thus given by p_n.

The contract state variables

At a given anniversary date t_n, the value of a GMIB contract, purchased by an x_0-year old policyholder at inception, is determined by three main state variables : the account value, the benefit base, and the a two-states variable determining if he is alive or dead at time t_n.

— **Account value** A_t **:** the value of the investment account which is indexed on the asset value S_t, and reduced by withdrawals and fees.

— **Benefit base** G_t **:** also referred as the guarantee account, is an "imaginary" wealth upon which annuities, guaranteed withdrawals and benefits are calculated. However, if the insured wants to lapse the contract, he will not be able to get this wealth.

— **Death Process** I_n **:** a two states variable in $\{0, 1\}$ informing if the policyholder died during $(t_{n-1}, t_n]$, or is still alive at t_n. The death probability in the interval $(t_{n-1}, t_n]$ is given by $q_n = \mathbb{P}(I_n = 0 \mid I_{n-1} = 1)$, which depends on the policyholder's age at inception.

More state variables need to be included if one needs to incorporate stochastic interest rate and/or volatility, take into account taxation or consider different benefit bases, i.e. evolving differently or for different riders.

We restrict our analysis to single premium contracts $A_0 = G_0$, i.e. one premium at inception with no additional contributions. The policyholder can either withdraw money or exercise the income benefit. Withdrawals include "zero" withdrawals, guaranteed ones, i.e. up to a limited amount fixed by the insurer, excess withdrawals, i.e. withdrawals that exceed the guaranteed withdrawal amount, or completely surrender the contract, i.e. lapse.

For the sake of simplicity, we assume that the policyholder can take withdrawals each policy anniversary t_n, and denote by γ_n the withdrawal amount. The income benefit also starts at anniversary years and, in case of a death benefit, the latter is paid out at these dates as well. Thus, the state variables described above may have discontinuities at times $t_1, ..., t_N$. Therefore, for a state variable Y, we distinguish between its value $Y_{t_n^-}$ before and $Y_{t_n^+}$ after events take place at the anniversary date t_n.

Development between two policy years $(t_{n-1}, t_n]$

Assuming that an annual guarantee fee α is continuously charged by the issuer, the value of the account value A_{t_n} evolves as :

$$A_{t_n^-} = A_{t_{n-1}^+} \times \frac{S_{t_n}}{S_{t_{n-1}}} \exp(-\alpha \Delta t), \quad n = 1, 2, ..., N,$$

where $\Delta t = t_n - t_{n-1}$, and S_t follows a geometric Brownian and has the closed formula :

$$S_{t_n} = S_{t_{n-1}} \exp\left(\left(r - \frac{1}{2}\sigma^2\right)\Delta t + \sigma\sqrt{\Delta t} z_n\right),$$

where $z_1, ..., z_N$ are independent and identically distributed standard Normal random variables.

In practice, the guaranteed fee is charged discretely and proportional to the account value that can easily be incorporated into the wealth process. Denoting the discretely charged fee with the annual basis as $\bar{\alpha}$, the wealth process becomes

$$A_{t_n^-} = A_{t_{n-1}^+} \frac{S_{t_{n-1}}}{S_{t_n}} (1 - \bar{\alpha}) \Delta t$$

The benefit base remains constant between two policy years, i.e :

$$G_{t_n^-} = G_{t_{n-1}^+}.$$

Remark 2. *For continuously charged fees, the evolution of the account value can be rewritten in the form of an SDE :*

$$dA_t = (r - \alpha) A_t dt + \sigma A_t dW_t,$$

where W_t *is the risky asset Brownian motion,* σ *its volatility, and* r *is the risk free rate.*

In Retirement Cornerstone ®, some of the fees are actually proportional to the benefit base. We denote by α^A *(resp.* α^G*) fees proportional to the account value (resp. benefit base). In this case, we rewrite the account dynamic between* t_{n-1}^+ *and* t_n^- *as :*

$$dA_t = (r - \alpha^A) A_t dt - \alpha^G G_{t_{n-1}^+} dt + \sigma A_t dW_t, \tag{2.1}$$

Transition at a policy year t_n

As mentioned earlier, the contract events take place at the discrete policy years. In the following, we denote by γ_n^{gua} the guaranteed withdrawal amount, and $\bar{f}_n$ the cash flow at time t_n. The guaranteed withdrawal amount at t_n is typically proportional to the the benefit base at time t_{n-1}^+ (or t_n^-), by a rate η fixed by the insurer at inception, i.e. $\gamma_n^{gua} = \eta G_{t_n^-}$. Each policy year can exhibit the following scenarios :

1. **The insured has died within the previous year** $(t_{n-1}, t_n]$ **:**
   If the insured has died within the previous year and no death benefit has been set in place we have $A_{t_n^+} = 0$, $G_{t_n^+} = 0$, $\gamma_n^{gua} = 0$ and $\bar{f}_n = 0$.
2. **The insured has survived the previous policy year and does not withdraw any money from the account at time** t_n **:**
   Different ratchet and roll-up mechanisms can be applied to the benefit base at t_n, thus changing the value of the guaranteed withdrawal amount. The different parameters develop as follows :
   — Roll-up only : $G_{t_n^+} = (1 + \eta) G_{t_n^-}$

— Ratchet : $G_{t_n^+} = max(G_{t_n^-}, A_{t_n^-})$

— Reset : $G_{t_n^+} = max\big((1+\eta)G_{t_n^-}, A_{t_n^-}\big)$

Here η represents the roll-up rate which determines the quantity credited annually to the benefit base. This quantity is also used to compute the guaranteed withdrawal amount each year by $\gamma_n^{gua} = \eta.G_{t_n^-}$. If no withdrawals are made from the contract, i.e $W_{t_n} = 0$, we have $A_{t_n^+} = A_{t_n^-}$ and the cash flows $\bar{f}_n = 0$.

3. **The insured has survived the previous policy year and at the policy anniversary withdraws an amount within the limits of the guaranteed withdrawal amount :**
Any withdrawal up to the guaranteed annual withdrawal amount reduces the account value by the withdrawn amount. Of course, we do not allow for negative policyholder account values and thus get $A_{t_n^+} = max(0, A_{t_n^-} - \gamma_n)$ and $\bar{f}_n = \gamma_n$. The transformations discussed in 2) occur simultaneously with the withdrawals resulting in :

— Roll-up only : $G_{t_n^+} = (1+\eta)G_{t_n^-} - \gamma_n$.

— Ratchet : $G_{t_n^+} = max(G_{t_n^-} - \gamma_n, A_{t_n^-})$.

— Reset : $G_{t_n^+} = max((1+\eta)G_{t_n^-} - \gamma_n, A_{t_n^-}$.

The guaranteed annual amount γ_n^{gua} needs to be recalculated using the formula presented immediately above. Note that if the annuity owner withdraws the maximum quantity γ_n^{gua}, the level of the benefit base remains stable when the roll-up is taken into account :

$$G_{t_n^+} = (1+\eta)G_{t_n^-} - \gamma_n^{gua} = (1+\eta)G_{t_n^-} - \eta.G_{t_n^-} = G_{t_n^-}.$$

4. **The insured has survived the previous policy year and the policy anniversary, and withdraws an amount exceeding the limit of the withdrawal guarantee :**
In this case the account value is again reduced by the withdrawal amount $A_{t_n^+} = \max(0, A_{t_n^-} - \gamma_n)$. The benefit base as of the last contract anniversary date is reduced pro-rata by the percentage of the excess withdrawal w.r.t the account value, i.e. $\frac{\gamma_n - \gamma_n^{gua}}{A_{t_n^-}} G_{t_n^-}$. Therefore we have :

$$G_{t_n^+} = G_{t_n^-}\Big(1 - \frac{\gamma_n - \gamma_n^{gua}}{A_{t_n^-}}\Big)$$

We then apply the ratchet if there is any, i.e $G_{t_n^+} = max(G_{t_n^+}, A_{t_n^-} - \gamma_n)$.

5. **The insured has survived the previous policy year and decides to activate the GMIB rider :**
In this case, the contract matures and lifetime payments begin the following policy anniversary date taking into account the state of the variables at time t_n. Details on annuitization are given in the following section.

We summarize the previous cases into the following :

— Roll-up only case :

$$\begin{aligned} A_{t_n^+} &= h^A(A_{t_n^-}, G_{t_n^-}, \gamma_n) := \max(0, A_{t_n^-} - \gamma_n) \\ G_{t_n^+} &= h^G(A_{t_n^-}, G_{t_n^-}, \gamma_n) := \begin{cases} \max(0, (1+\eta)G_{t_n^-} - \gamma_n) & \text{if} \quad \gamma_n \le \gamma_n^{gua} \\ G_{t_n^-}\Big(1 - \frac{\gamma_n - \gamma_n^{gua}}{A_{t_n^-}}\Big) & \text{if} \quad \gamma_n > \gamma_n^{gua} \end{cases} \end{aligned} \tag{2.2}$$

— Ratchet only case :

$$\begin{aligned} A_{t_n^+} &= h^A(A_{t_n^-}, G_{t_n^-}, \gamma_n) := max(0, A_{t_n^-} - \gamma_n), \\ G_{t_n^+} &= h^G(A_{t_n^-}, G_{t_n^-}, \gamma_n) := \begin{cases} \max(A_{t_n^-} - \gamma_n, G_{t_n^-} - \gamma_n) & \text{if} \quad \gamma_n \le \gamma_n^{gua} \\ \max(A_{t_n^-} - \gamma_n, G_{t_n^-}\Big(1 - \frac{\gamma_n}{A_{t_n^-}}\Big) & \text{if} \quad \gamma_n > \gamma_n^{gua}. \end{cases} \end{aligned}$$

— Reset (roll-up + ratchet) case :

$$\mathrm{A}_{t_n^+} = h^{\mathrm{A}}(\mathrm{A}_{t_n^-}, \mathrm{G}_{t_n^-}, \gamma_n) := max(0, \mathrm{A}_{t_n^-} - \gamma_n),$$
$$\mathrm{G}_{t_n^+} = h^{\mathrm{G}}(\mathrm{A}_{t_n^-}, \mathrm{G}_{t_n^-}, \gamma_n) := \begin{cases} \max(\mathrm{A}_{t_n^-} - \gamma_n, (1+\eta)\mathrm{G}_{t_n^-} - \gamma_n) & \text{if} \quad \gamma_n \leq \gamma_n^{gua} \\ \max\left(\mathrm{A}_{t_n^-} - \gamma_n, \mathrm{G}_{t_n^-}\left(1 - \frac{\gamma_n - \gamma_n^{gua}}{\mathrm{A}_{t_n^-}}\right)\right) & \text{if} \quad \gamma_n > \gamma_n^{gua}. \end{cases}$$

Remark 3. *The guaranteed rate is usually set equal to the ratchet rate, i.e.* $\gamma_n^{gua} = \eta \mathrm{G}_{t_n^-}$ *at time* t_n. *(ii) The ratchet case can easily be deduced from the reset case by setting the ratchet and guaranteed rate to* 0.

The income and death benefit

At maturity, the holder of a GMIB contract can select to take a lump sum of the account value $\mathrm{A}_{t_\mathrm{N}}$, annuitize this amount at an "actual" annuitization rate or annuitize the benefit base at pre-specified guaranteed annuitization rate. Annuity factors which give the annuitization rates, denoted by $\ddot{a}_{t_\mathrm{N}}^{act}$ for the actual, and $\ddot{a}_{t_\mathrm{N}}^{gua}$ for the guaranteed, are defined as the price of an annuity paying one dollar each year with either a the market's rates curve, or an internal guaranteed rates defined by the insurer. The calculations of the annuity factors takes into account the probability that the insurer is alive in the future. They are given by :

$$\ddot{a}_{t_\mathrm{N}}^{(.)} = \sum_{t_i = t_\mathrm{N}}^{\omega - x_0} p_i e^{-r_{t_i}^{(.)}(t_i - t_\mathrm{N})},$$

where $r^{(.)}$ are risk-free interest rate in case of annuitizing the account value, and based on hypothesis fixed by the insurer in case of annuitizing the benefit base. Therefore, annuitizing the account value is equivalent to a lump sum, and annuitizing a benefit base G is equivalent to the amount $\mathrm{G}\frac{\ddot{a}_{t_\mathrm{N}}^{gua}}{\ddot{a}_{t_\mathrm{N}}^{act}}$.

For GMIB contracts analyzed in this thesis, annuitization is not restricted to the maturity t_N. Indeed, t_N is actually the last anniversary date in which the insured is allowed to annuitize. Typically, the policyholder can exercise his income benefit starting the 10th year of the contract. An annuity factor is then defined for each date $t_n \in \{t_{10}, t_\mathrm{N}\}$. These factors are increasing since an older insurer will likely to have less annuities than a younger one.

Thus, the cash flow of the income benefit, based on a financially rational acting customer, is given by :

$$\mathrm{P}(t_n, \mathrm{A}_{t_n^-}, \mathrm{G}_{t_n^-}) = \max(\mathrm{A}_{t_n^-}, \mathrm{G}_{t_n^-} \frac{\ddot{a}_{t_n}^{gua}}{\ddot{a}_{t_n}^{act}}),$$

otherwise $\mathrm{P}(t_n, \mathrm{A}_{t_n^-}, \mathrm{G}_{t_n^-}) = 0$, P denotes the income benefit, $\mathrm{A}_{t_n^-}$ the level of the account value, and $\mathrm{G}_{t_n^-}$ the level of the benefit base.

The policyholder can subscribe to a GMDB along with the GMIB. In this case, if the policyholder's death occurs before or at the contract maturity, a death benefit is provided to the beneficiaries. Assuming the policyholder dies during $(t_{n-1}, t_n]$, the beneficiaries will receive the amount $\mathrm{D}(t_n, ., .)$ at t_n. There are several types of death benefits. The most famous one is the so-called return of premium death benefit (return of principle for Retirement Cornerstone ®AXA product). In the Retirement Cornerstone ®AXA product, the death benefit can also consist of the greater of the annual ratchet benefit base and the current account value. Insurers typically charge tenth of potential market growth for this

additional rider. In the case of the Retirement Cornerstone ®13, this additional charge is 0.35% for this particular death benefit.

At any case, we resume the death benefit cash flow $D(t_n,.,.)$ at t_n by :

$$D(t_n, A_{t_n^-}, G_{t_n^-}) = \begin{cases} A_0, & \text{return of principal death benfit (type 0),} \\ A_{t_n^-}, & \text{account value death benfit (type 1),} \\ G_{t_n^-}, & \text{benefit base the death benefit (type 2),} \\ \max(A_{t_n^-}, G_{t_n^-}), & \text{greater of the two death benefit (type 3).} \end{cases}$$

Our numerical analyses will be based on type 2, where he benefit base evolves as described in previous section.

2.4 Contract valuation

To model mortality, the standard way is to use official life tables to estimate the death probability $q_n = \mathbb{P}(I_n = 0 \mid I_{n-1} = 1)$ during $(t_{n-1}, t_n]$. They provide annual death probabilities for each age and gender in a given country. Some adjustments can be applied to these tables. In addition to life tables, other approaches can be considered such as the stochastic benchmark Lee-Carter model, see [114], which forecasts the required death probabilities accounting for systematic mortality risk.

For pricing purposes, we consider a pool of policyholders who hold identical contracts and in which each insured has the same age, gender and thus same probabilities of life and death[2]. We assume the number of policyholders to be large enough such that the assumption that deaths occur exactly according to probability q_n is justified. Given this set of conditions, mortality risk is fully diversified.

In the following, we set up the pricing framework of the GMIB contract. In particular, we are interested in the rational policyholder behavior which maximizes the expected value of his future cash flows. We will address a stochastic control problem as formulated in [150].

The stochastic control problem

Let $\boldsymbol{\gamma} = (\gamma_1, ..., \gamma_N)$ be a withdrawal strategy, $\mathbf{G} = (G_{t_0}, ..., G_{t_N})$ the state variable corresponding to the benefit base, $\mathbf{A} = (A_{t_0}, ..., A_{t_N})$ the account value, and $\mathbf{I} = (I_0, ..., I_N)$ the death state. We introduce the state vector before the withdrawal as $X_n = (A_{t_n^-}, G_{t_n^-}, I_n)$ at time t_n^- and $\mathbf{X} = (X_1, ..., X_N)$. The present value of the overall payoff of the GMIB contract is defined as :

$$H_0(\mathbf{X}, \boldsymbol{\gamma}) = B_{0,N} H_N(X_N) + \sum_{n=1}^{N-1} B_{0,n} f_n(X_n, \gamma_n), \tag{2.3}$$

where :

$$H_N(X_N) = P(t_N, A_{t_N^-}, G_{t_N^-}) \times \mathbb{1}_{I_N=1} + D(t_N, A_{t_N^-}, G_{t_N^-}) \times \mathbb{1}_{I_N=0}, \tag{2.4}$$

is the cash flow at maturity and

$$f_n(X_n, \gamma_n) = \bar{f}_n(A_{t_n^-}, G_{t_n^-}, \gamma_n) \times \mathbb{1}_{I_n=1} + D(t_n, A_{t_n^-}, G_{t_n^-}) \times \mathbb{1}_{I_n=0}, \tag{2.5}$$

2. Other criterion can be taken into account to compose an "homogeneous pool". Some Life Tables consider smoking, the policy value etc... However, these criterion are specific to other insurance products. To our best knowledge, the mortality table of variable annuities is restricted to age. The resulting death probability is a weighted average of same age male and female holding the policy.

is the cash flow at time t_n. Here $\mathbb{1}_{(.)}$ is the indicator function, and $B_{i,j}$ is the discount factor from t_j to t_i

$$B_{i,j} = exp\Big(-r(t_j - t_i)\Big),\ t_j > t_i$$

To simplify notations, we drop the mortality state variable I_n when the policyholder is alive, i.e. $I_n = 1$, in the function argument. We define $V(t_n, A, G)$ the price of the contract with a guarantee at the policy year t_n when $A_{t_n^-} = A$, $G_{t_n^-} = G$. We assume that the financial risk can be eliminated via continuous hedging, i.e. complete and frictionless market, and that mortality risk is fully diversified via selling the contract to a large number of insured of the same age. Thus, the average of the contract payoffs of M policyholders $H_0(\mathbf{X}, \boldsymbol{\gamma})$ converges to $\mathbb{E}^{\mathbb{I}}_{t_0}[H_0(\mathbf{X}, \boldsymbol{\gamma})]$ as $M \to \infty$, where $\mathbb{I}$ is the real probability measure corresponding to the mortality process $I_1, I_2, ..., I_N$. Then the price under the given withdrawal strategy $\boldsymbol{\gamma}$ can be calculated as :

$$V(t_0, A_{t_0}, G_{t_0}) = \mathbb{E}^{\mathbb{I},\mathbb{Q}}_{t_0}[H_0(\mathbf{X}, \boldsymbol{\gamma})], \tag{2.6}$$

where $\mathbb{E}^{\mathbb{I},\mathbb{Q}}_{t_0}[H_0(\mathbf{X}, \boldsymbol{\gamma})]$ denotes the expectation w.r.t the state vector $\mathbf{X}$, conditional on information available at time t_0, i.e w.r.t both the financial risky asset process under $\mathbb{Q}$, and the mortality process under the real probability measure $\mathbb{I}$.

Remark 4. *The fair fee $\bar{\alpha} = \alpha^*$ is defined as the fees charged so that the value of the contract at time t_0 is equal to the premium, i.e. $V(0, A_0, G_0) = A_0$. It is important to note that the strategy γ can change for different realizations of underlying wealth process and the control variable γ_n at t_n affects the transition law of the underlying wealth process from t_n to t_{n+1}, i.e calculating the contract price in this case is reduced to solving an optimal stochastic control problem.*

The withdrawal strategy $\boldsymbol{\gamma}$ can depend on the information available at time t_n through the state variable X at t_n and is assumed to be given when the price of the contract is calculated in (2.6). Withdrawal strategies are classified into three categories : *static, optimal* and *suboptimal*.

— **Static case.** Under a static strategy $\boldsymbol{\gamma}$, the policyholder's decisions are deterministic, fixed at the beginning of the contract, and independent the state variable value. Under this strategy, the price of the contract can be calculated as :

$$V(t_0, A_{t_0}, G_{t_0}) = \mathbb{E}^{\mathbb{I},\mathbb{Q}}[H_0(\mathbf{X}, \boldsymbol{\gamma})].$$

— **Optimal case.** Under the optimal withdrawal strategy, the withdrawal amount γ_n depends on the information available at time t_n through the state variable X_n. The optimal strategy is the strategy $\boldsymbol{\gamma}$ under which the contract price is maximized, i.e worst case scenario for the insurer/best case scenario for the insured :

$$\boldsymbol{\gamma}^*(\mathbf{X}) = \underset{\gamma \in \mathscr{A}}{\operatorname{argsup}} \mathbb{E}^{\mathbb{I},\mathbb{Q}}[H_0(\mathbf{X}, \boldsymbol{\gamma})] \tag{2.7}$$

where the supremum is taken over all admissible strategies $\boldsymbol{\gamma}$ and denoted by the set $\mathscr{A}$. That is for each time t_n we have $\gamma_n \in [0, A_{t_n^-}]$.

— **Suboptimal case.** Any other strategy $\boldsymbol{\gamma}$ different from $\boldsymbol{\gamma}^*$ is called suboptimal. It can also depend on the state variable.

In the following we will be interested in the optimal case. Given that the state variable $\mathbf{X} = (X_1, ..., X_N)$ is a Markov process, and the contract payoff is represented by Formula (2.3), the calculation of the contract value under optimal strategy given by Equation (2.7), is brought to a more general problem whereby the policyholder starts at an arbitrary time t_n. This falls within the framework of standard

optimal stochastic control problems for a controlled Markov process. Note that the control variable γ_n depends on the account value A and benefit base G.

Finding the contract value $V(t_n, x)$ at time t_n when $X_n = x$ for $n = N-1, ..., 0$ is done via a backward Bellman equation. Since the account value A evolves between two anniversary dates, and the benefit base is a constant piecewise function (i.e. changes at anniversary dates only), the required backward recursion is written between t_{n+1}^- and t_n^+ as

$$\begin{aligned} V(t_n^+, A, G) &= \mathbb{E}^{\mathbb{I}}\Big[B_{n,n+1}\Big(\mathbb{1}_{I_{n+1}=1} \times \mathbb{E}^{\mathbb{Q}}_{t_n^+}[V(t_{n+1}^-, A_{t_{n+1}^-}, G_{t_{n+1}^-}) \mid A, G] \\ &\qquad + \mathbb{1}_{I_{n+1}=0} \times \mathbb{E}^{\mathbb{Q}}_{t_n^+}[D(t_{n+1}^-, A_{t_{n+1}^-}, G_{t_{n+1}^-}) \mid A, G]\Big)\Big] \\ &= (1-q_{n+1})\mathbb{E}^{\mathbb{Q}}_{t_n^+}[V(t_{n+1}^-, A_{t_{n+1}^-}, G_{t_{n+1}^-}) \mid A, G] + q_{n+1}\mathbb{E}^{\mathbb{Q}}_{t_n^+}[D(t_{n+1}^-, A_{t_{n+1}^-}, G_{t_{n+1}^-}) \mid A, G], \end{aligned}$$

with jump condition

$$V(t_n^-, A, G) = \max_{\gamma_n \in \mathscr{A}_n}\Big(\bar{f}_n(A, G, \gamma_n) + V(t_n^+, h^A(A, G, \gamma_n), h^G(A, G, \gamma_n))\Big)$$

The recursion starts from the maturity condition $V(t_N^-, A, G) = P(t_N^-, A, G)$ goes backwards for $n = N-1, N-2, ..., 0$.

Remark 5. *Given that the mortality and financial asset processes are assumed independent, and the withdrawal decision does not affect the mortality process, we have :*

$$\sup_{\boldsymbol{\gamma}} \mathbb{E}^{\mathbb{Q},\mathbb{I}}_{t_0}[H_0(\mathbf{X}, \boldsymbol{\gamma})] = \sup_{\boldsymbol{\gamma}} \mathbb{E}^{\mathbb{Q}}_{t_0}[\mathbb{E}^{\mathbb{I}}_{t_0}[H_0(\mathbf{X}, \boldsymbol{\gamma})]].$$

One can calculate the expected value of the payoff (2.3) w.r.t the mortality process :

$$\tilde{H}_0(\mathbf{A}, \mathbf{G}) = \mathbb{E}^{\mathbb{I}}_{t_0}[H_0(\mathbf{X}, \boldsymbol{\gamma})],$$

and then calculate the price under the given strategy $\mathbb{E}^{\mathbb{Q}}_{t_0}[\tilde{H}_0(\mathbf{A}, \mathbf{G})]$*, or under the optimal strategy* $\sup_{\boldsymbol{\gamma}} \mathbb{E}^{\mathbb{Q}}_{t_0}[\tilde{H}_0(\mathbf{A}, \mathbf{G})]$.

Therefore we have :

$$\begin{aligned} \tilde{H}_0(\mathbf{A}, \mathbf{G}) = B_{0,N}\Big(&P(t_N, A_{t_N^-}, G_{t_N^-}) \times \mathbb{E}^{\mathbb{I}}_{t_0}[\mathbb{1}_{I_N=1}] + D(t_N, A_{t_N^-}, G_{t_N^-}) \times \mathbb{E}^{\mathbb{I}}_{t_0}[\mathbb{1}_{I_N=0}]\Big) \\ &+ \sum_{n=1}^{N-1} B_{0,n}\Big(\bar{f}_n(A_{t_n^-}, G_{t_n^-}, \gamma_n) \times \mathbb{E}^{\mathbb{I}}_{t_0}[\mathbb{1}_{I_n=1}] + D(t_n, A_{t_n^-}, G_{t_n^-})\Big) \end{aligned}$$

Moreover, since $\mathbb{E}^{\mathbb{I}}_{t_0}[\mathbb{1}_{I_n} = 1] = \mathbb{P}(\tau > t_n \mid \tau > t_0) = p_n$ *,and* $\mathbb{E}^{\mathbb{I}}_{t_0}[\mathbb{1}_{I_n} = 0] = \mathbb{P}(t_{n-1} < \tau < t_n \mid \tau > t_0) = p_{n-1}q_n$ *for random death time* τ*, i.e.* $p_n = p_{n-1}(1-q_n)$*, we can rewrite* $\tilde{H}_0(\mathbf{A}, \mathbf{G})$ *as the following :*

$$\begin{aligned} \tilde{H}_0(\mathbf{A}, \mathbf{G}) = B_{0,N}\Big(&p_N P(t_N, A_{t_N^-}, G_{t_N^-}) + q_N p_{N-1} D(t_N, A_{t_N^-}, G_{t_N^-})\Big) \\ &+ \sum_{n=1}^{N-1}\Big(p_n \bar{f}_n(A_{t_n^-}, G_{t_n^-}, \gamma_n) + p_{n-1}q_n D(t_n, A_{t_n^-}, G_{t_n^-})\Big). \end{aligned} \tag{2.8}$$

Note that, previously we defined $q_n = \mathbb{P}(t_{n-1} < \tau \leq t_n \mid \tau > t_{n-1})$.

The payoff (2.3) has the same general form as the payoff (2.8). Thus, the optimal stochastic control

problem $\Phi(t_0, \mathrm{A}_0, \mathrm{G}_0) = \sup_{\boldsymbol{\gamma}} \mathbb{E}[\tilde{\mathrm{H}}_0(\mathbf{A}, \mathbf{G})]$ *can be solved using Bellman equation. We describe the optimization problem at each policy anniversary date recursively by the two following equations*

$$\Phi(t_n^+, \mathrm{A}, \mathrm{G}) = \mathbb{E}^{\mathbb{Q}}_{t_n^+}\left[\mathrm{B}_{n,n+1}\Phi(t_{n+1}^-, \mathrm{A}_{t_{n+1}^-}, \mathrm{G}_{t_{n+1}^-}) \mid \mathrm{A}, \mathrm{G}\right], \tag{2.9}$$

and

$$\begin{aligned}\Phi(t_n^-, \mathrm{A}, \mathrm{G}) = \max_{\gamma_n \in \mathscr{A}_n} \Big(& p_n \bar{f}_n(\mathrm{A}, \mathrm{G}, \gamma_n) + p_{n-1} q_n \mathrm{D}(t_n, \mathrm{A}, \mathrm{G}) \\ & + \Phi(t_n^+, h^{\mathrm{A}}(\mathrm{A}, \mathrm{G}, \gamma_n), h^{\mathrm{G}}(\mathrm{A}, \mathrm{G}, \gamma_n))\Big),\end{aligned} \tag{2.10}$$

for $n = \mathrm{N}-1, \mathrm{N}-2, ..., 0$*, starting from the final condition :*

$$\Phi(t_{\mathrm{N}}^-, \mathrm{A}, \mathrm{G}) = p_{\mathrm{N}} \mathrm{P}(t_{\mathrm{N}}^-, \mathrm{A}, \mathrm{G}) + p_{\mathrm{N}-1} q_{\mathrm{N}} \mathrm{D}(t_{\mathrm{N}}^-, \mathrm{A}, \mathrm{G}). \tag{2.11}$$

As a consequence, the recursion leads to the same solution $\Phi(t_0, \mathrm{A}, \mathrm{G}) = \mathrm{V}(t_0, \mathrm{A}, \mathrm{G})$*, and the same optimal strategy* $\boldsymbol{\gamma}$*. Moreover, for each* t_n *we have* $\Phi(t_n, \mathrm{A}, \mathrm{G}) = p_n \mathrm{V}(t_n, \mathrm{A}, \mathrm{G}) + p_{n-1} q_n \mathrm{D}(t_n, \mathrm{A}, \mathrm{G})$*.*

2.4.1 Numerical scheme for the discrete withdrawal model

Realistic VA riders with discrete events such as ratchets, bonuses as set-up options and optimal withdrawals have no closed form solutions. Their fair price needs to be calculated numerically, even for a standard Brownian motion with constant interest rates and volatility.

The numerical solution of the backward recursion (2.9)-(2.10) is accomplished using PDEs, direct integration or regression-type Monte Carlo methods. Under the static strategy, one can always use standard Monte-Carlo to simulate state variables forward in time till the contract maturity or the policyholder death and average the payoff cash flows over a large number of independent realizations.

In the case of discrete withdrawal, following the procedure of deriving the Hamilton-Jacobi-Bellman (HJB) equations in stochastic control problems, the value of the annuity under optimal withdrawal is found to be governed by a one-dimensional PDE, similar to the Black-Scholes equation, with jump conditions at each withdrawing date to link the prices at the adjacent periods.

In the following, we provide detailed description of the algorithm used to compute the fair value of the VA riders and the optimal strategy.

General algorithm

The algorithm starts from a final condition for the contract value at t_{N}^-. Subsequently, solving the PDE gives solution for the contract value at $t_{\mathrm{N}-1}^+$. The PDE used to compute the expected value (2.9) under the assumed risk-neutral process for the risky asset S_t is easily derived using Feynman-Kac theorem. When the risky asset follows a geometric Brownian motion process, the governing PDE right after a withdrawal decision t_n^+ to right before the following one t_{n+1}^- for $n = \mathrm{N}-1, \mathrm{N}-2, ..., 0$ is expressed as the following

$$\partial_t \Phi + \frac{1}{2}\sigma^2 \mathrm{A}^2 \partial_{\mathrm{AA}} \Phi + (r - \alpha^{\mathrm{A}})\mathrm{A}_t \partial_{\mathrm{A}} \Phi - \alpha^{\mathrm{G}} \mathrm{G}_{t_n^+} \partial_{\mathrm{A}} \Phi - r\Phi = 0, \tag{2.12}$$

to which we add boundary conditions given in the next section.
Note that the benefit base changes only at the anniversary dates and is a constant parameter between

two anniversary dates. PDE (2.12) is solved using the Crank-Nicolson finite differences methods. [60], [99] used the scheme for pricing GMWB with discrete optimal withdrawals. Of course, if the volatility and/or interest rates are stochastic, then one needs to add extra dimensions to the PDE.

Applying the jump condition (2.10) to the solution at t_{N-1}^+ we obtain the solution at t_{N-1}^- from which further backward time stepping gives us solution at t_{N-2}^+, and so on. The numerical algorithm takes the following key steps :

1. Generate a finite grid for the account value A and benefit base G, i.e. $A_0 < A_1 < ... < A_J$ and $0 = G_0 < G_1 < ... < G_K$.
2. At t_N, define the final condition for each note point (A_j, G_k), $j = 1,2,...,J$ and $k = 1,2,...,K$ to get $\Phi(t_n^-, A, G)$ and the boundary conditions (2.17)-(2.18) for A_{min} and A_{max} for each potential $G_{k\in\{1,2,...,K\}}$.
3. For each potential benefit G_k, $k = 1,2,...,K$, solve the PDE using the Crank-Nicolson finite differences scheme to obtain $\Phi(t_{N-1}^+, A, G)$.
4. Apply the jump condition (2.10) by performing a linear search of the withdrawal amount γ_{N-1}^* that gives the maximum $\Phi(t_{N-1}^-, A, G)$. In general, this involves a two-dimensional interpolation in (A, G) since the $h^G(A, G, .)$ and $h^G(A, G, .)$ do not necessarily fall in the grid nodes.
5. Repeat (3) and (4) for $t = t_{N-2}, t_{N-3}, ..., t_1$.
6. Evaluate Equation (2.12) for the backward time step t_1 to t_0 to obtain solution $\Phi(t_0, A, G)$ at A_0 and G_0.

We can add more complexity to the model, for example by incorporating stochastic interest rates or stochastic volatility. In this case, the dimension of the pricing PDE (2.12). We can also add more constant path-wise state variables that evolve only at the anniversary (tax base, extra benefit base, etc...), which will affect only the jump condition, i.e. the search of the optimum will have to be performed based on the new variables as well.

The finite difference scheme will be discussed in more detail in the next section.

Description of the finite differences scheme

Recall that the value function Φ satisfies the following recursion

$$\Phi(t_n^+, A, G) = \mathbb{E}_{t_n}^{\mathbb{Q}}\left[B_{n,n+1}\Phi(t_{n+1}^-, A_{t_{n+1}^-}, G_{t_{n+1}^-}) \mid A, G\right], \tag{2.13}$$

$$\Phi(t_n^-, A, G) = \max_{\gamma_n \in \mathcal{A}_n}\Big(p_n \bar{f}_n(A, G, \gamma_n) + p_{n-1} q_n D(t_n^+, A, G) + \Phi(t_n^+, h^A(A, G, \gamma_n), h^G(A, G, \gamma_n))\Big). \tag{2.14}$$

Within each time interval (t_{n-1}, t_n), only the account value varies since all the benefit bases, death and life, remain constant. Thus, for $t \in (t_n^+, t_{n+1}^-]$, the annuity value $\Phi(t, A, G)$ solves the following linear PDE for each fixed value of the benefit base G

$$\partial_t \Phi + \mathcal{L}\Phi = 0, \tag{2.15}$$

where the operator $\mathcal{L}$ is

$$\mathcal{L}\Phi = \frac{1}{2}\sigma^2 A^2 \partial_{AA}\Phi + (r - \alpha^A) A \partial_A \Phi - \alpha^G G \partial_A \Phi - r\Phi. \tag{2.16}$$

Localization and boundary conditions

Equation (2.15) is originally posed on the domain $(t, \mathrm{A}) \in [0,\mathrm{T}] \times [0,\infty)$. For computational purposes, and because asset prices are finite and so is the account value, one needs to localize this domain to $[0,\mathrm{T}] \times [0,\mathrm{A}_{\max}]$ where $\mathrm{A}_{\max}$ is large enough not to be attained by the account value during the lifetime of the annuity. Thus, we need to add complementary boundary conditions. We consider that we are between two anniversary dates t_n^+ and t_{n+1}^- backwards.

— When $\mathrm{A} = 0$, the policyholder has no longer the possibility to make any withdrawal from his account. However, if the IB election is possible, then the income period begins, given the policyholder is alive, and the death benefit is activated if he is dead at t_{n+1}. Since the account value is equal to zero, then the annuitization will be indexed on the benefit base. Therefore, we have :

$$\Phi(t,0,\mathrm{G}) = e^{-r(t_{n+1}-t)}\Big(p_{n+1}\frac{\ddot{a}^{gua}_{t_{n+1}}}{\ddot{a}^{act}_{t_{n+1}}}\mathrm{G} + p_n q_{n+1}\mathrm{D}(t_{n+1}^-,0,\mathrm{G})\Big). \tag{2.17}$$

— When $\mathrm{A} = \mathrm{A}_{\max}$, we consider retrieving all the cash more interesting than any other strategy if the policyholder is alive. If he dies, the death benefit will be activated. Therefore, the Dirichlet boundary condition for this case is

$$\Phi(t,\mathrm{A}_{\max},\mathrm{G}) = e^{-r(t_{n+1}-t)}\Big(p_{n+1}\mathrm{A}_{\max} + p_n q_{n+1}\mathrm{A}_{\max}\Big) = e^{-r(t_{n+1}-t)}p_n\mathrm{A}_{\max}. \tag{2.18}$$

Let us define the solution domains

$$\bar{\Omega}_n = [t_{n-1}^+, t_n^-] \times [0,\mathrm{A}_{\max}]$$
$$\bar{\Omega} = \bigcup_{t_n}[t_{n-1}^+, t_n^-] \times [0,\mathrm{A}_{\max}].$$

The pricing problem for the GMIB variable annuity combined with DB under the discrete withdrawal scenario is then achieved in $\bar{\Omega}$ as follows : within each set $\bar{\Omega}_n$, $n = 1,...,\mathrm{N}-1$, the solution to the problem is the viscosity solution of a decoupled set of linear PDEs (2.12) with final condition (2.11) and boundary conditions (2.17)-(2.18) computed from the nonlinear algebraic Equation (2.14).

Construction of the scheme

Let $(\mathrm{A}_0,\mathrm{A}_1,...,\mathrm{A}_\mathrm{J})$ be the equally spaced grid in the direction of the account value with $\mathrm{A}_0 = 0$ and $\mathrm{A}_\mathrm{J} = \mathrm{A}^{\max}$. Analogously $(\mathrm{G}_0,...,\mathrm{G}_\mathrm{K})$ is an equally spaced grid for the benefit base with $\mathrm{G}_0 = 0$ and $\mathrm{G}_\mathrm{K} = \mathrm{G}^{\max} = \mathrm{A}^{\max}$. The spacial steps for both variables are considered to be equal. That is :

$$\Delta\mathrm{A} = \Delta\mathrm{G},$$

where $\Delta\mathrm{A} = \frac{\mathrm{A}_{\max}-\mathrm{A}_0}{\mathrm{J}}$ and $\Delta\mathrm{G} = \frac{\mathrm{G}_{\max}-\mathrm{G}_0}{\mathrm{K}}$.

Hence, $\mathrm{A}_j = j\Delta\mathrm{A}$ and $\mathrm{G}_k = k\Delta\mathrm{G}$, $\forall j,k$. The discrete time steps are denoted by $n\Delta t$ for $n = 1,...,\mathrm{N}$ where $\mathrm{T} = \mathrm{N}\Delta t$. Since, in our analysis, we consider that events occur only at anniversary dates which are yearly, $\Delta t = 1$ and each time t_n coincides with the discrete time step $t_n = n$.

The numerical procedure to solve the approximation in (2.15) is the standard finite difference approach. We employ the two-level implicit finite difference scheme to discretize the differential term $\mathscr{L}\Phi$ as given in (2.16). Let $\mathscr{L}_h\Phi^n_{j,k}$ denote the discrete value of the differential operator at the node $(\mathrm{A}_j,\mathrm{G}_k,t_n)$. The approximation is then given by

$$\mathscr{L}_h\Phi^n_{j,k} = \frac{\sigma^2}{2}\mathrm{A}_j^2\frac{\Phi^n_{j+1,k}-\Phi^n_{j,k}+\Phi^n_{j-1,k}}{\Delta\mathrm{A}^2} + \{(r-\alpha^{\mathrm{A}})\mathrm{A}_j - \alpha^{\mathrm{G}}\mathrm{G}_k\}\frac{\Phi^n_{j+1,k}-\Phi^n_{j-1,k}}{2\Delta\mathrm{A}} - r\Phi^n_{j,k}$$

The general theta-scheme for solving Equation (2.15) is given by

$$\frac{\Phi_{j,k}^{n+1}-\Phi_{j,k}^{n}}{\Delta t}=\theta\mathscr{L}\Phi_{j,k}^{n+1}+(1-\theta)\mathscr{L}_h\Phi_{j,k}^{n}$$

where θ is a weighting factor, $0<\theta\leq 0$. For $\theta=0$, this scheme is the explicit scheme, whereas $\theta=1$ corresponds to the implicit one. The error of the previous cases is of $O(\Delta A^2,\Delta t)$. The explicit scheme has stability issues while the implicit scheme is absolutely stable. The most popular scheme for approximating the solution of the Black-Scholes equation is the Crank-Nicolson scheme obtained for $\theta=1/2$. The latter is shown to be unconditionally stable and $O(\Delta A^2,\Delta t^2)$ convergent, see [68]. In particular, this scheme is used in [60] to solve the optimal pricing problem of the GMWB rider with rational behavior.

The discretization w.r.t the benefit base is not important here. However, since the PDE is solved backwards between t_{n+1}^- and t_n^+, we need to divide this time period, i.e. the period between two consecutive withdrawal dates, into finer time steps for a good accuracy due to the finite difference approximation to the partial derivatives.

Applying the jump condition

Recall that changes in the benefit base only occur at withdrawal dates. After withdrawing the amount γ_n at time t_n, the account value changes from $A_{t_n^-}$ to $A_{t_n^+}=h^A(A,G,\gamma_n)$, and the benefit base drops from $G_{t_n^-}$ to $G_{t_n^+}=h^G(A,G,\gamma_n)$. The jump condition of $\Phi(t_n,A,G)$ across t_n is given by

$$\begin{aligned}\Phi(t_n^-,A,G)=\max_{0\leq\gamma_n\leq A}\Big(&\Phi(t_n^+,h^A(A,G,\gamma_n),h^G(A,G,\gamma_n))\\&+p_n\bar{f}(A,G,\gamma_n)+p_{n-1}q_nD(t_n^-,A,G)\Big)\end{aligned}\tag{2.19}$$

For the optimal strategy, the withdrawal amount γ_n is chosen under the restriction $0\leq\gamma_n\leq A$ to maximize the value of $\Phi(t_n^-,A,G)$ in Equation (2.19).

The application of the jump condition decreases the account value and benefit base. For each G_j, a continuous solution from PDE (2.15) is associated. We can restrict the possible values for the withdrawal amount to multiples of ΔA. This implies, for a given account value A_j at time t_n^-, the withdrawal amount γ takes j possible values : $\gamma=A_j-A_l$, $l=1,2,...,j$. However, numerical tests showed that a finer grid is preferable for the withdrawal amount. Therefore, it is not guaranteed that the account value, nor the benefit base after the withdrawal, $A_{t_n^+}$and $G_{t_n^+}$, fall within their respective grid nodes. To solve this issue, a two-dimensional interpolation is required. In this work we adopted a bi-linear interpolation.

Suppose the jump condition requires the value $\Phi(.,A,G)$ at the point (A,G) located inside a grid $A_i\leq A\leq A_{i+1}$ and $G_j\leq G\leq G_{j+1}$, then the interpolation is performed as the following :

$$\Phi(.,A,G)\approx\frac{G_{i+1}-G}{G_{i+1}-G_i}\Phi(.,A,G_j)+\frac{G-G_i}{G_{i+1}-G_i}\Phi(.,A,G_j),\tag{2.20}$$

where :

$$\begin{aligned}\Phi(.,A,G_j)&\approx\frac{A_{i+1}-A}{A_{i+1}-A_i}\Phi(.,A_i,G_j)+\frac{A-A_i}{A_{i+1}-A_i}\Phi(.,A_{i+1},G_j),\\\Phi(.,A,G_j)&\approx\frac{A_{i+1}-A}{A_{i+1}-A_i}\Phi(.,A_i,G_{j+1})+\frac{A-A_i}{A_{i+1}-A_i}\Phi(.,A_{i+1},G_{j+1}).\end{aligned}$$

At last, the jump condition is achieved through combining (2.19) and (2.20) to find the optimal withdrawal and maximize the function Φ.

Similarity and dimension reduction

An important feature of the contract value is that it exhibits good scaling properties in the Black-Scholes case. We can easily verify that the solution $\Phi(t,A,G)$ of PDE (2.15) with boundary conditions (2.17) and event conditions (2.18) verifies

$$\Phi(t,\xi A,\xi G) = \xi\Phi(t,A,G)$$

for any scalar $\xi > 0$. Therefore, choosing $\xi = 1/G$ we obtain

$$\Phi(t,A,G) = G\Phi(t,\frac{A}{G},1) = G\phi(t,\tilde{A}),$$

where $\tilde{A} = \frac{A}{G}$. It means that we need only solve the corresponding equations for the one-dimensional function ϕ defined in the following :

— Between two consecutive withdrawal dates (t_{n-1}, t_n) ϕ follows the PDE

$$\partial_t\phi + \frac{1}{2}\sigma^2\tilde{A}^2\partial_{\tilde{A}\tilde{A}}\phi + (r-\alpha^A)\tilde{A}\partial_{\tilde{A}}\phi - \alpha^G\partial_{\tilde{A}}\phi - r\phi = 0 \tag{2.21}$$

— At the anniversary dates t_n, the jump condition is explicitly expressed as the following :

$$\phi(t_n^-,\tilde{A}_{t_n^-},\tilde{\gamma}_n) = \max_{\gamma}\Big(h_1(\tilde{\gamma})\phi(t_n^+,h_2(\tilde{A}_{t_n^+},\tilde{\gamma})) + p_n\tilde{\gamma}_n + p_{n-1}q_nD(t_n^+,\tilde{A}_{t_n^+},1)\Big)$$

where the functions h_1 and h_2 are a reduced version of the account value and benefit base evolution at the anniversary dates, and are defined according to different cases :

1. Roll-up only case :

$$h_1(\tilde{\gamma}) = \begin{cases} 1+\eta-\tilde{\gamma}_n & \text{if} \quad \tilde{\gamma}_n \le \eta \\ 1-\frac{\tilde{\gamma}_n-\eta}{\tilde{A}} & \text{if} \quad \tilde{A} \ge \tilde{\gamma}_n > \eta \end{cases}$$

$$h_2(\tilde{\gamma}) = \begin{cases} \frac{\tilde{A}-\tilde{\gamma}_n}{1+\eta-\tilde{\gamma}_n} & \text{if} \quad \tilde{\gamma}_n \le \eta \\ \frac{\tilde{A}-\tilde{\gamma}_n}{1-\frac{\tilde{\gamma}-\eta}{\tilde{A}}} & \text{if} \quad \tilde{A} \ge \tilde{\gamma}_n > \eta \end{cases}$$

2. Ratchet only case :

$$h_1(\tilde{\gamma}) = \max(\tilde{A}-\tilde{\gamma}_n, 1-\frac{\tilde{\gamma}_n}{\tilde{A}})$$
$$h_2(\tilde{\gamma}) = \min(1,\tilde{A})$$

3. Reset case (Roll-up + ratchet) :

$$h_1(\tilde{\gamma}) = \begin{cases} \max(\tilde{A}-\tilde{\gamma}_n, 1+\eta-\tilde{\gamma}_n) & \text{if} \quad \tilde{\gamma}_n \le \eta \\ \max(\tilde{A}-\tilde{\gamma}_n, 1-\frac{\tilde{\gamma}_n-\eta}{\tilde{A}}) & \text{if} \quad \tilde{A} \ge \tilde{\gamma}_n > \eta \end{cases}$$

$$h_2(\tilde{\gamma}) = \begin{cases} \min(1, \frac{\tilde{A}-\tilde{\gamma}_n}{1+\eta-\tilde{\gamma}_n}) & \text{if} \quad \tilde{\gamma}_n \le \eta \\ \min\left(1, \frac{\tilde{A}-\tilde{\gamma}_n}{1-\frac{\tilde{\gamma}-\eta}{\tilde{A}}}\right) & \text{if} \quad \tilde{A} \ge \tilde{\gamma}_n > \eta \end{cases}$$

It can be easily verified that PDE (2.21) does not depend on the benefit base since the latter is constant between two consecutive withdrawal dates t_{n-1}^+ and t_n^-. Therefore, the resolution of the two-dimensional problem can be reduced into a one-dimensional problem. This is particularly useful when adding more stochastic variables like stochastic volatility and/or interest rates.

2.5 Results

The main goal of this study is to assess the behavior risk of a given GMIB product, in case policyholders maximize their expected cash flows. Through optimal withdrawal amounts, or IB election, the insurer is concerned that his product may lead to "undesirable" policyholders behaviors. Given a state of the universe in a future time, and a set of up-front fixed hypotheses (management and guarantee fees, interest rates and volatility), the optimal framework allows us to predict these behaviors through the stochastic control problem detailed in previous sections. As of the product, we consider Retirement Cornerstone©commercialized by AXA U.S. The product hypotheses usually change to account for a new financial environment or customers needs.

We choose two close variations of the product launched in the last decade, which we call them respectively, Product A and Product B. These are two GMIB products to which a death benefit (DB) can be added, i.e. Product A-DB and Product B-DB. The parameters values are given in Table 2.5 :

Parameters	Product A	Product B
Policyholders initial age x_0	60	60
First date for IB election	10th	10th
Last age for IB election	85th	95th
Last age for DB if any	95	95
Interest rate r	2% and 4%	2% and 4%
Volatility σ	20%	20%
Roll-up rate η	6%	$r+1\%$
Initial premium A_0	\$100,000	\$100,000
Total fees $\bar{\alpha} = \alpha^A + \alpha^G$	3.5%	3.5%

TABLEAU 2.5 – Model parameters

We compare these variations for roll-up and reset, with and without death benefit. We conduct the experiments to illustrate the following :

— Policyholders rational behavior for Products A, B, A-DB and B-DB based on the two dimension approach, giving the withdrawal surface as a function of the account value A and benefit base G, for two different interest rate values, in four different periods in time. We also give withdrawals as a function of the moneyness A/G based on the dimension reduction approach.

— Policyholders rational behavior for different values of the volatility $\sigma = \{10\%, 20\%, 30\%, 40\%\}$.

— The contract initial value as a function of interest rates for Product A-DB.

— The contract initial value as a function of the total fees.

Overview of the policyholder behavior

Roll-up only case

First, in Figure 2.2, we present the withdrawal amount surface as a function of the account value A and benefit base G, for Products A and A-DB with roll-up only for fixed times $t = 3, 13, 23$ and $r = 2\%, 4\%$.

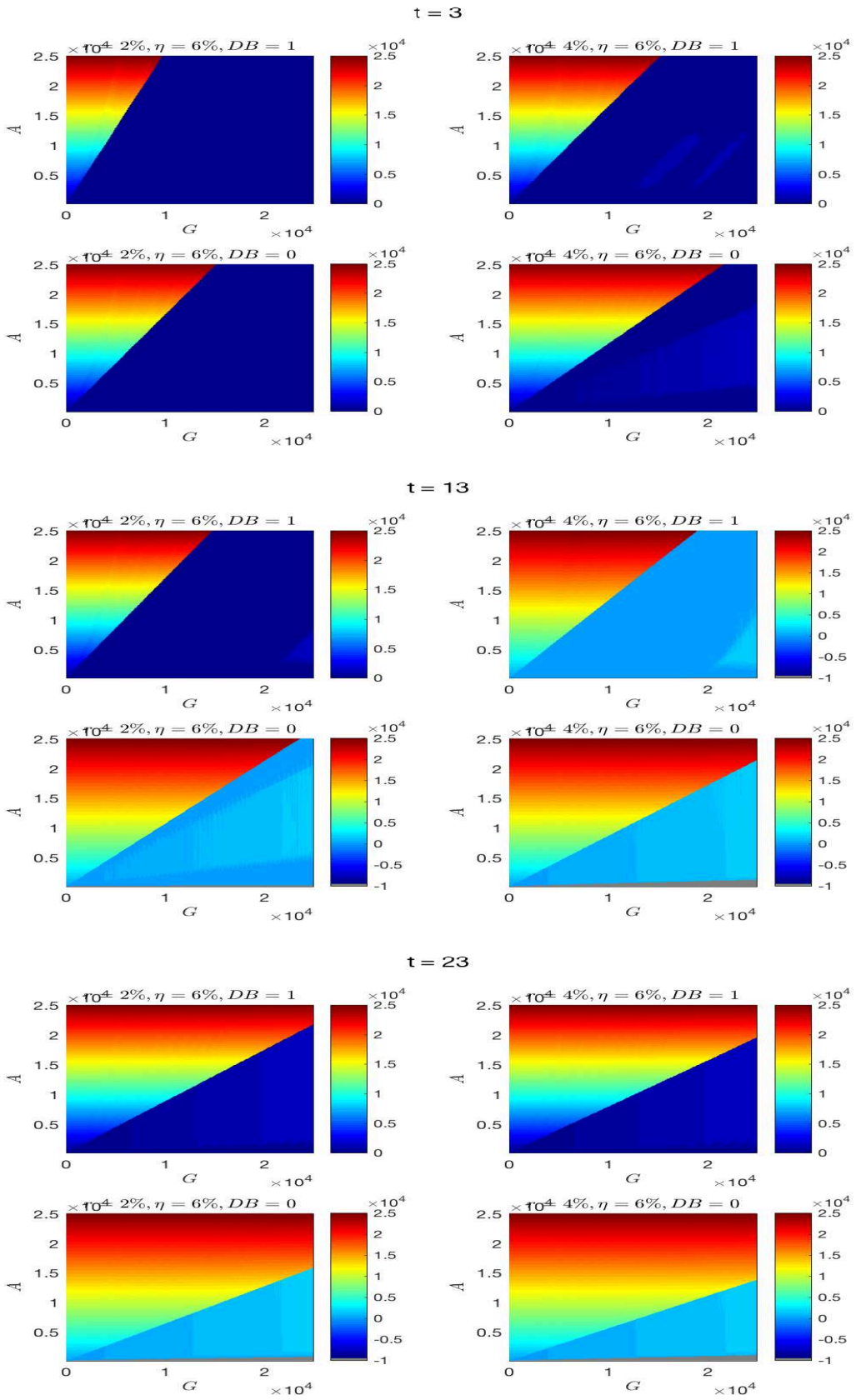


FIGURE 2.2 – Policyholder optimal withdrawal amount as a function of the account value A and benefit base G for Product A and A-DB

As a first remark, the guaranteed account must be higher than the account value for the policyholder to stay in the contract. Depending on the moneyness, he or she can choose not to withdraw, or withdraw the guaranteed amount. The IB election (using the guaranteed account) is left at maturity or for very small account values. These finding are actually confirmed in Figure 2.3 where we give the withdrawal strategy as a function of time t and moneyness A/G based on the dimension reduction scheme.

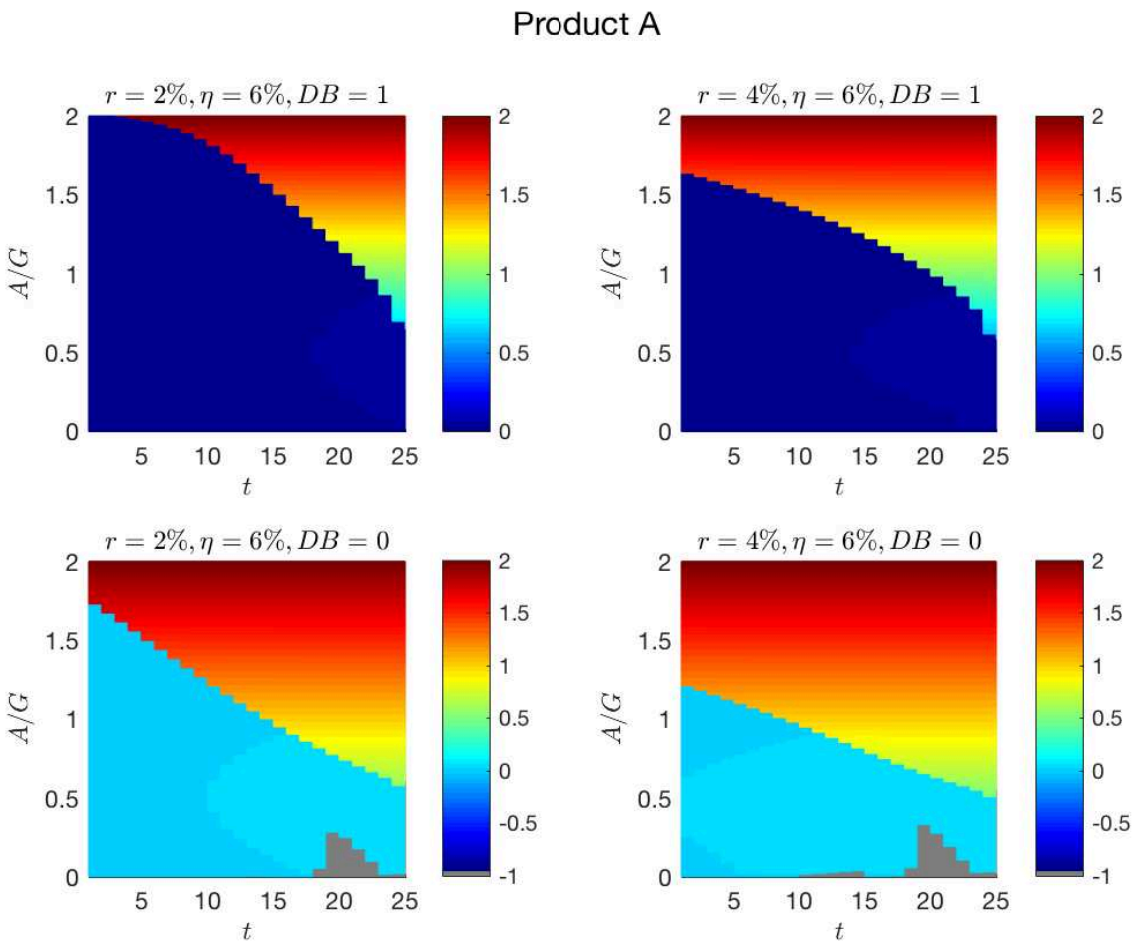


FIGURE 2.3 – The withdrawal strategy $\tilde{\gamma} = \frac{\gamma}{G}$ as a function of time t and moneyness A/G

Based on Figures 2.2-2.3, we can point out the following remarks :

— There is a wider range of guaranteed withdrawals for $r = 4\%$ compared to $r = 2\%$. It tells us that for a roll-up rate as high as 6%, the policyholder tends to wait for the benefit base to increase at this rate in a low interest rates environment instead of draining the subaccount and guaranteed account.

— The death benefit increases the expected cash flows in the future, which is also a motif for the policyholder to wait.

— The IB election indexed on the guaranteed account is only expected to happen in the absence of a death benefit. Even in such case, it only takes place closer to the maturity of the product and for small account values. Actually, the ratio $\frac{\ddot{a}^{act}}{\ddot{a}^{gua}}$ is not very favorable for the insured, and he or she would rather start withdrawing the guaranteed amount few years earlier.

— Actually, fees are quite high so the account value usually drops quickly. This restricts the analysis to A/G relatively small.

Note that reducing the dimension allows to increase the speed of calculations.

Product A was launched in a period where interest rates where around 4% which justifies the choice of a 6% roll-up rate. Given the behavior expressed in the Figures 2.3, the actual interest rate level may seem quite high and the product more interesting for the insured than the insurer.

Later, Product B was launched with reconsidered assumptions. The roll-up rate becomes indexed on the short term interest rates with a spread of 1%. The insurance company attracted the costumer by setting a longer limiting age to annuitize (until the policyholder's 95th anniversary). In Figure 2.4 we give the policyholder behavior in time as a function of the moneyness.

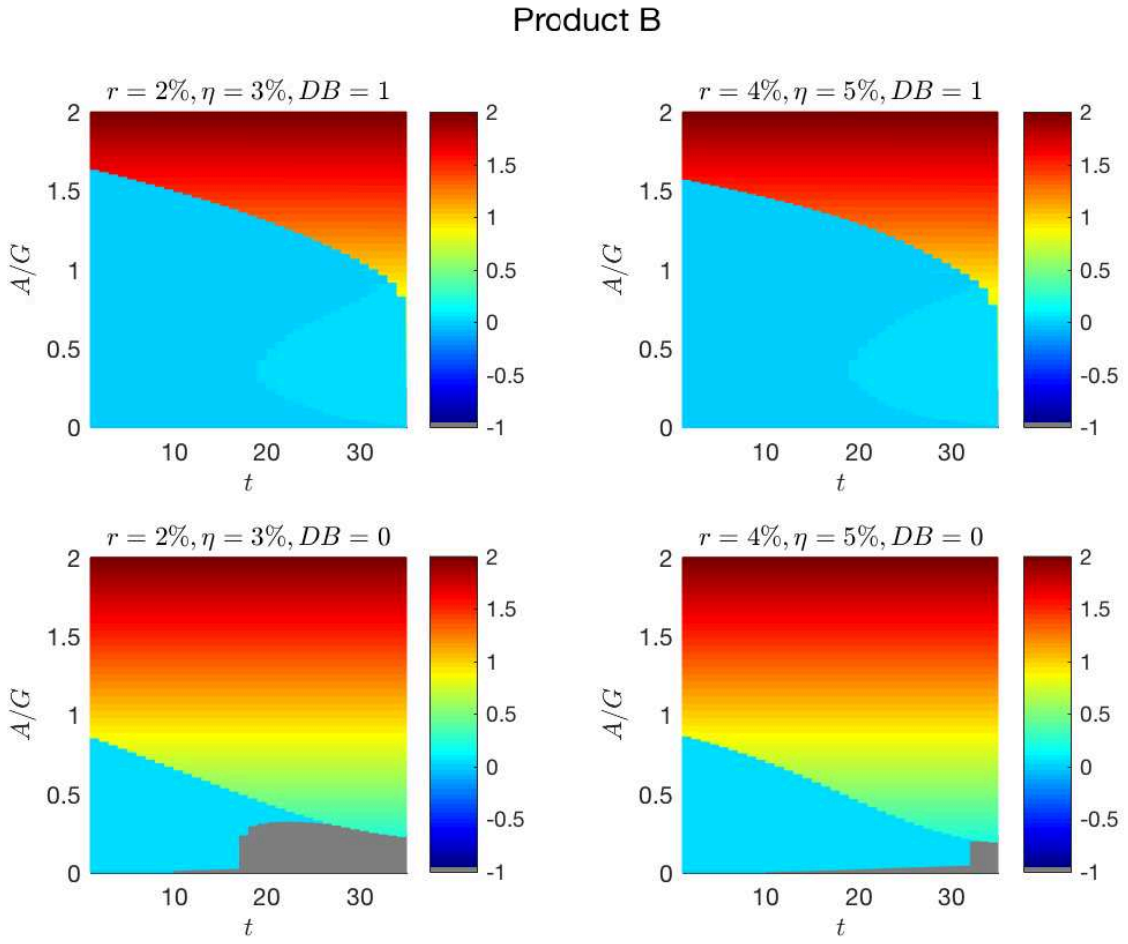


FIGURE 2.4 – The withdrawal strategy as a function of time t and moneyness A/G

While the behavior remains very close, we can however notice that in the absence of the death benefit, partial withdrawals disappear, the likelihood of the IB election is higher for low interest rate, and left until the last years for higher interest rate. Moreover, the likelihood of recovering the account value

is higher than in Product A. In the presence of a death benefit withdrawing the guarantee becomes interesting around the 20th anniversary of the contract, while an IB election is exercised at maturity.

Reset case

Ratchets allow for the benefit base to be set at the account value level when the latter is higher than the previous benefit base level. Combined with the roll-up, we have the reset which is a very attractive feature for policyholders who are interested in the stochastic performance of stock markets, but at the same time want to have a guaranteed minimum performance. We present in Figures 2.5 and 2.6 the results related to products A and B with and without death benefit for the previous interest rates and roll-up values. In this case, the insured sticks with guaranteed or zero withdrawals for most time, and

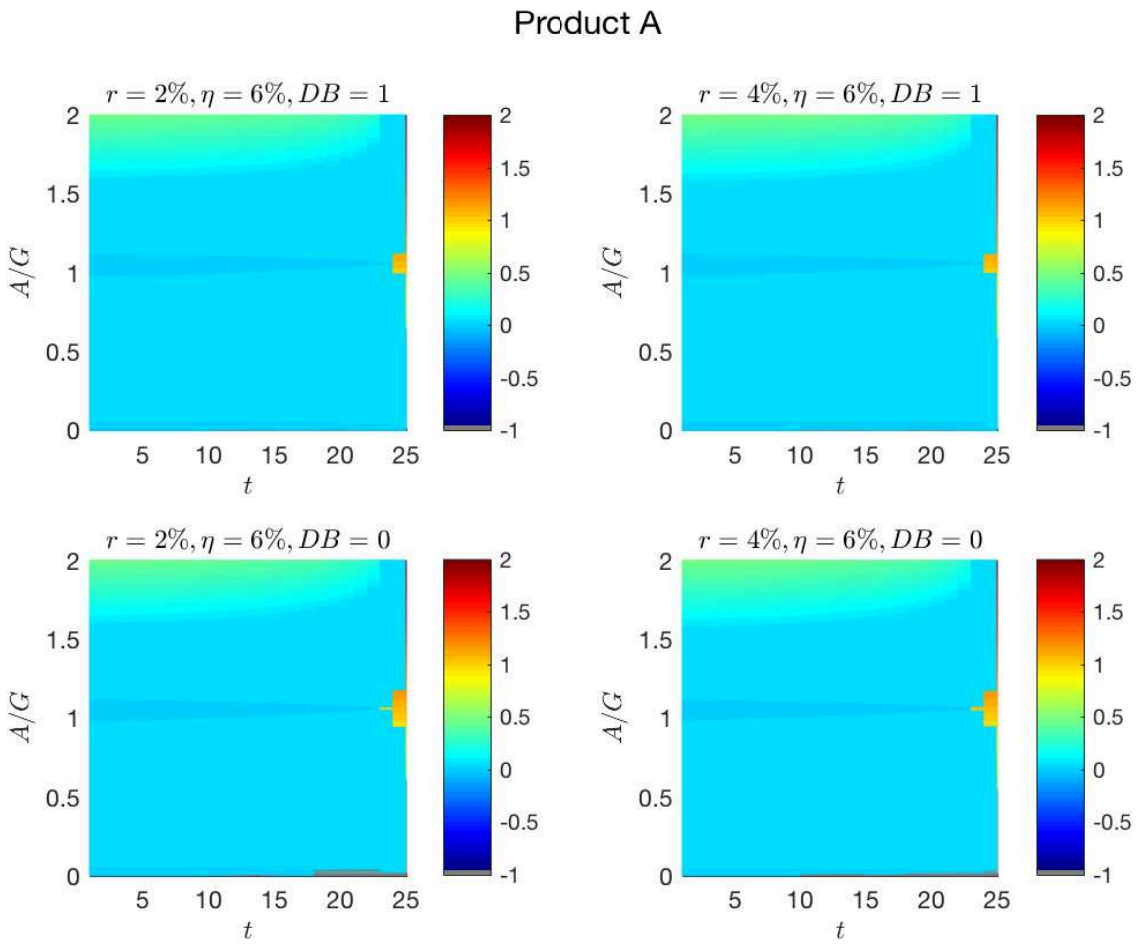


FIGURE 2.5 – The withdrawal strategy as a function of time t and moneyness A/G

tends to elect the income benefit in the last anniversary date if the account value is low ,and recover it otherwise. The reset is very costly for the insurance company, however, fees are quite high for this product and the ratchet takes place only at early dates since the account value is brought down by the fees rate.

In what follows, we will focus on the roll-up only case in an attempt to analyze the impact of some key parameters in the pricing and expected policyholders rational behavior for these products.

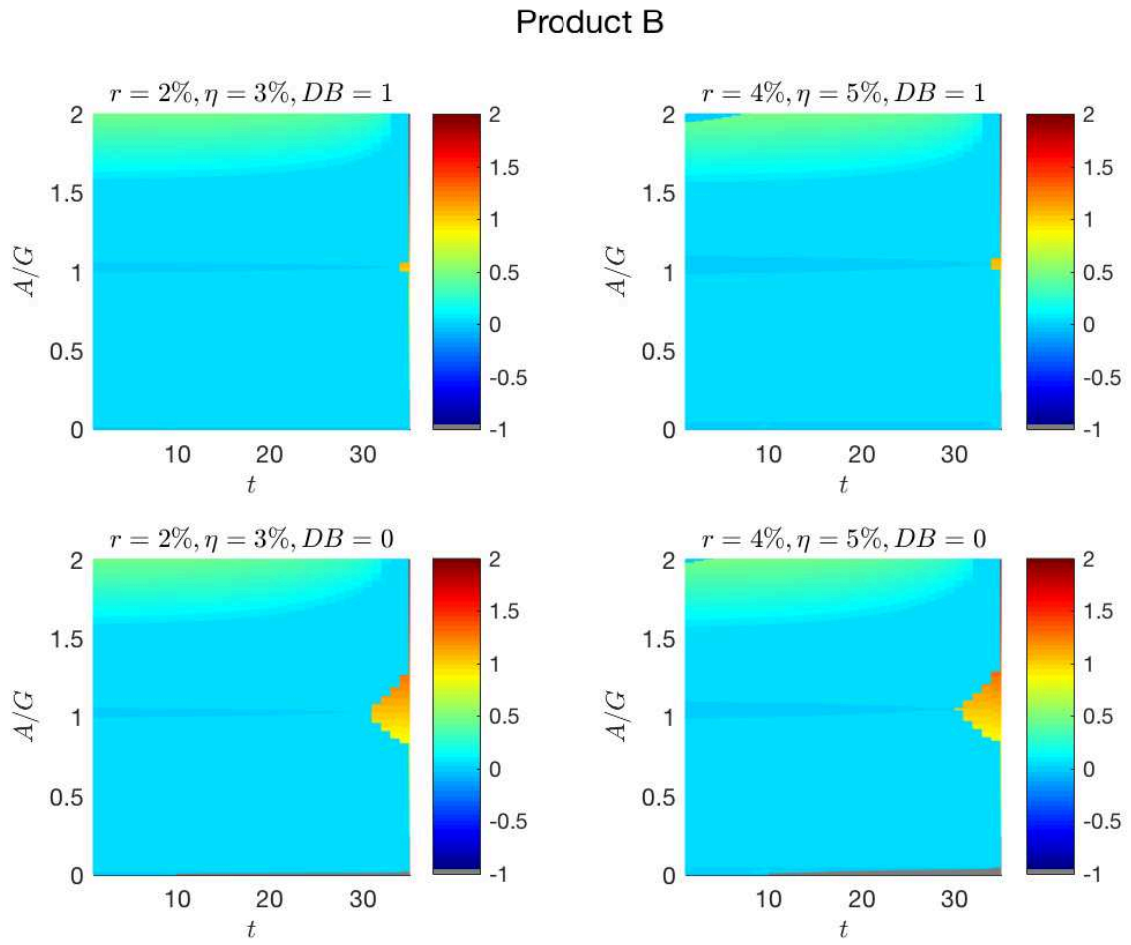


FIGURE 2.6 – The withdrawal strategy as a function of time t and moneyness A/G

The impact of volatility

The volatility level assumption is very important for variable annuities in general and the GMIB product in particular. Based on Product A for the hypothesis used above, we compare two levels of volatility (which we can compare to the 20% volatility case given in Figure 2.3). We can see in Figures 2.7 and 2.8 that the lower the volatility, the earlier guaranteed withdrawals start. Moreover, the lapsing likelihood is also higher. This means that, the more uncertain are markets, the more the policyholder tends to withdraw money from his account. On the other hand, the IB election does not seem to be affected.

Roll-up rate and fees

There is a trade-off between roll-up rate and fees. The roll-up is the mechanism that ensures the policyholder a minimum return, which can be higher than the money market. However, to be able to provide interesting roll-up rates, insurance companies need to be hedged from uncertain interest rates. In this sense, they use for example swaps. Therefore, they need to collect fees that at least allow for a fair pricing for the contract, i.e. such that the paid cash flows equal the premium. On the other hand, high fees can have a perverse effect. By decreasing the subaccount value, especially in periods of low performance, present collected fees reduce future potential ones. On the long run, combined with

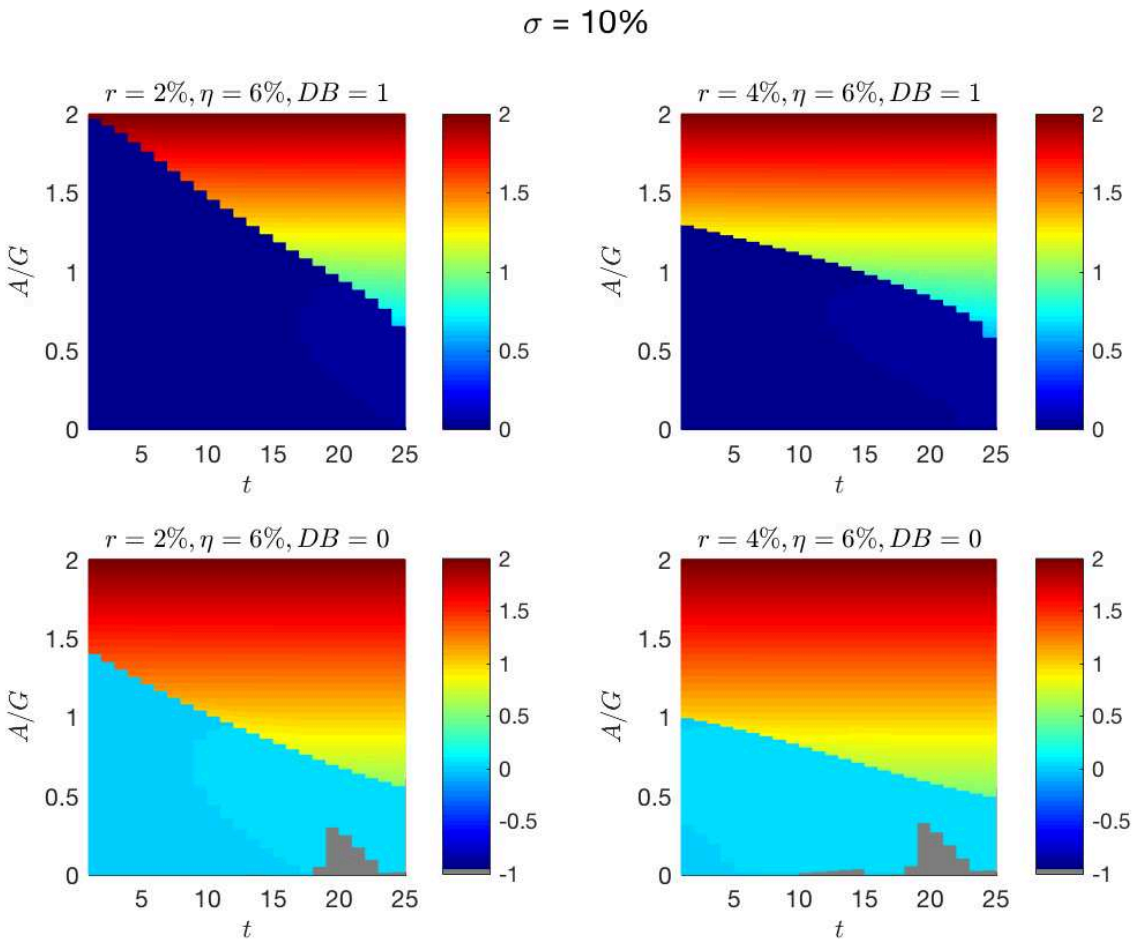


FIGURE 2.7 – The withdrawal strategy as a function of time t and moneyness A/G for σ = 10%

guaranteed withdrawals, the income benefit can be elected by bringing the account value to zero. Moreover, when the the account value falls to zero, the insurance company can no longer collect fees and starts to pay the guarantee.

In Figures 2.9, we compare the value of the contract at inception for different parameters of Product B as a function of total fees. The fair price corresponds to $\phi(0, \tilde{A}_0) = 1$. We see that the contract is under-priced with death benefit. The fair fees would be as high as 7%. Without the death benefit, they are around 3% which is close to the rates applied by the insurance company. In Figure 2.10, we conduct a similar test by varying the roll-up rate for Products A, A-DB, B and B-DB for $r = 2\%$. We notice that $\phi(0, \tilde{A}_0) = 1$ corresponds to a roll-up rate that is close to the interest rate except one of the products. Indeed, Products A, A-DB and B-DB are under-priced for the features they provide. On the other hand, the insurance company was conservative in the roll-up rate assumption for Product B which allows it to be profitable even for the worst case scenario. Of course, insurance companies do not expect (and hope not) that all policyholders follow an optimal behavior. However, prudent hypotheses can prevent from important losses. Including a proportion of policyholders that are likely to behave optimally is one of the solutions. Note that the GMIB product is less risky than the GMWB in that the annuity factor is defined with conservative assumptions.

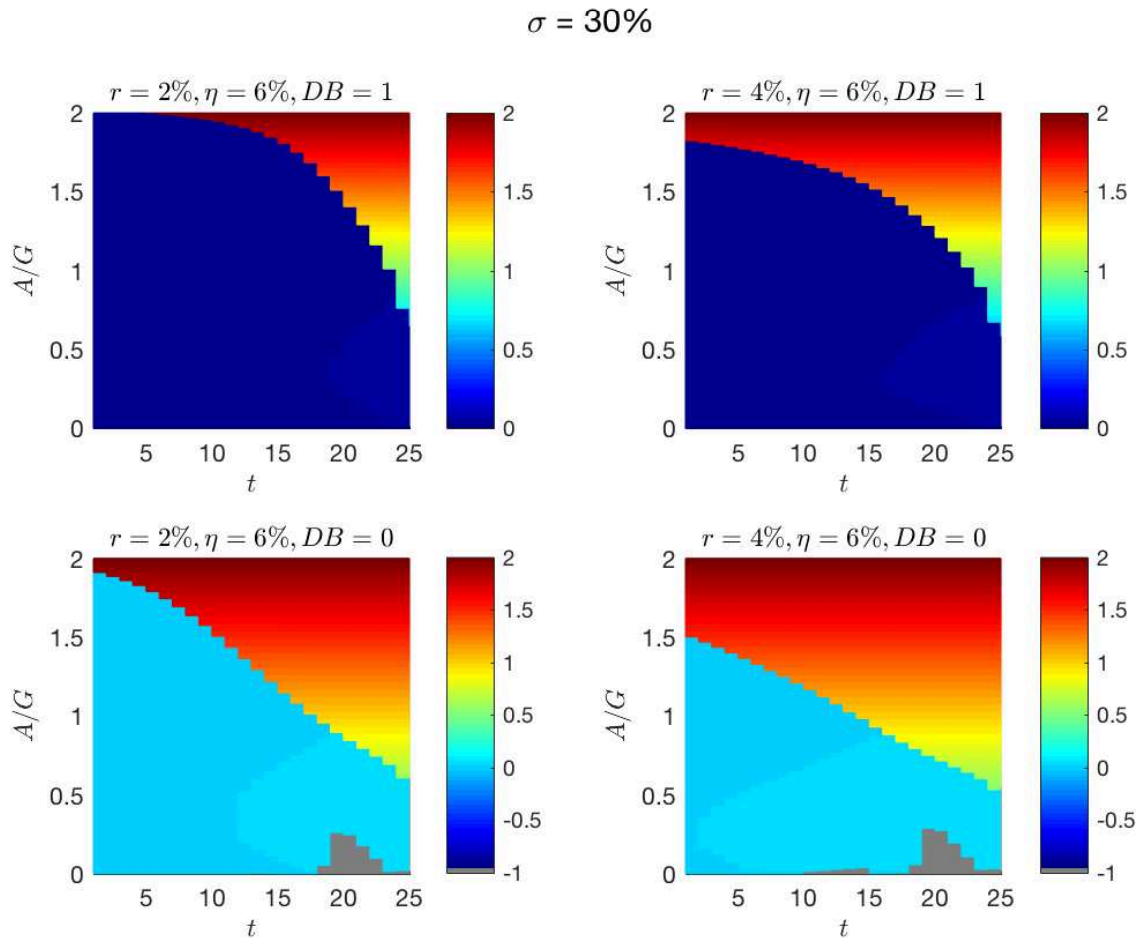


FIGURE 2.8 – The withdrawal strategy as a function of time t and moneyness A/G for σ = 30%

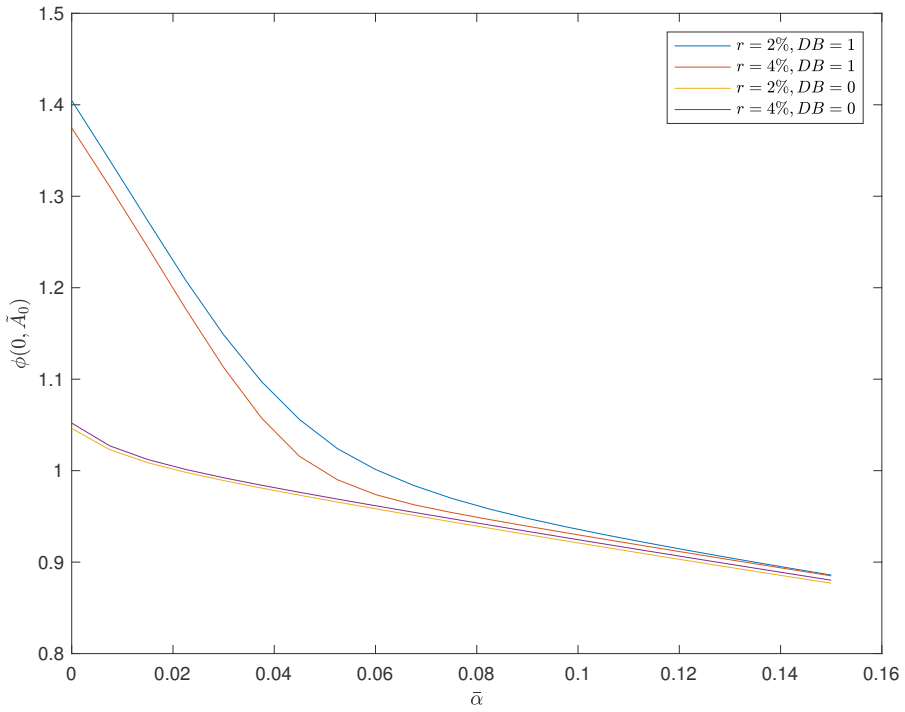


FIGURE 2.9 – The contract value at inception as a function of the total fees $\bar{\alpha}$ for Product A with and without DB for $r = 2\%, 4\%$

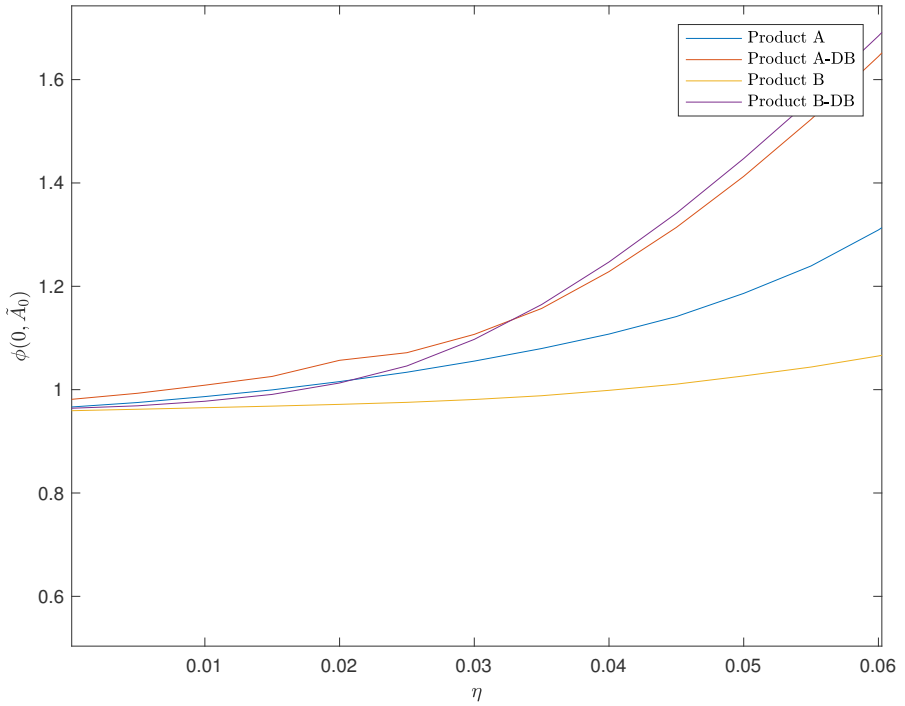


FIGURE 2.10 – The contract value at inception as a function of the roll-up rate η for Product A and B with and without DB for $r = 2\%$

2.6 Conclusion

In this work, we analyzed the optimal behavior of a policyholder entering a GMIB contract combined with a death benefit guarantee. The solution is based on a optimal stochastic control framework in a Black-Scholes framework, and solved numerically using recursive dynamic programming techniques. Considering only the account value varies between two anniversary dates, we used finite differences methods and a linear search for the optimal withdrawal to maximize the expectation of discounted future cash flows. Such calculations give an optimal withdrawal function that depends on time, account value and benefit base. Taking advantage of the good scaling properties provided by the contract payoff and the asset price, we are able to reduce the dimensionality of the problem to time and moneyness, making calculations faster and results interpretation easier.

The policyholder's optimal behavior corresponds to the worst case scenario for insurance companies. Therefore, even though insurers are not expected to behave optimally, a good insight of how they may act in case they do, given a market environment, can allow insurers to be more effective in pricing and hedging their products. We find that the optimality consisted mainly in four choices; zero withdrawals, guaranteed withdrawals, lapse and IB election. We presented these results for two different products before analyzing the impact of some of the contract key parameters.

Finally, these results can be used as a guide to practitioners in the design of new products where a particular client behavior is desired. The model can be used to compute the fair withdrawal fee with hedging purposes. In particular, we find that these products are under-priced in case of a optimal behaviors, as it was already mentioned by [130] in the case of GMWBs.We believe that understanding policyholder behavior is a critical concern to the future of life insurance business, and due to its importance, further research is required.

Chapitre 3

iCPPI and Gap Risk

***Abstract*— Individualized Constant Proportion Portfolio Insurance (iCPPI) products are attractive alternatives to traditional unit linked products offering a guaranteed minimum return, such as variable annuities. They offer high potential returns whilst limiting the downside risk by implementing a dynamic allocation strategy between risky and risk-free assets tailored to the risk attitude of the beneficiary. But performance evaluation of iCPPI products should not rely on the unrealistic assumptions of continuous market price variations and continuous rebalancing of asset allocations. We adopt a more general and realistic price jump model and examine several dynamic strategies as well as gap put options to mitigate the risk that the value of the product falls below the guaranteed minimum.

Keywords : CPPI ; dynamic multiplier ; jump processes and gap risk ; vanilla and gap options.

3.1 Introduction

Increased market volatility and falling interest rates triggered by the 2008-2009 financial crisis reduced the performance of traditional long-term investment products, increased their risks and, where applicable, their capital requirements. In this context the new iCPPI products provide an attractive alternative to many traditional long-term investment products offering a guaranteed minimum return, such as variable annuities, for several reasons : lower exposure to uncertain volatilities and extreme market price movements, lower costs, and lower regulatory capital requirements, to name a few.

Already, with rising life expectancy, current provisions for retirement may not be sufficient for many people to secure acceptable life standards after retirement. To achieve sufficiently high investment returns, together with low risks over the long term, funds should remain invested in stocks and other risky assets as well as in the safer bonds over an extended period well into retirement. The design of long-term investment products should also reflect the requirements and risk attitudes of individual investors.

Constant Proportion Portfolio Insurance (CPPI) is the name given to an investment strategy that provides a minimum guaranteed return, the “floor” (usually defined as the discounted value of a final capital guarantee) and aims to maintain, at all times, an exposure to a risky asset equal to a constant multiple of the “cushion” defined as the excess value of the fund above the floor. The final capital guarantee and the multiplier are chosen to satisfy the risk attitude of the investor.

Assuming the risky asset follows a geometric Brownian process, constant rate for the risk-free asset, and continuous relancing with no transaction costs, the CPPI payoff is optimal for an investor with

a risk tolerance coefficient varying linearly with wealth, see [29], [125], [140]. Specifically, the CPPI payoff is equal to the floor plus a cushion value which is proportional to asset price power the multiplier. The floor and multiplier are chosen according to the two parameters of a HARA utility function to maximize the expected utility of the investor. Additional advantages offered by CPPI strategies over more traditional investments with minimum guaranteed returns are : price transparency, open time-horizon, no early redemption penalty, wide range of alternative investments for the risky asset, and flexibility to add other guarantees such as ratchets (see II.1.4).

iCPPI is a CPPI strategy adapted to evolving individual needs and market conditions. The floor and multiplier are modified accordingly. Thus iCPPI may combine most of the advantages of CPPI with the need for flexibility and enhanced risk management.

ICPPI providers (typically, an insurance company) face many challenges in the implementation of the dynamic strategy that replicates the guaranteed payoff. Adjusting the risky asset/ risk-free asset allocation can take place at discrete times only, there are transaction costs, and risky asset prices may jump. Thus, there is a difference between the theorical value of an iCPPI strategy under hypothetical assumptions, and reality. In particular, there is a positive probability for the value of the fund to fall below the guaranteed floor. We call such shortfall the gap risk.

The analysis of the gap risk has often been limited to simple conditions to preserve analytical tractability :

— Unrealistic modeling of the risky asset price market including continuous price dynamics, zero-cost trading and unlimited liquidity.

— Simple parameterization of the CPPI strategy such as constant capital guarantee and multiplier.

— Simplistic rebalancing strategies such as constant frequency.

As a result, the iCPPI offers a mechanism that takes advantage of the specific advantages of both stocks and bonds, while complying with growing needs of flexibility as experienced by policyholders.

However, the implementation of iCPPIs at insurance companies levels suffers from a number of operational constraints on the asset management : the rebalancing occurs through regular albeit discontinuous (at most daily) checks between the insurance company and a bank; depending on the design of the iCPPI and the discontinuous rebalancing frequency, the magnitude of the earnings at extreme risk may require the externalization of the gap risk management to the bank. As a result the main issue experienced by the insurance company remains to minimize the downside risk and keep control of the gap risk, which involves three main challenges : This article extends previous analyses of the gap risk by introducing :

— Price jump dynamics

— A dynamically adjusted multiplier

— Advanced rebalancing strategies, vanilla and gap put options to mitigate the gap risk

3.1.1 A brief review of the literature

CPPI as a mechanism falls within portfolio management techniques which ensure a lower bound on the portfolio value at a given maturity. In theory, one can have protection against unfavorable market scenarios for some asset by investing in a put option with strike equal to the desired lower bound. This is known as the option based portfolio insurance (OBPI) and was introduced by [116] and [41]. However, the put option needed to perform this protection may not be available in the market, for example

if the investment horizon is long. And even though one can replicate the payoff of the put option by trading the asset and the cash, such replication is costly and imperfect.

The CPPI was introduced by [140] and [29]. Its wide use in the financial industry, see [138], brought a lot of attention. Typical buyers are large individual investors and institutional investors such as pension funds. The main topics the literature on the CPPI covers are : limiting its risk, developing hedging techniques to cover the remaining risk, and the behavior of the CPPI.
[26] defined an upper bound for the multiplier m such that the investment in the risky portfolio is maximized under the gap risk must stay under a certain limit. [45] extended the calculation of the VaR and GVar of the CPPI portfolio. [144] and [57] on the other hand, contributed in extending the CPPI in order to build a protection against the small but existing gap risk. The price and size of such protection is model-dependent and will depend on the probability of hitting the floor. Finally, [26] and [84] study in detail the behavior of the CPPI strategy under specific conditions for the underlying asset. [26] for example considered the case where the underlying risky asset of the CPPI fund is an index or a basket of indexes. They used a Multivariate Variance Gamma (MVG) model for a series of correlated spreads to price the CPPI.

3.1.2 Review of CPPI mecanism basics

Consider at time t a risky asset (e.g., a share) with price S_t and a risk-free asset (e.g., a Treasury bond) with price B_t returning a constant rate r. The CPPI fund is invested into these two assets so that part of its value, called the "floor" F_t, , is guaranteed whilst the excess value above the floor, called the "cushion" $C_t = V_t - F_t$, remains exposed to the risky asset price fluctuations. At any time, the exposure to the risky asset, e_t, is kept at a constant multiple, m, of the cushion, that is :

$$e_t = mC_t$$

The rest of the value of the fund is invested (or, if negative, borrowed) at the risk-free rate (Note that the exposure e_t may be acquired at no cost if using an off-balance sheet instrument such as a future, which may be advantageous because of liquidity and low transaction costs). The floor is often chosen to increase over time at the risk-free rate (it could not be made to increase faster indefinitely), that is :

$$F_t = F_0 e^{rt}$$

In theory, when the risky asset price follows a geometric Brownian motion, and with continuous, zero-cost rebalancing (Black-Scholes conditions), the value of the cushion is path independent and proportional to S_t^m. In other words, it is the value of a power option. It is convex when $m > 1$ (like a long call option), linear when $m = 1$, and concave when $m < 1$, like a short put option. But unlike standard call and put options there is no need to fix an expiry date, a CPPI strategy is open-ended. Under the above assumptions, the value of the cushion would never fall to zero; in practice, if it does fall to zero or below zero (e.g., because of a price jump or of discrete rebalancing), the entire fund is monetized, i.e., is entirely invested in the risk-free asset, and the product provider must make up the shortfall to deliver the floor value. In practice there may also be other constraints such as no borrowing or additional features such as ratcheting up the floor. In those cases, the path independency and open-endedness of the product are lost and the payoff profile becomes more complex.

3.2 Methodolody

3.2.1 CPPI in theory and practice

Continuous-time framework

The risky asset S is defined by the diffusion equation $dS_t = \mu S_t dt + \sigma S_t dW_t$ where W is a standard Brownian motion. The previous hypothesis for the risk-free asset are kept. In such context, and assuming continuous time CPPI, the cushion C is log-normally distributed with drift $m(\mu - r) + r$ and volatility $m\sigma$:

$$C_t = C_0 \exp\left(\left(m(\mu - r) + r - \frac{m^2\sigma^2}{2}\right)t + m\sigma W_t\right)$$

and the portfolio value V has the path independent expression :

$$V_t = F_t + (V_0 - F_0)\exp\left(\left(m(\mu - r) + r - \frac{m^2\sigma^2}{2}\right)t + m\sigma W_t\right)$$

However, such assumptions are unrealistic and not consistent with market practice. To remedy these unrealistic hypothesis, two alternatives are studied : modeling in a discrete-time framework and in a Lévy framework.

Discrete-time CPPI

In practice the CPPI is rebalanced in discrete time, where the shortfall probability is no longer equal to 0, which implies to monetize more often.
A sequence of equidistant refinements of the interval $[0, T]$ is defined :

$$\Theta = \{t_0 = 0 < \ldots < t_{N-1} < t_N = T\}$$

where $t_{k+1}^N - t_k^N = \frac{T}{N}$ for $k = 0, \ldots, N-1$. The number of shares is constant on the intervals $]t_i, t_{i+1}]$. Let $t_s := min\{t_k \in \Theta | V_{t_k} - F_{t_k} \leq 0\}$. The first time the portfolio value touches the floor. The discrete-time cushion follows the equation :

$$C_{t_{k+1}} = e^{r(t_{k+1} - \min\{t_s, t_{k+1}\})}(V_{t_0}^{\Theta} - F_{t_0}) \prod_{i=1}^{min\{s,k+1\}} \left(m\frac{S_{t_i}}{S_{t_{i-1}}} - (m-1)e^{r\frac{T}{N}}\right),$$

or recursively :

$$C_{t_{k+1}} = \begin{cases} C_{t_k}\left(m\frac{S_{t_{k+1}}}{S_{t_k}} - (m-1)e^{r\frac{T}{N}}\right) & \text{if } C_{t_k} > 0, \\ C_{t_k}e^{r\frac{T}{N}} & \text{if } C_{t_k} \leq 0. \end{cases} \tag{3.1}$$

V_{t_k} is given through the relation $V_{t_k} = C_{t_k} + F_{t_k}$.

To comply with the CPPI algorithm and respect practical constraints, the number of shares of the risky and safe assets (α and β) are as follows :

— $\alpha_{t_k} = min\left(max\left(\frac{mC_{t_k}}{S_{t_k}}, 0\right), \frac{V_{t_k}}{S_{t_k}}\right)$.

— $\beta_{t_k} = \frac{V_{t_k} - \alpha_{t_k}S_{t_k}}{B_{t_k}}$.

When adding transaction costs, they are taken as a proportion of the change in the risky exposure, i.e $\propto (\alpha_{t_k} - \alpha_{t_{k-1}}) \times S_{t_k}$. So at time t_k, the number of shares of the risky asset will be reduced to :

$$\tilde{\alpha}_{t_k} = \alpha_{t_k} - |\alpha_{t_k} - \alpha_{t_{k-1}}| \times \text{nb of bps}$$

The CPPI capital guarantee is ensured, as long as the bond floor is not breached through enabling to fully invest the portfolio into the non risky assets.The probability of breaching the floor is defined as the probability that the portfolio value falls below the floor, i.e $P^{BF} := \mathbb{P}(V_T \leq G) = \mathbb{P}(\exists t \in [0,T] : V_t \leq F_t)$. The local shortfall probability is the conditional probability defined as : $P^{LBF}_{t_i,t_{i+1}} = \mathbb{P}\left(V_{t_{i+1}} \leq F_{t_{i+1}} \mid V_{t_i} > F_{t_i}\right)$. The two are related as follows $P^{BF} = 1 - \prod_{i=1}^{i=N}(1 - P^{LBF}_{t_i,t_{i+1}})$.

This probability which was equal to zero in the continuous Black-Scholes model, is now greater than zero. Assuming the portfolio did not breach the floor up to t_k, the probability of breaching the floor at t_{k+1}, is that of a downside jump in the risky asset of more than about $1/m$. Its mathematical expression is :

$$P^{LSF}_{t_i,t_{i+1}} = \mathbb{P}\left(\frac{S_{t_{i+1}}}{S_{t_i}} \leq \frac{m-1}{m} e^{r\frac{T}{N}}\right),$$

where the evolution of the risk-free part with rate r is taken into account.

The backtesting is based on the period Q1-2006 to Q4-2010 on S&P500 index. Simulating paths (N=10,000) in the Black & Scholes model is made using the 3-month realized volatility based on the standard deviation (see Figure 3.1), a constant asset return $\mu = 8\%$. The rate of the risk-free asset is $r = 4\%$. Three rebalancing frequencies are being compared regarding the distribution of the final portfolio value (daily, weekly and monthly), with the following assumptions :

— Initial investment/Guarantee : \$100, and \$100

— Duration : 5 years

— Transaction costs : 10 bps

	Buy & Hold Strategy	CPPI with $m = 3$			CPPI with $m = 6$		
		Daily	Weekly	Monthly	Daily	Weekly	Monthly
Mean	126.97	123.31	122.39	119.75	124.10	124.87	125.01
Std-Dev	7.18	31.58	32.66	36.86	42.62	43.88	48.10
95% quantile	116.90	100.48	99.98	97.01	99.99	99.13	89.69
99.5% quantile	113.42	100.02	99.88	91.47	99.98	95.20	74.28
5% quantile	140.21	194.37	195.23	197.94	216.51	218.50	225.46
0.5% quantile	150.63	266.47	284.07	282.58	291.49	293.75	311.46
Rebalancing cost	0.01	0.91	0.44	0.26	0.78	0.46	0.31
P^{BF}	0	0.0018	0.0947	0.5289	0.2016	0.5730	0.6555

TABLEAU 3.1 – Final value metrics : Buy & Hold strategy vs CPPI with $m = 3$ vs CPPI with m = 6

The CPPI strategy under daily rebalancing performs better against a bear market than the weekly and monthly ones due to its reactiveness to decrease the risky exposure whenever needed. With such frequency, the guarantee is almost ensured; the less frequent we rebalance the more we are exposed to breaching the floor (as illustrated by fatter left tails (see Figure 3.2 bottom, right). The backtesting (Figure 3.2 top) and Table 3.5 illustrate the following remarks :

— In periods of mild market conditions, transaction costs negatively affect the performance of a daily rebalancing, although not to a significant extent.

— During a market crash, the three strategies monetize, with the daily rebalancing having less losses than the two others.

— The empirical probability of breaching the floor decreases when the rebalancing frequency increases.

— The cost of rebalancing increases with the frequency and with the multiplier. However, in our results, the cost of daily rebalancing for $m = 6$ is lower than the one with $m = 3$. This is explained by the fact that such a high multiplier allows for a total risky exposure and thus no rebalancing reducing the cost.

When comparing different strategies (Buy & Hold, CPPI with m = 3 and m = 6), we have the following results :

— The Buy & Hold strategy has higher expectation and lower standard deviation (table 3.5). This is mainly due to the low exposure to the risky asset. Its performance is highly correlated to the non-risky return (chosen to be 4%).

— The 5% and 0.5% quantiles show that the CPPI with $m = 6$ has a larger right tail and thus, performs better than the two others in bullish market. This remark is also illustrated in Figure 3.8.

Daily rebalancing almost prevents the bond floor from being breached, which ensures the capital guarantee at maturity. However, constant volatility and log-normal distribution modeling are not consistent with empirically observed jumps during extreme market moves likely to breach the bond floor. In order to relax these unrealistic assumptions, jumps are thus added through Lévy processes as developed in the next section.

Adding jumps

We assume that the process of the risky asset follows a Lévy process :

$$\frac{dS_t}{S_t} = dZ_t,$$

where Z is a Lévy process. The risk-free asset F_t is still deterministic.

Let $\tau = \inf\{t : V_t \leq B_t\}$ the time where the portfolio value is fully invested in the risk-free asset. Until τ the actualized cushion ($C^*_t = \frac{C_t}{F_t}$) is as follows : $C^*_t = C^*_0 \mathscr{E}(mL)_t$, where $\mathscr{E}$ denoting the stochastic exponential :

$$\mathscr{E}(Z)_t = Z_0 e^{Z_t - \frac{1}{2}<Z>_t} \prod_{s \leq t, \Delta Z_s \neq 0} (1 + \Delta Z_s) e^{-\Delta Z_s},$$

which gives us the portfolio value :

$$V_t = \begin{cases} V_t \left\{ 1 + \left(\frac{V_0}{F_0} - 1 \right) \mathscr{E}(mL)_t \right\} & \forall\, t \leq \tau, \\ V_\tau e^{r(t-\tau)} & \text{if } t > \tau. \end{cases}$$

The probability of breaching the floor can be expressed as :

$$P^{BF} = \mathbb{P}(\exists t \in [0,T], V_t \leq B_t) = 1 - \mathbb{P}\left(\forall\, t,\ \Delta L_t < \frac{1}{m} \right),$$

$$P^{BF} = 1 - \exp\left(-T \int_{-\infty}^{-1/m} \nu(dx) \right),$$

which is illustrated by the fact that the number of downside jumps of size more than $\frac{1}{m}$ follows a Poisson distribution with intensity $T\nu(-\infty, -1/m)$.

For computation tractability, we choose the double exponential Kou model, see [112]). Under the risk neutral probability, the risky asset is modeled as follows :

$$\frac{dS_t}{S_{t^-}} = \mu dt + \sigma dW + d\left(\sum_{i=1}^{N_t} e^{Y_i} - 1 \right),$$

where W is a standard brownian motion, N is a poisson process with rate λ, the constants μ and $\sigma > 0$ are drift and volatility of the diffusion part and the jump sizes $\{Y_1, Y_2, ...\}$ are i.d.d random variables with a common asymmetric double exponential distribution of density :

$$f_Y(y) = (1-p)\eta^+ e^{-\eta^+ y}\mathbb{1}_{y\geq 0} + p\eta^- e^{\eta^- y}\mathbb{1}_{y<0},$$

where η^+ is intensity of positive jumps while η^- and p are the intensity of negative jumps and the probability of their occurrence.

Under this jump model, and assuming a continuous rebalancing frequency, the probability of breaching the floor takes the following form :

$$P^{BF} = 1 - \exp\left(-T p\lambda(1-\frac{1}{m})^{\frac{1}{\eta^-}}\right).$$

We substitute the Black-Scholes framework with the Kou model which we calibrate on implied volatility smile (between 2006 and 2011 for a 1-month implied volatility on a weekly basis [1]). We carry out the calibration by minimizing the quadratic error :

$$\sum_{i=1}^{9}\left(C_{t_i}(T,K_i)^{\text{Market}} - C_{t_i}^{\text{Kou}}(T,K_i,\sigma,p,\eta^+,\eta^-,\lambda)\right)^2,$$

where T is 1-month maturity, nine strikes K_i from 80 to 110 and $(p,\eta^+,\eta^-,\lambda,\sigma)$ are the jump parameters (more details about . We give different statistics for these parameters in Table 3.2.
In order to avoid instability in parameters, we chose several starting points and set boundary condi-

	Average	5% percentile	Std-Dev
p	0.64	0.84	0.24
η^+	0.16	0.28	0.06
η^-	0.15	0.28	0.07
λ	0.62	2.44	0.12
σ	18.29%	29.64%	0.08

TABLEAU 3.2 – Average, 5% percentile and standard-deviation of the Kou model parameters estimated from the option data between 2006 and 2011

tions. An example of the result on the calibration is shown in Figure 3.4.
A few remarks on the calibration can be made :

— Since the upward-sloping part of the smile is very small, the positive jumps are hardly calibrated in a reliable manner. However, the pricing of the gap option (section II.2.2) only needs the negative jumps intensity (*i.e* the downward-sloping part of the smile).

— The calibration is better on close-to-maturity options (as mentioned in [152]). It allows a better capture of instantaneous jump.

— The calibrated parameters will be used for hedging gap risk in the last section.

Figure 3.5 compares different discrete rebalancing frequencies with a jump modeling :

— Even for daily rebalancing, breaching the floor is unavoidable with the same probability as the two other frequencies.

— The three rebalancing frequencies give similar results when taking transaction costs into account.

	Kou model		
	Daily	Weekly	Monthly
Mean	146.28	147.10	147.57
Std-Dev	52.84	52.93	53.11
95% quantile	92.19	92.21	92.03
99.5% quantile	59.38	59.08	59.23
5% quantile	238.13	238.67	239.41
0.5% quantile	349.41	350.92	350.37
Rebalancing cost	0.92	0.45	0.26

TABLEAU 3.3 – Performance of the CPPI for different models and rebalancing frequencies

The previous illustrations show that both the frequency of the rebalancing and the modeling affect the final value. The two metrics previously defined for different modeling assumptions

— The local probability of breaching the floor :

$$\mathrm{P}^{\mathrm{LBF}}_{t_i,t_{i+1}} := \mathbb{P}\left(\mathrm{V}_{t_{i+1}} \leq \mathrm{F}_{t_{i+1}} \,|\, \mathrm{V}_{t_i} > \mathrm{F}_{t_i}\right).$$

— The overall probability of breaching the floor :

$$\mathrm{P}^{\mathrm{BF}} := \mathbb{P}\left(\exists t \in [0,\mathrm{T}] : \mathrm{V}_t \leq \mathrm{F}_t\right) = \mathbb{P}\left(\mathrm{V}_{\mathrm{T}} \leq \mathrm{F}_{\mathrm{T}}\right).$$

— For Black-Scholes model in discrete-time rebalancing :

$$\mathrm{P}^{\mathrm{LBF}}_{t_i,t_{i+1}} = \mathcal{N}\left(-\frac{log(\frac{m}{m-1}) + (\mu - r)\frac{\mathrm{T}}{\mathrm{N}} - \frac{1}{2}\sigma^2\frac{\mathrm{T}}{\mathrm{N}}}{\sigma\sqrt{\frac{\mathrm{T}}{\mathrm{N}}}}\right).$$

and

$$\mathrm{P}^{\mathrm{BF}} = 1 - \prod_{i=0}^{\mathrm{N}-1}\left(1 - \mathrm{P}^{\mathrm{LBF}}_{t_i,t_{i+1}}\right).$$

— For Kou jump process in continuous time :

$$\mathrm{P}^{\mathrm{BF}} = 1 - \exp\left(-\mathrm{T}p\lambda(1 - \frac{1}{m})^{\frac{1}{\eta^-}}\right).$$

— Results depend on the model parameters and discretization time step :
  — Gap risk goes to 0 as the rebalancing tends to be more frequent
  — When considering a discontinuous path (jump models), even in continuous rebalancing the gap risk value > 0

Impact of the ratchet feature

The ratchet feature is used by insurance companies to attract investors as it periodically locks in profit, see [43] and [9] for more details : at anniversary dates the guarantee is set to the highest value so

1. Implied volatilities are collected from Bloomberg dataset. The calibration is performed weekly.

Model	Frequency	P^{BF}
B&S	Monthly	$9,07 \times 10^{-5}$
	Weekly	$1,2 \times 10^{-10}$
	Daily	~ 0
Kou	Continuous	0,000410

TABLEAU 3.4 – The probability of breaching the floor for different models and rebalancing frequencies.

far. The guarantee G becomes a time dependent function.

$$G_t = \begin{cases} V_0 & \text{if } t = 0 \\ \max(G_{t^*_{k-1}}, V_{t^*_k}) & \text{if } t = t^*_k \\ G_{t^*_k} & \text{if } t \in (t^*_k, t^*_{k+1}). \end{cases}$$

The bond floor is then defined as $F_t = G_t e^{-r(T-t)}$.
This feature has advantages and drawbacks. Locking-in the cash will ensure a higher guarantee but also reduces the cushion, the risky exposure and thus the upside potential risk.

The main results from figure 3.6 are :

— The mean and standard deviation of the final value increase with the rebalancing frequency (see table 3.5). This is justified by the path dependency of the guarantee which has a larger distribution with higher rebalancing frequency.

— The quantiles on the two tails of the final value distribution increase with the rebalancing frequency, while the distribution is shifted to the right with narrower body.

	Without ratchet			With ratchet		
	Daily	Weekly	Monthly	Daily	Weekly	Monthly
Mean	123.82	124.26	124.17	145.46	143.01	134.03
Std-Dev	41.96	43.29	47.25	100.08	81.75	45.60
95% quantile	99.99	99.68	90.88	100.61	100.53	99.99
99.5% quantile	99.99	97.57	77.84	99.99	99.94	98.27
5% quantile	214.18	216.42	222.23	268.74	261.73	219.52
0.5% quantile	289.15	292.33	314.58	700.97	603.28	359.59
P^{BF}	0.11	0.47	0.64	0.11	0.48	0.84

TABLEAU 3.5 – Final value metrics : Comparison between a CPPI without and with the ratchet feature

Consider the stopping time τ as the first time the portfolio value breaches the floor which does not depend on the bond floor level. The distribution of τ is the same in case of adding the ratchet, i.e. the probability of breaching the floor is not usually affected by the ratchet feature in theory. However, in our simulations, this probability in higher for the monthly rebalancing. This might be

3.3 Mitigating downside risk : Preventing from breaching the floor

3.3.1 Adjusting the multiplier to market conditions

By focusing on managing returns in downside markets, CPPI effectively manages portfolio volatility. Over the 5-year data (which included one bullish market, one bear market and a recovery), the CPPI strategy resulted in a slightly lower return – but also a significantly lower volatility. Additionally, the

worst one year return for the CPPI strategy was significantly less than that of the Index Portfolio.

The manager usually sets the multiplier at the beginning of the period. The risky exposure depends then on the evolving cushion. As the probability of breaching the floor may surge in market crash, or the manager might miss the subsequent market recovery, the multiplier needs to be adjusted accordingly with the market conditions.

A first approach to define a dynamic multiplier is the choice of the optimal m, deduced from the closed form solutions for optimal payoffs, and optimal certainty equivalent returns (CERs) using HARA utilities and log-normal distribution, see [142]. The authors give the following formula $m^* = \eta(\mu - r)/\sigma^2$ (η here is the investor's sensitivity of risk tolerance to wealth). A particular case is the growth optimal leverage with $\eta = 1$ which is resulted in optimizing the growth rate of the leveraged strategy (cushion).

An alternative to the optimal multiplier is a Value-At-Risk based multiplier where investors choose the confidence level according to their risk tolerance as well but focused on tail risks.
Based on the weight w_t^{R} of the Value-At-Risk Based Portfolio Insurance (VBPI) introduced by [106], and the expression of the risky exposure in both strategies ($\mathrm{E}_t = m_t \mathrm{C}_t = w_t^{\mathrm{R}} \mathrm{V}_t$), the expression for the multiplier at time t is :

$$m_t = \frac{1}{1 - \exp\left((\mu - r - \frac{1}{2}\sigma^2)(\mathrm{T} - t) - z_p \sigma \sqrt{\mathrm{T} - t}\right)}$$

Since the dynamic multiplier depends on both volatility and return estimates, in order to improve its efficiency, μ and σ can be re-estimated at each time step. However, the estimation of the drift is hardly accurate. Therefore, we will restraint the time dependency to the volatility. It will be re-estimated through a 3-month sliding window to take into account different market regimes.

The two approaches offer an interesting alternative to constant multiplier which lacks flexibility to market conditions. The comparison between these two approaches through a backtesting from 2006 to 2011, is illustrated in Figure 3.7. The focus on two periods (2006-2007 and post 2008 crisis), in Figure 3.8, illustrates that the VaR-based multiplier can perform better than the "optimal" one in bullish market and recovery (e.g 18% return Q2-2009 until Q1-2011 vs 11% in the post 2008 crisis). In contrast, during bear market, using the "optimal" multiplier (through $m < 1$) helps keep a relatively higher cushion but misses the recovery as it doesn't allow a high leverage.

In order to allow to participate in the market recovery to a greater extent, the multiplier is adjusted with a modified volatility estimator, either through a short-term exponentially weighted moving average (EWMA with $\lambda = 0.94$) realized volatility or an estimator based on implied volatility (of the strike consistent with the latest market returns). For example, if the underlying jumped 5% downward, the implied volatility with strike 95% will be chosen. For unavailable strikes, we use a linear interpolation. This strategy starts reinvesting into the risky asset as soon as Q3 2009, resulting in a higher performance by allowing the portfolio to capture more of the upside return when markets rebound. The backtesting in Figure 3.10 illustrates that the new multiplier is more reactive when adjusting with the implied volatility estimator. However, the 3-month realized volatility provides a higher multiplier and, when considering transaction costs, leads a lower cost of management.
Finally, the fixed frequency rebalancing is switched to a trigger rebalancing which occurs when the multiplier is out of a specific range chosen by the portfolio manager. In our case, on average the rebalancing frequency becomes every other day, which is consistent with the usual practice in CPPI asset management. At the same time, the cost of rebalancing is cut by half in comparison to a daily rebalancing (i.e. as low as a weekly or monthly rebalancing). Figure 3.11 illustrates the increasing performance

specifically under a range-bound high volatility regime, e.g. Q1-Q3 2008.

Adjusting the multiplier dynamically allows it to be more reactive to market conditions and explicitly dependent on the investor's risk aversion. However, it does not totally annihilate the downside risk in case of sudden jumps, where options may be useful to hedge those gap risks.

3.3.2 Hedging gap risks

The CPPI methodology will not necessarily protect the portfolio against a "black swan" event (such as a market crash of 20% in one day). To the extent that asset allocation shifts are implemented via underlying funds, the rebalancing trade can only occur at the end-of-day NAV. Even if futures are used to implement shifts intra-day, there can be gap movements in the futures markets. This is where a small gap risk protection sleeve can add value to the portfolio. To protect against such a "black swan" event, it is important to already have put options on market indices in the portfolio.

Vanilla Put option

A simple hedging strategy for the CPPI through embedded option can be constructed using short maturity put options. Touching the bond floor is mathematically equivalent to the cushion becoming negative. Assuming the event hasn't occurred up to time t_k, using equation (3.1), we have :

$$\mathrm{C}_{t_{k+1}} < 0 \Leftrightarrow m\frac{\mathrm{S}_{t_{k+1}}}{\mathrm{S}_{t_k}} - (m-1)e^{r\frac{\mathrm{T}}{\mathrm{N}}} < 0.$$

Hedging this risk is equivalent to forcing this quantity to be positive. This can be done by buying a put option at each of the CPPI rebalancing period with strike $(1-\frac{1}{m})e^{r\frac{\mathrm{T}}{\mathrm{N}}}\mathrm{S}_{t_k}$ and as a maturity the CPPI rebalancing frequency. To hedge the whole portfolio the manager needs a number of $m\frac{\mathrm{C}_{t_k}}{\mathrm{S}_{t_k}}$ puts, which is the risky asset exposure. The discounted payoff in this case is $e^{-r\frac{\mathrm{T}}{\mathrm{N}}}\mathrm{C}_{t_k}((m-1)e^{r\frac{\mathrm{T}}{\mathrm{N}}} - m\frac{\mathrm{S}_{t_{k+1}}}{\mathrm{S}_{t_k}})^+$. The hedging cost at time t_k can be written as :

$$\mathrm{Cost}_{t_k} = m\frac{\mathrm{C}_{t_k}}{\mathrm{S}_{t_k}}\mathbb{E}^{\mathcal{Q}}\left[\left((1-\frac{1}{m})e^{r\frac{\mathrm{T}}{\mathrm{N}}}\mathrm{S}_{t_k} - \mathrm{S}_{t_{k+1}}\right)^+\right].$$

Two approaches can be considered :

— The hedging costs (put prices) are deducted only afterwards from the portfolio value (which allows an estimation of how much the hedge would cost). In this case, the cushion follows the recursive relation :

$$\mathrm{C}_{t_{k+1}} = \mathrm{C}_{t_k}\left(m\frac{\mathrm{S}_{t_{k+1}}}{\mathrm{S}_{t_k}} + (1-m)e^{r\frac{\mathrm{T}}{\mathrm{N}}}\right)^+$$

The cost of hedging can be computed as the sum of all put options prices necessary for the hedging :

$$\mathrm{C} = \sum_{k=0}^{n-1} m\frac{\mathrm{C}_{t_k}}{\mathrm{S}_{t_k}}\mathbb{E}^{\mathcal{Q}}\left[\left(\left(1-\frac{1}{m}\right)e^{r\frac{\mathrm{T}}{\mathrm{N}}}\mathrm{S}_{t_k} - \mathrm{S}_{t_{k+1}}\right)^+\right].$$

— In practice, the price of the puts used for the hedge will be deducted from the portfolio value at each step. This is translated in the second approach where the cushion dynamics follows the

recursive equation :

$$\tilde{C}_{t_{k+1}} = e^{-r\frac{T}{N}}\tilde{C}_{t_k}\frac{\left(m\frac{S_{t_{k+1}}}{S_{t_k}} + (1-m)e^{r\frac{T}{N}}\right)^+}{\mathbb{E}^{\mathcal{Q}}\left[\left(m\frac{S_{t_{k+1}}}{S_{t_k}} + (1-m)e^{r\frac{T}{N}}\right)^+ |\mathcal{F}_{t_k}\right]}.$$

In Figure 3.13, we compare the effects of hedging using puts. We can notice the following :

— The guarantee is ensured and the manager no longer holds the risk of breaching the floor. However, once the put is exercised and the floor recovered, the manager needs to monetize in order to keep the guarantee until maturity.

— In terms of distributions, the CPPI distribution with a put hedging is a truncation of the classical CPPI where losses are cut (left tail limited by the guarantee).

Gap put option

An alternative risk mitigating action lies in the use of gap options which allow for a protection against sudden significant and persistent downside market moves : if a gap event occurs between two consecutive dates, the buyer receives the difference between the performance of the risky asset at gap $r = \frac{S_t}{S_{t-1}} - 1$ and the threshold J. In case of the CPPI, the proposed solution is a gap put option whose notional is the risky exposure with strike $J = 1/m$, where m is the multiplier. We give in the following the main results for pricing of a gap option from [152]. We let the reader refer to the main article for details.

Suppose that the time to maturity T of a gap option is subdivided onto N periods of length h (e.g. days) : $h = \frac{T}{N}$. The return of the kth period will be denoted by $R_k^{\Delta} = S_{kh}/S_{(k-1)h}$.

Let α denote the return level which triggers the gap event and k^* be the time of first gap expressed in the units of h : $k^* := \inf\{k : R_k^h \leq \alpha\}$. The gap option is an option which pays to its holder the amount $f(R_{k^*}^h)$ at time hk^*, if $k^* \leq N$ and nothing otherwise.

Assuming a deterministic interest rate r and an i.i.d log returns $(R_k^h)_{k=1}^N$ and denote the distribution of $log(R_1^h)$ by $p_h(dx)$. Then the price of a gap option is given by :

$$G_h = e^{rh}\int_{-\infty}^{\beta} f(e^x)p_h(dx)\frac{1-e^{-rT}(\int_{\beta}^{\infty}p_h(dx))^N}{1-e^{-rT}\int_{\beta}^{\infty}p_h(dx)},$$

with $\beta := log(\alpha) < 0$.
It is complicated to obtain numerical results using this expression. Therefore, an approximate formula is used.

Let us assume $S_t = S_0 e^{X_t}$, where X is a Lévy process. Considering the hypothesis $rh \sim 10^{-4}$ and $h \to 0$, the following formula is obtained :

$$G_h \simeq \int_{-\infty}^{\beta} f(e^x)\frac{1-e^{-rT}(\int_{\beta}^{\infty}p_h(dx))^N}{1-e^{-rT}\int_{-\infty}^{\beta}\nu(dx)}$$

Assuming a Kou model (for its tractability and simplicity in integration) and considering the put payoff (i.e $f(x) = (K-x)^+$). The price then becomes :

$$G_h \simeq \frac{\lambda p \eta^-}{1+\eta^-}K^{1+1/\eta^-}\frac{1-e^{-T(r+\lambda p e^{\beta/\eta^-})}}{r+\lambda p e^{\beta/\eta^-}}$$

with p the probability that a given jump is negative, η^- its intensity and λ the poisson process rate. Moreover, for the CPPI we are interested in the payoff $((m-1)e^{rh} - m\frac{S_{kh}}{S_{(k-1)h}})^+$ which is equivalent to $(\frac{m-1}{m}e^{rh} - x)^+$ and thus, $K = (1-1/m)e^{-rh}$

The gap put option allows to cut the loss compensates for the loss as the portfolio value breaches the bond floor. However, insurance investors holding a CPPI who want to hedge it with gap option may face the following issues :

- The price of the gap option is usually sold higher than its theoretical cost for several reasons :
  - The cost of the hedging the gap option for the bank may be quite higher because of the illiquidity of deep out of the money options that replicate it.
  - The replicating formula is tricky to implement and interpret, as significantly model dependent (jumps multiple parameters, lack of robustness).
- Actually, the gap option proposed by the bank might have a different design and payoff from the one considered for the hedge.
- The bank usually hedges the gap up to the first order only.
- The gap risk is borne by the bank only if there is some reconciliation by the insurance company within 24/48 hours, out of which the insurer bears herself the gap risk. As a result, operational risks are significant and represent a major part of the economic capital requirements (e.g. under Solvency II framework).

	Hedging strategies		
	Vanilla Put Hedge 1	Vanilla Put Hedge 2	Gap Option
Mean	136.97	133.35	134.98
5% quantile	218.70	215.40	217.00
0.5% quantilee	277.21	273.53	275.22
Hedging cost	N.C	2.26	1.08

TABLEAU 3.6 – Final value metrics : Comparison between different hedging strategies

3.4 Conclusion

In this article we have presented a study of the CPPI as an insurance contract, a review of its theory and practice as well as its modeling and hedging issues for a risk/return/cost perspective. The main conclusions are :

- Continuous CPPI is only theoretical : given market frictions and the probability of not ensuring the guarantee, all the more that jumps occur more than not.
- As a result, jump processes are a valuable input for the CPPI modeling : they allow to catch a probability of breaching the floor different than zero (even in the continuous-time framework; [83] and [84] came up with the same conclusion) and therefor, detect, define and hedge gap risk.
- Correctly choosing and adjusting the multiplier dynamically significantly reduce the downside risk according to a Value-At-Risk indicator : The multiplier decreases in period of turmoils reducing the risky exposure and increases back during market recovery.
- Hedging the gap risk is possible through two types of options : Vanilla Puts and Gap Put options. The first one is more common due to liquid assets, but the hedging cost may turn out to be too expensive and the maturity too limited. The second type of options is less liquid (bought only through an agreement) but is cheap.

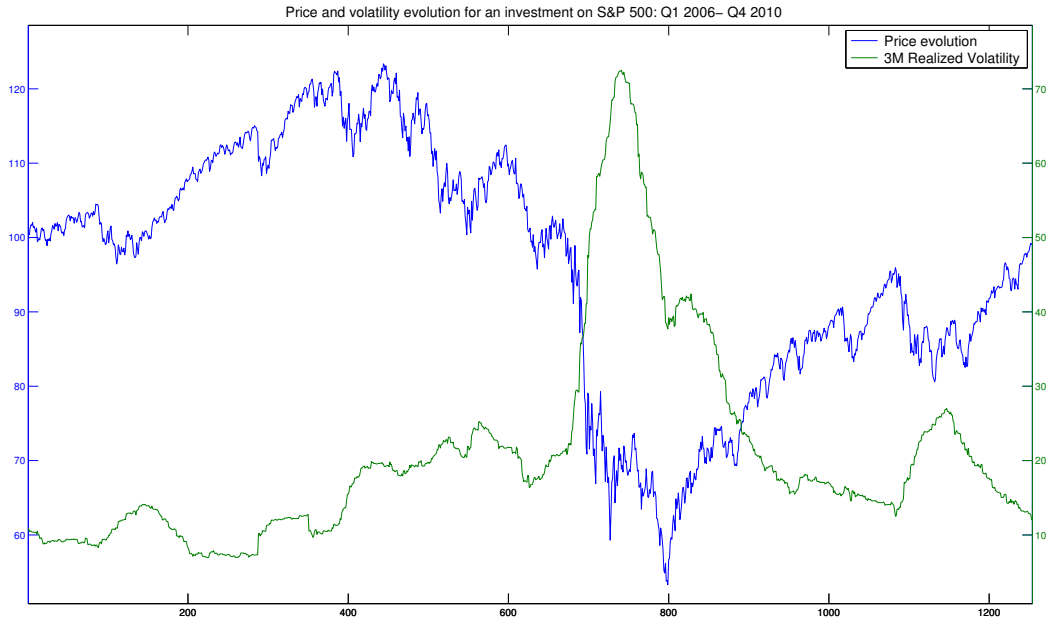


FIGURE 3.1 – Evolution of an investment in the S&P500 for the period Q1 2006 to Q4 2010

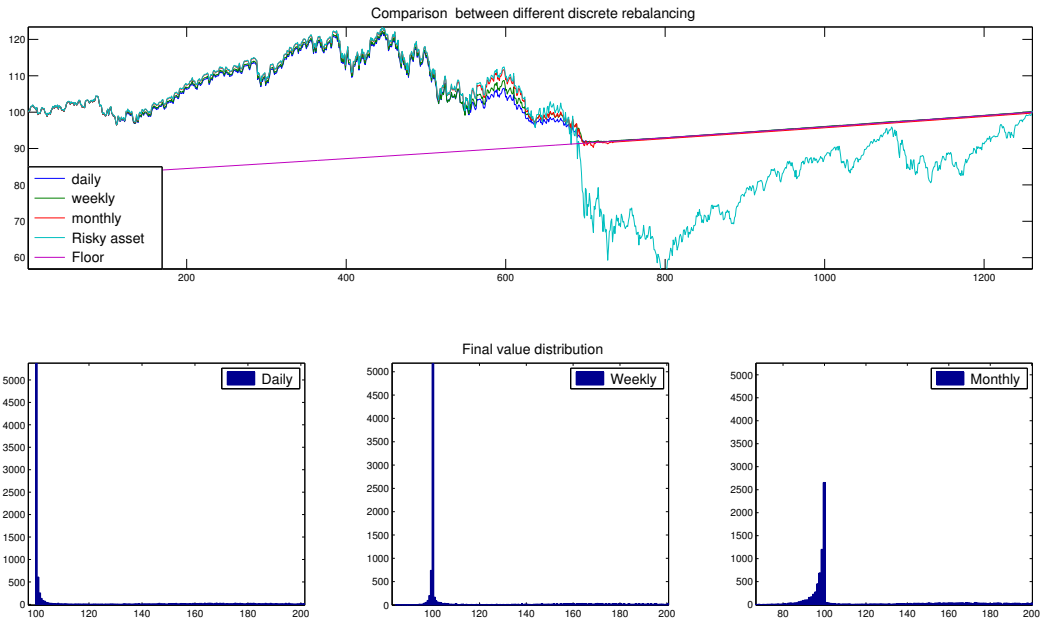


FIGURE 3.2 – Backtesting and distribution of the three various rebalancing frequencies under B&S model.

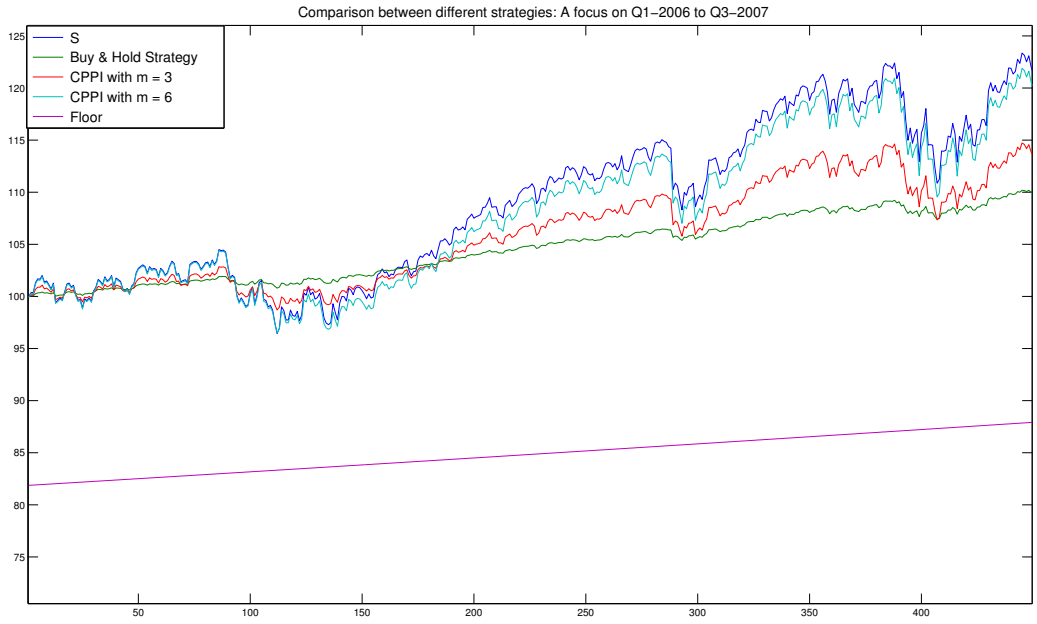


FIGURE 3.3 – Comparison between the Buy & Hold strategy, CPPI with m = 3 and CPPI with = 6 through backtesting (S&P500)

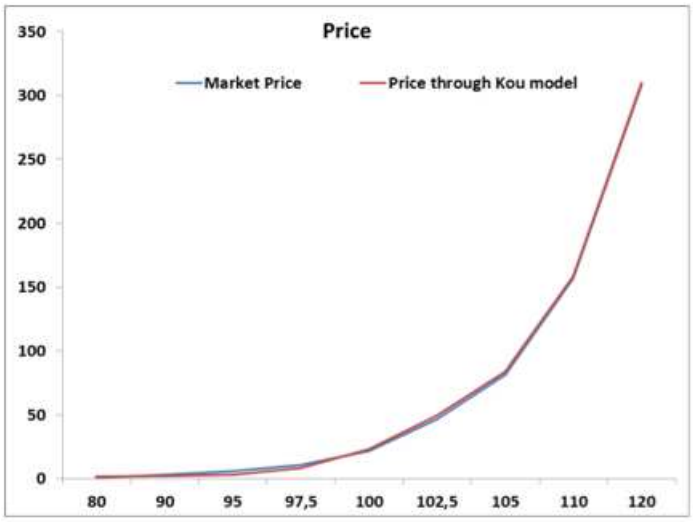


FIGURE 3.4 – Calibration of the Kou model using 1-month maturity call options price on the S&P500

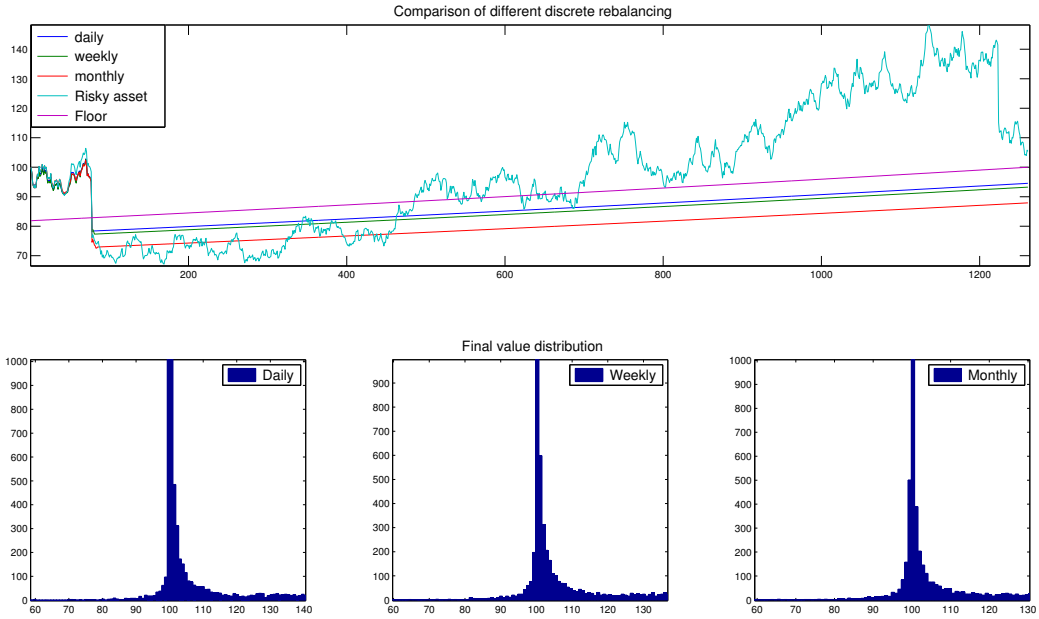


FIGURE 3.5 – Simulation and distribution of the three various rebalancing frequencies under Kou model.

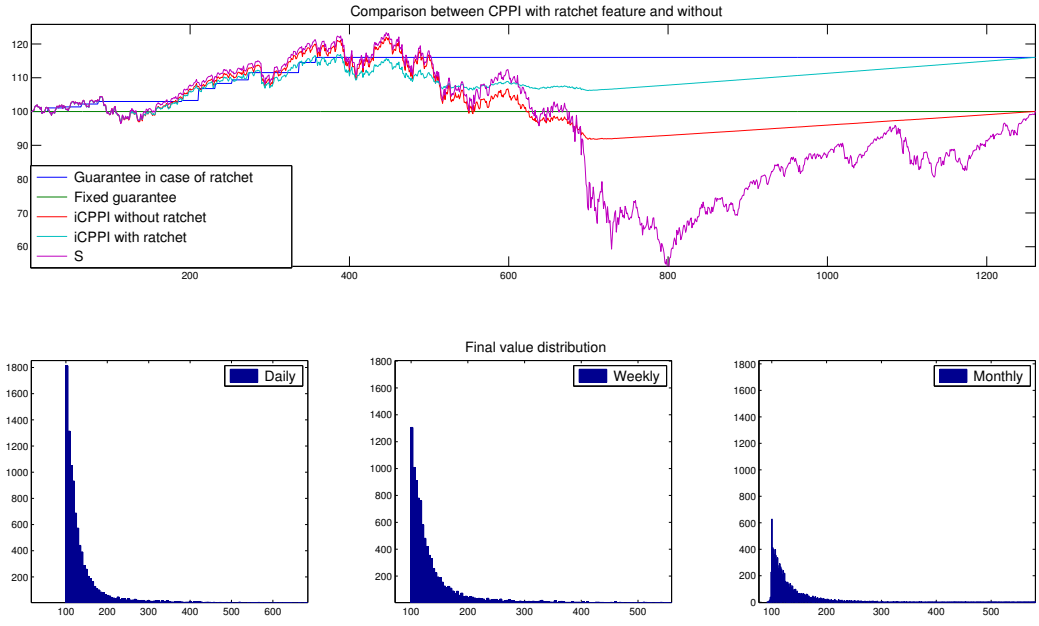


FIGURE 3.6 – The figure on the top is a backtesting on the previous set of data to compare a classical iCPPI and one with the ratchet feature. The three histograms on the bottom are those of the final value distribution for the three different rebalancing frequencies on the iCPPI with ratchet.

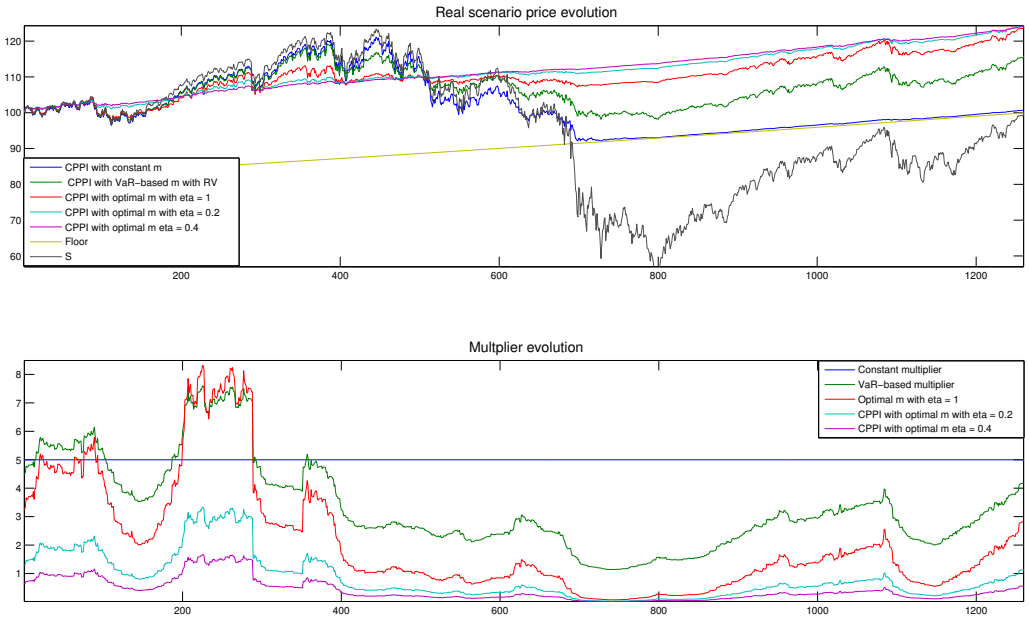


FIGURE 3.7 – Comparison between different multipliers (VaR-based with $p = 99.5\%$ and the optimal one with risk tolerance $\eta = 0.2, 0.4$ and 1) based on Realized Volatility

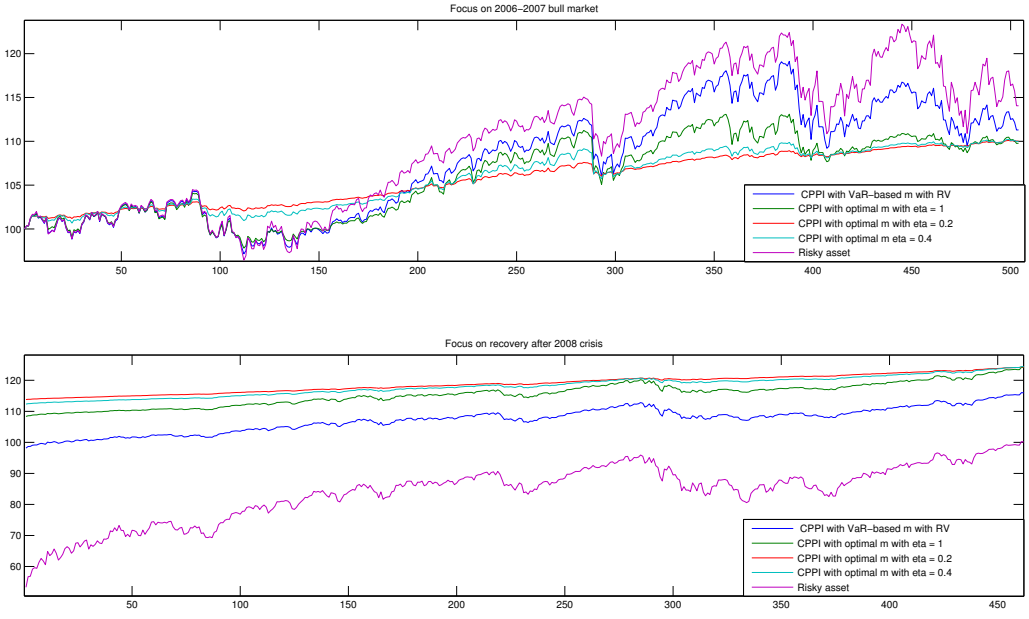


FIGURE 3.8 – Focus on two bullish market periods where the CPPI with the VaR-based m performs better than the optimal one

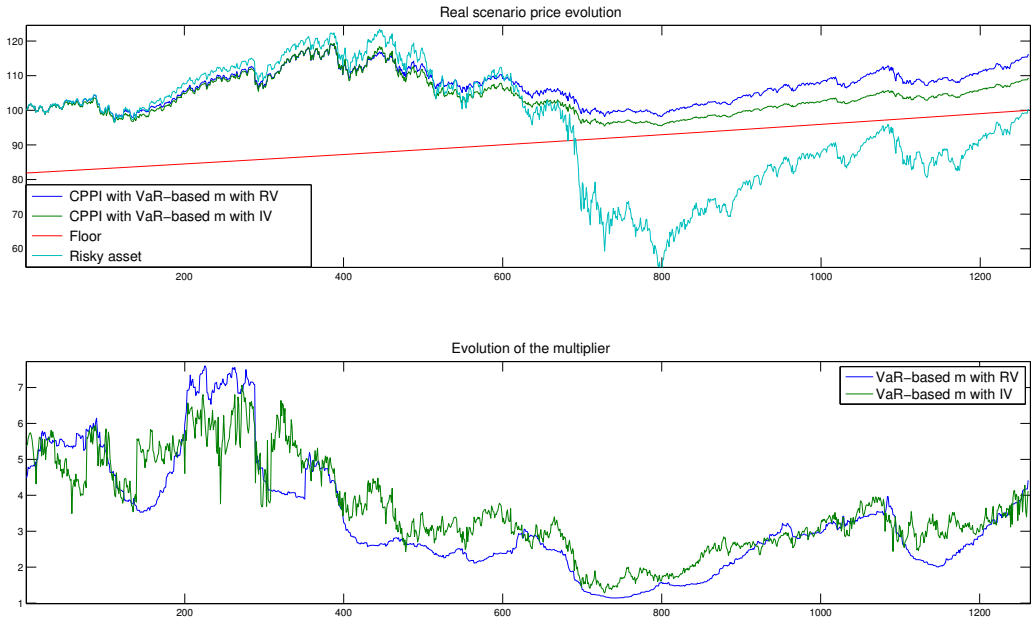


FIGURE 3.9 – Comparison between dynamic multiplier based on RV and on IV through backtesting

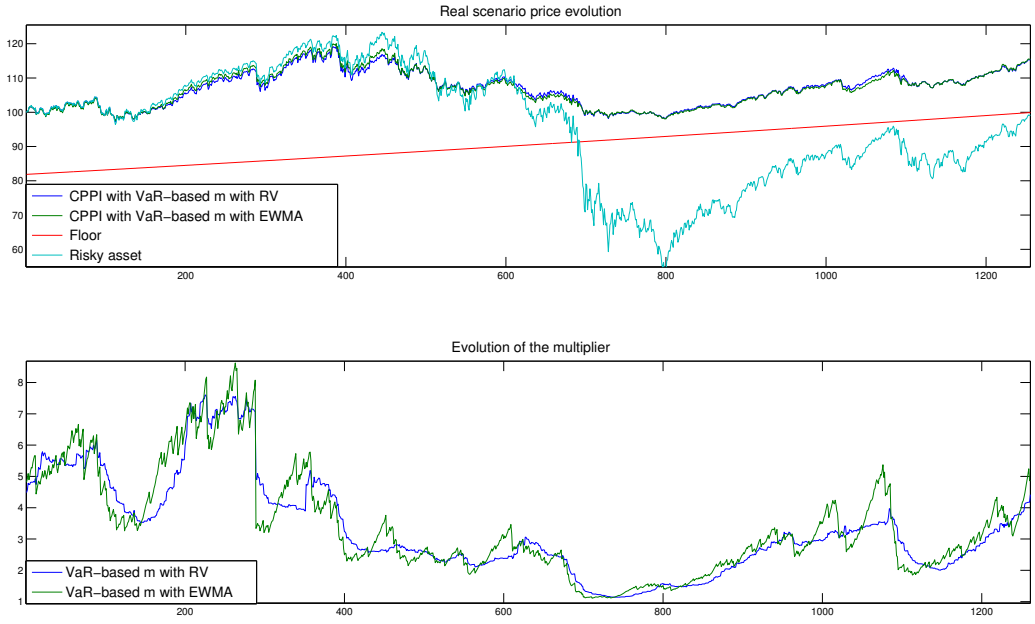


FIGURE 3.10 – Comparison between dynamic multiplier based on RV and on EWMA through backtesting

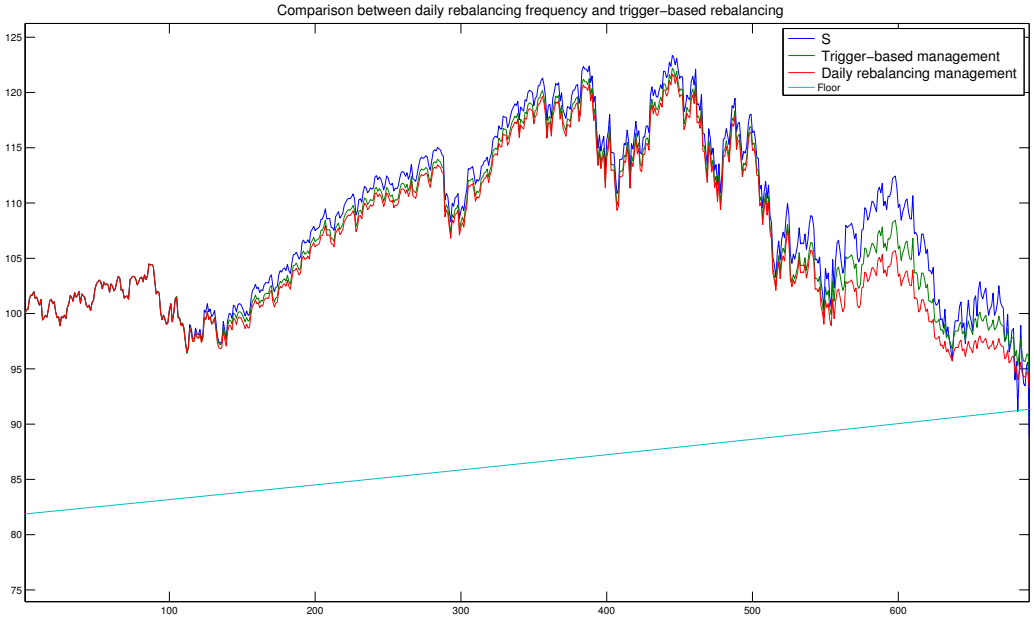


FIGURE 3.11 – Comparison between trigger rebalancing vs fixed frequency rebalancing

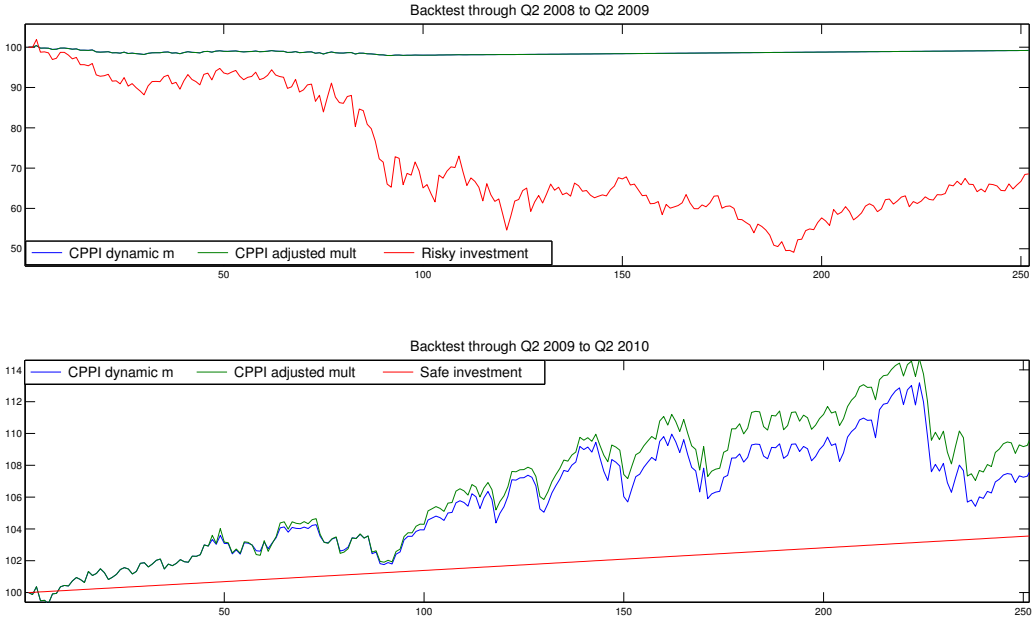


FIGURE 3.12 – Comparison between the dynamic multiplier and an adjusted one based on a manager decision depending on market recovery

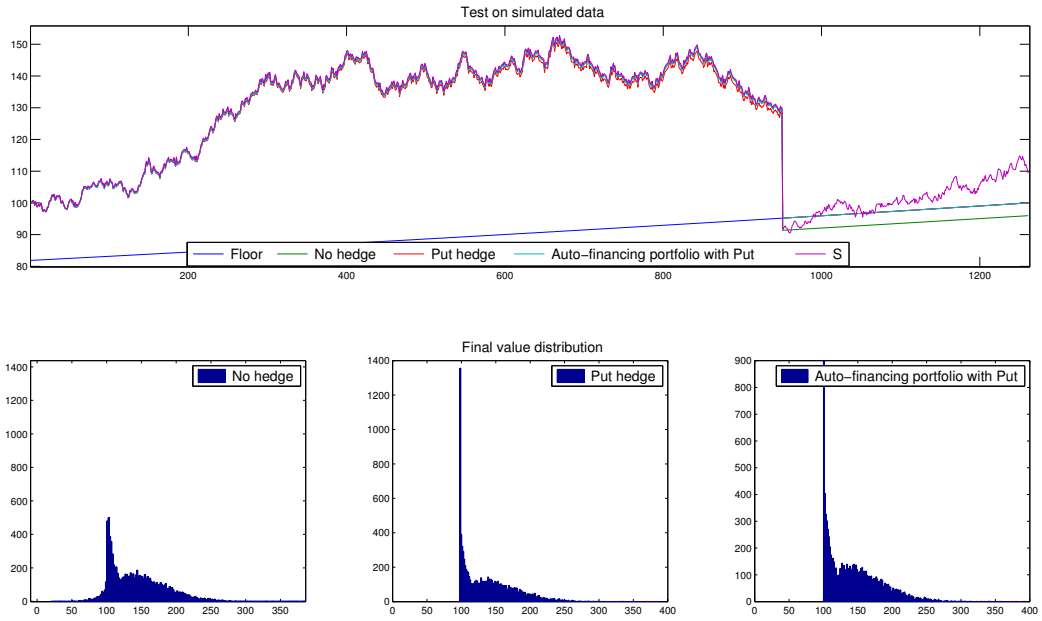


FIGURE 3.13 – Comparison between no hedging and put hedging in its two approaches

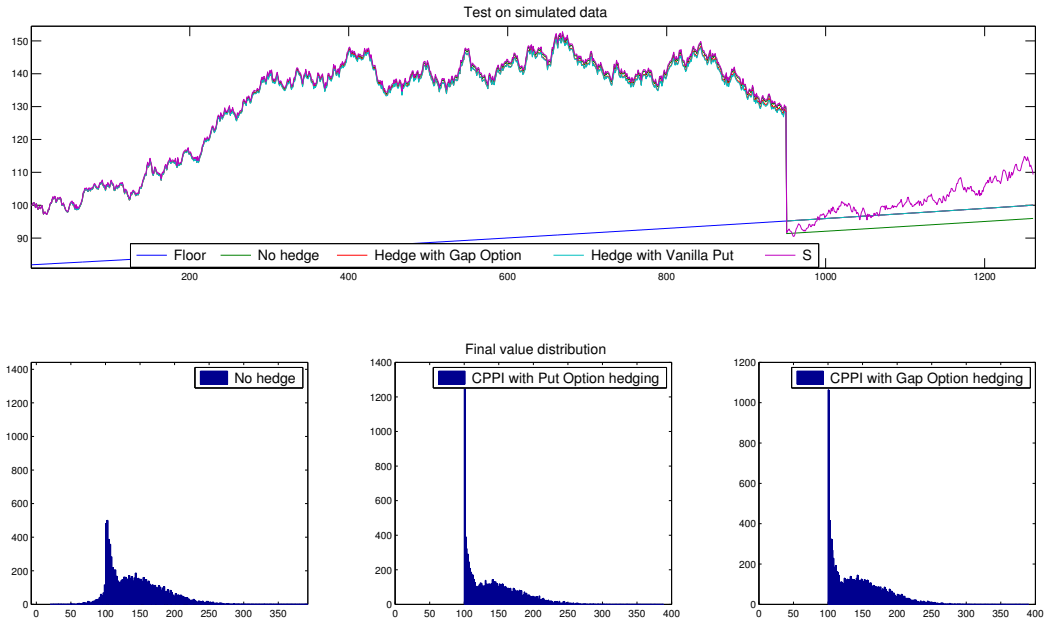


FIGURE 3.14 – Comparison between a vanilla and a gap option hedging

Deuxième partie

Market impact and volatility

Chapitre 4

On Optimal Options Book Execution Strategies with Market Impact

***Abstract*— We consider the optimal execution of a book of options when market impact is a driver of the option price. We aim at minimizing the mean-variance risk criterion for a given impact function. First, we develop a framework to justify the choice of our impact function. Our model is inspired from Leland's option replication with transaction costs. The option effective price is then expressed through a Black-Scholes like PDE with a modified volatility the depends on the size of the trade. We set up a stochastic control framework and solve an Hamilton-Jacobi-Bellman equation using finite differences methods. The simple expected cost problem suggests that the strategy is characterized by a convex increasing trading speed, in contrast to the equity case where the optimal strategy results in a constant trading speed. However, in this framework, the underlying price does not seem to affect the agent's decision. By taking the agent risk aversion into account through the variance, the strategy seems to be more sensitive to the underlying price evolution, urging the agent to trade faster at the beginning of the strategy.**

Keywords : Market impact; option pricing; optimal execution; mean-variance; stochastic control; HJB equation

4.1 Introduction

This chapter addresses the optimal execution of a large portfolio of options, from a starting composition to a specified final one, within a specified period of time. This problem was first introduced by Bertsimas and Lo [27] in the context of a large block of equity portfolio. They define best execution as the dynamic trading strategy that minimizes the expected cost of trading over a fixed period. Given a fixed block of shares and a price-impact function that drives the execution price of an individual trade – that depends on the traded shares and market conditions –, they obtain the optimal *sequence* of trades. Their approach focuses on the expected cost and ignores the volatility of revenues for different trading strategies. Later on, Almgren and Chriss [7] work on a more general framework, using the variance as a suitable penalty for the uncertainty of the cost. They developed the fundamentals of the optimal execution under *market impact* constraints. Seventeen years later, this problem is still attracting the interest of both practitioners and academics. Since then, there was a tremendous growth in optimal trading algorithms, fueled by high frequency trading and powerful computer units which increased market efficiency and decreased the free-lunch opportunities.

4.1.1 Motivation

Derivatives can be used to hedge or to speculate, and many financial institutions include them as a considerable part of their books, occupying a large proportion in their trading activity. Options are common derivatives that allow making profit when market goes down or even sideways. However, the most useful function of options remains the hedging against undesirable movements of an asset. Let us take the example of an insurance company. Insurance liabilities are characterized by three main features : long term time horizon, large volumes and significant market exposure. Given the market uncertainties, the insurance company needs protection strategies in order to hedge its significant risk exposure. In that respect, the use of put options is very common. Buying such hedging portfolios on a significant scale requires taking into account the size of the transaction explicitly.

There is a large literature which develops various parametric models of derivative prices according to the no-arbitrage theory, see [31] and [126]. Bates who surveyed this literature in [17], emphasizes that it cannot fully capture the empirical properties of option prices. He concludes that there is a need for a new approach to pricing derivatives that focuses on the "financial intermediation of the underlying risks by option market-makers", see [17]. In [34], Bondarenko addresses the "overpriced puts puzzle" through studying the historical prices of the S&P 500 put options. He finds that their price has been too high and incompatible with canonical asset-pricing models. Such mismatch is illustrative of a potential equilibrium premium stemming from supply/demand imbalances, which is not explicitly considered by traditional models. Trying to solve this dilemma, Gârleanu and co-authors, see [85], conduct an empirical work to develop a demand-based theory. In their approach, market makers who incur higher unhedgeable risks will move the price up if the net demand is positive and down if it is negative. Even though the equity options market is in zero net supply, negative net demand by the end users in the equity options market, can be viewed as economically equivalent to the positive net supply in the equity market. Another empirical study was conducted in [52] to determine the effect of option and stock illiquidity on delta-hedged equity option returns. They present strong evidence on illiquidity premium in option markets, using effective spreads on a large number of underlying firms with intraday trades and quotes.

The theoretical literature also studied the impact of illiquidity of the underlying stocks on option prices. In a frictionless, complete-market model, the price of the option can be replicated by trading in the underlying asset and risk free bond. If the asset is subject to additional transaction/market impact costs, this should affect the return on options. In [115], Leland provides a theoretical analysis of this effect using a hedging argument. Because option market makers are net long in equity option markets, they need to create a synthetic short option using the underlying stock. Another substantial amount of literature on option pricing and hedging in the presence of market impact has been developed. We mention, among others, [1, 118, 119] in which the problem of option hedging for a large trader who experiences market impact is considered. An Hamilton-Jacobi-Bellman (HJB) equation is derived and a fully non-linear pricing partial differential equation (PDE) is solved. This is known as the feedback effect and yields to an option pricing with the adjusted underlying price. Other references on the subject deserve to be mentioned, see [77, 78, 117].

To our best knowledge, the optimal execution of a portfolio of options under market impact constraints has been forsaken, and our work is the first one to address this problem. Our paper seeks to incorporate the market impact on options into the framework of optimal execution. To do so, the Leland's transaction costs approach will be used to derive the option effective price, and the Almgren expected mean-variance optimization framework will be adopted to define what is the best execution strategy.

4.1.2 A review of the linear market impact model for equity

A market impact model aims at describing the quantitative feedback of a large order, one that affects the execution price and quantifies illiquidity.

In [5] and [7], the price impact is presented as a combination of two components :

— A permanent component that reflects the information transmitted to the market by the buy/sell imbalance.

— A temporary component that reflects the price concession needed to attract counter-parties within a specified short time interval.

We consider an asset S which is traded continuously. The number of shares of the traded asset is described by an absolutely continuous trajectory $t \to x_t$, $\dot{x}_t$, its derivative w.r.t time corresponds to the speed of trading of the security. In the absence of market impact, the asset is modeled by a geometric Brownian motion (GBM)

$$S_t = S_0 e^{-\frac{1}{2}\sigma^2 t + \sigma W_t}. \tag{4.1}$$

When market impact is taken into account, the execution price $\tilde{S}_t$ is defined by

$$\tilde{S}_t = S_t(1 + \eta \dot{x}_t + \gamma(x_t - x_0)), \tag{4.2}$$

where S is the unaffected stock price process, and η and γ are constants.
The term $\eta \dot{x}_t$ corresponds to the temporary or instantaneous impact of trading $\dot{x}_t dt$ shares at time t and only affects this current order. The term $\gamma(x_t - x_0)$ is the permanent price impact which was accumulated by all transactions until time t.

Both impacts in the previous model are linear on the trading volume. In practice, however, even though strong evidence from [100] argues that permanent impact is linear, the linearity of the instantaneous impact is an unrealistic assumption. Perold and Salomon [141] argue that the liquidity premium per share demanded by the market will be either a convex or a concave function of the block size. Other independent empirical studies have demonstrated that the price change induced by the sequential execution of a total volume X follows an approximate $\sqrt{X}$ law, see [16, 44], or of a 3/5 power law, see [8]. In this paper, we focus exclusively on linear price-impact function since our main interest is the optimal execution strategy.

4.1.3 Main results and organization of the paper

In Section 4.2 we derive a put pricing model that takes into account market impact using Leland's transaction costs framework, see [115]. We consider the point of view of a market-maker who sells put options and immediately takes position on the underlying asset to hedge his options. The hedging comes with an additional cost due to trading the asset. The additional cost is then incorporated into the option price. Our option execution price can be expressed through a Black-Scholes like PDE with an enlarged volatility such that

$$\tilde{\sigma}^2 = \sigma^2 + f(t, \dot{x}_t, x_t, \sigma),$$

where σ is the asset volatility and f is the market impact function (depends on time, volatility, inventory and trading speed).

We can either consider the PDE directly or express the put price using the Black-Scholes closed formula with the enlarged volatility when σ is constant. We can also extend the formula using a Taylor

approximation w.r.t the volatility. By doing so, we can recover an additive formulation of the impact function. Either way, we conclude that the impact on the option drives its volatility and that the impact is a function of the option vega.

Once given the put price in presence of market impact, we will set up the optimal execution framework in Section 4.3. We consider an agent who has a fundamental need for option exposure. Such agent is denoted as "end-user" and wishes to acquire a large quantity of put options but needs to minimize the cost of its acquisition. The solution to such optimization leads to a closed formula in the expected cost minimization framework. The trading speed in this case is of the inverse of a 3/2 power law, in contrast to the equity case which is known to result in a constant trading rate strategy.

In Section 4.4 we choose the mean-variance framework as in previous works, see [6, 7, 121]. An approximation to the variance is performed as its drift part is complicated to compute explicitly. Such approximation was already used in the Almgren-Chriss framework. Taking the impact as an additional term in the option price process, we find an HJB equation which leads to a quasi-linear PDE. Using a proper parametrization of the state variable, we are able to reduce the dimension of the problem and solve the PDE numerically through finite differences methods in Section 4.5. Results are presented in Section 4.5.2

Finally, in Section 4.6 we use a semi-Lagrangian approach for the mean-variance framework when the option price follows the PDE formulation. References on this method can be found in [49, 64, 75]. This allows to use a local volatility model and consider permanent impact as well.

4.2 Market impact model : A transaction costs approach

The main objective of the following section is to express the option execution price in the presence of market impact. We build our model upon two areas of the cited literature. Market impact in equity market, and option pricing under transaction costs. The main result of the section is the option price PDE expressed through Formula (4.5).

4.2.1 The transaction costs approach revisited

We consider a market with two agents ; the "option market-maker" (or sell-side trader) who is issuing the vanilla put option and dynamically hedging his position, and the "end-user" (or buy-side trader) who buys the option in order to hedge her fundamental risk. If end-users were able to hedge perfectly – as in a Black-Scholes framework – then option prices are determined by the classical no-arbitrage theory without market impact constraints. In reality, however, perfect hedging is not possible (because of the impossibility of trading continuously, stochastic volatility, jumps, transaction costs, etc...).

In his seminal paper [115], Leland suggests a modified Black-Scholes approach to a contingent claim pricing with proportional transaction costs. Fortunately, we can assimilate market impact to transaction costs. We exploit this framework in order to develop an option pricing model that allows to incorporate the traded quantity as an additional variable.

The dynamic hedging of the option consists in investing a proportion Δ on the underlying and the remaining on a non-risky asset. Let $\hat{T}$ be the maturity of the option, issued at time 0, and let us consider a discrete grid with n revision dates. The non-risky asset is the *numeraire* S_0 and the risky asset is

given under the martingale measure by the SDE

$$dS_t = \sigma S_t dW_t, \quad 0 \le t \le \hat{T}.$$

From the option market-maker perspective, the trading involves proportional impact rate I_0, fixed at the date on which she sells the option (and starts delta-hedging it). This is similar to a transaction cost in the Leland framework. The current value of the portfolio process at time t is defined by

$$V_t^n = V_0^n + \int_0^t \Delta_u^n dS_u - \sum_{t_i \le t} I_0 S_{t_i} \mid \Delta_{i+1}^n - \Delta_i^n \mid, \quad t \le \hat{T} \tag{4.3}$$

where :

— $t_i = t_i^n = i/n$, $0 \le i \le n$, $t_0 = 0$, $t_n = \hat{T}$ are the revision dates.

— $\Delta^n = \Delta_i^n$ on the interval $]t_{i-1}, t_i]$, $\Delta_{n+1}^n := \Delta_n^n$.

— Δ_i^n is $\mathscr{F}_{t_{i-1}}$-measurable.

Δ^n corresponds to the trading strategy. The number of shares of the risky asset that the holder possesses in the period i is then Δ_i^n. The dynamics (4.3) means that the portfolio process V^n is self-financed and in presence of market impact (which is proportional to the traded volume).

When market is complete without friction, the option price is exactly replicated by the terminal value of the self-financing portfolio :

$$V_t = \mathbb{E}[(K - S_{\hat{T}})^+ \mid \mathscr{F}_t] = P(t, S_t) = V_0 + \int_0^t \partial_S P(u, S_u) dS_u, \quad t \le \hat{T},$$

where P is solution of the PDE :

$$\begin{cases} \partial_t P(t,S) + \frac{1}{2}\sigma^2 S^2 \partial_{SS} P(t,S) = 0, & (t,S) \in [0,\hat{T}[\times]0,\infty] \\ P(\hat{T}, s) = (K-s)^+, & s \in]0,\infty[. \end{cases} \tag{4.4}$$

Following Leland's approach, we construct a strategy which can be perceived as a modified-Delta Black-Scholes replication formula. The constant volatility σ is replaced by $\tilde{\sigma}$ in order to compensate for the market impact cost. The "enlarged volatility" $\tilde{\sigma}$ is defined by

$$\tilde{\sigma}^2 = \sigma^2 + \sigma I_0 n^{1/2} \sqrt{\frac{8}{\pi}}.$$

Therefore, the modified option price under market impact follows the PDE :

$$\begin{cases} \partial_t \tilde{P}(t,S) + \frac{1}{2}\tilde{\sigma}^2 S^2 \partial_{SS} \tilde{P}(t,S) = 0, & (t,S) \in [0,\hat{T}[\times]0,\infty] \\ \tilde{P}(\hat{T}, s) = (K-s)^+, & s \in]0,\infty[. \end{cases} \tag{4.5}$$

Let us give, in the following, the intuition behind the result inspired from Leland strategy. Assuming the solution $\tilde{P}$ to (4.5) is smooth enough, we have

$$\tilde{P}(t,S_t) = \tilde{P}(0,S_0) + \int_0^t \partial_S \tilde{P}(u,S_u) dS_u + \frac{1}{2}\int_0^t \left[\sigma^2 - \tilde{\sigma}^2\right] S_u^2 \partial_{SS} \tilde{P}(u,S_u) du.$$

Therefore, $\tilde{P}(t,S_t)$ can be seen as the continuous version of a portfolio process (4.3), provided that $\Delta_i^n = \partial_S \tilde{P}(t_{i-1}, S_{t_{i-1}})$ and the drift term in the formula above corresponds to the cumulative market impact cost. We want to equate the two following increments :

$$\frac{1}{2}\left[\sigma^2 - \tilde{\sigma}^2\right] S_u^2 \partial_{SS} \tilde{P}(u,S_u) du,$$

and

$$-\mathrm{I}_0 \mid \partial_{\mathrm{S}}\tilde{\mathrm{P}}(u+\Delta u, \mathrm{S}_{u+\Delta u}) - \partial_{\mathrm{S}}\tilde{\mathrm{P}}(u, \mathrm{S}_u) \mid \mathrm{S}_{u+\Delta u}.$$

To do so, we use the Taylor expansion

$$\begin{aligned}\partial_{\mathrm{S}}\tilde{\mathrm{P}}(u+\Delta u, \mathrm{S}_{u+\Delta u}) - \partial_{\mathrm{S}}\tilde{\mathrm{P}}(u, \mathrm{S}_u) &\approx \partial_{\mathrm{S},t}\tilde{\mathrm{P}}(u, \mathrm{S}_u)\Delta u + \partial_{\mathrm{SS}}\tilde{\mathrm{P}}(u, \mathrm{S}_u)(\mathrm{S}_{u+\Delta u} - \mathrm{S}_u) \\ &\approx \partial_{\mathrm{SS}}\tilde{\mathrm{P}}(u, \mathrm{S}_u)(\mathrm{S}_{u+\Delta u} - \mathrm{S}_u),\end{aligned}$$

where

$$\mathrm{S}_{u+\Delta u} - \mathrm{S}_u \approx \sigma \mathrm{S}_u(\mathrm{W}_{u+\Delta u} - \mathrm{W}_u).$$

Since $\partial_{\mathrm{SS}}\tilde{\mathrm{P}} \geq 0$, we should look for $\tilde{\sigma}$ such that

$$\frac{1}{2}[\sigma^2 - \tilde{\sigma}^2]\Delta u \approx -\mathrm{I}_0\sigma \mid \mathrm{W}_{u+\Delta u} - \mathrm{W}_u \mid \frac{\mathrm{S}_{u+\Delta u}}{\mathrm{S}_u}.$$

Then, considering the conditional expectation knowing $\mathscr{F}_u$, and the equalities

$$\mathbb{E} \mid \mathrm{W}_{u+\Delta u} - \mathrm{W}_u \mid = \sqrt{\Delta u}\sqrt{\frac{2}{\pi}}, \quad \frac{\mathrm{S}_{u+\Delta u}}{\mathrm{S}_u} = 1 + \sigma(\mathrm{W}_{u+\Delta u} - \mathrm{W}_u),$$

we deduce that

$$\tilde{\sigma}^2 = \sigma^2 + \mathrm{I}_0 n^{1/2}\sqrt{\frac{8}{\pi}}\sigma. \tag{4.6}$$

Remarks 1. *:*

— *The mathematical justification of the approximate replication principle in the Leland transaction costs framework brought a lot of attention and turned out to be quite difficult to obtain. The replication principle fails to be true for constant transaction costs. In particular, the agent should choose and fix the number of revision dates n large enough so that the mean square hedging error is controlled as there is a limit error as $n \to \infty$. In [122], Lott gives the first rigorous result : the approximation error tends to zero in probability if the transaction costs coefficient depends on n and decreases on a rate proportion to $n^{-1/2}$ (in this case, $\tilde{\sigma}$ does not depend on n).*

— *To solve the previous issues, we will fix the number of revision dates to be of a reasonable order to provide a reasonable price. We do not aim at reviewing the foundations of the replication issues of the Leland framework, instead, our main interest is finding a simple and intuitive pricing formula for the option market-maker, one that takes into account additional fees due to market impact. As a consequence, the end-user will pay the option at a higher price than what the Black-Scholes pricing model suggests. The quantity of his trades will be an important factor.*

— *In the Black-Scholes framework, the implied volatility is constant, and so is the "enlarged volatility". In practice, however, the implied volatility is not constant, i.e depends on the maturity and the moneyness. One of the common market practices is to take local volatility models. Fortunately, the Leland approach is still valid in this case as showed in [117]. One just needs to replace the constant volatility σ by a local volatility one $\sigma(t, \mathrm{S}_t)$. The justification above remains the same.*

4.2.2 The option price under market impact and the market impact function

In the following, we properly link the market impact framework to the transaction costs one. We consider a universe with the two agents; option market-makers and end-users. The end-user acquires, at time $t < \hat{T}$ (where $\hat{T}$ is the maturity of the option at time 0, and $\tau = \hat{T} - t$ is the residual time to maturity at time t), x_t options with a trading speed $\dot{x}_t$. The option market-maker, on the other hand, meets the end-user's demand by selling the exact quantity at the exact same trading speed on the "option market". He then hedges these options on the "equity market" [1]. In doing so, he expects the asset price to be affected. The market impact (on the asset) is decomposed, as exposed in the introduction through Equation (4.2), into a temporary I^{temp} and a permanent I^{perm} impact, and both are linear on the quantity. At the infinitesimal time between t and $t + dt$, the temporary impact resulting in selling $dx_t = \dot{x}_t dt$ options is expressed by

$$I_t^{temp} dt = \eta \dot{x}_t dt.$$

The permanent impact is a result of all previous trades up to time t. Note that at time t, the market-maker has already sold $(x_t - x_0)$ options and took position on the underlying asset to hedge them. Therefore, between t and $t + dt$ we have

$$I_t^{perm} dt = \gamma(x_t - x_0) dt$$

The combination of the two impacts gives the total market impact rate at time t

$$I_t = \eta \dot{x}_t + \gamma(x_t - x_0). \tag{4.7}$$

Without adding complexity, the reasoning in Section 4.2.1 is straightforward in this case. We adapt the formula of the augmented volatility expression in Equation (4.6) at the arbitrary time $t \in [0, T]$, where $T < \hat{T}$ is the end time of the trading strategy. The revision step h is fixed instead of the number of revision dates, regardless of the time-to-maturity $\tau = \hat{T} - t$ of the option. The number of revision dates becomes $n = \frac{\hat{T}-t}{h}$ and the augmented volatility at an arbitrary time t can be expressed as

$$\tilde{\sigma}_t^2 = \sigma^2 + (\tilde{\eta}\dot{x}_t + \tilde{\gamma}(x_t - x_0))\sqrt{\hat{T} - t}\sigma, \tag{4.8}$$

where $\tilde{\eta} = \eta\sqrt{\frac{8}{h\pi}}$ and $\tilde{\gamma} = \gamma\sqrt{\frac{8}{h\pi}}$ are constants depending on the underlying impact factors η and γ and the revision step h is fixed to be one day (meaning that the option market-maker will revise her hedging position on a daily basis). Finally, the option price under market impact $\tilde{P}$ follows the PDE :

$$\begin{cases} \partial_u \tilde{P}(u, S) + \frac{1}{2}\tilde{\sigma}_t^2 S^2 \partial_{SS} \tilde{P}(u, S) = 0, \quad (u, S) \in [t, \hat{T}[\times]0, \infty] \\ \tilde{P}(\hat{T}, s) = (K - s)^+. \end{cases} \tag{4.9}$$

Remark 6. *The term $\tilde{\gamma}(x_t - x_0)$ comes from previous trades (hedging options sold before 0 and t). By doing so, to the value of the "imaginary" asset price becomes $\tilde{S}_{t^-} = S_t + \gamma(x_t - x_0)S_t$. Then, by selling $\dot{x}_t dt$ at time t, and taking position of the underlying asset, the affected asset price becomes $\tilde{S}_t = \tilde{S}_{t^-} + \eta\dot{x}_t S_t$. Since the replication strategy at time t is based on the asset price S_t at time t, the impact term in the enlarged volatility is proportional to $\tilde{S}_t - S_t$, and therefore we have Equation (4.8).*

Proposition 1. *The execution price $\tilde{P}$ of an option sold at time t can be written using Black-Scholes closed-formula with $\tilde{\sigma}_t$ as the volatility parameter :*

$$\tilde{P}(t, S_t) = BS(t, S_t, K, r = 0, \hat{T}, \tilde{\sigma}_t),$$

1. We consider that the option market, in which the option is traded, is separated from the equity market, in which the option is hedged. This allows to justify the fact that the buy-side trader does not observe any changes in the asset price.

where BS *corresponds to the Black-Scholes closed-formula of the put option price,* S_t *is the option spot,* K *its strike,* r *the risk-free rate chosen to be* 0, $\hat{T}$ *the option maturity and* $\tilde{\sigma}_t$ *the enlarged volatility defined by*

$$\tilde{\sigma}_t^2 := \sigma^2 + (\tilde{\eta}\dot{x}_t + \tilde{\gamma}(x_t - x_0))\sqrt{\hat{T} - t}\sigma.$$

Using a simple Taylor approximation to the first order around the spot variance $V = \sigma^2$, and assuming the impact term is negligible to σ^2, we can rewrite the above expression as a sum of the Black-Scholes option price with volatility σ and an additional term corresponding to the option market impact :

$$\begin{aligned}\tilde{P}(t,S_t) &\approx P(t,S_t) + (\tilde{\sigma}_t^2 - \sigma^2)\partial_V P(t,S_t) \\ &\approx P(t,S_t) + \frac{1}{2}\left\{\tilde{\eta}\dot{x}_t + \tilde{\gamma}(x_t - x_0)\right\}\sqrt{\hat{T} - t}\nu(t,S_t),\end{aligned} \tag{4.10}$$

where $\nu(t,S_t) = \partial_\sigma P$ is the Black-Scholes vega of the option, calculated on the constant volatility σ

$$\nu(t,S_t) = \sqrt{\hat{T} - t}S_t N'(d_1) = \sqrt{\hat{T} - t}KN'(d_2),$$

where

$$\begin{aligned}N'(x) &= \frac{1}{\sqrt{2\pi}}e^{-x^2/2} \\ d_1 &= \frac{\log\frac{S_t}{K} + \frac{1}{2}\sigma^2(\hat{T} - t)}{\sigma\sqrt{\hat{T} - t}} \\ d_2 &= \frac{\log\frac{S_t}{K} - \frac{1}{2}\sigma^2(\hat{T} - t)}{\sigma\sqrt{\hat{T} - t}} = d_1 - \sigma\sqrt{\hat{T} - t}.\end{aligned}$$

It follows that the impact function is defined by

$$f(t,S_t,\dot{x}_t,x_t) := \tilde{P}(t,S_t) - P(t,S_t) = \frac{1}{2}\left\{\tilde{\eta}\dot{x}_t + \tilde{\gamma}(x_t - x_0)\right\}\sqrt{\hat{T} - t}\nu(t,S_t). \tag{4.11}$$

Remarks 2. *:*

— *The impact term on the underlying asset is not observed by the end-user. This assumption is made for the sake of deriving the market impact function on the option only. In fact, the sell-side trader bases her hedging portfolio on an "imaginary" asset whose price is subject to market impact. This effect will not be seen in the "real" asset and will not be perceived by the end-user. We can see it as if the market in which options are sold is separated from the market where the underlying asset is traded for the hedging. Such an assumption is not absurd given that the end-user is assumed to trade the option exclusively.*

— *Equation (4.10) allows to write the option execution price as the sum of martingale (the option price in the absence of market impact, i.e* $\eta = \gamma = 0$*) plus a positive term reflecting the additional cost due to market impact. The end-user trades the option exclusively and pays the additional hedging costs fixed by the sell-side trader and given by Equation (4.11). Of course the PDE (4.9) is more precise and allows to deal with both the Black-Scholes case and a local volatility model. However, the additive formula is more convenient to build an Almgren-like optimal execution framework.*

— *The vega-dependent impact function is a result of the agent adjusting the replicating portfolio according the "imaginary" price movement.*

— *The number of revision dates n can be fixed instead of the revision step h, which has the advantage of suppressing the term* $\sqrt{\hat{T}-t}$ *in the formula. However, by doing so, the sell-side trader needs to adjust the revision step* $h = \frac{\hat{T}-t}{n}$ *which decreases as time increases. The first option seems, however, more logical. The sell-side trader will revise her position on a daily basis for example (i.e* $h = 1/252$*).*

— *Using the Vega-Gamma relationship (i.e* $\nu = \sigma\tau S^2\Gamma$*), we can rewrite the impact term as a function of the option Gamma :*

$$f(t,S_t,\dot{x}_t,x_t) = \frac{1}{2}\left\{\tilde{\eta}\dot{x}_t + \tilde{\gamma}(x_t - x_0)\right\}\sigma S_t^2\left(\hat{T}-t\right)^{3/2}\Gamma(t,S_t), \tag{4.12}$$

where $\Gamma(t,S_t) = \frac{N'(d_1)}{S_t\sigma\sqrt{\hat{T}-t}} = \frac{KN'(d_2)}{S_t^2\sigma\sqrt{\hat{T}-t}}$.

Proposition 2. *In the Black & Scholes framework, the put option effective price is written as the following :*

$$\tilde{P}(t,S_t,\dot{x}_t,x_t) = P(t,S_t) + \frac{1}{2}\left\{\tilde{\eta}\dot{x}_t + \tilde{\gamma}(x_t - x_0)\right\}\sigma S_t^2\left(\hat{T}-t\right)^{3/2}\Gamma(t,S_t), \tag{4.13}$$

where :

— $\tilde{\eta}$ *controls the temporary impact strength in* $\$ \times hour/N$ *of options.*

— $\tilde{\gamma}$ *controls the permanent impact strength and is in* $\$N$ *shares.*

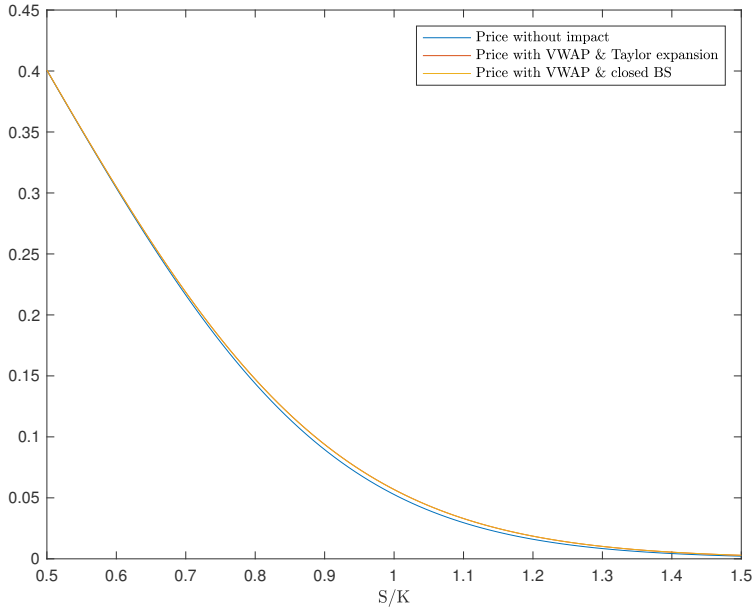
— x_t *is the quantity held at time* t *and* $\dot{x}_t$ *is the speed of trading in number of options per time unit.*

— Γ *is the delta sensitivity w.r.t to the asset price (*$\Gamma(t,S_t) > 0$*).*

Remark 7. *Buying the option (i.e* $\dot{x}_t > 0$*) will typically lead to increasing its price, hence the execution price* $\tilde{P}_t \geq P_t$ *for* $t \in [0,T]$.

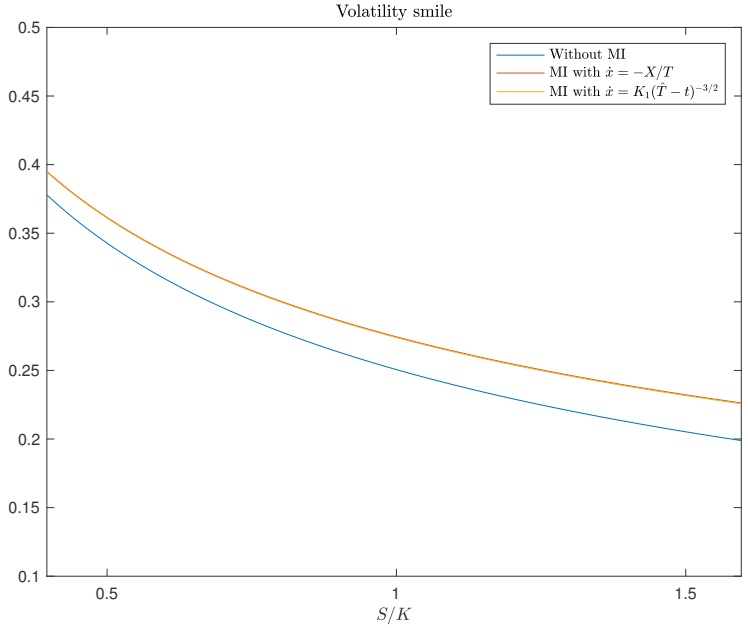
Numerical experiment :
We illustrate in Figure 4.1 the put price as a function of the moneyness for the Black-Scholes case. We seek to compare the price with and without market impact and justify the use of the Taylor expansion. We consider that the rate of trading is constant. The figure shows that the price when the option is subject to market impact is greater than the price without market impact. This is what we expect when we have a demand pressure on the option. On the other hand, the approximation by the Taylor expansion is almost equal to the closed formula. This will be very convenient to solve the optimal trade execution problem.

FIGURE 4.1 – Put price as a function of the moneyness.



The Black-Scholes case results in a shift in the implied volatility (but a constant one in both cases). Solving PDE 4.9 with a Constant Elasticity Volatility model (CEV) gives the illustrations in Figure 4.2. In addition to a higher volatility across the moneyness, the increase in the implied volatility is larger for out of the money options.

FIGURE 4.2 – The CEV model volatility smile and put price evolution



4.3 The optimal execution problem

In the previous section, we managed to define an effective price for an option under market impact. We assumed that the issuer prices the option by taking its replication price and an adjustment to his position when the underlying price is impacted by the order. We pointed out an important assumption; we described the impact on the asset as being "imaginary", meaning that the end-user does not see the asset price moving because of the strategy. The option price under market impact was found to verify a Black-Scholes like PDE with an "enlarged" volatility. The dependence to the position x and trading speed $\dot{x}$ is linear and increasing (the higher the trading speed and quantity, the higher the volatility and thus the option price).

Particularly, an agent who is willing to trade a large quantity of options will see the impact as an important dilemma. If he trades rapidly, then the actual cost of her strategy will be different from the one she anticipated. In real life, the agent is exposed to price manipulation, liquidity issues and market impact, especially for large trades; the cost of placing one large order to close his position will be far greater than the sum of infinitely small orders differed in time.

Many works have been held, in the equity market, to show that because of the trades size of typical institutional investors, they are usually broken up into smaller ones and executed over the course of several days, see for example [48, 108]. Chan and Lakonishok [48] for example, show that only 20% of the market value of the trades splits in their set of data are completed within a day, and that over 53% are spread over four trading days or more. For this reason, best execution can not be defined as a single number in the context of a single trade. It is a strategy that unfolds over the course of several days and which ought to adapt to changing market conditions. This is intuitively true for market options as well, even though very little literature deals with this issue.

Taking this into account, instead of executing his orders at once, the agent has to split them over the time interval $[0,\mathrm{T}]$ by means of a dynamic order execution strategy.

4.3.1 The general framework

Let us consider a *trade execution strategy* in which an initial long or short position of X options with fixed strike K and maturity $\hat{\mathrm{T}}$ is liquidated by a fixed time horizon $[0,\mathrm{T}]$, where $\mathrm{T} < \hat{\mathrm{T}}$ is the end time. We describe such a strategy by the asset position x_t held at time $t \in [0,\mathrm{T}]$. The initial position x_0 is positive for sell strategy and negative for buy strategy. The condition $x_{\mathrm{T}^+} = 0$ assures that the initial position has been unwound by time T. The path $x = (x_t)_{t\in[0,\mathrm{T}]}$ will be nonincreasing for pure sell strategy and nondecreasing for a pure buy strategy.

We restrict our framework to pure buy strategies. The end-user's purpose is to hedge the risk of a complex product (structured product, Variable Annuity, etc...) indexed on an underlying asset, by acquiring vanilla put options on that same underlying asset. Let $(\Omega,\mathscr{F},\mathbb{P})$ be the usual probability space on the filtration $(\mathscr{F}_t)_{t\in\mathbb{R}_+}$ satisfying the usual assumptions. In the absence of market impact and under a null risk-free rate, the no-arbitrage price of a put option is defined by :

$$\mathrm{P}_t = \mathbb{E}_{\mathbb{Q}}\left[(\mathrm{K} - \mathrm{S}_{\hat{\mathrm{T}}})^+ \mid \mathscr{F}_t\right]$$

where $\mathbb{E}_{\mathbb{Q}}$ is the expectation under the risk-neutral probability measure $\mathbb{Q}$ in which the asset price is a martingale. At each time t, $\dot{x}_t dt$ options are bought at price $\tilde{\mathrm{P}}_t$. The effective price $\tilde{\mathrm{P}}$ is the option impact price and can be either defined by PDE (4.9) or through the reduced to Equation (4.13). Thus,

the cost arising from the strategy x is

$$\mathscr{C}(x) := \int_0^{\mathrm{T}} \tilde{\mathrm{P}}_t \dot{x}_t dt$$

The agent's objective is then to minimize a certain objective function, which takes into account his risk aversion, and may involve both cost and risk terms, over the class of admissible trading strategies x with side conditions $x_0 = \mathrm{X}$ and $x_{\mathrm{T}} = 0$. This is known as the *optimal trade execution* problem.

In this paper we will treat two cases :

— The expected cost $\mathbb{E}[\mathscr{C}(x)]$

— The mean-variance case $\mathbb{E}[\mathscr{C}(x)] + \lambda \mathbb{V}ar[\mathscr{C}(x)]$

Except from Section 4.6, we will develop the framework under the Black & Scholes case. $\tilde{\mathrm{P}}_t$ will be defined by Equation (4.10), and permanent impact will be excluded, i.e $\tilde{\gamma} = 0$.

4.3.2 The Black-Scholes framework under a temporary market impact

The effective price $\tilde{\mathrm{P}}_t$ is given by Equation (4.13) with $\tilde{\gamma} = 0$ is reduced to :

$$\tilde{\mathrm{P}}_t = \mathrm{P}_t + \frac{1}{2}\tilde{\eta}\dot{x}_t \sigma \mathrm{S}_t^2 \left(\hat{\mathrm{T}} - t\right)^{3/2} \Gamma(t, \mathrm{S}_t). \tag{4.14}$$

We can rewrite the cost function as the following :

$$\mathscr{C}(x) = \int_0^{\mathrm{T}} \mathrm{P}_t \dot{x}_t dt + \frac{1}{2}\tilde{\eta} \int_0^{\mathrm{T}} \dot{x}_t^2 \sigma \mathrm{S}_t^2 \left(\hat{\mathrm{T}} - t\right)^{3/2} \Gamma(t, \mathrm{S}_t) dt.$$

Using a simple integration by part and Ito's formula, the cost arising from the strategy x becomes :

$$\mathscr{C}(x) = -\mathrm{XP}_0 - \int_0^{\mathrm{T}} \sigma x_t \mathrm{S}_t \Delta(t, \mathrm{S}_t) d\mathrm{W}_t + \frac{1}{2}\tilde{\eta}\sigma \int_0^{\mathrm{T}} \dot{x}_t^2 \mathrm{S}_t^2 \left(\hat{\mathrm{T}} - t\right)^{3/2} \Gamma(t, \mathrm{S}_t) dt, \tag{4.15}$$

where Δ is the Black-Scholes delta of the option calculated on σ.

The expected cost of strategy x is then

$$\mathbb{E}[\mathscr{C}(x)] = -\mathrm{XP}_0 + \frac{1}{2}\tilde{\eta}\mathbb{E}\left[\int_0^{\mathrm{T}} \dot{x}_t^2 \mathrm{S}_t^2 \left(\hat{\mathrm{T}} - t\right)^{3/2} \Gamma(t, \mathrm{S}_t) dt\right]. \tag{4.16}$$

Theorem 1. *The optimal strategy x^* resulting in minimizing the expected cost under the Black & Scholes framework given by Equation (4.16) is characterized by :*

$$\dot{x}^*(t) = \frac{\mathrm{K}_1}{(\hat{\mathrm{T}} - t)^{3/2}}$$
$$x^*(t) = \frac{\mathrm{K}_1}{(\hat{\mathrm{T}} - t)^{1/2}} + \mathrm{K}_2$$

where $\mathrm{K}_1 = \frac{\mathrm{X}}{2\left(\hat{\mathrm{T}}^{-\frac{1}{2}} - (\hat{\mathrm{T}} - \mathrm{T})^{-\frac{1}{2}}\right)}$ *and* $\mathrm{K}_2 = -2\mathrm{K}_1(\hat{\mathrm{T}} - \mathrm{T})^{-1/2}$.

Démonstration. Under the Black-Scholes framework, $S_t^2\Gamma(t,S_t)$ is a martingale which allows to write

$$\min_x \mathbb{E}[\mathscr{C}(x)] = \frac{1}{2}\tilde{\eta}S_0^2\Gamma(0,S_0)\min_x \mathbb{E}\left[\int_0^T (\hat{T}-t)^{3/2}\dot{x}_t^2 dt\right].$$

Assuming the strategy x is deterministic, the problem is reduced to solving

$$\min_x \int_0^T (\hat{T}-t)^{3/2}\dot{x}(t)^2 dt. \tag{4.17}$$

Using the calculus of variations, we find that the solution to (4.17) verifies the Euler-Lagrange equation

$$\frac{d}{dt}(2(\hat{T}-t)^{3/2}\dot{x}(t)) = 0,$$

along with the boundary conditions

$$x(0) = X, \quad x(T) = 0.$$

Which gives

$$\dot{x}^*(t) = \frac{K_1}{2(\hat{T}-t)^{3/2}}$$
$$x^*(t) = \frac{K_1}{(\hat{T}-t)^{1/2}} + K_2,$$

where $K_1 = \frac{X}{2(\hat{T}^{-1/2}-(\hat{T}-T)^{-1/2})}$ and $K_2 = -2K_1(\hat{T}-T)^{-1/2}$.
Moreover, using the Cauchy-Schwarz inequality

$$\int_0^T f^2(t)dt \int_0^T g^2(t)dt \geq \left(\int_0^T f(t)g(t)dt\right)^2,$$

where $f(t) = \frac{1}{(\hat{T}-t)^{3/4}}$ and $g(t) = (\hat{T}-t)^{3/4}\dot{x}_t$, we have

$$\mathbb{E}\left[\int_0^T (\hat{T}-t)^{3/2}\dot{x}_t^2 dt\right] \geq \frac{X^2}{\int_0^T (\hat{T}-t)^{-3/2}dt}.$$

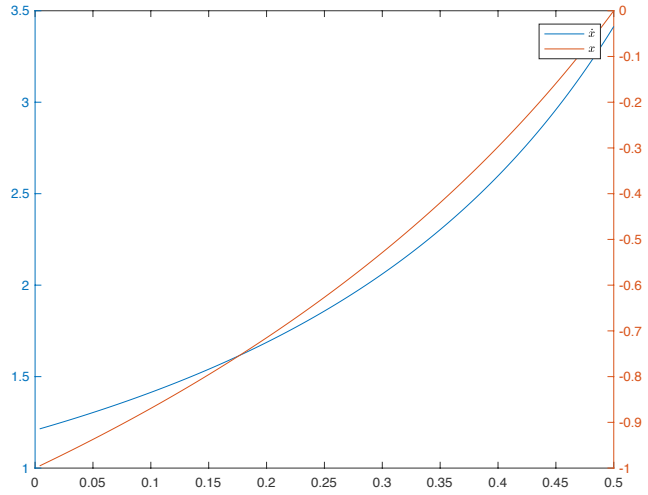
And the equality holds for $(\dot{x}^*, x^*)$. □

Remarks 3. *:*

— *We recall that the expected cost optimal strategy for the equity case is characterized by having a constant trading rate $\dot{x}_t^* = -\frac{X}{T}$, as shown in [27] in a discrete-time setting. In the option framework under the impact function we select, the trading speed is an increasing convex function of time.*

— *The expected cost under the Black & Scholes framework is the only case where a closed solution can be found. This strategy is illustrated in Figure 4.3 for $\hat{t} = 1$, $T = 0.5$ and $X = -1$. We see that both the trading rate and the inventory are increasing and convex. It is to mention that the optimal strategy arising from the expected cost framework does not depend on the underlying asset price.*

FIGURE 4.3 – The trading strategy for an expected cost framework under the Black-Scholes setting.



In the following sections, we develop the optimal execution framework under a risk/reward criterion under the Black & Scholes framework and the impact price defined by Equation (4.10).

4.4 Adding the agent risk aversion : A mean-variance framework

The expected cost is usually used for an agent who does not monitor the risk of his strategy. Investors, however, usually takes into account their risk aversion, through utility functions or using risk/reward criterion such as the mean-variance. The literature on these problems is rich for optimal execution of a book of equity shares. For example Almgren and Chriss [7], and Forsyth [75], study the mean-variance optimal execution problem. In [90], Gatheral and Schied take the time-average value-at-risk associated with the P&L of the position, while Forsyth and co-authors, see [76], use a quadratic variance as a risk criterion. In this paper, we focus on the mean-variance criterion in light of [6, 7]. The mean-variance of the cost of trading is defined by :

$$\mathbb{E}\left[\mathscr{C}(x)\right] + \tilde{\lambda}\mathbb{V}\left[\mathscr{C}(x)\right],$$

where $\tilde{\lambda} > 0$ is the variance penalty. The choice of this coefficient can not be explained in terms of fundamental investment preferences. The value is chosen in order to obtain solutions that bring out a certain meaning to the optimization problem.

The variance of the cost function term can be written as the following

$$\begin{aligned}
\mathbb{V}\left[\mathscr{C}(x)\right] &= \mathbb{E}\Big[\Big(\int_0^{\mathrm{T}} \tilde{\mathrm{P}}_t \dot{x}_t dt - \mathbb{E}\Big[\int_0^{\mathrm{T}} \tilde{\mathrm{P}}_t \dot{x}_t dt\Big]\Big)^2\Big] \\
&= \mathbb{E}\Big[\Big(\int_0^{\mathrm{T}} \frac{1}{2} x_t(\tilde{\sigma}_t^2 - \sigma^2)\mathrm{S}_t^2 \partial_{\mathrm{SS}}\tilde{\mathrm{P}}(t,\mathrm{S}_t)dt - \int_0^{\mathrm{T}} x_t \sigma \mathrm{S}_t \partial_{\mathrm{S}}\tilde{\mathrm{P}}(t,\mathrm{S}_t)d\mathrm{W}_t\Big)^2\Big] \\
&= \mathbb{E}\left[\int_0^{\mathrm{T}} x_t^2 \sigma^2 \mathrm{S}_t^2 \partial_{\mathrm{S}}\tilde{\mathrm{P}}^2(t,\mathrm{S}_t)dt\right] \\
&\quad + \{\text{terms arising from uncertainty in the drift part}\}.
\end{aligned}$$

The exact expression of the variance of $\mathscr{C}(x)$ is complicated since all terms are random. A reasonable assumption is that the largest source of uncertainty arises from the stochastic integral part.

In [121, 153], the authors explore this approximation. They argued that the terms in the drift part are small compared with market dynamics. In fact, we verify numerically that such approximation makes sense in our case. We consider a time discretization of the trading interval $[0,\mathrm{T}]$ and compare the "true" variance to the approximation by taking the stochastic integral part. The error is indeed small due to the coefficient $\tilde{\eta}^2$ in the drift part.

In this section we are interested in the price impact formulation of Equation (4.10) with temporary impact only. That is, we can easily deduce that the mean-variance objective function can be approximated as the following :

$$\mathbb{E}\left[\mathscr{C}(x)\right]+\lambda\mathbb{E}\left[\mathscr{C}(x)\right]\approx$$
$$\mathbb{E}\Big[\int_0^{\mathrm{T}}\frac{1}{2}\tilde{\eta}\sigma\dot{x}_t^2\mathrm{S}_t^2\big(\hat{\mathrm{T}}-t\big)^{3/2}\Gamma(t,\mathrm{S}_t)dt+\tilde{\lambda}\int_0^{\mathrm{T}}x_t^2\sigma^2\mathrm{S}_t^2\Delta^2(t,\mathrm{S}_t)dt\Big],$$

where $\Delta(t,\mathrm{S}_t)$ is the Black-Scholes delta.

We define $\mathscr{X}(\mathrm{T},\mathrm{X})$ the set of all adapted and absolutely continuous strategies that satisfy the boundary conditions $x_0=\mathrm{X}<0$, $x_{\mathrm{T}}=0$, the integrability conditions $\mathbb{E}\Big[\int_0^{\mathrm{T}}\dot{x}_t^2\mathrm{S}_t^2\big(\hat{\mathrm{T}}-t\big)^{3/2}\Gamma(t,\mathrm{S}_t)dt\Big]<\infty$ and $\mathbb{E}\Big[\int_0^{\mathrm{T}}x_t^2\mathrm{S}_t^2\Delta^2(t,\mathrm{S}_t)dt\Big]<\infty$, and consider the following minimization problem

$$\mathrm{U}(0,\mathrm{S}_0,\mathrm{X}):= \tag{4.18}$$
$$\inf_{x\in\mathscr{X}(\mathrm{T},\mathrm{X})}\mathbb{E}\Big[\int_0^{\mathrm{T}}\Big\{\dot{x}_t^2\mathrm{S}_t^2\big(\hat{\mathrm{T}}-t\big)^{3/2}\Gamma(t,\mathrm{S}_t)+\lambda x_t^2\sigma^2(\mathrm{S}_t)\mathrm{S}_t^2\Delta^2(t,\mathrm{S}_t)\Big\}dt\Big], \tag{4.19}$$

where $\lambda:=2\tilde{\lambda}/\sigma\tilde{\eta}$.

This minimization problem does not admit a closed-form solution. In order to solve it, a proper stochastic dynamic programming framework needs to be set up.

Proposition 3. *We parameterize strategies x by their speed of trading and define the control α such that $\alpha_t:=-\dot{x}_t$. We introduce $\mathscr{A}(\mathrm{T},\mathrm{X})$ the class of all progressively measurable processes $(\alpha_t)_{0\le t\le\mathrm{T}}$, for which the parameterized strategy x^α defined by*

$$x_t^\alpha:=\mathrm{X}-\int_0^t\alpha_s ds,\ \ 0\le t\le\mathrm{T},$$

belongs to the set $\mathscr{X}(\mathrm{T},\mathrm{X})$. We restrict our framework to Markovian controls and thus, solving such optimal stochastic control problem at time 0 is brought to a more general case where the agent starts buying at any arbitrary time $t\in[0,\mathrm{T}]$ with an initial quantity x without losing the optimality.

$$\alpha_t=\alpha(t,\mathrm{S}_t,x_t)$$

Remark 8. *The agent's optimal trading speed α_t at time t is completely determined by the current state (t,S_t,x_t), i.e time t, current stock price S_t and current quantity x_t.*

For a given strategy $\alpha(.,.,.)$, the value function $\mathrm{U}(t,\mathrm{S},x)$ is defined as

$$\mathrm{U}(t,\mathrm{S},x)=$$
$$\inf_{\alpha\in\mathscr{A}(\mathrm{T},\mathrm{X})}\mathbb{E}_t\Big[\int_t^{\mathrm{T}}\Big\{\alpha_u^2\mathrm{S}_u^2\big(\hat{\mathrm{T}}-u\big)^{3/2}\Gamma(u,\mathrm{S}_u)+\lambda\sigma^2(x_u^\alpha)^2\mathrm{S}_u^2\Delta^2(u,\mathrm{S}_u)\Big\}du\Big],$$

where $\mathbb{E}_t$ is the expectation conditional to $\mathrm{S}_t=s$ and $x_t^\alpha=x$. Note that this problem fits into the *finite-fuel* framework because of the state dependence of the class $\mathscr{A}(\mathrm{T},\mathrm{X})$ of admissible control processes.

Remarks 4. *(i) We can rewrite the state variables* S_t *and* x_t^α *in terms of stochastic differential equations (SDE) :*

$$dS_t = \sigma S_t dW_t$$
$$dx_t^\alpha = -\alpha_t dt$$

where the initial conditions are $S_0 = s_0$ *and* $x_0 = X$.
(ii) Using the standard procedure of deriving the Hamilton-Jacobi-Bellman (HJB) equation in stochastic control problems, see [157], $U(t,S,x)$ *should satisfy the following HJB equation*

$$\partial_t U + \frac{1}{2}\sigma^2 S^2 \partial_{SS} U + \tilde{\lambda} x^2 S^2 \Delta^2(t,S) + \inf_{\alpha\in\mathbb{R}} \left\{ \alpha^2 S^2 \left(\hat{T}-t\right)^{3/2} \Gamma(t,S) - \alpha \partial_x U \right\} = 0. \tag{4.20}$$

(iii) Assuming U *is sufficiently smooth, Bellman's principle of optimality for stochastic minimization problems suggests that the value function of the dynamic programming is always a submartingale and a martingale under the optimal strategy.*
(iv) The so-called finite-fuel constraint required from strategies in $\mathscr{A}(T,X)$ *(i.e* $\int_0^T \alpha_t dt = X$*), suggests the value function* U *should satisfy a singular terminal condition of the form*

$$\lim_{t\to T} U(t,S,x) = \begin{cases} 0 & if \quad x = 0 \\ +\infty & if \quad x \neq 0. \end{cases} \tag{4.21}$$

The intuition of Equation (4.21) is that a state with a non zero option position with no time left for its liquidation means that the liquidation task has not been performed, and so it should receive an infinite penalty.

In what follows, we will substitute the infinite penalty problem with a finite terminal condition. We index the value function by ε such that

$$U_\varepsilon(t,s,x) = \inf_{\alpha\in\mathscr{A}(T,X)} \mathbb{E}_t \left[\int_t^T \left\{ \alpha_u^2 S_u^2 \left(\hat{T}-u\right)^{3/2} \Gamma(u,S_u) + \lambda\sigma^2 (x_u^\alpha)^2 S_u^2 \Delta^2(u,S_u) \right\} du + \frac{1}{\varepsilon}\psi(x_T^\alpha) \right]. \tag{4.22}$$

With terminal condition

$$U_\varepsilon(T,s,x) = \frac{1}{\varepsilon}\psi(x) \begin{cases} 0 & \text{if} \quad x = 0 \\ \gg 1 & \text{if} \quad x \neq 0. \end{cases} \tag{4.23}$$

Remarks 5. *(i) The function* ψ *should equal to 0 when* $x = 0$ *and at the same time we need to be able to penalize a final state with remaining inventory through* ε*. A proper* ψ *will be more convenient than another one. In our case, the dependence on the state variable* x *suggests* $\psi(x) = x^2$*. This ansatz is used in [104] for example.*
(ii) The solution to (4.22)-(4.23) is ε*-dependent. For different values* ε *and* ε' *we find different solutions* U_ε *and* $U_{\varepsilon'}$*. However, we can construct a time series* $(U_{\varepsilon_n})_n$ *that converges to the original problem when* $\varepsilon_n \underset{n\to+\infty}{\longrightarrow} +\infty$.
(iii) Finally, we can easily verify that U_ε *satisfies the HJB equation (4.20) with terminal condition (4.23).*

In the following, we seek to solve the dynamic programming problem (4.22) with terminal condition (4.23). Our first main result is the following :

Theorem 2. *Let* U_ε^* *be a regular function which solves the PDE :*

$$\begin{cases} \partial_t U_\varepsilon^* + \frac{1}{2}\sigma^2 S^2 \partial_{SS} U_\varepsilon^* + \lambda x^2 \sigma^2 S^2 \Delta^2(t,S) - \frac{\left(\partial_x U_\varepsilon^*\right)^2}{4(\hat{T}-t)^{3/2}\Gamma(t,S)} = 0 \\ U_\varepsilon^*(T,S_T,x_T) = \frac{1}{\varepsilon}\psi(x_T^\alpha). \end{cases} \tag{4.24}$$

Then U_ε^* *is the unique solution to the optimal execution problem (4.22). Moreover, the optimal execution rate* $\alpha_t^* = -\dot{x}_t^*$ *is such that :*

$$\alpha_t^* = \frac{\partial_x U_\varepsilon^*(t,S_t,x_t^*)}{4(\hat{T}-t)^{3/2}S_t^2\Gamma(t,S_t)}. \tag{4.25}$$

The proof of this theorem uses the following lemma :

Lemma 1. *Let* U_ε^* *be a regular function that solves PDE (4.24). Then* U_ε^* *is a classical solution of the minimization problem (4.20), (4.23), and the optimal execution speed* $\alpha_t^* = -\dot{x}_t^*$ *is defined by :*

$$\begin{aligned}\alpha_t^* &= \frac{\partial_x U^*(t,S_t,x_t^*)}{4(\hat{T}-t)^{3/2}S_t^2\Gamma(t,S_t)} \\ &= \arg\min\left\{\alpha^2(\hat{T}-t)^{3/2}S_t^2\Gamma(t,S_t) - \alpha\partial_x U^*(t,S_t,x_t^*)\right\}.\end{aligned} \tag{4.26}$$

Démonstration. The proof to the lemma is straightforward and uses the fact that the function $h(\alpha) = \alpha^2(\hat{T}-t)^{3/2}S_t^2\Gamma(t,S_t) - \alpha\partial_x U_\varepsilon^*(t,S_t,x_t^*)$ is convex and attains its minimum for $h'(\alpha) = 0$. This gives (4.26) and

$$h(\alpha^*) = -\frac{\left(\partial_x U^*(t,S_t,x_t^*)\right)^2}{4(\hat{T}-t)^{3/2}S_t^2\Gamma(t,S_t)}.$$

Thus, we obtain the lemma through substituting α^* back into Equation (4.20). □

Proof of Theorem 2 : Let $\alpha \in \mathscr{A}(T,X)$ be an arbitrary control process, and define the stopping time

$$\theta_n := T \wedge \inf\left\{s > t : \left|\int_t^s \alpha_u du\right| \geq n\right\} \tag{4.27}$$

We first need to show that

$$U_\varepsilon^*(t,S,x) \leq \mathbb{E}\Big[\int_t^T \left\{\alpha_u^2(\hat{T}-u)^{3/2}S^2\Gamma(u,S_u) + \lambda\sigma^2(x_u^\alpha)^2S_u^2\Delta^2(u,S_u)\right\}du + \frac{1}{\varepsilon}\psi(x_T^\alpha)\Big],$$

where U_ε^* is the solution of the PDE (4.24).

First, we have that the right hand side of the inequality is finite. This is justified by the admissibility of α which includes the integrability conditions $\mathbb{E}[\int_t^T \alpha_u^2(\hat{T}-u)^{3/2}S_u^2\Gamma(u,S_u)du] < \infty$ and $\mathbb{E}\Big[\int_t^T (x_u^\alpha)^2S_u^2\Delta^2(u,S_u)du\Big] < \infty$. In addition, the fuel constraint $\int_0^T \alpha_s ds = X$, and Cauchy-Schwarz inequality imply

$$|x_t^\alpha| = \left|\int_t^T \alpha_s ds\right| \leq \sqrt{(T-t)\int_0^T \alpha_s^2 ds} \in L^2(\mathbb{P}).$$

By Ito's formula, we have

$$\begin{aligned}U_\varepsilon^*(t,S,x) = U_\varepsilon^*(\theta_n,S_{\theta_n},x_{\theta_n}) &- \int_t^{\theta_n}\left(\partial_t + \frac{1}{2}\sigma^2S_t^2\partial_{SS} - \alpha_r\partial_X\right)U_\varepsilon^*(r,S_r,x_r)dr \\ &- \int_t^{\theta_n}\sigma S_r\partial_S U_\varepsilon^*(r,S_r,x_t)dW_r.\end{aligned}$$

Observe that

$$\begin{aligned}&\partial_t U_\varepsilon^* + \frac{1}{2}\sigma^2S_r^2\partial_{SS}U_\varepsilon^* - \alpha_r\partial_X U_\varepsilon^* + \alpha_r^2(\hat{T}-r)S^2\Gamma(r,S_r) + \lambda\sigma^2(x_r^\alpha)^2S_r^2\Delta^2(r,S_r) \\ &\quad\geq \partial_t U_\varepsilon^* + \frac{1}{2}\sigma^2S^2\partial_{SS}U_\varepsilon^* + \lambda\sigma^2(x_r^\alpha)^2S_r^2\Delta^2(r,S_r) + \inf_{\alpha\in\mathscr{X}(X,T)}\left\{\alpha^2(\hat{T}-r)S^2\Gamma(r,S_r) - \alpha\partial_X U_\varepsilon^*\right\},\end{aligned}$$

and that the integrand in the stochastic integral is bounded on $[t,\theta_n]$, a consequence of the regularity of U_ε^*. It follows that :

$$U_\varepsilon^*(t,S,x) \le \mathbb{E}\Big[U_\varepsilon^*(\theta_n,S_{\theta_n},x_{\theta_n})\Big] + \mathbb{E}\Big[\int_t^{\theta_n}\{\alpha_r^2(\hat{T}-r)S^2\Gamma(r,S_r)+\lambda\sigma^2(x_r^\alpha)^2\Delta^2(r,S_r)\}dr\Big].$$

We now take the limit as n increases to infinity. Since $\theta_n \to T$ a.s we have

$$\lim_{n\to\infty}\mathbb{E}\Big[\int_t^{\theta_n}\{\alpha_r^2(\hat{T}-r)S^2\Gamma(r,S_r)+\lambda\sigma^2(x_r^\alpha)^2\Delta^2(r,S_r)\}dr\Big] = \mathbb{E}\Big[\int_t^{T}\{\alpha_r^2(\hat{T}-r)S^2\Gamma(r,S_r)+\lambda\sigma^2(x_r^\alpha)^2\Delta^2(r,S_r)\}dr\Big].$$

On the other hand

$$\begin{aligned}\big|U_\varepsilon^*(\theta_n,S_{\theta_n},x_{\theta_n})\big| &\le \frac{1}{\varepsilon}\psi(x_T^\alpha) + K_1\max_{t\le r\le T}(S_r^2) + K_2\max_{t\le r\le T}(S_r^2)\int_t^T (x_r^\alpha)^2 dr + K_3\\ &\le \frac{1}{\varepsilon}\psi(x_T^\alpha) + K_1\max_{t\le r\le T}(S_r^2) + \frac{1}{2}K_2\max_{t\le r\le T}(S_r^2)T^2\int_0^T \alpha^2 dr + K_3.\end{aligned}$$

It follows from the dominated convergence

$$U_\varepsilon^*(t,S,x) \le \mathbb{E}\left[\int_t^T\Big\{\alpha_u^2(\hat{T}-u)S^2\Gamma(u,S_u)+\lambda\sigma_0^2(x_u^\alpha)^2S_u^2\Delta^2(u,S_u)\Big\}du + \frac{1}{\varepsilon}\psi(x_T^\alpha)\right].$$

The control given by 4.25 is well defined (T<$\hat{T}$) and the solution U_ε regular.

Remark 9. *Setting the control variable as the speed of trading $\dot{x}$ is very common. In [104] for example, the authors use this parametrization to obtain the HJB equation. The problem is then reduced to one dimension using an Ansatz. An alternate solution is to adopt an exponential growth parametrization. This way, reducing the dimension is straightforward .*

4.4.1 Deriving the HJB with multiplicative state variable in the Black-Scholes case

Let us consider the controlled state variable x_t^κ and the control κ such that

$$dx_t^\kappa = -\kappa_t x_t^\kappa dt$$

where $\kappa_t > 0$, $x_0 = X$ and x_t increasing and bounded by 0. And let us define $\mathcal{K}(T,X)$ the set of admissible control processes κ such that x belongs to $\mathcal{X}(T,X)$.

Remark 10. *Ideally, x_t vanishes as t goes to T to verify the finite fuel constraint of total acquisition (i.e liquidation of a short position) at T. This is only possible if κ is infinite at a certain time. As in the previous case, the limit $\varepsilon \to 0$ forbids to trade a large quantity at the end time by imposing a penalty for the final state. Substituting the infinite limit by a large penalty on the value function as $t \to T$ allows to have a regular control variable which can be solved numerically. The remaining inventory will be acquired as a last additional trade at $t = T$.*

Using the multiplicative parameterization we have :

$$\begin{aligned}&\mathbb{E}_t\Big[\int_t^T\Big\{\kappa_u^2(x_u^\kappa)^2(\hat{T}-u)^{3/2}S_u^2\Gamma(u,S_u)+\lambda(x_u^\kappa)^2\sigma^2S_u^2\Delta^2(u,S_u)\Big\}du\Big]\\ &\qquad = x^2\mathbb{E}_t\Big[\int_t^T e^{-\int_t^u 2\kappa_s ds}\Big\{\kappa_u^2(\hat{T}-u)^{3/2}S_u^2\Gamma(u,S_u)+\lambda\sigma^2S_u^2\Delta^2(u,S_u)\Big\}du\Big]. \qquad (4.28)\end{aligned}$$

Moreover, by taking the terminal condition $U_\varepsilon(T,S_T,x_T)=\frac{1}{\varepsilon}(x_T^\kappa)^2=x^2\frac{1}{\varepsilon}e^{-\int_t^T 2\kappa_s ds}$ we have :

$$U_\varepsilon(t,s,x)=:x^2u_\varepsilon(t,s)$$

where u_ε is a reduced value function.

Definition 1. *The reduced value function u_ε is defined by :*

$$u_\varepsilon(t,S)=\inf_{\kappa\in\mathcal{K}}\mathbb{E}_t\left[\int_t^T e^{-\int_t^u 2\kappa_s ds}\left\{\kappa_u^2(\hat{T}-u)^{3/2}S_u^2\Gamma(u,S_u)+\lambda\sigma^2S_u^2\Delta^2(u,S_u)\right\}du+\frac{1}{\varepsilon}e^{-\int_t^T\kappa_s ds}\right]. \quad (4.29)$$

Proposition 4. *By means of Ito's formula, u_ε verifies the HJB equation :*

$$\partial_t u_\varepsilon+\frac{1}{2}\sigma^2S^2\partial_{SS}u_\varepsilon+\inf_\kappa\left\{\kappa^2(\hat{T}-t)^{3/2}S^2\Gamma(t,S)-2\kappa u_\varepsilon\right\}+\lambda\sigma^2S^2\Delta^2(t,S)=0 \text{ and } u_\varepsilon(T,s)=\frac{1}{\varepsilon}. \quad (4.30)$$

Besides, $h(\kappa)=\kappa^2(\hat{T}-t)^{3/2}S^2\Gamma(t,S)-2\kappa u_\varepsilon(t,S)$ attains its minimum for $h'(\kappa)=0$ and

$$\kappa^*(t,S)=\frac{u_\varepsilon(t,S)}{(\hat{T}-u)^{3/2}S^2\Gamma(t,S)} \text{ and } h(\kappa^*)=-\frac{u_\varepsilon^2}{(\hat{T}-u)^{3/2}S^2\Gamma(t,S)}.$$

By injecting the previous expression into the HJB equation (4.30), we deduce the PDE for u_ε.

Theorem 3. *Let u_ε^* be a regular function verifying the following PDE*

$$\begin{cases}\partial_t u_\varepsilon^*+\frac{1}{2}\sigma^2S^2\partial_{SS}u_\varepsilon^*+\lambda\sigma^2S^2\Delta^2(t,S)-\frac{1}{(\hat{T}-t)^{3/2}S^2\Gamma(t,S)}u_\varepsilon^2=0\\ u_\varepsilon^*(T,s)=\frac{1}{\varepsilon}.\end{cases} \quad (4.31)$$

Then u_ε^* is the unique solution to the reduced optimization problem (4.29). The optimal trading rate κ_t^* is defined by :

$$\kappa^*(t,S)=\frac{u_\varepsilon(t,S)}{(\hat{T}-t)^{3/2}S^2\Gamma(t,S)}.$$

Remarks 6. *(i) The proof of Theorem 3 uses the same verification arguments as developed for Theorem 2. (ii) The multiplicative form gives an interesting interpretation for the control variable. By linearizing the quadratic term in PDE (4.31), we can see $\frac{1}{(\hat{T}-t)^{3/2}S^2\Gamma(t,S)}u_\varepsilon$ as a discount factor of the value function. In the numerical scheme, this term will be taken explicitly in order to solve the PDE. (iii) There is a feedback between the value function u_ε and the control κ. (iv) Using the analogy to equity markets, the denominator $(\hat{T}-t)^{3/2}S^2\Gamma(t,S)$ in the optimal trading rate can be interpreted as a market depth. In our case, the depth is a function of time and asset price.*

4.4.2 Localisation and boundary conditions

The original reduced problem (4.29) is posed on the domain $(t,s)\in[0,T]\times[0,\infty]$. For computational purposes, and because asset prices are finite, one needs to localize this domain to $[0,T]\times[0,S_{max}]$. Thus, we need to add the following complementary conditions :

— When $S=0$, the put price tends to the strike : $P(t,S)\approx K$, $S^2\Gamma(t,S)\approx 0$ and $S^2\Delta^2(t,S)\approx 0$. We simply need to solve

$$\partial_t u_\varepsilon+\inf_\kappa\{-2\kappa u_\varepsilon\}=0 \text{ and } u_\varepsilon(T)=\frac{1}{\varepsilon}.$$

This limit condition is a singular control problem which can also be expressed as a variational equation

$$\min\{\partial_t u_\varepsilon,u_\varepsilon\}=0 \text{ and } u_\varepsilon(T)=\frac{1}{\varepsilon}.$$

— When $S = S_{max} \gg K$, the put price becomes insignificant. We have $S^2\Gamma(t,S) \approx 0$ and $S^2\Delta^2(t,S) \approx 0$, which makes both the impact and the risk aversion term vanish. The optimization reduces to solving

$$\partial_t u_\varepsilon + \frac{1}{2}\sigma^2 S^2 \partial_{SS} u_\varepsilon + \inf_\kappa\{-2\kappa u_\varepsilon\} = 0 \text{ and } u_\varepsilon(T) = \frac{1}{\varepsilon}.$$

The variational inequality arising from this condition is :

$$\min\left\{\partial_t u_\varepsilon + \frac{1}{2}\sigma^2 S^2 \partial_{SS} u_\varepsilon, u_\varepsilon\right\} = 0 \text{ and } u_\varepsilon(T) = \frac{1}{\varepsilon}.$$

By taking these approximations on the boundaries, we are able to solve the problem numerically using finite differences methods as we will see in the next section.

4.5 Numerical solution and results

4.5.1 A finite differences scheme

The PDE (4.31) with additional boundary conditions of Section 4.4.2 is a Riccati equation. In [146], the authors solve the Algmren-Chriss optimal execution problem in terms of the Riccati equation using two methods. They find the exact solution using a time "reparametrization". Unfortunately, our case does not allow for a closed-form solution. Thus, finite differences methods is a convenient way to solve the problem.

Restricting the problem to the domain $[0,T] \times [0,S_{max}]$, we can use the argument of Lipschitz functions on a compact to solve the semilinear PDE (4.31) numerically. We discretize time and space and define $\tau = T - t$ as the time to the strategy end time. We fix ε and denote by u_j^n the numerical approximation to $u_\varepsilon(n\Delta\tau, j\Delta S)$, where $j = 1,2,...,J-1$ is the space grid index, and $n = 1,2,...,N-1$ the time grid index taken backwards. ΔS is the spacial step size and $\Delta\tau = \Delta t$ the time step.

Let $\mathscr{L}_h u_j^n$ denote the spatial discretization of the differential term $\mathscr{L}u = \frac{1}{2}\sigma^2\partial_{SS}u$, where :

$$\mathscr{L}_h u_j^n = \frac{1}{2}\sigma^2 S_j^2 \frac{u_{j+1}^n - u_j^n + u_{j-1}^n}{\Delta S^2} \tag{4.32}$$

We linearize the quadratic term by decomposing it into the product of an explicit and an implicit form. Thus, the general family of the two-level implicit schemes for solving the equation is given by :

$$\frac{u_j^{n+1} - u_j^n}{\Delta\tau} = \theta\mathscr{L}_h u_j^{n+1} + (1-\theta)\mathscr{L}_h u_j^n + \frac{1}{(\hat{T} - t_n)^{3/2} S_j^2 \Gamma_j^n} u_j^n u_j^{n+1} + \lambda\sigma^2 (S_j)^2 (\Delta_j^n)^2,$$

where $\Gamma_j^n = \Gamma(n\Delta\tau, j\Delta S)$ and $\Delta_j^n = \Delta(n\Delta\tau, j\Delta S)$.
For terminal condition $n = 0$ and boundary conditions $j = 0$ and $j = J$ we have the following :

— $u_j^0 = \frac{1}{\varepsilon}$ where $\frac{1}{\varepsilon} \gg 1$ which translates the penalty related to the finite fuel constraint.

— u_0^n verifies :

$$\min\left(-\frac{u_0^{n+1} - u_0^n}{\Delta\tau}, {u_\varepsilon}_0^n\right) = 0.$$

Parameter	Value
σ	30%
T (the strategy horizon)	1/12 (years)
$\hat{T}$ (the option maturity)	1 (years)
μ	0
r	0
S_0	1
K	S_0
Action	Buy
x_0	-1
$\tilde{\eta}$	0.05
Trading frequency	4 trades per day
λ	0, 1, 10, 100

TABLEAU 4.1 – Parameters for buying options under market impact over 1 month horizon

— And u_J^n verifies :

$$\min\left(-\frac{u_J^{n+1}-u_J^n}{\Delta\tau}-\theta\mathscr{L}_h u_J^{n+1}-(1-\theta)\mathscr{L}_h u_J^n, u_J^n\right)=0.$$

The inventory x_n at τ_n is expressed by

$$x_n^j = x_{n+1}^j e^{\kappa_n^j \Delta\tau}$$

where $x_N = X$, $x_0 = 0$ and the control is given by

$$\kappa_{t_n}^j = \frac{u_n^j}{(\hat{T}-t_n)^{3/2} S_j^2 \Gamma_j^n}.$$

Thus, the quantity traded at τ_n at the asset price node S_j is

$$x_n^j - x_{n+1}^j = x_{n+1}^j\left(e^{\kappa_n^j\Delta\tau}-1\right).$$

4.5.2 Results

Usually end-users have a net long positions in OTM or ATM puts. These positions are explained by the fact that end-users suffer from "crashophobia" as explained in [147]. In our numerical experiment, we present results for a long position on ATM put options. The moneyness is fixed w.r.t the asset price at time 0. The strike K is set to be equal S_0 and remains the same until the strategy ending date. The parameters for this example are given in Table 4.1.

Figure 4.4 gives the optimal execution strategy through the rate of trading κ as a function of the underlying price S and time t. The strategy does not depend on the trader position. However, as time increases the trading rate increases as well. The dependence of κ on the underlying price direction is barely notable. First, the case $\lambda = 0$ corresponds to the expected cost which was found to give a strategy that does not depend on the underlying price. The trading speed $\dot{x}_t$ of the strategy is increasing and convex in time, with contrast to the equity case where it was found to be constant. The mean-variance adds some dependency on the asset level but the surface representation does not allow to see it.

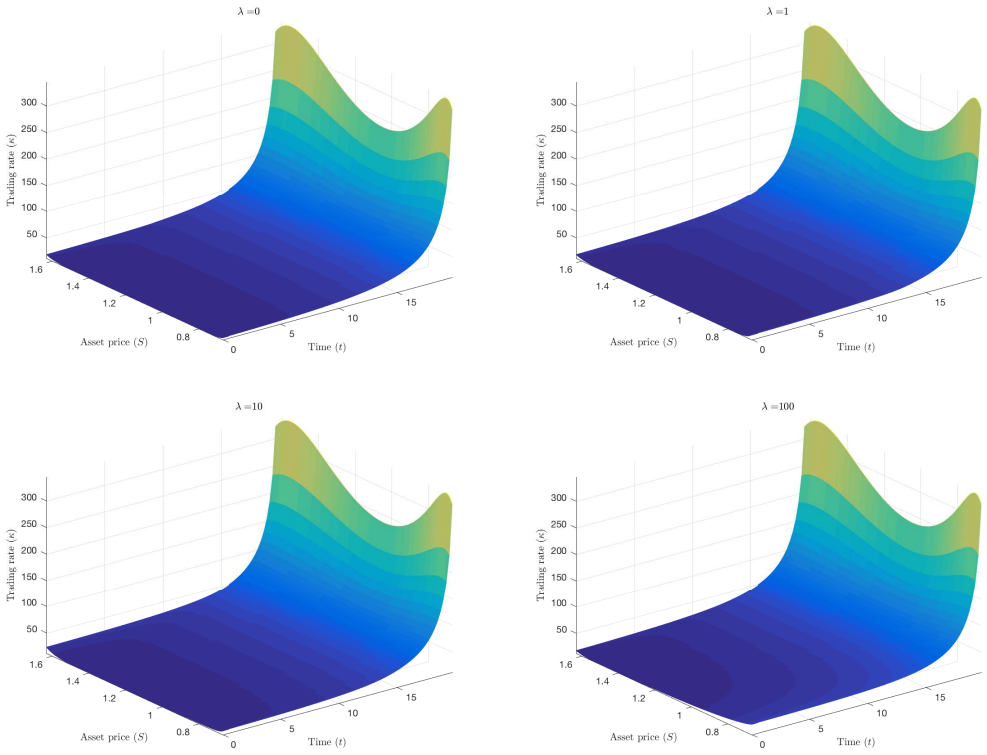


FIGURE 4.4 – The rate of trading κ as a function of the underlying price S and time t for different values of λ (λ = 0 top left, λ = 1 top right, λ = 10 bottom left, λ = 100 bottom right). The strike K = S_0 is fixed at time 0.

Figure 4.4 is misleading in a way; when the agent runs the strategy, the inventory and fundamental price will all evolve, hence this representation is to consider carefully. To gain additional insight into the dynamic behavior, we plot in Figures 4.5, 4.6 and 4.7 4 paths of the underlying price together with the rate of trading κ, the inventory x and quantity to be traded Δx. We can see clearly that adding the variance pushes the agent to adapt the strategy to the underlying level (Figure 4.6). Furthermore, when as the risk aversion parameter λ increases, the traded quantity tends to be larger at the beginning (Figure 4.7). Finally, we plot in Figure 4.8 a heat map of the distribution of the trading rate κ, trading speed $\dot{x}$, and inventory x over 10,000 simulations. This representation allows to see that the mean-variance with a high risk aversion is most sensitive to price movements. The case λ = 0 is the least affected by the spot variation.

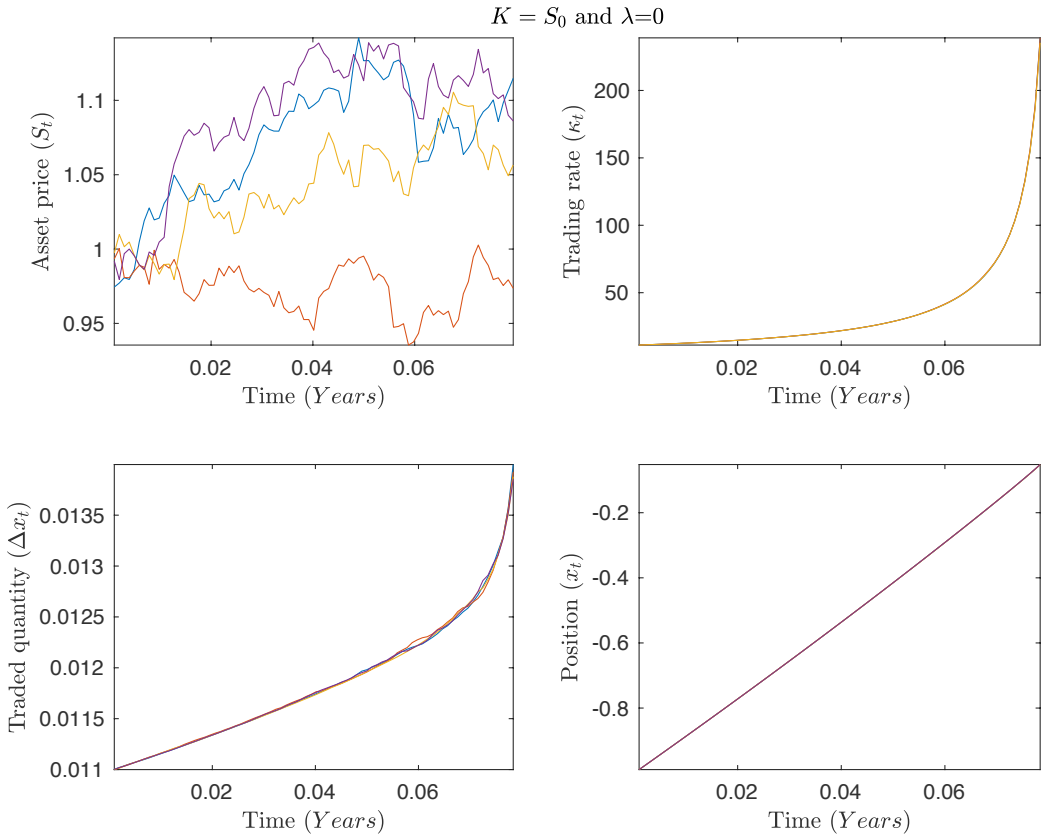


FIGURE 4.5 – Sample paths of the evolution of the fundamental price, trading rate, inventory and traded quantity throughout the execution for $\lambda = 0$.

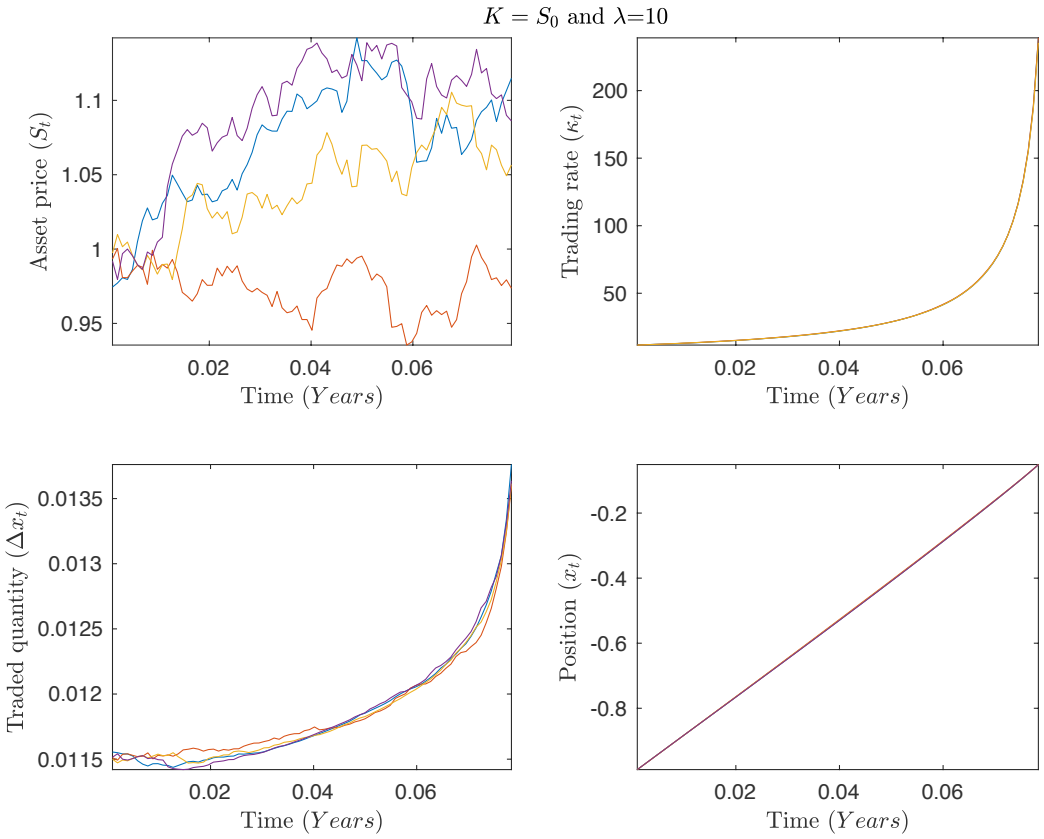


FIGURE 4.6 – Sample paths of the evolution of the fundamental price, trading rate, inventory and traded quantity throughout the execution for $\lambda = 10$.

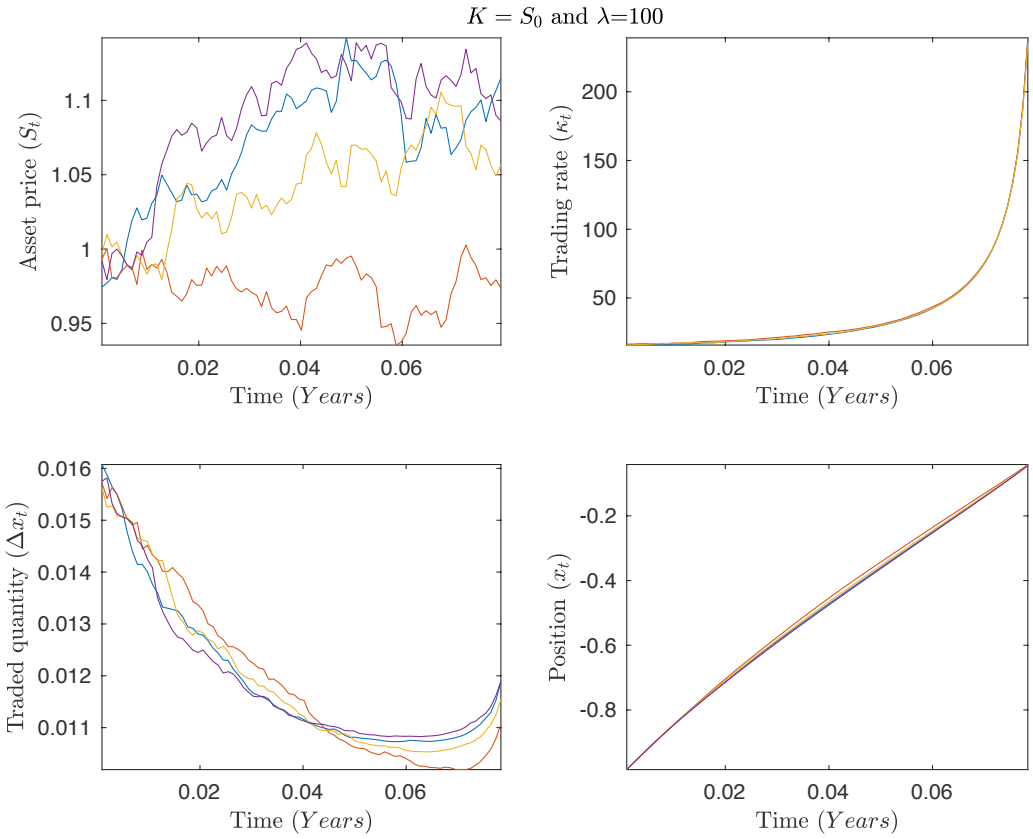


FIGURE 4.7 – Sample paths of the evolution of the fundamental price, trading rate, inventory and traded quantity throughout the execution for λ = 100.

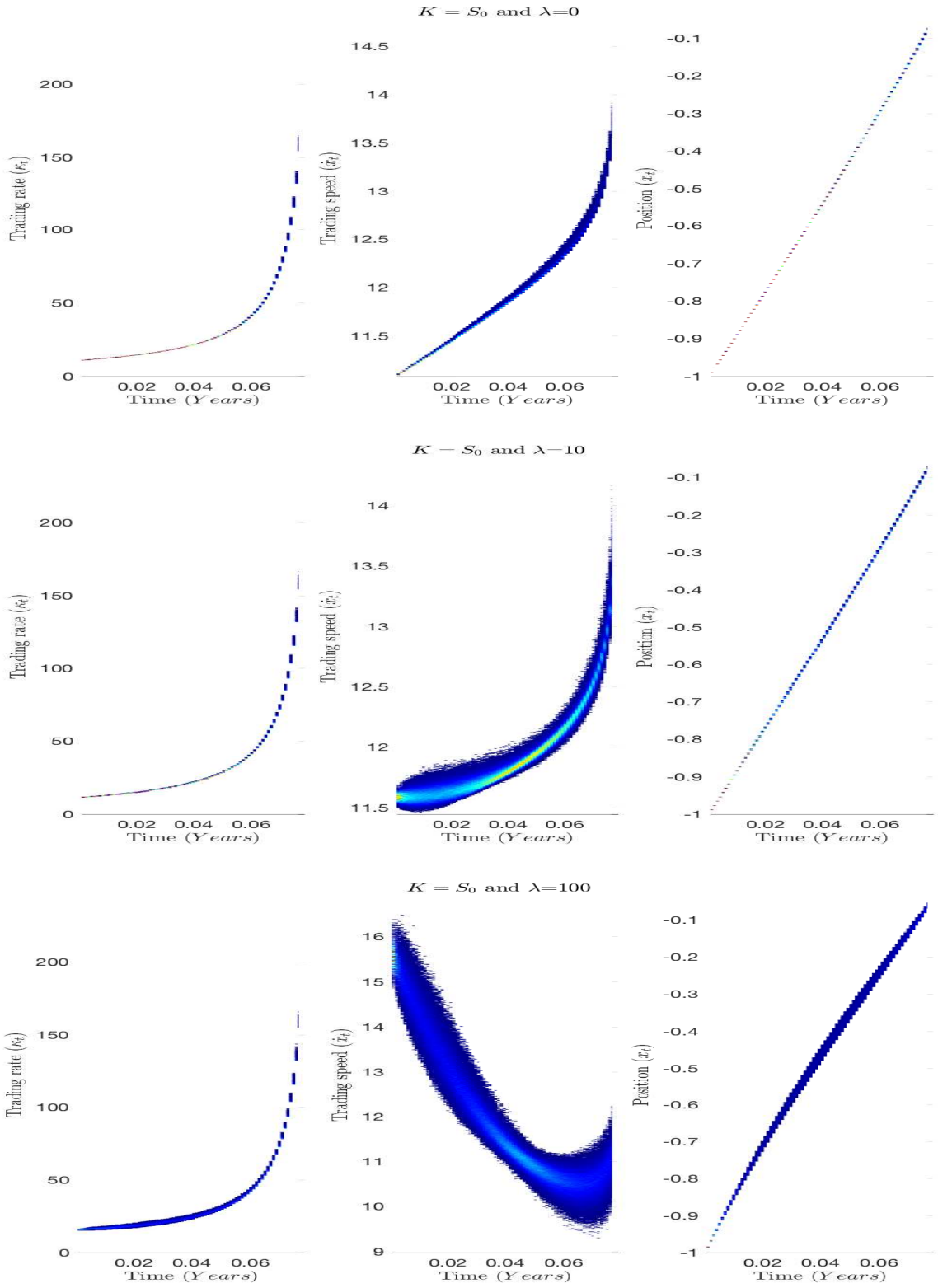


FIGURE 4.8 – Heat maps showing the density of inventory and trading speed throughout the execution for λ = 0, 10, and 100.

4.6 Extension to a local volatility model : A numerical method for the general case

In previous sections we exploited a closed-formula of the impact option price to derive the HJB equation. Under instantaneous impact, we were able to reduce the problem dimension to obtain a one-dimensional PDE which we solved numerically through finite differences methods. This framework was possible by linearizing the impact term in the volatility using a Taylor expansion.

In what follows, we would like to keep the nonlinear formulation of the impact in the asset price through the "enlarged volatility". PDE (4.4) actually provides the price of the option at time t given an inventory x_t and trading speed $\dot{x}_t$ and allows to take a local volatility model as pointed out in Remark 1 and developed in [117]. We will build a numerical scheme in this general case, under a constant elasticity volatility (CEV) model. We rewrite PDE (4.4) of $\tilde{\mathrm{P}}$ as the following

$$\begin{cases} \partial_u \tilde{\mathrm{P}}(u,\mathrm{S}) + \frac{1}{2}\tilde{\sigma}2(t,\mathrm{S})\mathrm{S}^2\partial_{\mathrm{SS}}\tilde{\mathrm{P}}(u,\mathrm{S}) = 0, & (u,\mathrm{S}) \in [t,\hat{\mathrm{T}}[\times]0,\infty] \\ \tilde{\mathrm{P}}(\hat{\mathrm{T}},s) = (\mathrm{K}-s)^+, & s \in]0,\infty[, \end{cases} \tag{4.33}$$

where

$$\tilde{\sigma}^2(t,\mathrm{S}) = \sigma^2(\mathrm{S}) + (\tilde{\eta}\dot{x}_t + \tilde{\gamma}(x_t - x_0))\sqrt{\hat{\mathrm{T}} - t}\sigma(\mathrm{S}),$$

and

$$\sigma(\mathrm{S}) = \sigma_0 \mathrm{S}^{\beta/2-1}.$$

Recall that we can always obtain the Black-Scholes case by setting $\beta = 2$. In this case, one can directly take the Black-Scholes closed formula with the enlarged volatility.

We are again interested in minimizing the mean-variance of the cost arising from strategy x, defined by

$$\mathscr{C}(x) = \int_0^{\mathrm{T}} \tilde{\mathrm{P}}_t \dot{x}_t dt.$$

We neglect the drift term in the variance which leads to the following approximation

$$\mathbb{V}ar\big[\mathscr{C}(x)\big] \approx \mathbb{E}\Big[\int_0^{\mathrm{T}} x_t^2 \sigma^2(\mathrm{S}_t)\mathrm{S}_t^2 \partial_{\mathrm{S}}\tilde{\mathrm{P}}^2(t,\mathrm{S}_t,x_t,\dot{x}_t)dt\Big].$$

The mean-variance objective function is thus

$$\begin{aligned} &\mathbb{E}[\mathscr{C}(x)] + \lambda \mathbb{V}ar[\mathscr{C}(x)] \approx \\ &\qquad \mathbb{E}\Big[\int_0^{\mathrm{T}} \dot{x}_t \tilde{\mathrm{P}}(t,\mathrm{S}_t,x_t,\dot{x}_t)dt + \lambda \int_0^{\mathrm{T}} x_t^2 \sigma^2(\mathrm{S}_t)\mathrm{S}_t^2 \partial_{\mathrm{S}}\tilde{\mathrm{P}}^2(t,\mathrm{S}_t,x_t,\dot{x}_t)dt\Big]. \end{aligned}$$

We would like to find the trading strategy x for the following minimization problem

$$\inf_{x \in \mathscr{X}(\mathrm{T},\mathrm{X})} \mathbb{E}\left[\int_0^{\mathrm{T}} \left\{\dot{x}_t \tilde{\mathrm{P}}(t,\mathrm{S}_t,x_t,\dot{x}_t) + \lambda\sigma^2(\mathrm{S}_t)x_t^2\mathrm{S}_t^2\partial_{\mathrm{S}}\tilde{\mathrm{P}}^2(t,\mathrm{S}_t,x_t,\dot{x}_t)\right\}dt\right],$$

where $\tilde{\mathrm{P}}$ follows PDE (4.33). Thus, we develop the dynamic optimization framework for V.

For initial fixed $(t,S,x) = (t,S_t,x_t)$ where $t < T$, trading strategy α such that $dx_t^\alpha = -\alpha_t dt$ and risk aversion λ, we define the value function V at time t

$$\hat{V}(t,S,x) = \inf_\alpha \mathbb{E}\left[\int_t^T \left\{-\alpha_t \tilde{P}(t,S_t,x_t^\alpha,-\alpha_t) + \lambda\sigma^2(S_t)(x_t^\alpha)^2 S_t^2 \partial_S \tilde{P}^2(t,S_t,x_t^\alpha,-\alpha_t)\right\} dt\right]. \tag{4.34}$$

For $t < T$, let $V = V(\tau = T-t,S,x) = \hat{V}(t,S,x)$. It is easy to find that the optimal control α^* can be obtained by solving the following HJB equation

$$\partial_\tau V = \frac{1}{2}\sigma^2(S)S^2\partial_{SS}V + \inf_\alpha\left\{-\alpha\partial_x V + -\alpha_t\tilde{P}(t,S,x,-\alpha) + \lambda\sigma^2(S)x^2S_t^2\partial_S\tilde{P}^2(t,S,x,-\alpha)\right\}. \tag{4.35}$$

We restrict our variables τ, $S = S(\tau)$ and $x = x(\tau)$ to the domain

$$\Omega = [0,T] \times [0,S_{max}] \times [X,0].$$

And set the initial condition and boundary conditions for V as suggested in Section 4.4. That is :

— For $\tau = 0$

$$V(\tau = 0,S,x)\begin{cases} 0 & \text{if} \quad x = 0 \\ \gg 1 & \text{if} \quad x \neq 0. \end{cases} \tag{4.36}$$

— For $S = 0$ we simply solve

$$\partial_\tau V = \inf_\alpha\{-\alpha\partial_x V\}.$$

— For $S = S_{max}$ the function g vanishes as the put price and delta tends to zero, which leads to solving the following PDE

$$\partial_\tau V = \frac{1}{2}\sigma(S)^2S^2\partial_{SS}V + \inf_\alpha\{-\alpha\partial_x V\}.$$

Finally, we give a brief outline of the numerical method used to solve the coupled PDEs (4.33)-(4.35) along with the corresponding initial and boundary conditions. We follow [75] to provide an informal discretization of the latter using a semi-Lagrangien approach. We refer the reader to the reference [49] for more details concerning the semi-Lagrangian method for HJB equations.

Along the trajectory $x = x(\tau)$ defined by

$$dx^\alpha = \alpha d\tau \tag{4.37}$$

equation (4.35) can be written as

$$\inf_{\alpha\leq 0}\left\{\frac{DV}{D\tau}(\alpha) - \mathscr{L}V - g(t,S,x,-\alpha)\right\} = 0,$$

where the operator $\mathscr{L}V$ is given by

$$\mathscr{L}V = \frac{1}{2}\sigma^2(S)S^2\partial_{SS}V,$$

and where the Lagrangian derivative $\frac{DV}{D\tau}(\alpha) = \partial_\tau V + \alpha\partial_x V$ The Lagrangian derivative is the rate of change of V along the trajectory (4.37). At the same time we keep in mind that $\tilde{P}$ is the solution of the PDE (4.33) where we replace $\dot{x}$ by $-\alpha$ in the enlarged volatility.

Define a set of nodes $[s_0, s_1, ..., s_{i_{\max}}]$, $[x_0, x_1, ..., x_{j_{\max}}]$, discrete times $\tau^n = n\Delta\tau$, and localize the control candidates to values in finite interval $[\alpha_{\min}, \alpha_{\max}]$. Let $V(\tau^n, s_i, x_j)$ denote the exact solution to Equation (4.35), $\tilde{P}(\tau, s_i, x_j, \alpha)$ is the solution to Equation (4.33) when the control value is $\alpha \in [\alpha_{\min}, \alpha_{\max}]$ and $\tilde{\Delta}(\tau, s_i, x_j, \alpha) = \partial_S \tilde{P}(\tau, s_i, x_j, \alpha)$ its partial derivative w.r.t the asset price. Let $V^n_{i,j}$, $\tilde{P}^n_{i,j}(\alpha)$, $\tilde{\Delta}^n_{i,j}(\alpha)$ and $\mathscr{L}_u V^n_j$ denote, respectively, the discrete approximation to the exact solution, price, its first derivative and the differential operator as in (4.32).

Let $\alpha^n_{i,j}$ denote the approximate value of the control variable α at mesh node (τ^n, s_i, x_j). Then the approximate $\frac{DV}{D\tau}(\alpha)$ at (τ^{n+1}, s_i, x_j) by the following :

$$\left(\frac{DV}{D\tau}(\alpha)\right)^{n+1}_{i,j} \approx \frac{1}{\Delta\tau}(V^{n+1}_{i,j} - V^n_{i,\hat{j}}) \tag{4.38}$$

where $V^n_{i,\hat{j}}$ is an approximation of $V(\tau^n, s^n_i, x^n_{\hat{j}})$ obtained by linear interpolation of the discrete values $V^n_{i,j}$, with $(s^n_i, x^n_{\hat{j}})$ given by solving Equation (4.37) backwards in time for fixed $x^n_{i,j}$ to give

$$x^n_{\hat{j}} = x_j - \alpha^{n+1}_{i,j}\Delta\tau.$$

Our final discretization is then

$$V^{n+1}_{i,j} = \Delta\tau(\mathscr{L}_h V)^{n+1}_{i,j} + \inf_{\alpha^{n+1}_{i,j} \in [\alpha_{\min}, \alpha_{\max}]} \left\{ V^n_{i,\hat{j}} + \Delta\tau\Big(-\alpha^{n+1}_{i,j}\tilde{P}^{n+1}_{i,j}(\alpha^{n+1}_{i,j}) + \lambda\sigma_0(x_j)^2 S^{\beta}_j \big(\tilde{\Delta}^{n+1}_{i,j}(\alpha^{n+1}_{i,j})\big)^2 \Big) \right\}. \tag{4.39}$$

We need to solve a local optimization problem at each node at each time step in Equation (4.39). In fact, we are seeking the global minimum of the local optimization problem. If the set of controls $[\alpha_{\min}, \alpha_{\max}]$ is discretized with spacing h, then a linear search of the control space will converge to the viscosity solution of the HJB Equation (4.35) as argumented in [154]. However, the uniqueness of the solution is not guaranteed with such method.

Results are presented in Figures 4.9 and 4.10 for the set of parameters given in Table 4.1 and a CEV parameter $\beta = 0.1$. The first figure gives the surface of the control variable by fixing one of the variables as a function of the remaining two others. The expected cost case does not seem to be sensitive to the moneyness as confirmed by the Heat Map in 4.10. The mean-variance case, on the contrary incites the agent to acquire more at the beginning of the strategy.

Remarks 7. *(i) The numerical experiment is very sensitive to the option pricing. To lead the optimization, one needs both the option price and delta over each time and space grid and for each potential control* α*. Both are computed using finite differences methods, which turns out to be quite slow. (ii) The option price variation over small time periods is very small. This leads to some discrete pattern of the control. (iii) More sophisticated methods need to be developed for solving such an optimization problem where the controlled function* f *is the solution of a PDE. In particular, forward-backward BSDEs can be very useful. (iv) Finally, we can conclude from the expected cost case that, even though the impact is nonlinear in the option price, the strategy seems to be independent of the asset price evolution. This, however, needs to be invested further. The risk criterion (here the variance) is what makes the agent sensitive to the market uncertainty.*

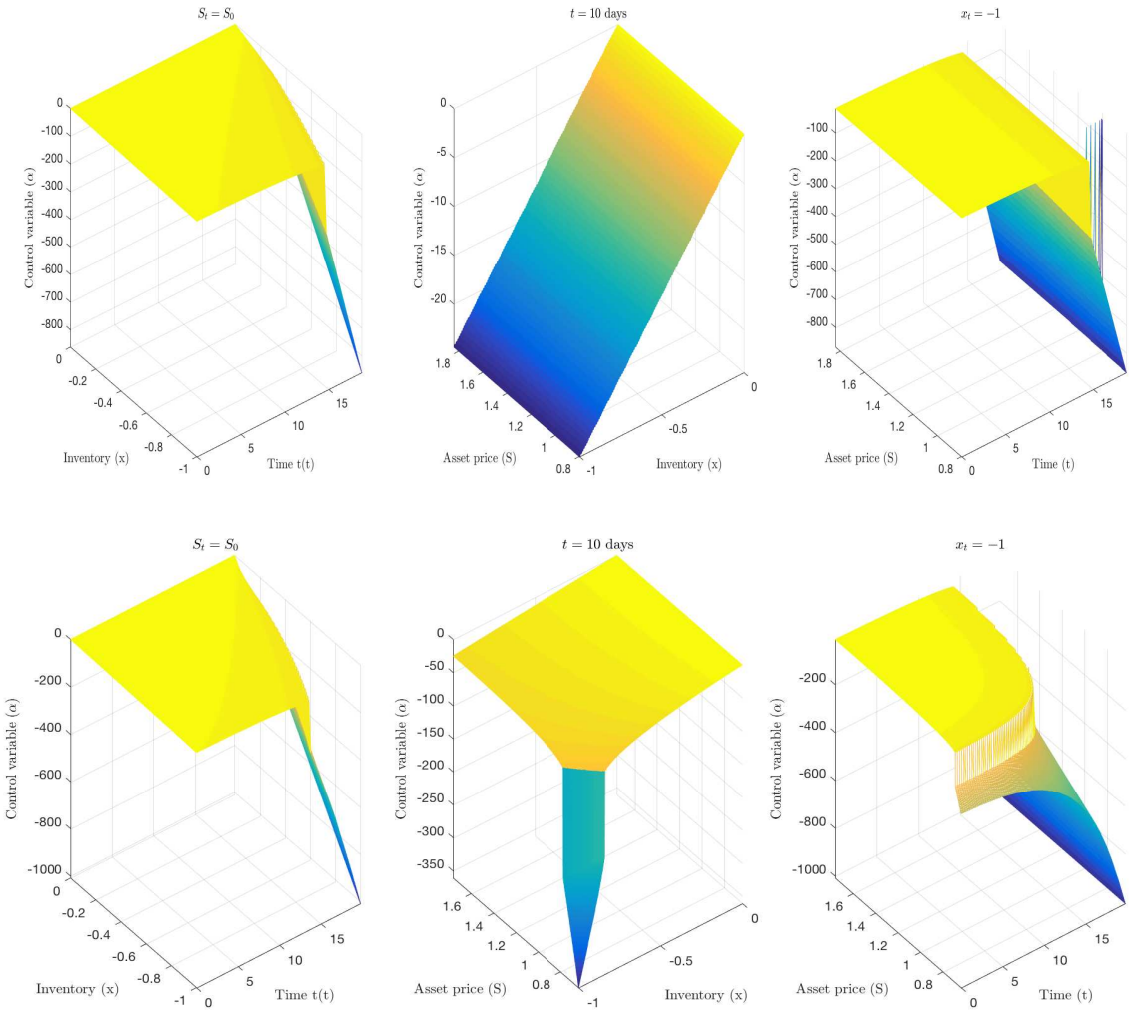


FIGURE 4.9 – The surface of the control variable α by fixing one of the directions (time t, asset price S and inventory x) as a function of the two remaining ones : top for $\lambda = 0$, bottom for $\lambda = 100$.

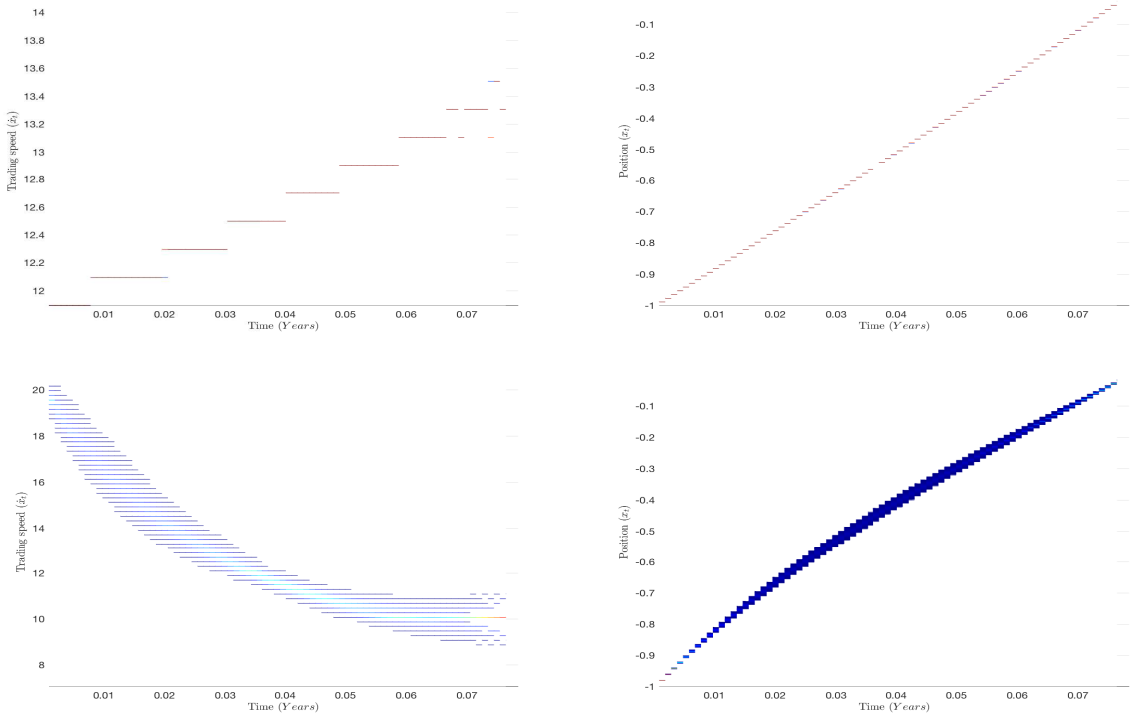


FIGURE 4.10 – Heat maps showing the density of the trading speed and inventory : top for $\lambda = 0$, bottom for $\lambda = 100$.

Chapitre 5

Range-based proxies and rough volatility

***Abstract*— In [89], it has been shown that volatility exhibits a fractional behavior with a Hurst exponent $H < 0.5$, changing the typical perception of volatility. In their study, Gatheral and co-authors used the realized volatility. In our analysis, we explore range-based proxies of the volatility process to confirm their findings on more available data (range-based) and non-standard assets. We find that the log-volatility based on range-based estimators behaves like a fractional Brownian motion with H lower that 0.1. We also find that rough fractional stochastic volatility model (RFSV) is a relevant volatility model. Moreover, the prediction power of this model outperforms that of the AR, HAR and GARCH models in most cases.**

Keywords : Range-based volatility; Garman-Klass; Parkinson; volatility scaling; fractional Brownian motion; fractional Ornstein-Uhlenbeck; volatility forecasting.

5.1 Introduction

Volatility plays a crucial role in many areas of finance and economics such as risk management and portfolio selection. It is known to be both time-varying and predictable, and stochastic volatility models are one way to deal with these features. As a consequence, its modeling and forecasting spurs the interest of many authors, academics and practitioners alike.

In the financial markets, the common practice is to represent asset prices by a continuous semi-martingale. A given log-price $\log S_t$ is defined by

$$d\log S_t = \mu_t dt + \sigma_t dW_t$$

where μ_t is a drift term, W_t a Brownian motion and σ_t, the key ingredient is the process volatility. The Black-Scholes framework assumes the volatility to be either constant or deterministic. Such specification proved to be inadequate in the late eighties. The main reason is the inconsistence of the Black-Scholes model with the observed European options. This gave rise to alternatives such as local volatility models, such as Dupire's, see [69], and Derman and Kani's, see [63]. These models consider σ_t as a deterministic function of time and asset price. Even though they enable us to perfectly fit a given implied surface, its dynamic is quite unrealistic. An other alternative is to model the volatility σ_t by a continuous Brownian semi-martingale, typically correlated with W. These so-called stochastic volatility models have been the center of interest of many authors. We cite amongst such stochastic volatility models, the Hull and White model [101], the Heston model [97] and the SABR model [95]. However, generated option prices are still not consistent with observed European option prices. The reader can refer to [87] for a review of different approaches. More recently market practice is to use so-called local-stochastic volatility models which both fit the market exactly and generate reasonable

dynamics.

Since the volatility is a latent variable, the first issue one faces when trying to exhibit its statistical properties is its estimation. One can only estimate it using the underlying asset prices or quoted options. For example, when we have daily stock returns, the squared variance is a well known volatility proxy, also known as realized volatility. It measures changes on the asset return over a specified period of time. If high frequency data is available (the whole price process during the day), this proxy is more precise and is used as the daily volatility.

Access to high frequency data is sometimes costly and/or unavailable for certain assets. Therefore, other proxies are used to estimate daily volatility. If we only have closing prices and need to estimate volatility on a daily basis, we can use the squared daily returns. A compromise can be found as, in addition to closing prices, open, high and low daily prices are available for most financial data sets. In [139], Parkinson was first to introduce an advanced volatility estimator using these so-called range prices instead of just closing prices, and this enables us to overcome the issues of the first two approaches, and present a way to better estimate daily volatility.

Statistical properties of volatility estimators raise interesting questions, particularly in relation to the smoothness of the volatility process. Researchers aim to uncover the underlying mechanisms that generate the data, using the empirical scaling evidence as a stylized fact that any theoretical model should also reproduce. For example, it is common belief that volatility exhibits what is commonly known as long range dependence. The implication of this it that volatility shocks today will influence its expectation in the same direction, see [33, 72] among others.

Stochastic or local volatility models mentioned earlier assume that the smoothness of the sample path of the volatility is that of a Brownian motion ($1/2-\varepsilon$ Hölder continuous for any $\varepsilon > 0$). In [56], Comte and Renault choose to address the question of long range dependence in terms of the regularity of the driving process. Their idea was to exploit the fractional Brownian motion. Recall that a fractional Brownian motion W^H with Hurst parameter $H \in (0,1)$ is an a.s. continuous, centered, self-similar Gaussian process with stationary increments and a covariance satisfying :

$$Cov(W_t^H, W_s^H) = \frac{1}{2}(|t|^{2H} + |s|^{2H} - |t-s|^{2H}), t, s \in \mathbb{R}$$

They proposed the fractional Ornstein-Uhlenbeck volatility model with a Hurst parameter greater than $1/2$ named fractional volatility model (FSV). Such model is $H-\varepsilon$ Hölder continuous with $H > 1/2$. More interest grew from this model and others develop deeper analysis and calibration, see [51, 55] among others. Later, Gatheral and co-authors, see [89], challenged the previous results and established that the log-volatility process is very close to that of a fractional Brownian motion with Hurst parameter around 0.1 ($< 1/2$). They also developed the rough fractional stochastic volatility model (RFSV) which operates with different parameter properties than the FSV, and justified that their model better respects the volatility smile and the data properties.

In this paper, we conduct a similar study to [89] using range-based estimators to find the best forecasting model. We replicate their analysis step by step in order to revisit their finding on less standard assets. We actually find that the Hurst parameter for range-based proxies on our set of data is even lower than 0.1 and sometimes even close to 0, confirming that rough volatility hypothesis can not be refuted at this point, while further analyses allow to dismiss the hypothesis that it is generated from the FSV model ($H > 1/2$).

Our paper is organized as follows. In Sections 5.2 we give an overview of range-based volatility estimation before choosing one volatility proxy to work with. We conduct our statistical study in Section 5.3 where we find that log-volatility from range-based proxies behave like a fractional Brownian motion with Hurst exponent lower than 0.1. We validate the rough fractional stochastic volatility model (RFSV) introduced in [89] on our data in Section 5.4, compare its prediction power with other common models in Section 5.5, and finally conclude in Section 5.6.

5.2 Overview on range-based volatility estimation

In this section we review a few range-based volatility estimators and compare them to realized volatility.

We assume that the asset price over a one day period of time, S_u, follows a geometric Brownian motion :

$$dS_u = \mu S_u du + \sigma S_u dW_u,$$

where $u \in (t-1, t]$ is the time index between two consecutive days $t-1$ and t, μ is the drift, σ the volatility considered constant along one day, and W_u a standard Brownian motion. By Ito's lemma the log price $\log(S_u)$ follows a Brownian motion with drift $\mu^* = \mu - \frac{\sigma^2}{2}$ and volatility σ. During a day, it is common practice to assume that the drift is equal to zero (i.e. $\mu^* \approx 0$).

The volatility being a latent variable, one needs to estimate it at each period of time with the available asset prices. We denote by H, L, O, C the high, low, open and close prices respectively. The log-returns r_t are defined by $r_t = \log(C_t) - \log(O_t)$ (or $\log(C_t) - \log(C_{t-1})$ when taking only close-to-close prices). Volatility changes, and our first interest is to be able to estimate it in a precise way and on a daily basis. Taking the squared return r_t^2 is one possible solution. It is an unbiased estimator of σ^2 under the normal log-returns assumption with zero mean. However, this estimator is quite noisy.

Range prices bring more consistency and information about the entire process than the close-to-close prices. In this context, Beckers, see [21], shows that volatility estimators can be improved by incorporating high and low prices, along with closing prices. Of course, range-based volatility estimators are not as efficient as realized volatility under ideal conditions, i.e. estimated from high frequency data, but remain a good alternative when this data is not available. It was shown in [4] and [151] that these range-based estimators are robust to microstructure noise and prove to be efficient and simple to compute.

In [139], Parkinson was first to develop a classical range-based estimator using high and low prices information expressed by the following formula :

$$\sigma^2_{\text{Parkinson}} = \frac{1}{4\log 2}\left(\log \frac{H_t}{L_t}\right)^2.$$

The correctness of this formula relies on a constant volatility assumption during each one day time period. It exploits the extreme value method, and the coefficient $\frac{1}{4\log 2}$ is nothing but the variance of the range variable (i.e the different between the minimum and maximum). The reader can refer to [139] for details and to [73] for the asymptotic distribution of the range.

The Parkinson estimator is asymptotically unbiased under the assumption that a geometric Brownian motion without drift can describe the path of the asset price changes.

Later, Garman and Klass, see [86], established a more efficient estimator that takes the following form :

$$\sigma^2_{GK} = 0.511\left(\log\frac{H_t}{L_t}\right)^2 - 0.019\left(\log\frac{C_t}{O_t}\left(\log\frac{H_t}{O_t} - \log\frac{L_t}{O_t}\right) - 2\log\frac{H_t}{O_t}\log\frac{L_t}{O_t}\right) - 0.383\left(\log\frac{C_t}{O_t}\right)^2.$$

This estimator combines the squared return and Parkinson volatility estimators into a new estimator with smaller variance. Garman and Klass actually proved that this estimator is optimal in a mean-variance sense among a certain class of estimator, see [86].

A more practical estimator is recommended with nearly the same efficiency but eliminates the small cross-product terms expressed as :

$$\sigma^2_{GK} = \frac{1}{2}\left(\log\frac{H_t}{L_t}\right)^2 - (2\log 2 - 1)\left(\log\frac{C_t}{O_t}\right)^2. \tag{5.1}$$

We rely on this estimator when using the Garman-Klass volatility proxy.

Because in [86] and [139] log-prices are assumed to follow geometric Brownian motion with no drift, many authors tried to correct this mismatch for securities with non-zero mean. In [145], more sophisticated drif-independent measures of volatility are introduced. The Rogers-Satchell estimator for example takes the form :

$$\hat{\sigma}^2_{RS} = \log\frac{H_t}{O_t}\left(\log\frac{H_t}{O_t} - \log\frac{C_t}{O_t}\right) + \log\frac{L_t}{O_t}\left(\log\frac{L_t}{O_t} - \log\frac{C_t}{O_t}\right).$$

Kunimoto [113] and Yang-Zhang [155] also deserve to be mentioned. Yang-Zhang estimator, however, can only be used over multiple days and therefore won't be interesting for our analysis.

The previously mentioned estimators are unbiased estimators of σ^2. When applying the square root and estimating the volatility, all σ estimators are biased. This was expected since $\mathbb{E}[\sigma^2]$ and $\mathbb{E}[\sigma]^2$ are generally different.

In term of efficiency, all previous estimators exhibit very substantial improvements compared to the close-to-close estimator. Efficiency measure of a volatility estimator $\hat{\sigma}^2_i$ is defined as the ratio of the variance of this estimator and the variance of the close-to-close estimator $\hat{\sigma}^2_{CC}$:

$$\text{Eff}(\hat{\sigma}^2_t) = \frac{\mathbb{V}ar(\hat{\sigma}^2_{CC})}{\mathbb{V}ar(\hat{\sigma}^2_i)}.$$

By definition, the squared return estimator has efficiency 1. Parkinson reported that his estimator is 2.5 to 5 times more efficient than simple close-to-close variance estimator. Garman-Klass reports 7.4 while Rogers-Satchell efficiency is 6.0 and Kunimoto is 10.

In light of [13], we compare the performance and distributional properties of different range-based volatility estimators on the S&P 500 over a 3786 period from January 2000 to April 2015. Our benchmark is the realized volatility from the Oxford-Man Institute of Quantitative Finance Realized Library. To perform the comparison we use the following measures :
The mean squared error defined by :

$$\text{MSE}(\sigma_{\text{estimated}}) = \mathbb{E}\left[(\sigma_{\text{estimated}} - \sigma_{\text{benchmark}})^2\right].$$

The mean absolute bias given by :

$$\text{MAD}(\sigma_{\text{estimated}}) = \mathbb{E}\left[\mid \sigma_{\text{estimated}} - \sigma_{\text{benchmark}} \mid\right].$$

The proportional bias expressed as :

$$\text{Prop.Bias}(\sigma_{\text{estimated}}) = \mathbb{E}\left[\left(\frac{\sigma_{\text{estimated}}}{\sigma_{\text{benchmark}}} - 1\right)\right].$$

Recall that the benchmark volatility $\sigma_{\text{benchmark}}$ here corresponds to realized volatility.

The results are given in Figures 5.1 and 5.2 and Table 5.1.

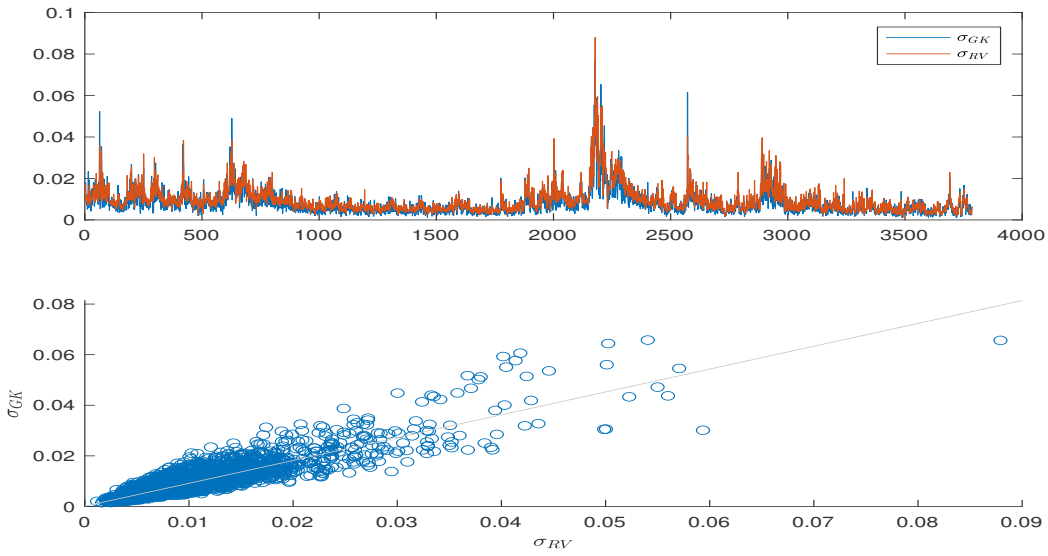


FIGURE 5.1 – Comparison between the GK proxy and the RV proxy as a benchmark. Top is the evolution of the volatility time series for both estimators. Bottom graph is a scatter plot of the GK estimator to the RV estimator.

	MSE	MAD	Prop.Bias	Std.Dev
$\hat{\sigma}_{\text{CC}}$	0.533×10^{-4}	0.0051	-0.1275	0.0092
$\hat{\sigma}_{\text{P}}$	0.092×10^{-4}	0.0021	-0.0849	0.0063
$\hat{\sigma}_{\text{GK}}$	0.094×10^{-4}	0.0021	-0.1313	0.0058
$\hat{\sigma}_{\text{RS}}$	0.206×10^{-4}	0.0028	-0.1762	0.0062

TABLEAU 5.1 – Comparison measures for different volatility estimators

We can see in Table 5.1, Figures 5.1 and 5.2 the following :

— Range-based estimators are lower than the benchmark.

— Range-based estimators reduce the variance compared to squared returns.

— The Garman-Klass estimator is the closest to intraday realized variance and has the smallest variance.

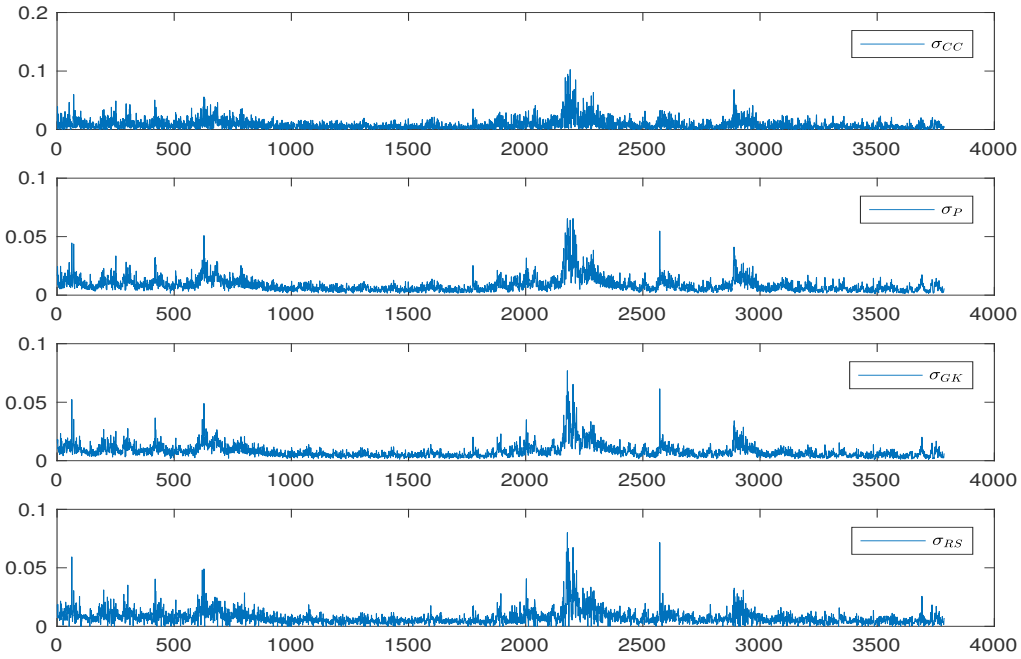


FIGURE 5.2 – Comparison between different range-based estimators.

— Range-based estimators seem to exhibit the same regularity as that of realized volatility.

We would like to confirm the last point. In our study, we focus on the volatility based on the Garman-Klass proxy. Our goal is to confirm that the scaling properties discussed in [89], are also satisfied by the volatility when using these proxies.

5.3 Range-based volatility as spot volatility proxy : empirical results

5.3.1 The scaling of the Garman-Klass proxy

We carry out our analysis on the volatility proxy for a variety of assets. In [89], Gatheral and co-authors use common indexes for which high frequency based realized volatility is available on the Oxford-Man Institute of Quantitative Finance Realized Library (S&P 500, Bund ...). In this paper, we choose to apply our analysis on more "exotic" assets (S&P 400, IBEX 35, IBOV, S&P 100, INDU, SHSZ300, MEXBOL, FTSE 100, XIN9I, HSI), and some stocks (TOTAL, ASX200, GOOGLE and MICROSOFT). Most of these assets are not available on the Library, and more importantly, since many financial institutions still do not have access to high frequency based proxies, the choice of range-based is very convenient. Range data availability allows for their computation for any class of assets and any assets tickers.

We choose to present the analysis for the S&P 100 [1] and IBEX 35 [2], and give numerical results for the

1. The S&P 100 is a subset of the S&P 500 and includes 102 leading U.S. stocks with exchange-listed options. Constituents of the S&P 100 represent about 63% of the market capitalization of the S&P 500 and almost 51% of the market capitalization of the U.S. equity markets as of January 2017.

2. IBEX 35 is the benchmark stock market index of the Bolsa de Madrid, Spain's principal stock exchange. It is a market capitalization weighted index comprising the 35 most liquid Spanish stocks traded in the Madrid Stock Exchange General Index

remaining ones. We focus mainly on the Garman-Klass proxy and present results on the Parkinson estimator for verification.

The set of data corresponds to 2521 trading days from April 19, 2005 to April 22, 2015. Volatility proxies are based on range-data extracted from Bloomberg database. Let $\sigma_{t_0}, \sigma_{t_1}, ..., \sigma_{t_N}$, be the time series of the GK proxy computed for this period, where $t_{i+1} - t_i$ corresponds to one business day.

Our scaling measure $m(q, \Delta)$ is the q-th absolute moment of the increments of log-volatility and are defined by :

$$m(q, \Delta) := \frac{1}{N} \sum_{k=1}^{\lfloor N/\Delta \rfloor} | \log(\sigma_{k\Delta}) - \log(\sigma_{(k-1)\Delta}) |^q,$$

for different $q > 0$ and lags Δ going from 1 to about 400 days. Our goal is to revisit the finding in [89] that the spot log-volatility has the same scaling properties as a fractional Brownian motion with Hurst exponent $H < 1/2$, and therefore one can model it with such process. It is worthy to mention to following remarks :

Remark 11. — *The quantity $m(q, \Delta)$ is the discrete equivalent of $\mathbb{E}[| \log(\sigma_\Delta) - \log(\sigma_0) |^q]$. Recall that the fractional Brownian motion W^H verifies :*

$$\mathbb{E}[| W^H_{t+\Delta} - W^H_t |^q] = \tilde{K}_q \Delta^{qH},$$

— *We are expecting the volatility to behave closely to the fractional Brownian motion. As a result, we would observe the following relationship :*

$$m(q, \Delta) \sim K_q \Delta^{qH}. \tag{5.2}$$

— *Given that data is finite and not time-equidistant (unavailable data on weekends for example), the measure $m(q, \Delta)$ is the result of averaging over all possible increments by taking a rolling window, and selecting only increments that correspond to the chosen lag Δ (log-vol increments between two successive volatility measures between Friday through Monday are considered as 3 days lag).*

To verify the validity of (5.2), we plot $\log(m(q, \Delta))$ against $\log(\Delta)$ for different values of q. Depending on the results of this first regression, we can write, for a given q,

$$m(q, \Delta) \sim K_q \Delta^{\zeta_q}, \tag{5.3}$$

where ζ_q defines a general scaling function.

Results are displayed in Figures 5.3 and 5.4. We can notice the following :

— The values of $m(q, \Delta)$ for different q against $\log(\Delta)$ lie within a straight line (left figures). This confirms that both S&P 100 and IBEX 35 exhibit a scaling property given by Equation (5.3).

— The R-squared values given in Table 5.2 confirm that the data is close to the fitted regression line for all values of q.

— Plotting ζ_q as a function of q (right), confirms our expectation; the scaling is linear in q and verifies Equation (5.2), with $H = 0.081$ for S&P 100 and $H = 0.072$ for IBEX 35.

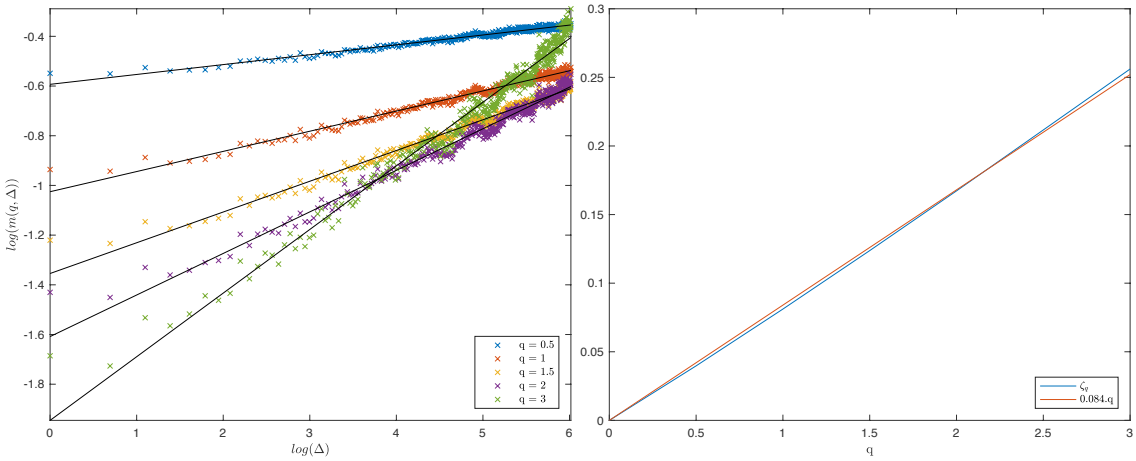


FIGURE 5.3 – $\log m(q,\Delta)$ as a function of $\log\Delta$ (left), ζ_q (blue) and $0.084\times q$ (green) (right), S&P 100 (Garman Klass volatility).

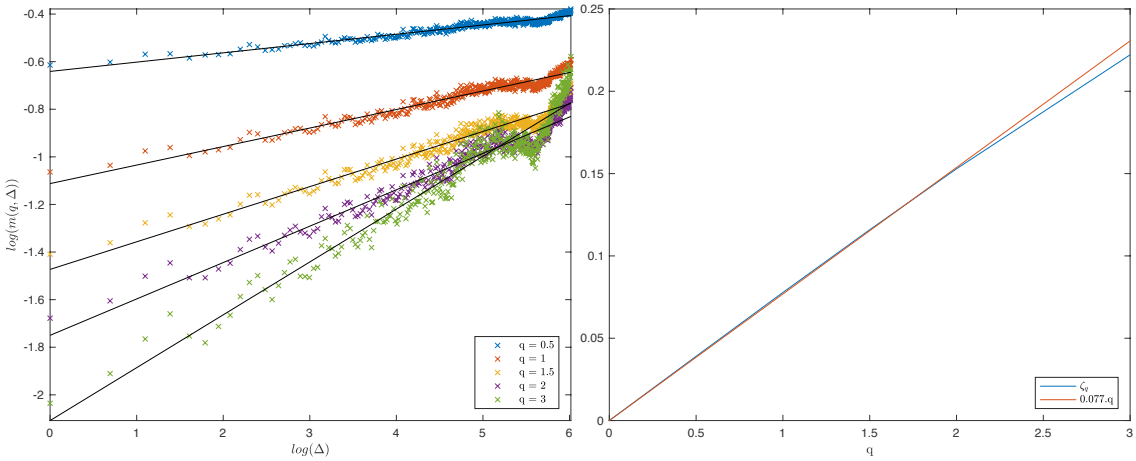


FIGURE 5.4 – $\log m(q,\Delta)$ as a function of $\log\Delta$ (left), ζ_q (blue) and $0.072\times q$ (green) (right), IBEX35 (Garman Klass volatility).

q	0.5	1	1.5	2	3
S&P 100	0.9640	0.9760	0.9800	0.9797	0.9717
IBEX 35	0.9225	0.9290	0.9287	0.9252	0.9111

TABLEAU 5.2 – The R-squared measure of the regression $\log(m(q,\Delta)) \sim \log(\Delta)$

Of course the ζ_q and Hq are not perfectly matched. One possible reason for this mismatch is using discrete samples. In fact, simulating fBm using the same number of points results in a slight concave figure, and ends up in recovering a Hurst parameter that slightly overestimates the real one.

In order to make sure our estimations of H do not depend on the time interval, we split the data into two periods with the same lag and re-estimate H for each period separately. The aim of this experience is to confirm that the scaling is time independent for all assets. Regressing $\log(m(q,\Delta))$ on $\log(\Delta)$ for each ticker and for $q = 0.5, 1, 1.5, 2, 3$ and $\Delta = 1, ..., 410$, we find that ζ_q is linear on q for the GK proxy. As we can see in Table 5.6, the Hurst parameter lies between 0.01 and 0.082 confirming that the volatility process is rough. One might however notice that splitting the data resulted in a first half with a slightly greater H than the second half. We think that it is due to the presence of the 2008 crisis in the first half.

Ticker	H for the whole period	H (first half)	H (second half)
SP100	0.0841	0.0897	0.0714
IBEX35	0.072	0.0753	0.071
HSI	0.0516	0.0605	0.0394
MEXBOL	0.0627	0.0737	0.0463
FTSE100	0.0751	0.0708	0.0728
ASX200	0.0489	0.0476	0.0415
TOTAL	0.0774	0.0835	0.0687
XIN9I	0.0674	0.0649	0.069
SHSZ300	0.0689	0.0718	0.0636
BCOM	0.014	0.0099	0.0238
INDU	0.0804	0.0838	0.067
USDEUR	0.0353	0.0393	0.0321
IBOV	0.0685	0.0724	0.0609
MICROSOFT	0.06	0.0717	0.0401
GOOGLE	0.0656	0.0724	0.0542
SP400	0.0715	0.0753	0.0592

TABLEAU 5.3 – Estimates of H on the whole period and over two different time intervals for different indexes and stocks (Garman Klass volatility)

5.3.2 The scaling of the Parkinson volatility proxy

To ensure that the results apply to other range-based estimators, we reproduce the same analysis to the the Parkinson volatility proxy, based on the same data and on the same period. Results are expressed in Figures 5.5 and 5.6.

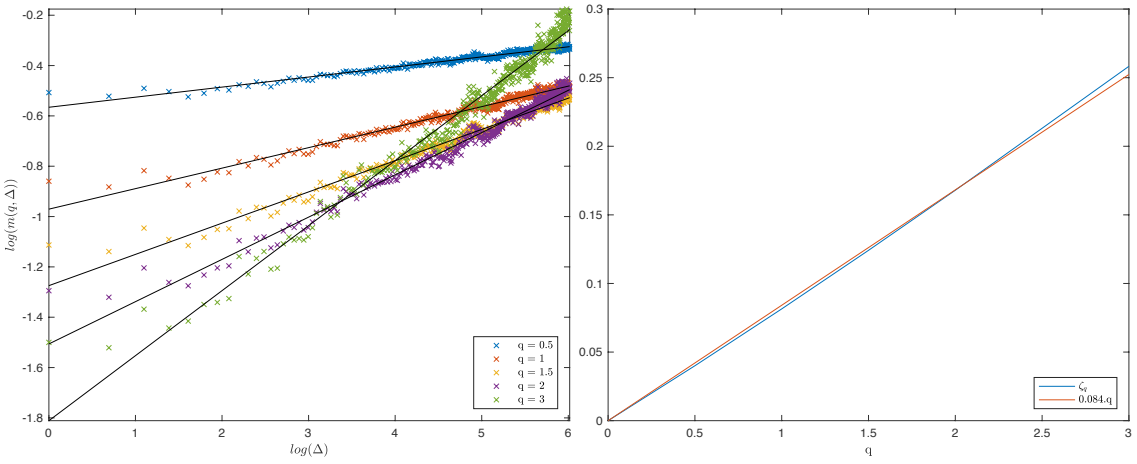


FIGURE 5.5 – $\log m(q,\Delta)$ as a function of $\log\Delta$ (left), ζ_q (blue) and $0.082\times q$ (green) (right), S&P 100 (Parkinson volatility).

The scaling is again linear in q with H = 0.082 for the S&P 100 and H = 0.064 for the IBEX 35. Values are very close to those found within the GK proxy. This confirms again that volatility based on the Parkinson proxy is rough.

The smoothness parameter H is detailed in Table 5.4 for all the assets for the Parkinson volatility. One more time, we give the value for the whole period, split the period into two halves and compute H for each half.

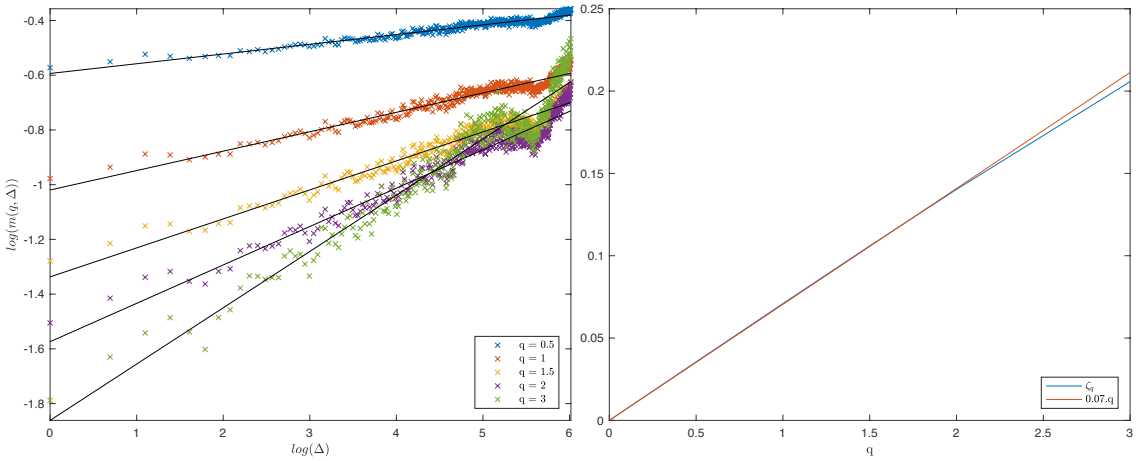


FIGURE 5.6 – $\log m(q,\Delta)$ as a function of $\log\Delta$ (left), ζ_q (blue) and $0.064 \times q$ (green) (right), IBEX 35 (Parkinson volatility).

The same conclusions apply to the Parkinson volatility :

Ticker	H for the whole period	H (first half)	H (second half)
SP100	0.0822	0.0888	0.0737
IBEX35	0.0644	0.0682	0.0648
HSI	0.0452	0.0555	0.0336
MEXBOL	0.0638	0.0738	0.0489
FTSE100	0.0774	0.0823	0.0669
ASX200	0.0513	0.0511	0.0422
TOTAL	0.0738	0.0856	0.0608
XIN9I	0.0595	0.0592	0.0593
SHSZ300	0.0591	0.0668	0.0499
BCOM	0.0127	0.00623	0.0251
INDU	0.08	0.08	0.0707
USDEUR	0.0265	0.0276	0.0261
IBOV	0.0694	0.0746	0.0603
MICROSOFT	0.0584	0.0685	0.0414
GOOGLE	0.0603	0.063	0.0542
SP400	0.0757	0.0822	0.0623

TABLEAU 5.4 – Estimates of H on the whole period and over two different time intervals for different indexes and stocks (Parkinson volatility)

— H remains between 0.01 and 0.09 with most assets around 0.07. This fact confirms that volatility is rough.

— H is higher for the first period. This might be explained by the fact that this period (period 2005-2010) contains the 2008 crisis.

— H for Parkinson proxy of the BCOM [3] asset is almost 0.

3. BCOM corresponds to the B Communications Ltd, which is a publicly traded holding company, headquartered in Israel, whose sole asset is a controlling interest in Israeli telecommunications provider Bezeq.

5.3.3 Distribution of the increments of the log-volatility

Now that we have established the common scaling behavior for our volatility proxies on the given data, we will focus on the S&P 100 for the following results. Certainly, we will ensure that the same results are common for other assets, but unless specified otherwise, all the plots concern the S&P 100.

It is well-known that the increments of log-volatility distribution is very close to the normal distribution, see for example [10]. This is also what we find in our data, see Figure 5.7. Moreover, rescaling the density by Δ^{H} for any given lag Δ recovers the 1-day increments density. This is consistent with the fractional Brownian motion with Hurst parameter H as seen in the previous section.

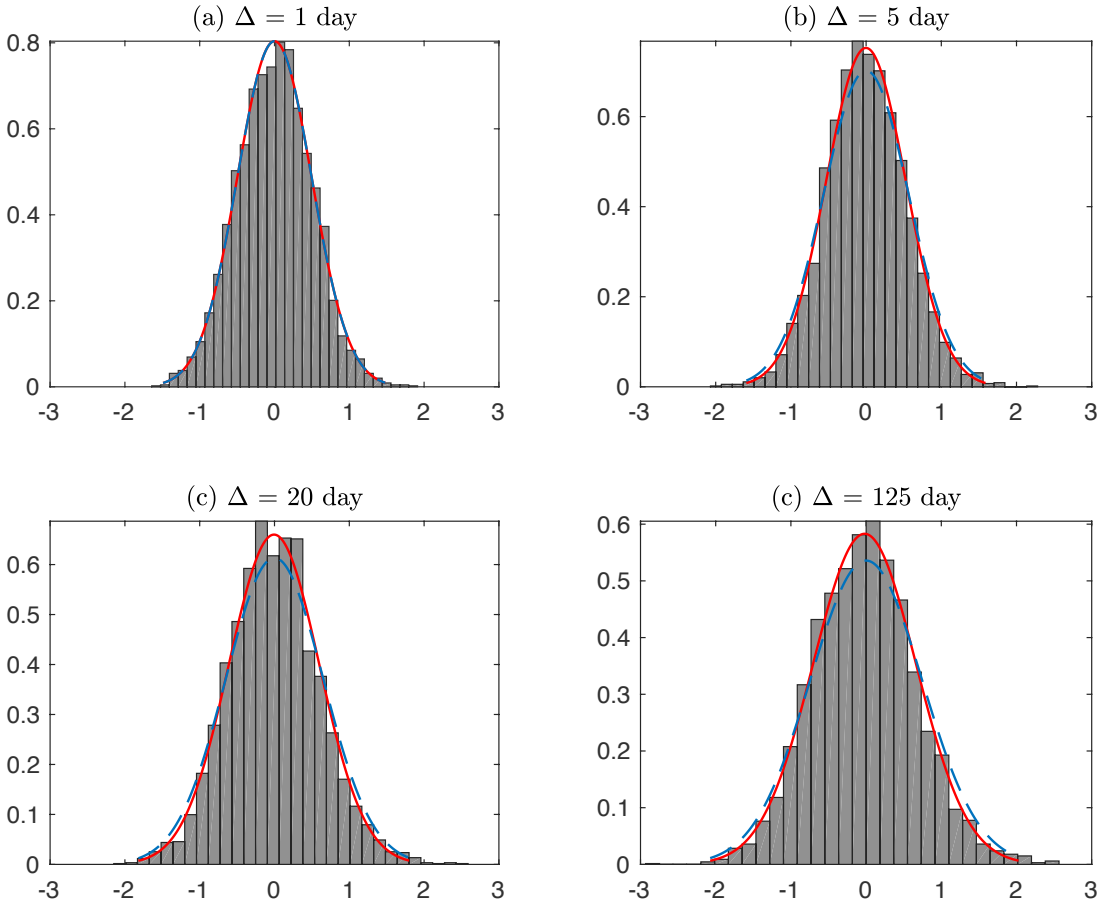


FIGURE 5.7 – Histograms for various lags Δ of the increments $\log(\sigma_{t+\Delta}) - \log(\sigma_t)$ of the Garman-Klass S&P 100 log-volatility; the normal fit to distribution of the Δ-days increments (red); normal fit to the 1-day increments rescaled by Δ^{H} (dashed blue)

5.4 RFSV model validation using range-based proxies

In the following section, we test the accuracy of the Rough fractional volatility model (RFSV) introduced in [89] using the volatility data based on the Garman-Klass proxy. We will ensure that the model reproduces the same behavior as the data.

5.4.1 The model

Empirical results of Section 5.2, show that the increments of log-volatility based on range proxies for various assets appear to have a scaling property with constant Hurst parameter. We also made sure, through Section 5.3.3 that their distribution is close to a normal distribution. Based on these results, the log-volatility increments can be modeled by the increments of a fBm as the following :

$$\log\sigma_{t+\Delta} - \log\sigma_t = \nu\left(W^H_{t+\Delta} - W^H_t\right), \tag{5.4}$$

where W^H is a fractional Brownian motion with the Hurst parameter estimated through the scaling of the volatility and ν is a positive constant corresponding to the volatility of the increments. We can rewrite Equation (5.4) under the form :

$$\sigma_t = \sigma\exp(\nu W^H_t),$$

where σ is a positive constant.

One of the drawbacks of this model is that it is not stationary. As a matter of fact, stationarity is a property that is desirable and useful for modeling time series. A possible model that keeps this property along with the fractional scaling is the fractional Ornstein-Uhlenbeck (fOU in short) process with a very long mean-reversion.

The fractional Ornstein-Ublenbeck process X_t is a stochastic process satisfying the stochastic differential equation :

$$dX_t = -\alpha X_t dt + \nu dZ_t,\ X_0 = 0$$

where both ν and α are positive parameters. When Z_t is the standard Brownian motion, we get the standard Ornstein-Uhlenbeck, see [136]. Our interest, however, is the when $Z_t = W^H_t$. We also consider an arbitrary initial point $X_0 = m$ instead of 0. The SDE followed by the the process of log-volatility becomes :

$$dX_t = -\alpha(X_t - m)dt + \nu dW^H_t,\ \ x_0 = m \tag{5.5}$$

where $m \in \mathbb{R}$ and (W^H_t) is the fBm with Hurst parameter H.

Equation (5.5) is then solved using the following explicit representation :

$$X_t = \nu\int_{-\infty}^{t} e^{-\alpha(t-s)}dW^H_t + m \tag{5.6}$$

where the stochastic integral with respect to fBM is simply a path-wise Riemann-Stieljes integral ([51]). Lastly, we recover the volatility, and thus define the RFSV model on the time interval [0,T] :

$$\sigma_t = \exp(X_t),\ \ t \in [0,T],$$

where (X_t) satisfies equation (5.6) for some $\nu > 0$, $\alpha > 0$, $m \in \mathbb{R}$ and $H < 1/2$ the measured smoothness of the volatility. In addition to the stationarity of such a model, choosing $\alpha \ll 1/T$ allows the log-volatility to behave locally (at time scales smaller than T) as a fBm. This observation is formalized by Proposition 3.1 in [89] we recall below :

Proposition 5. *Let* W^H *be a fBm and* X^α *defined by (5.6) for a given* $\alpha > 0$. *As* α *tends to zero,*

$$\mathbb{E}[\sup_{t\in[0,T]} |X^\alpha_t - X^\alpha_0 - \nu W^H_t|] \to 0$$

Remark 12. *:*

— *Proposition (5) implies that within the interval [0,T], and under the condition* $\alpha \ll 1/T$*, we can proceed as if the log-volatility process were a fBm. Setting* $\alpha = 0$ *allows to recover the simple non-stationary fBm (5.4).*

— *The RFSV differs from the classical FSV model of Comte and Renault, see [56], in that, instead of taking* $H > 1/2$ *and* α *large in FSVr, the RFSV model is defined for* $H < 1/2$ *and* α *small (actually* α *is chosen not to be equal to* 0 *only so that the volatility satisfies stationarity).*

— *The choice* $H < 1/2$ *is consistent with both the statistical properties of the data and generates a term structure of volatility skew that matches the observations.*

— *The choice of the fOU is for convenience and simplicity. Other models that imitate the fBm behavior at reasonable time scales and are stationary can be considered as well.*

— *The RFSV process reproduces approximately the exact scaling property as the fBm. This is a consequence of the following corollary :*

Corollary 1. *Let* $q > 0$*,* $t > 0$*,* $\Delta > 0$*. As* α *tends to zero, we have :*

$$\mathbb{E}[|\, X^{\alpha}_{t+\Delta} - X^{\alpha}_{t}\,|^{q}] \to \nu^{q} K_{q} \Delta^{qH}.$$

A detailed comparison between the RFSV and FSV can be found in [89].

5.4.2 Model validation

In previous sections, we come to the conclusion that RFSV model seems to be a relevant volatility model based on empirical results. But the question that arises is whether the estimated range-based volatility proxies, from simulated data with RFSV as the volatility process, behaves like the underlying process (in terms of the scaling properties). Our goal is to investigate this question. To do so, we simulate the spot volatility process using the RFSV model, simulate intraday prices, recover range prices (open, close, high and low), and finally estimate the range volatility from the simulated range prices. We will be able, on the one hand, to compare the behavior of the known real spot volatility process to that of the proxy used for its approximation. We will also be able to estimate the realized volatility and compare it to the Garman-Klass and to real data.

Spot volatility is simulated using RFSV model for 2,521 days. Since range-based volatility assumes the volatility process to be constant within the day, we take into account the randomness of the intraday prices P_u where u frequency is of the order of a few seconds, and simulate the spot volatility σ_t on a daily basis. We choose parameters that are consistent with the S&P 100 Garman-Klass volatility (even though a precise estimation for ν α and m is not simple) and consistent with our empirical estimates from Section 1.3, i.e. $H = 0.08$, $\nu = 0.3$, $m = X_0 = -5$ and $\alpha = 5 \times 10^{-4}$.

To simulate the volatility and price paths, we proceed as the following :

— Simulate fBm using a wavelet-based synthesis, see [2].

— Simulate the log-volatility process X for each day n according to a discrete scheme :

$$X_{n+1} - X_n = \nu \left(W^{H}_{n+1} - W^{H}_{n}\right) + \alpha(m - X_n).$$

— Simulate the asset price P by taking :

$$P_{n+(j+1)\delta} - P_{n+j\delta} = P_{n+j\delta}\sigma_n \sqrt{\delta} U_j,$$

where the U_j are iid standard Gaussian variables.

— Extract range prices (open, close, high and low prices) for each day, and compute the realized volatility and the Garman-Klass range-volatility.

We present in Figure 5.8 a plot of the Garman-Klass proxy from S&P 100 data along with the simulated spot volatility described above.

We compare the plots of the S&P 100 Garman-Klass proxy with the simulated path. Graphically, it seems that estimated volatility exhibits the same behavior as the simulation, at least to a visual extent. This was already verified in [89]. A zoom in or a zoom out gives typically the same kind of qualitative properties. To compare the smoothness of the real spot volatility (simulated paths), with that of the

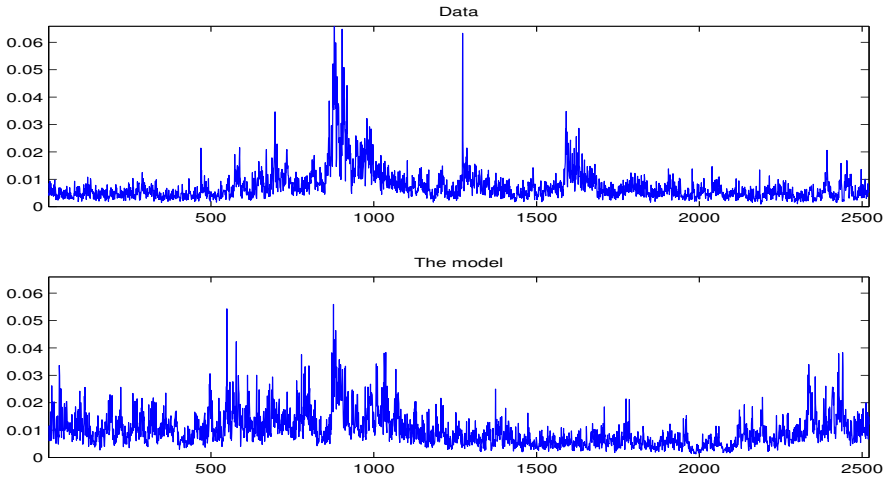


FIGURE 5.8 – Garman Klass volatility of S&P 100 (above) and simulated paths (below)

Garman-Klass proxy recovered from simulated prices, we repeat the analysis of Section 5.2. We plot in Figure 5.9 $\log(m(q,\Delta))$ as a function of $\log(\Delta)$ for both the real volatility and volatility proxy. The Hurst exponent of the estimated Garman-Klass proxy is relatively close to the true one (H = 0.079 compared to H = 0.056). The mismatch can in particular be due to simulation bias.

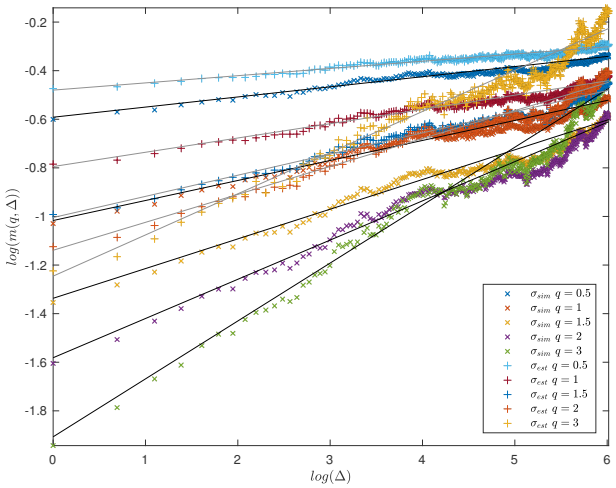


FIGURE 5.9 – $\log(m(q,\Delta))$ as a function of $\log(\Delta)$ for the real spot volatility with H = 0.08 and the Garman Klass proxy based on simulations.

5.4.3 FSV vs. RFSV

We have shown from our analysis that the RFSV model hypothesis can not be rejected. In light of empirical results on data, we would like to test if we can reject the FSV model through simulating the volatility process. To do so, we analyze the behavior of the smoothing function $m(q,\Delta)$ for small and large lags, using simulated volatility and asset prices, and see if the Garman-Klass and realized-volatility based on these simulations, behave like the one found for real data.

The first basic difference between RFSV and FSV is the range of the Hurst exponent values for the fractional Brownian motion. First, we consider the simple non mean-reverting fractional volatility model with $H > 0.5$, i.e. $\sigma_t = \sigma_0 e^{\nu W_t^H}$ with $H = 0.7$ and $\eta = 0.25$. We take a look at the scaling behavior of the realized-volatility and Garman-Klass volatility proxies based on this model in Figure 5.10. In Figure

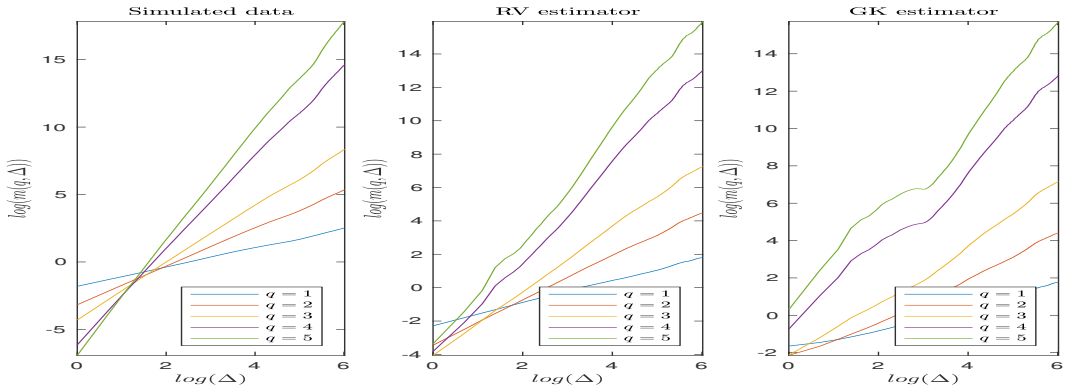


FIGURE 5.10 – $\log(m(q,\Delta))$ as a function of $\log(\Delta)$ for the data (left) Realized Volatility (center) and Garman-Klass volatility (right) over the simulated paths for $q = 1,2,3,4,5$

5.10, we see that the scaling of $\log(\Delta)$ is very close to a straight line for the RV and GK estimators. The resulting smoothing parameter found in these figures is close to the original one ($H = 0.69$ for RV estimator and $H = 0.64$ for GK estimator).

We are aware that such model, leading to crazy volatility values, does not make sense without mean-reversion. However, this allows us to exclude a fractional volatility model of the form $\sigma_t = \sigma_0 e^{\nu W_t^H}$ with $H > 0.5$.

Unlike the RSFV model where the mean-reversion is intrinsic to the model for $\alpha = 0$, Comte and Renault impose α to be large enough, i.e. $\alpha \gg 1/T$ where T is the time horizon of interest, to verify this property. We would like to test to what extent this model can be misleading in estimating the smoothness of the diffusion process. We compare the FSV and RFSV models for the set of parameters given in Table 5.5, which leads to the simulated time series given in Figure 5.11.

	FSV	RFSV
H	0.7	0.08
α	0.25	5×10^{-4}
ν	0.25	0.45
m	-4.5	-5
X_0	-4.5	-5
$\mathbb{E}[\log(\sigma)]$	-4.6	-4.7
$\mathbb{V}ar[\log(\sigma)]$	0.21	0.33

TABLEAU 5.5 – Parameters values used for simulating the FSV and RFSV models and the mean and variance of the simulated time series

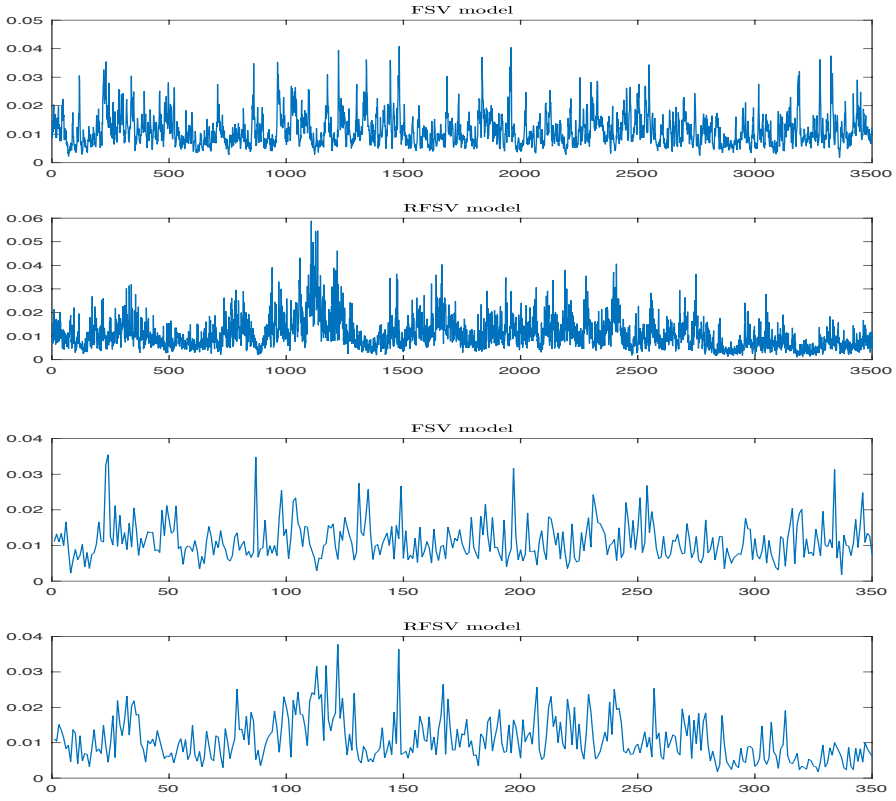


FIGURE 5.11 – Plot of one path of the FSV and RFSV models for H = 0.7 and H = 0.08 observed daily (top two), and every 10 days (bottom two).

In Figure 5.11, we see that both processes satisfy mean-reversion. The roughness of the RFSV model is quite clear when we observe the process daily, however, when observed every 10 days, it is less obvious to say which is the FSV and which is the RFSV.

In Figure 5.12, we check one more time the scaling behavior of the two processes and their RV and GK proxies estimated on a 24 hours windows and 1 second observed prices. We confirm through Figure 5.12 the following key results :

— The smoothing function ζ_q for the the RFSV model keeps the same pattern for short and long time scales, ($\log(m(q,\Delta))$ is close almost linear w.r.t $\log(\Delta)$).

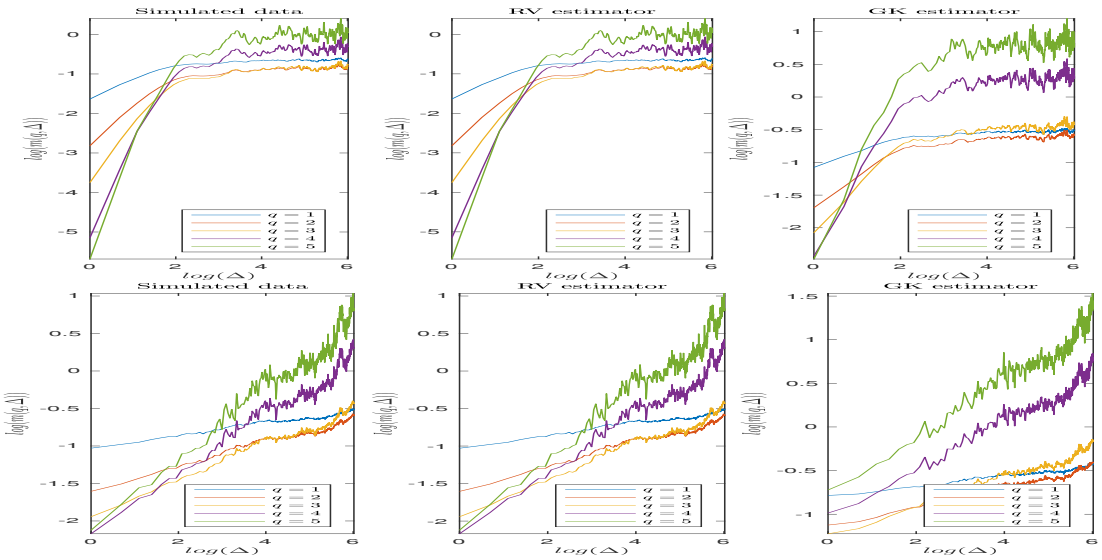


FIGURE 5.12 – $\log(m(q,\Delta))$ as a function of $\log(\Delta)$ for the simulated paths (left), Realized Volatility (center) and Garman Klass volatility (right) over the simulated paths for $q = 1,2,3,4,5$, for FSV model (top) and RFSV model (bottom)

— The FSV model seems to exhibit two slopes. At small scales, the slope is close to that of the Hurst exponent of the fractional Brownian motion that drives the process, i.e. $H \approx 0.7$. At large scales, the slope gives a value close 0. Actually, the stationarity of the process at large scales is responsible for such estimation.

Finally, in order to verify the impact of discretization on the estimators, we consider that prices are observed on an 8 hours time window every 1, 5 or 10 minutes. We compute the RV and GK estimators based on these observations. Results are shown in Figure 5.13 for the FSV (this test on the RFSV model does not bring more information).

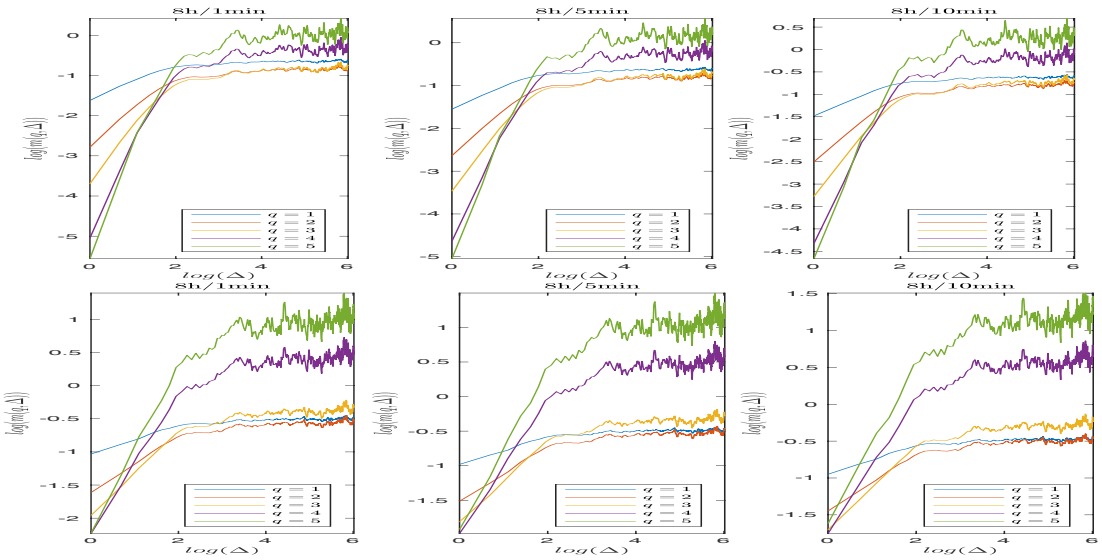


FIGURE 5.13 – $\log(m(q,\Delta))$ as a function of $\log(\Delta)$ for realized volatility (top) and Garman-Klass volatility (bottom) based on an 8 hours window with 1min (right), 5min (center) and 10min (left) discretization for the mean-reverting FSV model with H = 0.07.

We observe an other phenomena for the GK estimator. The slope on very small scales ($\Delta = 1,2$) gives a relatively smaller smoothing parameter (H around 0.19) than for intermediate lags (H around 0.24). We believe that this is due to the noise of this estimator. Moreover, the value of the smoothing parameter is still very small compared the one used to simulate the process.

Finally, we can conclude that FSV volatility with H > 0.5 behaves differently than the data. It may therefore be excluded from being a good volatility model. Rough fractional model on the other hand, seems at this point, the most plausible to model the volatility. Even though it was not illustrated in here, mean-reversion models with standard Brownian diffusion is even less plausible.

Quantitative justification

In this section, we would like to quantify the phenomena encountered previously. Our goal is to see how estimating the smoothing parameter is affected by the lag range. We denote by X_t the asset log price $X_t = \log(P_t)$. Since spot volatility does not really make sense at the intraday level (beyond seasonality), and to avoid the smoothing issue of realized volatility estimation, we assume that the volatility is constant within a day, i.e. for day i and for each time $t \in [i, i+1)$ $\sigma_t = \sigma_i =$ constant , with $\sigma_i = e^{\nu W_i^H}$ (although this model does not make much sense for H > 0.5).

Assume we have n observations for each day i for the price process $P_i^{j=1,\dots,n}$, and let $\Delta_j^n X$ the log price increments of day i :

$$\Delta_j^n X = \log(P_i^j) - \log(P_i^{j-1}).$$

Using the central limit theorem (CLT) for realized volatility we have :

$$\begin{aligned}(\sigma_i^{RV})^2 &= \sum_{j=1}^{n} (\Delta_j^n X)^2 \\ &\simeq \int_i^{i+1} e^{2\nu W_i^H} dt + \frac{1}{\sqrt{n}} \sqrt{2 \int_i^{i+1} e^{4\nu W_i^H} dt}\xi \text{ where } \xi \sim \mathcal{N}(0,1) \\ &\simeq e^{2\nu W_i^H} \big(1 + \sqrt{\frac{2}{n}}\xi\big),\end{aligned}$$

which leads to the approximation :

$$\log(\sigma_i^{RV}) \simeq \nu W_i^H + \sqrt{\frac{1}{2n}}\xi \text{ where } \xi \sim \mathcal{N}(0,1).$$

Taking the increments of the log volatility between i and $i+\Delta$ for a given time lag Δ, we have :

$$\log(\sigma_{i+\Delta}^{RV}) - \log(\sigma_i^{RV}) \simeq \underbrace{\nu\big(W_{i+\Delta}^H - W_i^H\big)}_{O\left(\nu\Delta^H\right)} + \underbrace{\sqrt{\frac{1}{2n}}(\xi + \xi')}_{O\left(\sqrt{\frac{1}{n}}\right)},$$

where ξ and ξ' are i.i.d Gaussian variables.

This equivalence leads to the following observations :

— When the vol of vol ν is small,i.e. such $\nu^2 \Delta^{2H} \ll \frac{1}{n}$, the noise takes the upper hand. As a result, we would observe a slope close to horizontal when plotting $\log(m(q,\Delta))$ against $\log(\Delta)$.

— On the contrary, when ν is large, the first term is predominant and the slope is close to qH which allows for a rather more precise estimation of the true Hurst exponent.

— In the intermediate case, two slopes can be observed; one corresponding to estimation noise, and the other proportional to the true Hurst exponent.

We illustrate these cases in Figure 5.14 below.

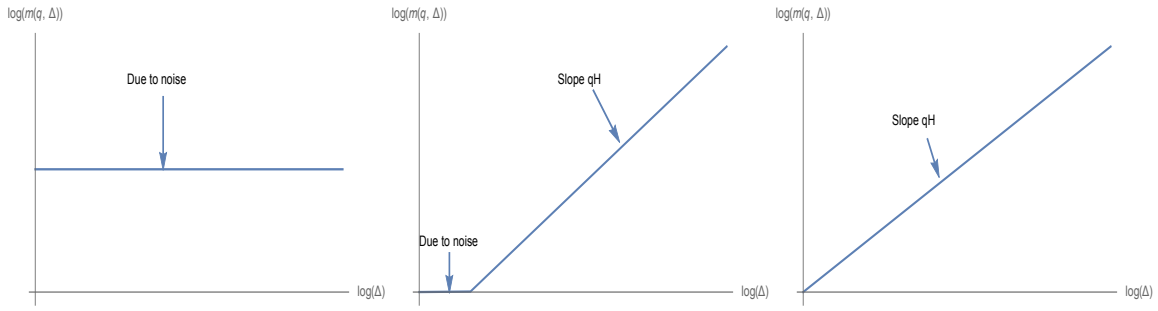


FIGURE 5.14 – Different cases encountered for the estimation of the slope of $\log(m(q,\Delta))$ against $\log(\Delta)$.

5.5 Forecasting range-based volatility using the RFSV model

The purpose of this section is to compare the predictability of the RFSV model with other commonly used models such as the AR, HAR or GARCH models.

5.5.1 Forecasting log-volatility

The key formula on which the prediction method is based is the following one :

$$\mathbb{E}[W^H_{t+\Delta} \mid \mathscr{F}_t] = \frac{\cos(H\pi)}{\pi} \Delta^{H+1/2} \int_{-\infty}^{t} \frac{W^H_s}{(t-s+\Delta)(t-s)^{H+1/2}} ds,$$

where W^H is a fBm with $H < 1/2$ and $\mathscr{F}_t$ the filtration it generates, see Theorem 4.2 of [135]. By construction, over any reasonable time scale of interest, as formalized in Corollary 1, we may approximate the fOU volatility process in the RFSV model as $\log\sigma^2_t \approx 2\nu W^H_t + C$ for some constants ν and C. Our prediction formula for the log-variance then follows :

$$\mathbb{E}[\log\sigma^2_{t+\Delta} \mid \mathscr{F}_t] = \frac{\cos(H\pi)}{\pi} \Delta^{H+1/2} \int_{-\infty}^{t} \frac{\log\sigma^2_s}{(t-s+\Delta)(t-s)^{H+1/2}} ds. \tag{5.7}$$

This formula, or rather its approximation through a Riemann sum (we assume in this section that volatilities are perfectly observed, although they are in fact estimated), is used to forecast the log-volatility 1,5 and 20 days ahead ($\Delta = 1, 5, 20$).

In the spirit of [58], we compare the predictive power of Formula (5.7) with that of AR, HAR and GARCH forecasts. Recall that for a given integer $p > 0$, the AR(p) and HAR predictors take the following form (where the index i runs over the series of daily volatility estimates) :

— AR(p) :

$$\widehat{\log(\sigma^2_{t+\Delta})} = K_0^{\Delta} + \sum_{i=0}^{p} C_i^{\Delta} \log(\sigma^2_{t-i}).$$

— HAR :

$$\widehat{\log(\sigma^2_{t+\Delta})} = K_0^{\Delta} + C_0^{\Delta}\log(\sigma^2_t) + C_5^{\Delta}\frac{1}{5}\sum_{i=0}^{p} C_i^{\Delta}\log(\sigma^2_{t-i}) + C_{20}^{\Delta}\frac{1}{20}\sum_{i=0}^{20}\log(\sigma^2_{t-i}).$$

We estimate AR and HAR using a rolling time window of 500 days. For the HAR case, we use standard linear regression to estimate the coefficients as explained in [58]. In the sequel, we consider $p = 5$ and $p = 10$ in the AR formula. Indeed, these parameters essentially give the best results for the horizons at which we wish to forecast the volatility (1, 5 and 20 days). For each day, we forecast volatility for five different indexes.

We then assess the quality of the various forecast by computing the ratio P between the mean squared error of our predictor and the approximated variance of the log-variance :

$$P = \frac{\sum_{k=500}^{N-\Delta}\left(\log(\sigma^2_{k+\Delta}) - \widehat{\log(\sigma^2_{t+\Delta})}\right)^2}{\sum_{k=500}^{N-\Delta}\left(\log(\sigma^2_{k+\Delta}) - \mathbb{E}[\log(\sigma^2_{k+\Delta})]\right)^2},$$

where $\mathbb{E}[\log(\sigma^2_{k+\Delta})]$ denotes the empirical mean of the log-variance over the whole period.

We present in Table 5.6 the ratio P for different models in order to compare the RFSV prediction power with other autoregressive models for predicting the log variance.

Ticker	AR(5)	AR(10)	HAR(3)	RFSV
SP100 $\Delta = 1$	0.451	0.446	0.443	0.466
SP100 $\Delta = 5$	0.644	0.635	0.546	0.557
SP100 $\Delta = 21$	0.897	0.894	0.734	0.718
IBEX35 $\Delta = 1$	0.594	0.594	0.582	0.622
IBEX35 $\Delta = 5$	0.843	0.824	0.728	0.728
IBEX35 $\Delta = 21$	1.18	1.17	0.943	0.908
HSI $\Delta = 1$	0.529	0.523	0.513	0.52
HSI $\Delta = 5$	0.647	0.633	0.575	0.577
HSI $\Delta = 21$	0.805	0.801	0.665	0.671
MEXBOL $\Delta = 1$	0.572	0.567	0.553	0.589
MEXBOL $\Delta = 5$	0.731	0.709	0.648	0.645
MEXBOL $\Delta = 21$	0.922	0.917	0.757	0.764
FTSE100 $\Delta = 1$	0.474	0.465	0.463	0.476
FTSE100 $\Delta = 5$	0.627	0.614	0.545	0.545
FTSE100 $\Delta = 21$	0.859	0.855	0.699	0.688
ASX200 $\Delta = 1$	0.536	0.524	0.524	0.527
ASX200 $\Delta = 5$	0.658	0.652	0.577	0.573
ASX200 $\Delta = 21$	0.806	0.793	0.707	0.688
TOTAL $\Delta = 1$	0.540	0.534	0.527	0.558
TOTAL $\Delta = 5$	0.720	0.704	0.640	0.636
TOTAL $\Delta = 21$	1.008	1.015	0.809	0.789
XIN9I $\Delta = 1$	0.587	0.58	0.568	0.582
XIN9I $\Delta = 5$	0.712	0.695	0.637	0.641
XIN9I $\Delta = 21$	0.913	0.918	0.762	0.758
SHSZ300 $\Delta = 1$	0.574	0.568	0.56	0.572
SHSZ300 $\Delta = 5$	0.707	0.695	0.634	0.634
SHSZ300 $\Delta = 21$	0.896	0.904	0.772	0.753

BCOM $\Delta = 1$	0.846	0.838	0.805	0.83
BCOM $\Delta = 5$	0.876	0.854	0.821	0.825
BCOM $\Delta = 21$	0.956	0.937	0.874	0.854
INDU $\Delta = 1$	0.451	0.446	0.444	0.458
INDU $\Delta = 5$	0.617	0.612	0.532	0.541
INDU $\Delta = 21$	0.858	0.857	0.716	0.699
USDEUR $\Delta = 1$	0.530	0.514	0.507	0.521
USDEUR $\Delta = 5$	0.611	0.581	0.532	0.544
USDEUR $\Delta = 21$	0.755	0.728	0.618	0.638
IBOV $\Delta = 1$	0.602	0.595	0.587	0.617
IBOV $\Delta = 5$	0.779	0.754	0.68	0.691
IBOV $\Delta = 21$	1.010	1.008	0.843	0.836
MICROSOFT $\Delta = 1$	0.579	0.576	0.566	0.603
MICROSOFT $\Delta = 5$	0.749	0.737	0.668	0.673
MICROSOFT $\Delta = 21$	0.936	0.931	0.807	0.79
GOOGLE $\Delta = 1$	0.500	0.497	0.492	0.529
GOOGLE $\Delta = 5$	0.683	0.672	0.581	0.595
GOOGLE $\Delta = 21$	0.864	0.861	0.729	0.722
SP400 $\Delta = 1$	0.454	0.451	0.445	0.464
SP400 $\Delta = 5$	0.616	0.601	0.525	0.538
SP400 $\Delta = 21$	0.816	0.81	0.668	0.67

TABLEAU 5.6 – Ratio P for AR, HAR and RFSV predictors for $\log(\sigma^2_{t+\Delta})$

As we can see in Table 5.6, even though RFSV sometimes underperform AR, and HAR for $\Delta = 1$, it performs at least as good as the HAR when predicting more days ahead ($\Delta = 5, 21$) and outperforms the AR model.

Compared to AR, HAR whose parameters change through time, depend on the time horizon, need to be re-calibrated and even encounter calibration issues for some periods, the RFSV is more parsimonious since it only requires the parameter H to forecast the log-variance. In addition to that, the smoothness typically does not change over time or very slightly.

We notice that prediction through the RFSV can be linked to that of [67], where the issue of the prediction of the log-volatility in the multifractal random walk model of [12] is tackled. In this model,

$$\mathbb{E}[\log(\sigma^2_{t+\Delta}) \mid \mathscr{F}_t] = \frac{1}{\pi}\sqrt{\Delta}\int_{-\infty}^{t} \frac{\log(\sigma^2_s)}{(t-s+\Delta)\sqrt{t-s}}\,dt,$$

which is the limit of our predictor when H tends to zero.

The prediction formula for the RFSV model can also be rewritten as

$$\mathbb{E}[\log(\sigma^2_{t+\Delta}) \mid \mathscr{F}_t] = \frac{\cos(\mathrm{H}\pi)}{\pi}\int_0^{+\infty} \frac{\log(\sigma^2_{t-\Delta u})}{(u+1)u^{\mathrm{H}+1/2}}\,du,$$

for a given small $\varepsilon > 0$, let r be the smallest real number such that

$$\int_r^{+\infty} \frac{1}{(u+1)u^{\mathrm{H}+1/2}} \leq \varepsilon.$$

Then we have, with an error of order ε

$$\mathbb{E}[\log(\sigma^2_{t+\Delta}) \mid \mathscr{F}_t] \approx \frac{\cos(\mathrm{H}\pi)}{\pi} \int_0^r \frac{\log(\sigma^2_{t-\Delta u})}{(u+1)u^{\mathrm{H}+1/2}} du.$$

This prediction formula says that future volatility depends on the whole path of the volatility process. However, since the weights decrease with time, one does not need to go to $-\infty$. It suffice to consider a time to go down to in order to forecast the future. This is roughly defined by setting the error margin ϵ. For example, in order to forecast Δ in the future, it is common practice by practitioners to take Δ in the past. This corresponds to $r = 1$, and $\epsilon = 0.35$ which is not so unreasonable.

5.5.2 Predicting the variance

Based on the same approximation of the fOU volatility process in the RFSV model, we rewrite $\sigma_t^2 = \exp(2\nu W_t^{\mathrm{H}} + C)$ for some constants ν and C. The prediction of the variance knowing the information at time t is :

$$\begin{aligned}\widehat{\sigma^2_{t+\Delta}} &= \mathbb{E}\left[\sigma^2_{t+\Delta} \mid \mathscr{F}_t\right] \\ &= \mathbb{E}\left[\exp(2\nu W^{\mathrm{H}}_{t+\Delta} + C) \mid \mathscr{F}_t\right].\end{aligned}$$

Since $W^{\mathrm{H}}_{t+\Delta}$ is conditionally Gaussian (as shown by [135]) with conditional variance $Var[W^{\mathrm{H}}_{t+\Delta} \mid \mathscr{F}_t] = c\Delta^{2\mathrm{H}}$ (where $c = \frac{\Gamma(3/2-\mathrm{H})}{\Gamma(\mathrm{H}+1/2)\Gamma(2-2\mathrm{H})}$) and using the fact that $\widehat{\log(\sigma_t^2)} \approx \mathbb{E}\left[2\nu W_t^{\mathrm{H}} + C \mid \mathscr{F}_t\right] = \mathbb{E}\left[\log(\sigma^2_{t+\Delta}) \mid \mathscr{F}_t\right]$, we have :

$$\widehat{\sigma^2_{t+\Delta}} = \exp\left(\widehat{\log(\sigma^2_{t+\Delta})} + 2c\nu^2\Delta^{2\mathrm{H}}\right)$$

Note that this expression uses the estimation of $\widehat{\log\left(\sigma^2_{t+\Delta}\right)} = \mathbb{E}\left[\log(\sigma^2_{t+\Delta}) \mid \mathscr{F}_t\right]$ which we have seen in Section 5.5.1 and ν^2 which is the exponential of the intercept in the linear regression of $\log(m(2,\Delta))$ on $\log(\Delta)$.
Once again, we compare the performance of the RFSV predictor to the AR, HAR and the GARCH predictors expressed as the following :

— AR(p) :

$$\widehat{\sigma^2_{t+\Delta}} = \mathrm{K}_0^{\Delta} + \sum_{i=0}^{p} \mathrm{C}_i^{\Delta} \sigma^2_{t-i}$$

— HAR :

$$\widehat{\sigma^2_{t+\Delta}} = \mathrm{K}_0^{\Delta} + \mathrm{C}_0^{\Delta}\sigma_t^2 + \mathrm{C}_5^{\Delta}\frac{1}{5}\sum_{i=0}^{p}\mathrm{C}_i^{\Delta}\sigma^2_{t-i} + \mathrm{C}^{\Delta}_{20}\frac{1}{20}\sum_{i=0}^{20}\sigma^2_{t-i}$$

— GARCH(1,1) :

$$\widehat{\sigma^2_{t+\Delta}} = \alpha_0\left(1 + \sum_{i=1}^{\Delta-1}(\alpha_1+\beta_1)^i\right) + (\alpha_1+\beta_1)^{\Delta}\sigma_t^2$$

Results on the variance prediction are given in Table 5.7. We can see that the RFSV model outperforms other predictors on all the considered time horizons. GARCH model performs poorly on the other hand.

Ticker	AR(5)	AR(10)	HAR(3)	GARCH(1,1)	RFSV
SP100 $\Delta = 1$	0.901	1.01	0.769	0.873	0.655
SP100 $\Delta = 5$	1	0.96	1.06	1.14	0.76
SP100 $\Delta = 21$	1.42	1.33	0.989	1.72	0.898
IBEX35 $\Delta = 1$	0.62	0.632	0.587	0.675	0.694
IBEX35 $\Delta = 5$	1	1.01	0.846	1.27	0.808
IBEX35 $\Delta = 21$	1.43	1.45	1.03	2.03	0.975
HSI $\Delta = 1$	0.993	1.16	0.847	1.37	0.794
HSI $\Delta = 5$	0.875	1.16	0.919	1.59	0.851
HSI $\Delta = 21$	1.1	1.56	0.977	1.81	0.932
MEXBOL $\Delta = 1$	0.58	0.591	0.566	0.686	0.659
MEXBOL $\Delta = 5$	0.938	0.866	0.808	1.26	0.767
MEXBOL $\Delta = 21$	1.32	1.31	0.965	1.77	0.928
FTSE100 $\Delta = 1$	0.65	0.67	0.618	0.776	0.646
FTSE100 $\Delta = 5$	0.833	0.908	0.808	1.25	0.721
FTSE100 $\Delta = 21$	1.15	1.21	0.926	1.69	0.872
ASX200 $\Delta = 1$	0.789	0.834	0.688	1.08	0.656
ASX200 $\Delta = 5$	0.826	0.845	0.721	1.25	0.725
ASX200 $\Delta = 21$	1	1.05	0.851	1.8	0.837
TOTAL $\Delta = 1$	0.519	0.554	0.497	0.587	0.568
TOTAL $\Delta = 5$	0.804	0.855	0.77	1.01	0.695
TOTAL $\Delta = 21$	1.29	1.39	0.997	1.59	0.885
XIN9I $\Delta = 1$	0.847	0.86	0.816	1.24	0.787
XIN9I $\Delta = 5$	0.943	0.945	0.86	1.54	0.862
XIN9I $\Delta = 21$	1.1	1.11	0.932	1.87	0.926
SHSZ300 $\Delta = 1$	0.841	0.853	0.806	1.17	0.767
SHSZ300 $\Delta = 5$	0.964	0.963	0.859	1.51	0.849
SHSZ300 $\Delta = 21$	1.06	1.07	0.918	1.96	0.903
BCOM $\Delta = 1$	0.823	0.82	0.776	1.28	0.824
BCOM $\Delta = 5$	0.884	0.851	0.818	1.52	0.812
BCOM $\Delta = 21$	0.967	0.961	0.86	1.95	0.834
INDU $\Delta = 1$	1.22	1.4	0.883	0.938	0.677
INDU $\Delta = 5$	1.04	1.05	1.35	1.22	0.779
INDU $\Delta = 21$	1.43	1.42	0.986	1.76	0.904
USDEUR $\Delta = 1$	0.711	0.73	0.663	0.945	0.673
USDEUR $\Delta = 5$	0.878	0.855	0.762	1.25	0.716
USDEUR $\Delta = 21$	1.05	1.03	0.816	1.46	0.823
IBOV $\Delta = 1$	0.899	1.04	0.687	0.894	0.686
IBOV $\Delta = 5$	1.05	0.896	0.908	1.11	0.782
IBOV $\Delta = 21$	1.74	1.33	1.06	1.75	0.932
MICROSOFT $\Delta = 1$	0.656	0.67	0.644	0.925	0.659
MICROSOFT $\Delta = 5$	0.926	0.894	0.91	1.29	0.746
MICROSOFT $\Delta = 21$	1.12	1.15	0.95	1.61	0.866
GOOGLE $\Delta = 1$	0.652	0.661	0.611	0.887	0.586
GOOGLE $\Delta = 5$	0.715	0.737	0.66	1.08	0.651
GOOGLE $\Delta = 21$	1.09	1.11	0.87	1.43	0.831
SP400 $\Delta = 1$	0.704	0.761	0.634	0.916	0.626
SP400 $\Delta = 5$	0.961	0.882	0.778	0.986	0.725

SP400 $\Delta = 21$	1.27	1.19	0.882	1.56	0.866

TABLEAU 5.7 – Ratio P for the AR, HAR and RFSV predictors for $\sigma^2_{t+\Delta}$

5.6 Conclusion

We aimed through this analysis to investigate the scaling behavior of the volatility range-based proxy. Gatheral and co-authors have already studied high-frequency based realized-volatility in this sense in [89]. They found that it exhibits a rough scaling behavior and that its logarithm behaves like a fractional Brownian motion with Hurst exponent of order 0.14. We applied their analysis using the range-based volatilities as proxies for spot volatility. The latter use only information about range prices (open, close, high, low) for their estimation. We also find that the volatility process is monofractal with a small Hurst exponent (lower than 0.1) and can indeed be as low as 0.014.

Further tests justify that log-volatility increments are approximately Gaussian. This allows us to model them using the RFSV model. To ensure that the RFSV is consistent with the results on data, we simulate intraday and range prices using this model, recover the realized-volatility, and compute again the range-based estimators. We perform again the statistical checking only to find that simulations give similar results to the data. This reinforces that the hypothesis of rough volatility can not be rejected at this point, while other models like FSV seem to be misleading and, to a certain extent, wrong.

Finally, to measure its prediction power, we compare RFSV with other models such as AR or HAR for the log-volatility prediction and with AR, HAR and GARCH for the variance prediction. The RFSV shows good performance compared to other models and its prediction power is at least comparable to that of the HAR. GARCH, on the other hand, shows very weak performance for these estimators.

Chapitre 6

Rough volatility : evidence from option prices

***Abstract*— It has been recently shown that spot volatilities can be very well modeled by rough stochastic volatility type dynamics. In such models, the log-volatility follows a fractional Brownian motion with Hurst parameter smaller than 1/2. This result has been established using high frequency volatility estimations from historical price data. We revisit this finding by studying implied volatility based approximations of the spot volatility. Using at-the-money options on the S&P500 index with short maturity, we are able to confirm that volatility is rough. The Hurst parameter found here, of order 0.3, is slightly larger than that usually obtained from historical data. This is easily explained from a smoothing effect due to the remaining time to maturity of the considered options.**

Keywords : Rough volatility; fractional Brownian motion; implied volatility; Medvedev-Scaillet approximation.

6.1 Introduction

Since the seminal work of Black and Scholes [30], the most classical way to model the behavior of the price S_t of a financial asset is to use continuous semi-martingale dynamics of the form

$$d\log S_t = \mu_t dt + \sigma_t dW_t,$$

with μ_t a drift process and W_t a Brownian motion. The coefficient σ_t is referred to as the volatility process. As is well-know, it is the key ingredient in the model when one is interested in derivatives pricing and hedging.

Historically, following the pioneering approach of [30], practitioners have first considered the case where the process σ_t is constant or deterministic, that is the Black and Scholes model. However, in the late eighties, it became clear that such specification for the volatility is inadequate. In particular, the Black and Scholes model is inconsistent with observed prices for liquid European options. Indeed the implied volatility, that is the volatility parameter that should be plugged into the Black-Scholes formula to retrieve a market option price, depends in practice on the strike and maturity of the considered option, whereas it is constant in the Black-Scholes framework.

Hence more sophisticated models have been introduced. A first possible extension, proposed by Dupire [69] and Derman and Kani [63], is to take σ_t as a deterministic function of time and asset price. Such models, called local volatility models, enable us to perfectly reproduce a given implied volatility

surface. However, its dynamic is usually quite unrealistic under local volatility. Another approach is to consider the volatility σ_t itself as an Ito process driven by an additional Brownian motion, typically correlated to W. Doing so one obtains less accurate static fits for the implied volatility surface but more suitable dynamics. Among the most famous of these stochastic volatility models are the Hull and White model [101], the Heston model [97] and the SABR model [95]. More recent market practice is to use so-called local-stochastic volatility models, see for example [24], which both fit the market exactly and generate reasonable dynamics.

In all the Brownian volatility models mentioned above, the smoothness of the sample path of the volatility is the same as that of a Brownian motion, namely $1/2-\varepsilon$ Hölder continuous, for any $\varepsilon > 0$. However, it is shown in [88] that in practice, spot volatility is much rougher than this. This result in [88] is based on a statistical analysis of historical data using sophisticated high frequency estimation methods. More precisely, it is established in [88] that the dynamic of the log-volatility process is very close to that of a fractional Brownian motion with Hurst parameter smaller than 1/2. Recall that a fractional Brownian motion W^H with Hurst parameter $H \in (0,1)$ is a Gaussian process with stationary increments such that

$$\mathrm{Cov}[W_t^H, W_s^H] = \frac{1}{2}\left(|t|^{2H} + |s|^{2H} - |t-s|^{2H}\right).$$

The Hölder regularity of W^H is $H-\varepsilon$ for any $\varepsilon > 0$ and for $H = 1/2$ we retrieve the classical Brownian motion. Therefore models where the volatility is driven by a fractional Brownian motion with $H < 1/2$ are called rough volatility models. Beyond fitting almost perfectly historical volatility time series, rough volatility models enable us to reproduce important stylized facts of liquid option prices that local or stochastic (or local-stochastic) volatility models typically fail to generate. In particular, the exploding term structure when maturity goes to zero of the at-the-money skew (the derivative of the implied volatility with respect to strike) is readily obtained, see [20, 79]. Other developments about rough volatility models can be found in [22, 23, 70, 71, 74, 80, 94, 105, 134].

The goal of this paper is to revisit the finding in [88] using implied volatility data. Indeed in [88], the authors work with historical price data from underlyings to estimate spot volatility. Here we use a spot volatility proxy which is not based on historical data, but on implied volatility. More precisely, we approximate the spot volatility by the implied volatility of an at-the-money liquid option with short maturity (or a refined version of it). This idea can be justified by the fact that in most models, the at-the-money implied volatility tends to the spot volatility as maturity goes to zero, see for example [132]. Our main result is a confirmation of that in [88] : When using alternate spot volatility measurement methods based on option prices, we can still conclude that volatility is rough.

The paper is organized as follows. We investigate in Section 6.2 the roughness of time series of spot volatility approximations given by implied volatilities of at-the-money options on the S&P500 index, with maturity one month. In Section 6.3, instead of using raw implied volatilities, we compute spot volatilities from implied ones through a correction formula due to Medvedev and Scaillet, see [124]. We then carry the same analysis as in Section 6.2. The results in Sections 6.2 and 6.3 are very similar to those in [88]. However, the estimated values for the Hurst parameter, although smaller than 1/2, are actually larger than those obtained in [88]. We show numerically and analytically in Section 6.4 that this upward bias comes from a regularizing effect due to the remaining time to maturity of the considered options.

6.2 At-the-money implied volatility with short maturity as spot volatility proxy

As explained in the introduction, our goal is to study the behavior of the spot volatility and to show that it is well approximated by a rough process. Of course this is a difficult task since volatility is a latent, unobserved variable. In [88], the authors use recent estimation methods based on ultra high frequency price data to estimate spot volatility. In this work, instead of using historical data as in [88], we wish to use option price data. This idea is reasonable if we use at-the-money options for which the time to maturity is short. Indeed, it is well-known that in most models, the at-the-money implied volatility converges to the spot volatility as maturity goes to zero, see for example [132].

6.2.1 Data description

In this section, we use a data set from Bloomberg [1], made of daily observations of the implied volatility of the option with maturity one month on the S&P500 index, from January 5, 2006 to May 5, 2011 [2]. Note that the data are in fact already interpolated internally by the data provider (using quoted options at 4 PM) and do not necessarily exactly correspond to transaction data, see [32]. In Section 6.3, we present a method enabling us to derive spot volatilities from observed option prices with various maturities. Here we rely on the data provider approach to get option prices with the same maturity. This is not an issue since our aim in this work is to show that a rough dynamic for the volatility is obtained from any reasonable spot volatility proxy.

6.2.2 Scaling property

Reminder about the statistical methodology

We first recall the strategy used in [88] to investigate the smoothness of volatility sample paths. There is no novelty in term of statistical device here since we take the very same approach as in [88], but based on implied volatilities. Our contribution in this section is on the empirical side.

Let $\sigma^{imp}_{t_0},...,\sigma^{imp}_{t_N}$ be the time series of implied volatilities extracted from our data base. Here for $i \geq 0$, $t_{i+1} - t_i$ corresponds to one business day. In the spirit of [88], we wish to review the behavior of the so-called structure function $m(q,\Delta)$ given by

$$m(q,\Delta) = \frac{1}{N} \sum_{k=0}^{\lfloor (N-1)/\Delta \rfloor} |\log(\sigma^{imp}_{t_{(k+1)}\Delta}) - \log(\sigma^{imp}_{t_k\Delta})|^q$$

for various $q > 0$ and lags Δ going from 1 to about 40 days [3]. Through the quantity $m(q,\Delta)$, our goal is to revisit the finding in [88] that the (spot) log-volatility is well approximated by a fractional Brownian motion with Hurst parameter H smaller than 1/2. In this case, assuming spot and implied volatilities coincide, we should observe the following relationship :

$$m(q,\Delta) \sim c_q \Delta^{qH}, \tag{6.1}$$

with c_q a constant depending on q. Indeed, we have for $t \geq 0$ and $\Delta > 0$

$$\mathbb{E}[|W^H_{t+\Delta} - W^H_t|^q] = \tilde{c}_q \Delta^{qH},$$

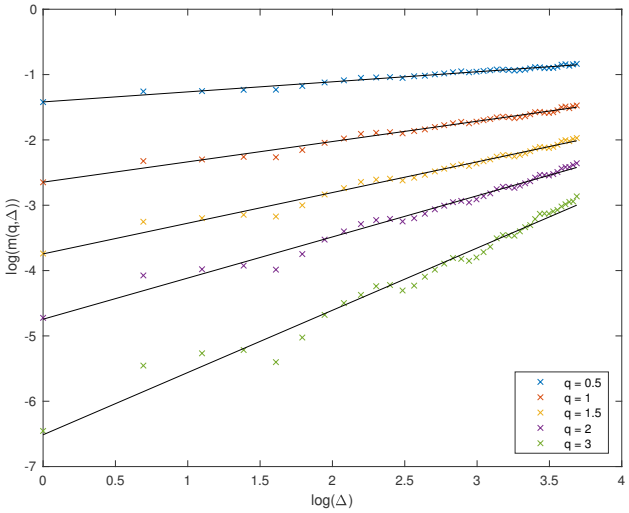
1. Data obtained from AXA Group Risk Management.
2. Data around the third Friday of each month (settlement date) are removed from the data base. We have 1166 points in total.
3. Of course when computing $m(q,\Delta)$ we in fact also average over the possible starting points $t_0,...,t_{\Delta-1}$.

with $\tilde{c}_q$ the absolute moment of order q of a standard Gaussian random variable.

Results

To investigate the validity of (6.1), we plot in Figure 6.1 the logarithm of $m(q,\Delta)$ against the logarithm of Δ, for several values of q.

FIGURE 6.1 – Scaling property of log-volatility increments.



For every q, the points with coordinates $(\log(\Delta), \log(m(q,\Delta)))$ are almost perfectly on the same line, and this for a wide range of Δ. Figure 6.1 is actually very similar to that obtained from historical volatility measurements in [88]. Thus we can deduce that indeed, for a given q,

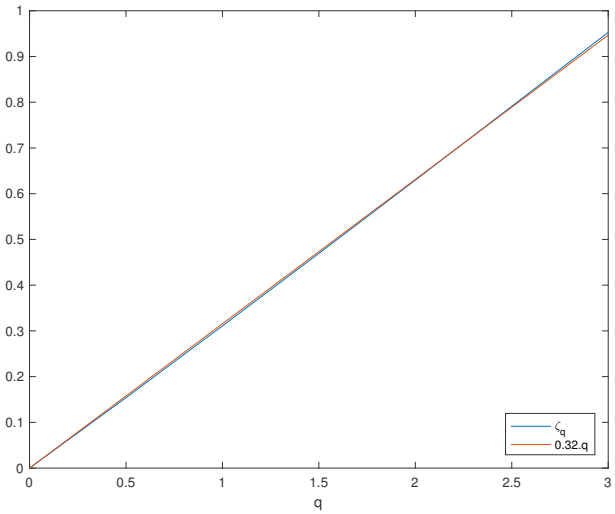
$$m(q,\Delta) \sim c_q \Delta^{\zeta(q)},$$

for some $\zeta(q)$.

Now we want to check whether $\zeta(q)$ can be taken of the form $q\mathrm{H}$ for some H, as suggested in [88]. This would lead to the same monofractal scaling as that of the fractional Brownian motion with Hurst parameter H. To answer this, we plot in Figure 6.2 the points with coordinates $(q,\zeta(q))$, where $\zeta(q)$ is taken as the slope of the line in Figure 6.1 corresponding to the power q, and the points with coordinates $(q, 0.32q)$ [4].

4. The value 0.32 is simply obtained from a standard linear fit of the points given by the slopes of the lines in Figure 6.1

FIGURE 6.2 – Monofractal scaling.

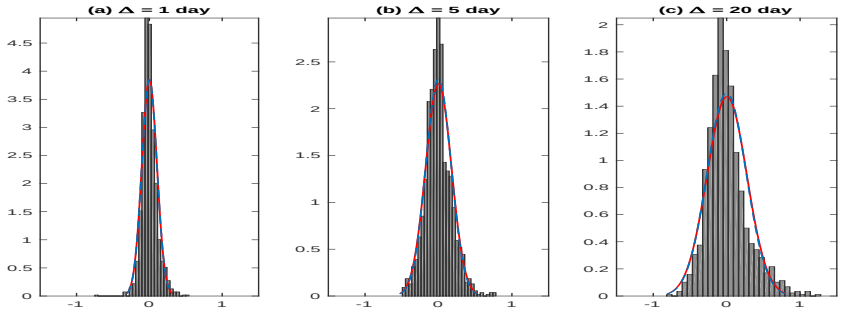


We see that the two graphs on Figure 6.2 can hardly be distinguished. This means that (6.1) almost perfectly holds, with H around 0.32. Note that such value for H corresponds to rough volatility since it is smaller than 1/2. However, it is larger than those reported in [88]. This is actually due to the fact that our options have a significant remaining time to maturity of one month. This induces a smoothing phenomenon in the estimation of the Hurst parameter. This effect is of the same nature as that described and explained in [88] caused by the discrepancy between spot and integrated volatility over a short time interval. We quantify this measurement bias numerically and analytically in Section 6.4.

6.2.3 Distribution of log-volatility increments

Recall that it is suggested in [88] that the log-volatility process is well modeled by a fractional Brownian motion with Hurst parameter smaller than 1/2. This implies monofractal scaling as investigated above but also a Gaussian behavior of the log-volatility increments. This feature is indeed satisfied when using historical estimates as measurements for spot volatility, see [88]. Here we wish to study whether such property also holds when the volatility proxies are given by our short term at-the-money implied volatilities. To this end, we display in Figure 6.3 histograms of log-volatility increments over different time intervals, together with a Gaussian density fit and the Gaussian density associated to the increments of a fractional Brownian motion with Hurst parameter equal to 0.32.

FIGURE 6.3 – Distribution of the log-volatility increments when using implied volatility as spot volatility proxy. The Gaussian fit is in blue and the density associated to the increments of a fractional Brownian motion with Hurst parameter equal to 0.32 is in red.



From these graphs, we obtain that empirical distributions of log-volatility increments are reasonably

approximated by Gaussian laws. However, we can remark that the empirical distributions are slightly more concentrated around their center. Finally, the Gaussian fits almost exactly coincide with those associated to the fractional Brownian motion with Hurst parameter equal to 0.32. Note that this would of course probably no longer be true if considering higher frequencies than the daily scale.

In conclusion, using at-the-money implied volatilities with maturity one month as spot volatility proxies, we obtain that log-volatility is well approximated by a rough fractional Brownian motion. This confirms the finding in [88].

6.3 A refined implied volatility based proxy for the spot volatility

In this section, we wish to study the robustness of the results obtained in Section 6.2. To do so, we work with another spot volatility proxy based on at-the-money options with short maturity. More precisely, we use the approximation formula from Medvedev and Scaillet, see [124]. This correction formula enables us to compute a spot volatility proxy from an at-the-money implied volatility with any (short) maturity. This is an advantage compared to what is done in Section 6.2 where only options with one month maturity are considered [5]. The drawback of Medvedev-Scaillet formula is that it is proved to be valid only within a restricted class of stochastic volatility models, which does not include rough volatility models. However our goal here is to see whether a proxy obtained from a Brownian volatility model still exhibits a rough behavior.

6.3.1 Data description and processing

Here our data set is provided by OptionMetrics and consists in daily close bid/ask prices of European puts and calls on the S&P500 index, from September 5, 2001 to January 31, 2012, for various strikes and maturities, together with the daily traded volumes. We discard options with price less than 2.5 cents of dollars or with zero trading volume. Besides, as in Section 6.2, prices corresponding to settlement dates are removed, so as obvious outliers.

We then want to compute implied volatilities from put and call prices. Thus we have to invert (everyday) the Black-Scholes formula. Therefore we need to fix for any time to maturity τ an underlying forward price $F(\tau)$ and a zero coupon bond price $D(\tau)$. To do so, we use the following classical approach based on put-call parity. The values of $F(\tau)$ and $D(\tau)$ are taken as solutions of the minimization problem

$$\underset{D,F}{\arg\min}\left\{\sum_i w_i\left(\frac{1}{2}(C_i^a - P_i^b) + \frac{1}{2}(C_i^b - P_i^a) - D(\tau)(F(\tau) - K_i)\right)\right\},$$

where $C_i^{a,b}$ and $P_i^{a,b}$ are respectively the call and put market prices (a standing for ask, b for bid) quoted at strike level K_i. The weights w_i are given by

$$w_i = \frac{\sqrt{\min\{V_i^C, V_i^P\}}}{\frac{1}{2}(C_i^a - C_i^b) + \frac{1}{2}(P_i^a - P_i^b)},$$

with V_i^C and V_i^P the trading volumes of call and put options at strike K_i. Finally, our implied volatility is taken as that of a call whose price would be the midprice between the bid and ask prices.

Recall that for our approximations to be valid, we focus on at-the-money implied volatilities with short maturity. Following [124], we only select implied volatilities of options with time to maturity ranging

5. Mixing various maturities without any correction would have been very arguable.

from 15 to 60 days. Shorter term options are discarded because quotes can be noisy. Moreover, we restrict our data to log forward moneyness belonging to the interval $[-0.03, 0.03]$. Such procedure yields a total number of 34842 implied volatilities over 2569 days.

6.3.2 The Medvedev-Scaillet correction formula

In [124], the authors consider a general modeling framework encompassing most of the classical parametric price models. They use a two factors jump-diffusion stochastic volatility model of the form

$$\begin{cases} dS_t = \big(r - \mu(\sigma_t)\big) S_t dt + \sigma_t S_t \, dZ_t + S_t \, dJ_t \\ d\sigma_t = a(\sigma_t) dt + b(\sigma_t)\big(\rho \, dZ_t + \sqrt{1-\rho^2}\, dW_t\big), \end{cases} \tag{6.2}$$

where Z_t and W_t are two independent Brownian motions and J_t is a Poisson-type jump process, independent of Z_t and W_t. Both r and the correlation coefficient ρ are assumed to be constant. The expected jump size $\mathbb{E}[\Delta J]$ is also constant, but the jump intensity $\lambda(\sigma_t)$ may depend on the volatility in a deterministic way. Here, as in the numerical experiments in [124], we consider the following parametric forms :

$$b(\sigma_t) = \beta \sigma_t^{\phi}, \quad \lambda(\sigma_t) = \lambda_0 \sigma_t^{\psi},$$

for some non-negative constants β, ϕ, λ_0 and ψ.

Let σ be the spot volatility and $\hat{\sigma} = \hat{\sigma}(\tau)$ be the at-the-money implied volatility of an option with time to maturity τ. Following [124], we build up our option-based spot volatility proxy in two steps. First, the chosen model is calibrated from the approximation formula in Proposition 7 in [124] using all our option prices over the entire time period. To retrieve the proxy for the spot volatility, we then consider the following expansion as τ goes to zero shown in [124] :

$$\sigma = \hat{\sigma} - I_1(0,\hat{\sigma})\sqrt{\tau} + \Big(I_1(0,\hat{\sigma})\frac{\partial I_1(0,\hat{\sigma})}{\partial \sigma} - I_2(0,\hat{\sigma}) + \frac{1}{2}\rho b(\hat{\sigma})\mathbb{E}[\Delta_J]\frac{\partial \lambda(\hat{\sigma})}{\partial \sigma}\Big)\tau + O(\tau\sqrt{\tau}). \tag{6.3}$$

The functions I_1 and I_2 are explicitly defined in [124] and depend only on β, ρ, ϕ, λ_0, ψ and $\mathbb{E}[\Delta_J]$.

6.3.3 The scaling property revisited

We now wish to study the scaling property of spot volatility proxies based on the approximation formula (6.3). We consider two cases : The Heston case, where $\phi = 0$ and $\lambda_0 = 0$, and the general case, where all the parameters are calibrated. The calibration results are given in Table 6.1.

TABLEAU 6.1 – Parameters calibrated on quoted S&P500 option prices, from September 5, 2001 to January 31, 2012.

PARAMETER	HESTON	GENERAL CASE
$\beta\rho$	-0.18 (0.00)	-3.27 (0.08)
ρ	-0.48 (0.00)	-0.39 (0.00)
ϕ	0	1.79 (0.02)
$\lambda_0\mathbb{E}(\Delta J)$	0	-0.6924 (0.03)
$\mathbb{E}(\Delta J)$	$--$ $--$	-0.17 (0.00)
ψ	$--$ $--$	1.11 (0.01)

Once the parameters are obtained, we can implement Equation (6.3) to compute everyday a spot volatility proxy. Note that in Equation (6.3), we take for $\hat{\sigma}$ the implied volatility with shortest time to maturity. Then we conduct the same analysis as in Section 6.2.2. The results are given in Figure 6.4 for the Heston model and Figure 6.5 for the general case (notations are the same as in Section 6.2.2).

FIGURE 6.4 – Scaling property of log-volatility increments when based on Heston proxy. In the second graph H is taken equal to 0.33.

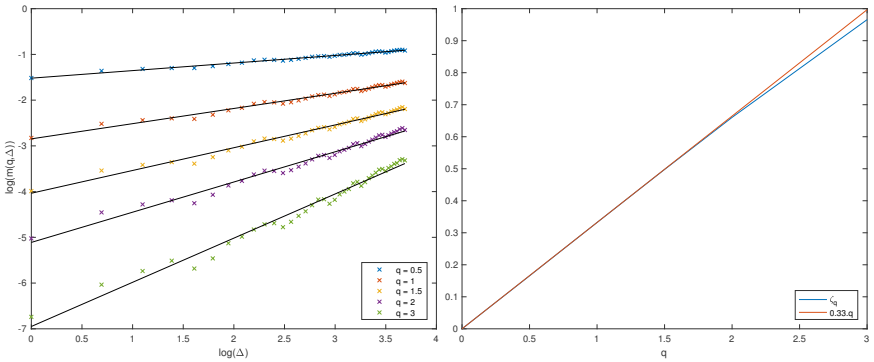
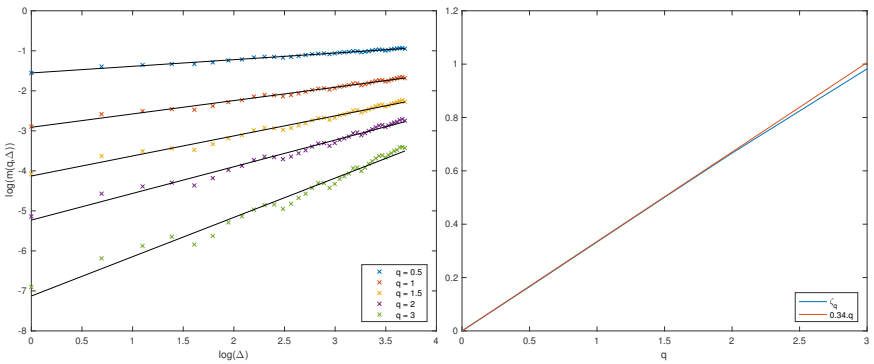


FIGURE 6.5 – Scaling property of log-volatility increments when based on the general case proxy. In the second graph H is taken equal to 0.34.



The results are very similar to those in Section 6.2.2. Here again we can confirm the fact that volatility is rough. This is even obtained although in the models in which the proxies are computed, volatility is of Brownian type and therefore not rough.

6.3.4 A control experiment

We have shown in the previous sections that when using market implied volatilities (directly or through Medvedev-Scaillet's formula in a Heston framework), we deduce that volatility is rough. However, one may wonder whether this effect is not just mechanical when considering implied volatilities. In other words, do implied volatility based estimators of the smoothness of volatility sample paths always lead to the conclusion that volatility is rough. This would of course question the validity of this finding.

To investigate this point, we consider the following Heston model for the stock price S_t and instantaneous variance v_t :

$$dS_t = rS_t dt + \sqrt{v_t} S_t dZ_t$$
$$dv_t = \kappa(\theta - v_t)dt + \sigma_V \sqrt{v_t} dW_t,$$

where Z and W are two Brownian motions with constant correlation ρ. We use the arbitrary set of parameters $\kappa = 6.4$, $r = 0.04$, $v_0 = \theta = 0.25^2$, $\sigma_V = 0.5$ and $\rho = -0.53$, and simulate the path of the stock price and variance over the time interval $[0, T]$, where T corresponds to 2520 days.

We first consider the scaling property of the (true) spot log-volatility increments in Figure 6.6.

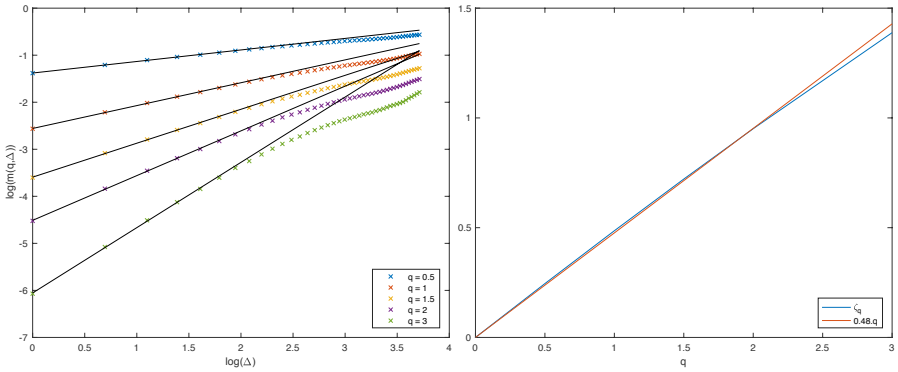
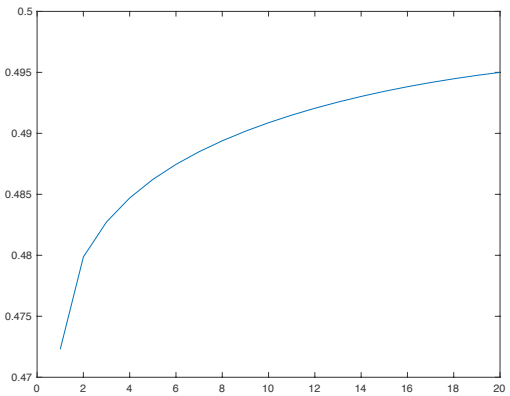


FIGURE 6.6 – The scaling property of log-volatility increments for a simulated Heston model. In the second graph H is taken equal to 0.48, estimation based on the five smaller lags for each q.

Contrary to what is observed on data, we observe two different behaviors for the empirical moments in the Heston case. For small lags, a fractional Brownian motion type scaling is observed whereas for large lags, stationarity kicks in. Still, estimation of the smoothness of volatility sample paths based on the first lags leads to a value of H around 0.48 which is close to 0.5 the theoretical one in the Brownian volatility case.

We now consider implied volatilities. Our goal is to see whether our methodology provides good estimates of the smoothness of volatility sample paths or it spuriously leads to the conclusion that volatility is rough in this Brownian volatility case. To do so, we build implied volatility time series for times to maturity from 1 to 20 days (using the Fourier pricing approach to Heston model, see [47]), for each day in the interval $[0, T]$. We then compute the scaling parameter for each time to maturity in the same way as previously. The results are given in Figure 6.7.

FIGURE 6.7 – Estimated values of the Hurst parameter using implied volatilities as a function of time to maturity (in days).



We see that the estimated smoothness is always around 0.5 (slightly below, notably for smaller times

to maturity). Thus, in the context of a Brownian volatility model, based on our methodology, one concludes that the smoothness of volatility sample paths is around 0.5. This means that the result obtained in this work that volatility is rough is not just a mechanical effect due to our statistical device or the nature of implied volatilities.

6.4 On the upward bias when estimating the Hurst parameter

We explain in this section why using implied volatility measures as spot volatility proxies induces an upward bias in the estimation of the Hurst parameter. We start with a numerical investigation of this phenomenon.

6.4.1 Monte Carlo study

To understand the extend of the bias when estimating the Hurst parameter, we simulate option prices in a rough volatility model. Then we compute the Hurst parameter based on these simulated data. Let $T > 0$. We consider the following model without leverage effect over the time interval $[0, T]$:

$$d\log S_t = \sigma_t dZ_t, \;\; d\log\sigma_t = \eta dW_t^H.$$

Here Z_t is a Brownian motion, W_t^H a fractional Brownian motion independent of Z_t and $\eta > 0$.

Simulation of fractional Brownian motion

We consider a time interval $[0, T]$ and fix an equidistant partition $0 = t_0 < t_1 < ... < t_n = T$. We first wish to simulate $(W_{t_1}^H, \dots, W_{t_n}^H)$. For $i, j \in \{1, ..., n\}$, we have

$$\mathbb{E}[W_{t_i}^H W_{t_j}^H] = \frac{1}{2}\left(t_i^{2H} + t_j^{2H} - |\, t_i - t_j \,|^{2H}\right).$$

Then we can use the Cholesky decomposition of the covariance matrix Σ of $(W_{t_1}^H, \dots, W_{t_n}^H)$: $\Sigma = LL^T$, where $L = (l_{ij})_{i,j\in\{1,n\}}$ is lower-triangular. Thus simulating a sample path of the fractional Brownian motion at times (t_i) can be done generating a vector $X = (X_1, ..., X_n)$ of independent standard Gaussian random variables and setting $(W_{t_1}^H, ..., W_{t_n}^H) = LX$, see for example [65] for details.

Simulating option prices under rough volatility

We place ourselves at time $t_i > 0$ and assume past spot volatilities and prices have been observed at times $t_1, \dots, t_i$. We want to compute the price at time t_i of an option with expiration date $t_k = t_i + \tau$ for some $\tau > 0$. The procedure goes as follows :

— We generate M paths of the volatility process on the interval $[t_{i+1}, t_k]$. This is done simulating $(W_{t_j}^H)_{t_{i+1} \le t_j \le t_k}$ conditional on past information, that is the filtration generated by $(X_{t_1}, ..., X_{t_i})$. Using the lower triangular form of L, these new values for the fractional Brownian motion at times $t_{i+1} \le t_j \le t_k$ can be obtained writing

$$W_{t_j}^H = \sum_{p=1}^{i} l_{jp} X_p + \sum_{p=i+1}^{j} l_{jp} X_p.$$

The i first variables X_p are those used to simulate the fractional Brownian motion up to time t_i, whereas $(X_{i+1}, \dots, X_j)$ is a sample of independent standard Gaussian random variables, independent from past values. Taking the exponential, we get our spot volatility sample path. We write σ^m for the m-th volatility trajectory.

— The price at time t_i of an at-the-money option with time to maturity τ is obtained computing

$$\frac{1}{M}\sum_{m=1}^{M} C_{BS}\left(S_{t_i}, \tau, \sqrt{\frac{1}{\tau}\sum_{p=i+1}^{k}(\sigma_{t_p}^{m})^2}\right),$$

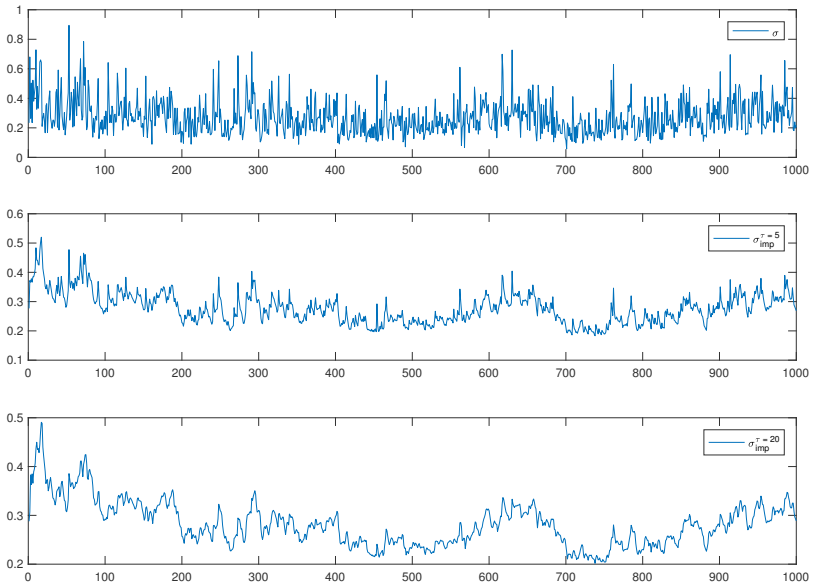
where $C_{BS}(S_{t_i}, \tau, \sigma)$ is the price of an at-the-money option with time to maturity τ in a Black-Scholes model with volatility σ, zero interest rate, and underlying value S_{t_i}.

— Eventually we invert Black-Scholes formula to obtain the implied volatility.

Results

We consider the following set of parameters : $H = 0.04$, $\eta = 1.0$ and $T = 1000$ days. Such parameters are consistent with [20, 88]. We take $\tau \in \{1, \dots, 20\}$ days and run $M = 10^4$ simulations. Figure 6.8 displays the sample path of the spot volatility together with those of the implied volatilities associated to 5 and 20 days.

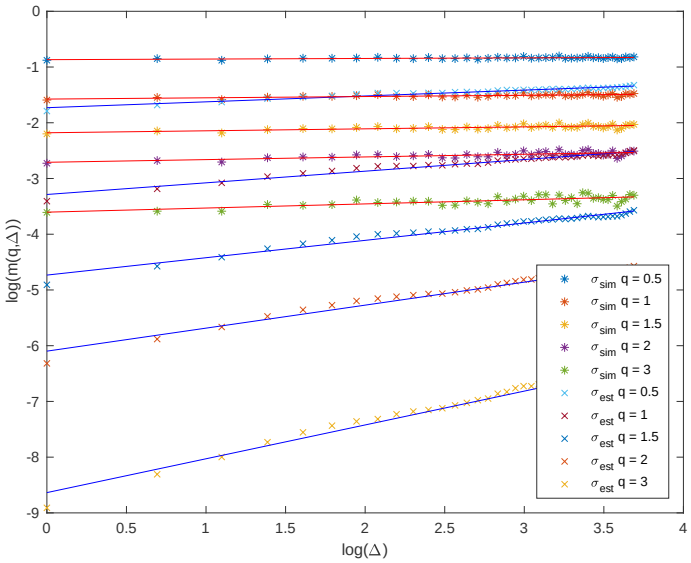
FIGURE 6.8 – Sample paths of spot volatility and implied volatilities for $\tau = 5$ and $\tau = 20$.



At the visual level, it is already clear that implied volatility trajectories are not as rough as that of the spot volatility. Furthermore, the longer the time to maturity, the larger the smoothing effect.

As in Sections 6.2 and 6.3, we now consider Equation (6.1). Based on our simulation, for several values of q, we plot in Figure 6.9 the logarithm of $m(q, \Delta)$ against the logarithm of Δ. This is done in two cases : when m is obtained from spot volatility values and when m is derived from implied volatility values, with $\tau = 5$ days.

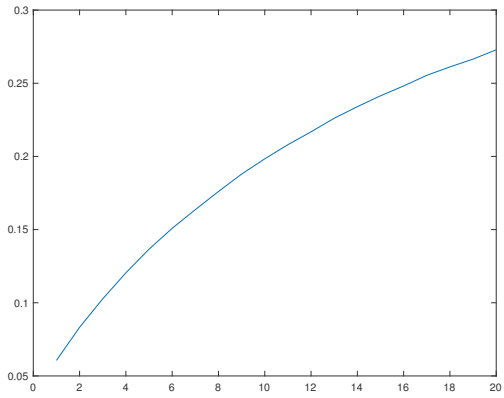
FIGURE 6.9 – Scaling property of log-volatility increments : spot volatility and implied volatility with $\tau = 5$.



We see that for a given q, when $m(q,\Delta)$ is computed from implied volatilities, the points with coordinates $(\log(\Delta), \log(m(q,\Delta)))$ remain on the same line. However, the slope of this line is larger than that obtained when $m(q,\Delta)$ is computed from spot volatilities (which provides the true underlying H up to small statistical error). Hence there is indeed a smoothing effect due to the remaining time to maturity of the considered options.

Finally, we give in Figure 6.10 the estimated values of H when using implied volatilities from the simulation, for different times to maturity.

FIGURE 6.10 – Estimated values of the Hurst parameter using implied volatilities as a function of time to maturity (in days).



Under our simulation framework, we see that using options with maturity 1 day, we obtain a quite accurate value for H of 0.06, while the true parameter is equal to 0.04. Taking longer maturities leads to an increasing bias. With 20 days maturity, one gets an estimated Hurst parameter of about 0.27. These results are in line with those in Sections 6.2 and 6.3.

6.4.2 Analytical illustration of the upward bias

In the spirit of Appendix C in [88], we finally want to provide a more quantitative understanding of the observed upward bias when estimating the Hurst parameter from implied volatilities. To do so, we consider a very crude approximation. Indeed we suppose that the at-the-money implied variance at time t of an option with time to maturity $\tau > 0$, denoted by $\hat{v}^\tau(t)$, is given by

$$\hat{v}^\tau(t) = \frac{1}{\tau}\int_t^{t+\tau} \mathbb{E}_t[v_u]\,du,$$

where v_u is the spot variance at time u and $\mathbb{E}_t[.]$ the conditional expectation operator with respect to information up to time t. Furthermore, we take a simplified rough volatility model assuming that for $u > 0$,

$$v_u = v_0 + \nu \mathrm{W}_u^{\mathrm{H}},$$

for some $v_0 > 0$ and $\nu > 0$. These approximations are actually probably enough to shed light on the bias phenomenon. Indeed it is due to the effects of the conditional expectation and integral operators appearing in the implied volatility.

In this simplified setting, our goal is to illustrate the smoothing effect leading to the upward bias. To do so, we compute a quantity very related to $m(2,\Delta)$, namely

$$\hat{m}^\tau(2,\Delta) = \mathbb{E}[(\hat{v}^\tau(\Delta) - \hat{v}^\tau(0))^2].$$

Indeed, under our assumptions, if the implied volatility were equal to the spot one, this quantity would be proportional to $\Delta^{2\mathrm{H}}$. However, we now show that because of the use of implied volatility in $\hat{m}(2,\Delta)$, this relationship no longer holds, particularly for large τ/Δ.

We recall the Mandelbrot and Van Ness representation of fractional Brownian motion :

$$\mathrm{W}_t^{\mathrm{H}} = c_{\mathrm{H}}\Big(\int_0^t (t-s)^{\mathrm{H}-1/2} d\mathrm{W}_s + \int_{-\infty}^0 \big((t-s)^{\mathrm{H}-1/2} - (-s)^{\mathrm{H}-1/2}\big) d\mathrm{W}_s\Big),$$

where W_t is a two-sided Brownian motion and c_{H} is so that the variance of $\mathrm{W}_1^{\mathrm{H}}$ is equal to 1. We easily have

$$\begin{aligned}\hat{v}^\tau(\Delta) = v_0 &+ \frac{\nu}{\tau} c_{\mathrm{H}} \int_0^\tau \int_{-\infty}^0 \big((\Delta+u-s)^{\mathrm{H}-1/2} - (-s)^{\mathrm{H}-1/2}\big) d\mathrm{W}_s du \\ &+ \frac{\nu}{\tau} c_{\mathrm{H}} \int_0^\tau \int_0^\Delta (\Delta+u-s)^{\mathrm{H}-1/2} d\mathrm{W}_s du.\end{aligned}$$

Using stochastic Fubini theorem, this gives

$$\begin{aligned}\hat{v}^\tau(\Delta) - \hat{v}^\tau(0) = &\frac{\nu}{\tau} c_{\mathrm{H}} \int_{-\infty}^0 \int_0^\tau \big((\Delta+u-s)^{\mathrm{H}-1/2} - (u-s)^{\mathrm{H}-1/2}\big) du d\mathrm{W}_s \\ &+ \frac{\nu}{\tau} c_{\mathrm{H}} \int_0^\Delta \int_0^\tau (\Delta+u-s)^{\mathrm{H}-1/2} du d\mathrm{W}_s.\end{aligned}$$

Hence we easily deduce from Ito isometry that

$$\hat{m}^\tau(2,\Delta) = \mathrm{A}\big(h_1(\Delta,\tau) + h_2(\Delta,\tau)\big),$$

with

$$\mathrm{A} = \frac{c_{\mathrm{H}}^2 \nu^2}{(\mathrm{H}+1/2)^2},$$
$$h_1(\Delta,\tau) = \frac{1}{\tau^2}\int_{-\infty}^{0}\left((\Delta+\tau-s)^{\mathrm{H}+1/2} - (\Delta-s)^{\mathrm{H}+1/2} - (\tau-s)^{\mathrm{H}+1/2} + (-s)^{\mathrm{H}+1/2}\right)^2 ds,$$
$$h_2(\Delta,\tau) = \frac{1}{\tau^2}\int_{0}^{\Delta}\left((\Delta+\tau-s)^{\mathrm{H}+1/2} - (\Delta-s)^{\mathrm{H}+1/2}\right)^2 ds.$$

We write $h_1(\Delta,\tau)$ under the form

$$\frac{1}{\tau^2}\Delta^{2\mathrm{H}+2}\int_{-\infty}^{0}\left((1+\frac{\tau}{\Delta}-s)^{\mathrm{H}+1/2} - (1-s)^{\mathrm{H}+1/2} - (\frac{\tau}{\Delta}-s)^{\mathrm{H}+1/2} + (-s)^{\mathrm{H}+1/2}\right)^2 ds.$$

Setting $\theta = \tau/\Delta$, we obtain

$$h_1(\Delta,\tau) = \Delta^{2\mathrm{H}} f_1(\theta),$$

where

$$f_1(\theta) = \frac{1}{\theta^2}\int_{-\infty}^{0}\left((1+\theta-s)^{\mathrm{H}+1/2} - (1-s)^{\mathrm{H}+1/2} - (\theta-s)^{\mathrm{H}+1/2} + (-s)^{\mathrm{H}+1/2}\right)^2 ds.$$

Similarly, we have

$$h_2(\Delta,\tau) = \Delta^{2\mathrm{H}} f_2(\theta),$$

where

$$f_2(\theta) = \frac{1}{\theta^2}\int_{0}^{1}\left((1+\theta-s)^{\mathrm{H}+1/2} - (1-s)^{\mathrm{H}+1/2}\right)^2 ds.$$

So

$$\hat{m}^{\tau}(2,\Delta) = \mathrm{A}\Delta^{2\mathrm{H}}\left(f_1(\theta)+f_2(\theta)\right).$$

Now remark that

$$\lim_{\theta\to 0} f_1(\theta) = (\mathrm{H}+1/2)^2\int_{-\infty}^{0}\left((1-s)^{\mathrm{H}-1/2} - (-s)^{\mathrm{H}-1/2}\right)^2 ds$$

and

$$\lim_{\theta\to 0} f_2(\theta) = (\mathrm{H}+1/2)^2\int_{0}^{1}(1-s)^{2\mathrm{H}-1}.$$

Consequently,

$$\lim_{\theta\to 0}(f_1(\theta)+f_2(\theta)) = (\mathrm{H}+1/2)^2\frac{1}{c_{\mathrm{H}}^2}.$$

Thus, when θ is small,

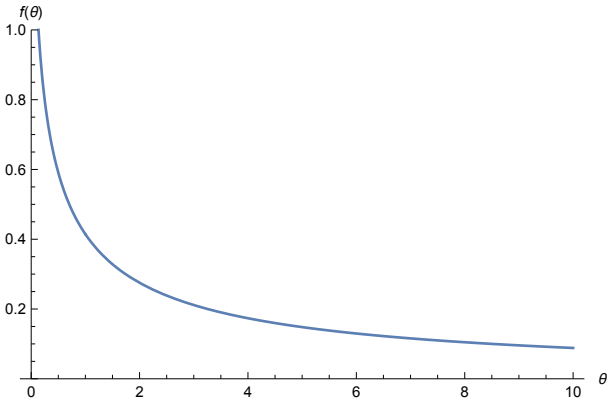
$$\hat{m}^{\tau}(2,\Delta) \sim \nu^2\Delta^{2\mathrm{H}}.$$

This means that the same scaling relationship as that associated to the spot volatility is approximately satisfied when considering implied volatilities with small enough times to maturity. Otherwise, one should add the multiplicative factor

$$f(\theta) = \frac{c_{\mathrm{H}}^2}{(\mathrm{H}+1/2)^2}\left(f_1(\theta)+f_2(\theta)\right)$$

on the right hand side of the above relationship. This disrupts the scaling property and implies biased estimations for the Hurst parameter. We draw in Figure 6.11 the graph of the function f for $\mathrm{H} = 0.04$.

FIGURE 6.11 – The function f for H = 0.04.



For fixed τ (as in Section 6.2), the function f is increasing with Δ. Therefore, when doing a regression analysis of the cloud of points with coordinates $(\log(\Delta), \log(\hat{m}^{\tau}(2,\Delta)))$, this implies an upward bias in the estimation of H due to a higher slope.

6.5 Conclusion

Using implied volatility data to approach spot volatility, we were able to confirm that volatility is rough. First using the one month at-the-money implied volatility on the S&P500 index as a volatility proxy. Then through computing spot volatilities from implied ones using a correction formula given by Medvedev and Scaillet. Following [88] we uncovered the monofractal scaling with a Hurst exponent H of order 0.32. This value corresponds to rough paths, but is larger than the one obtained in [88]. We also found that the distribution of the increments of log-volatility is close to Gaussian.

Given the value of H, we conducted a numerical analysis which consisted in estimating implied volatilities with different time to maturities, for a simulated rough volatility path with fixed H. We then estimated and compared the Hurst exponent for each time series. We found that the longer the time to maturity, the larger the smoothing effect. This upwards bias comes from a regularizing effect due to the remaining time to maturity. Using a crude but reasonable approximation, we were able to confirm this bias analytically.

Bibliographie

[1] F. Abergel and G. Loeper. Pricing and hedging contingent claims with liquidity costs and market impact. Available at SSRN : http ://ssrn.com/abstract=2239498 or http ://dx.doi.org/10.2139/ssrn.2239498, April 2013. 98

[2] P. Abry and F. Sellan. The wavelet-based synthesis for fractional Brownian motion proposed by F. Sellan and Y. Mayer : Remarks and fast implementation. *Applied Computational Harmonic Analysis*, 3 :377–383, 1996. 141

[3] M. Albizzati and H. Geman. Interest rate risk management and valuation of the surrender option in life insurance policies. *The Journal of Risk and Insurance*, 61(4) :616–637, 1994. 31, 32

[4] S. Alizadeh, M. W. Brandt, and F. X. Diebold. Range-based estimation of stochastic volatility models. *Journal of Finance*, 57(3) :1047–1091, 2002. 131

[5] R. Almgren. Optimal execution with nonlinear impact functions and trading-enhanced risk. *Applied Mathematical Finance*, 10 :1–18, 2003. 99

[6] R. Almgren. Optimal trading with stochastic liquidity and volatility. *SIAM Journal on Financial Mathematics*, 3 :163–181, 2012. 100, 110

[7] R. Almgren and N. Chriss. Optimal execution of portfolio transactions. *Journal of Risk*, 3 :5–39, 2000. 15, 97, 99, 100, 110

[8] R. Almgren, C. Thum, E. Hauptmann, and H. Li. Direct estimation of equity market impact. *Risk*, 18(7) :58–62, May 2005. Working paper. 99

[9] L. Andersen and V. Piterbarg. Interest rate modeling. *Atlantic Financial Press*, 2010. 82

[10] T. G. Andersen, T. Bollerslev, F. X. Diebold, and H. Ebens. The distribution of realized stock return volatility. *Journal of Financial Economics*, 61 :43–76, 2001. 139

[11] 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. 50

[12] E. Bacry and J. F. Muzy. Log-infinitely divisible multifractal processes. *Communications in Mathematical Physics*, 236(3) :449–475, 2003. 149

[13] T. G. Bali and D. Weinbaum. A comparative study of alternative extreme-value volatility estimators. *Journal of Futures Markets*, 25(9) :873–892, 2005. 132

[14] G. Barles, C. Daher, and M. Romano. Convergence of numerical schemes for parabolic equations arising in finance theory. *Mathematical Models and Methods in Applied Sciences*, 5(1) :125–143, 1995. 39

[15] G. Barone-Adesi and R. Whaley. Efficient analytic approximation of American option value. *Journal of Finance*, XLII(2), 1987. 35, 36, 38

[16] Barra. Market impact model handbook. Technical report, Barra Inc., 1997. 99

[17] D. S. Bates. Empirical option pricing : a retrospection. *Journal of Econometrics*, 116 :387–404, 2003. 98

[18] D. Bauer, A. Kling, and J. Russ. A universal pricing framework for Guaranteed Minimum Benefits in Variable Annuities. Working paper, 2006. 28

[19] D. Bauer, A. Kling, and J. Russ. A universal pricing framework for guaranteed minimum benefits in variable annuities. *Astin Bulletin*, 38(2) :621–651, 2008. 50

[20] C. Bayer, P. Friz, and J. Gatheral. Pricing under rough volatility. *Quantitative Finance*, 16(6) :887–904, 2016. 19, 154, 163

[21] S. Beckers. Variance of security price returns based on high, low and closing prices. *Journal of Business*, 1983. 131

[22] M. Bennedsen, A. Lunde, and M. S. Pakkanen. Hybrid scheme for Brownian semistationary processes. *arXiv preprint arXiv1507.03004*, 2015. 19, 154

[23] M. Bennedsen, A. Lunde, and M. S. Pakkanen. Decoupling the short-and long-term behavior of stochastic volatility. *Available at* SSRN 2846756, 2016. 19, 154

[24] L. Bergomi. Local, stochastic, local/stochastic, volatility models–and non-models. 2016. 154

[25] C. Bernard and M. Kwak. Semi-static hedging of variable annuities. *Insurance : Mathematics and Economics*, 67 :173–186, September 2014. 14

[26] P. Bertrand and J. Prigent. Portfolio Insurance : The extreme value approach to the CPPI method. *Finance*, 23 :68–86, 2002. 77

[27] D. Bertsimas and A. W. Lo. Optimal control of execution costs. *Journal of Financial Markets*, 1 :1–50, 1998. 16, 97, 109

[28] P. Bjerksund and G. Stensland. Closed form valuation of american options. Working paper, 2002. 35

[29] F. Black and R. Jones. Simplifying portfolio insurance. *Journal of Portfolio Management*, 14(1) :48–51, 1987. 76, 77

[30] F. Black and M. Scholes. The pricing of options and corporate liabilities. *The Journal of Political Economy*, 81 :637–654, 1973. 18, 153

[31] F. Black and M. Scholes. The pricing of options and corporate liabilities. *The Journal of Political Economy*, 81(3) :637–654, 1973. 98

[32] Bloomberg. Introduction into the new Bloomberg implied volatility calculations. 2008. 155

[33] T. Bollerslev, D. Osterrieder, N. Sizova, and G. Tauchen. Risk and return : long-run relations, fractional cointegration, and return predictability. *Journal of Financial Economics*, 108 :409–424, 2013. 130

[34] O. Bondarenko. Why are put options so expensive? *Quarterly Journal of Finance*, 4(3) :1–50, April 2014. 98

[35] B. Bouchard, A. Kalife, X. Tan, and L. Wang. Understanding guaranteed minimum withdrawal benefit : a study on financial risks and rational lapse strategy. Working paper, 2010. 26, 33, 34

[36] B. Bouchard and N. Touzi. Weak dynamic programming principle for viscosity solutions. *SIAM Journal on Control and Optimization*, 49(3) :948–962, 2011. 7

[37] B. Bouchard and X. Warin. Monte-Carlo valorisation of American options : facts and new algorithms to improve existing methods. preprint submitted to Elsevier, 2010. 27, 34, 39, 40

[38] P. Boyle and W. Tian. The design of equity-indexed annuities. *Insurance : Mathematics and Economics*, 43(3) :303–315, 2008. 50

[39] P. E. Boyle and M. R. Hardy. Reserving for maturity guarantees. *Insurance : Mathematics and Economics*, 21 :113–127, 1997. 3

[40] P. P. Boyle and M. R. Hardy. Reserving for maturity guarantees, institute for insurance and pensions research. University of Waterloo, 1996. 25

[41] M. Brennan and E. Schwartz. The pricing of equity-linked life insurance policies with an asset value guarantee. *Journal of Financial Economics*, 3 :195–213, 1976. 76

[42] M. J. Brennan and E. S. Schwartz. The valuation of American put options. *The Journal of Finance*, 1977. 29

[43] D. Brigo and F. Mercurio. Interest rate models - theory and practice : with smile, inflation and credit. *Springer Finance*, 2006. 82

[44] X. Brokmann, E. Sérié, J. Kockelkoren, and J.-P. Bouchaud. Slow decay of impact in equity markets. Technical report, Capital Fund Management, 2014. 99

[45] J. Brun and L. Prigneaux. VaR approach for credit CPPI and counterparty risk. *Quant Congress*, 2007. 77

[46] D. Bunch and H. Johnson. The American put option and its critical stock price. Journal Finance WFA Meetings, 1999. 35

[47] P. Carr and D. Madan. Option valuation using the fast Fourier transform. *Journal of Computational Finance*, 2(4) :61–73, 1999. 161

[48] L. Chan and J. Lakonishok. The behavior of stock prices around institutional trades. *Journal of Finance*, 50 :1147–1174, 1995. 107

[49] Z. Chen and P. Forsyth. A semi-Lagrangian approach for natural gas storage valuation and optimal operation. *SIAM Journal on Scientific Computing*, 30(1) :339–368, 2007. 100, 124

[50] Z. Chen, K. Vetzal, and P. A. Forsyth. The effect of modelling parameters on the value of gmwb guarantees. *Insurance : Mathematics and Economics*, 42(1) :165–173, 2008. 50

[51] P. Cheridito, H. Kawaguchi, and M. Maejima. Fractional Ornstein-Uhlenbeck processes. *Electronic Journal of Probability*, 8(3) :1–14, 2003. 130, 140

[52] P. Christoffersen, R. Goyenko, K. Jacobs, and M. Karoui. Illiquidity premia in the equity options market. *Available at SSRN 1784868*, 2014. 98

[53] T. F. Coleman, Y. Kim, Y. Li, and M. Patron. Robustly hedging variable annuities with guarantees under jump and volatility risks. *The Journal of Risk and Insurance*, 74(2) :347–376, 2007. 14

[54] T. F. Coleman, Y. Li, and M.-C. Patron. Hedging guarantees in variable annuities under both equity and interest rate risks. *Insurance : Mathematics and Economics*, 38 :215–228, 2006. 14

[55] F. Comte, L. Coutin, and E. Renault. Affine fractional stochastic volatility models. *Annals of Finance*, 8 :337–378, 2012. 130

[56] F. Comte and E. Renault. Long memory in continuous-time stochastic volatility models. *Mathematican Finance*, 8(4) :291–323, 1998. 20, 130, 141

[57] R. Cont and P. Tankov. Constant proportion portfolio insurance in the presence of jumps in asset prices. *Mathematical Finance*, 19(3) :379–401, July 2009. 12, 77

[58] F. Corsi. A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, 7(2) :174–196, 2009. 147, 148

[59] J. Crank and P. Nicolson. A practical method for numerical evaluation of solutions of partial differential equations of the heat conduction type. *Advanves in Computational Mathematics*, 6(1) :207–226, 1996. 39

[60] M. Dai, Y. Kuen Kwok, and J. Zong. Guaranteed minimum withdrawal benefit in variable annuities. *Mathematical Finance*, 18(4) :595–611, 2008. 60, 62

[61] M. Dai, Y. K. Kwok, and J. Zong. Guaranteed minimum withdrawal benefit in variable annuities. *Mathematical Finance*, 18(4) :595–611, 2007. 8, 50

[62] G. Deelstra and G. Rayée. Pricing variable annuity guarantees in a local volatility framework. *Insurance : Mathematics and Economics*, 53(3) :650–663, 2013. 51

[63] E. Derman and I. Kani. Riding on a smile. *Risk*, 7(2) :139–145, 1994. 18, 129, 153

[64] Y. D'Halluin, P. A. Forsyth, and G. Labahn. A semi-Lagrangian approach for American Asian options under jump diffusion. *SIAM Journal on Scientific Computing*, 27 :315–345, 2005. 100

[65] T. Dieker. Simulation of fractional brownian motion. *MSc theses, University of Twente, Amsterdam, The Netherlands*, 2004. 162

[66] G. Dionne, editor. *Handbook of insurance*. Springer, 2000. 1

[67] J. Duchon, R. Robert, and V. Vargas. Forecasting volatility with the multifractal random walk model. *Mathematical Finance*, 1 :83–108, 2012. 149

[68] D. J. Duffy. A critique of the Crank-Nicolson scheme, strengths and weakness for financial instrument pricing. *Wilmott*, 4 :68–76, 2004. 62

[69] B. Dupire. Pricing with a smile. *Risk*, 7(1) :18–20, 1994. 18, 129, 153

[70] O. El Euch, M. Fukasawa, and M. Rosenbaum. The microstructural foundations of leverage effect and rough volatility. *arXiv preprint arXiv :1609.05177*, 2016. 19, 154

[71] O. El Euch and M. Rosenbaum. The characteristic function of rough Heston models. *Mathematical Finance*, to appear, 2016. 19, 154

[72] R. F. Engle and A. J. Patton. What good is a volatility? *Quantitative Finance*, 1 :237–245, 2001. 130

[73] W. Feller. The asymptotic distribution of the range of sums of random variables. *Annals of Mathematical Statistics*, 3(22) :427–432, 1951. 131

[74] M. Forde and H. Zhang. Asymptotics for rough stochastic volatility and Lévy models. *Preprint*, 2015. 19, 154

[75] P. A. Forsyth. A Hamilton Jacobi Bellman approach to optimal trade execution. *Applied Numerical Mathematics*, 61(2) :241–265, 2010. Working paper. 100, 110, 124

[76] P. A. Forsyth, J. S. Kennedy, S. T. Tse, and H. Windcliff. Optimal trade execution : A mean quadratic variation approach. *Journal of Economic Dynamics and Control*, 36(12) :1971–1991, December 2012. 110

[77] R. Frey. Perfect option hedging for a large trader. *Finance and Stochastics*, 2 :115–141, 1998. 98

[78] R. Frey and A. Stremme. Market volatility and feedback effects from dynamic hedging. *Mathematical Finance*, 7(4) :351–374, October 1997. 15, 98

[79] M. Fukasawa. Short-time at-the-money skew and rough fractional volatility. *Quantitative Finance, to appear*, 2016. 19, 154

[80] H. Funahashi and M. Kijima. Does the Hurst index matter for option prices under fractional volatility? *Annals of Finance*, pages 1–20, 2015. 19, 154

[81] M. C. Fung, K. Ignatieva, and M. Sherris. Systematic mortality risk : An analysis of guaranteed lifetime withdrawal benefits in variable annuities. *Insurance : Mathematics and Economics*, 58 :103–115, 2014. 51

[82] J. Gao and E. R. Ulm. Optimal consumption and allocation in variable annuities with guaranteed minimum death benefits. *Insurance : Mathematics and Economics*, 51 :586–598, 2012. 50

[83] J. Garcia and S. Goosens. The art of credit derivatives : Demystifying the Black Swan. *The Wiley Finance Series*, 2009. 87

[84] J. Garcia, S. Goosens, and W. Schoutens. Let's jump together : Pricing credit derivatives. *Risk*, pages 130 – 133, September 2008. 77, 87

[85] N. Gârleanu, L. H. Pedersen, and A. M. Poteshman. Demand-based option pricing. *The Review of Financial Studies*, 22(10) :4259–4299, 2009. 98

[86] M. B. Garman and M. J. Klass. On the estimation of security price volatility from historical data. *Journal of Business*, 53 :67–78, 1980. 19, 132

[87] J. Gatheral. *The volatility surface : A practitioner's guide*. John Wiley & Sons, 2006. 129

[88] J. Gatheral, T. Jaisson, and M. Rosenbaum. Volatility is rough. *Available at* SSRN 2509457, 2014. 19, 20, 21, 154, 155, 156, 157, 158, 163, 165, 167

[89] J. Gatheral, T. Jaisson, and M. Rosenbaum. Volatility is rough. *Available at* SSRN 2509457, 2014. 129, 130, 131, 134, 135, 139, 140, 141, 142, 152

[90] J. Gatheral and A. Schied. Optimal trade execution under geometric Brownian motion in the Almgren and Chriss framework. *The International Journal of Theoretical and Applied Finance*, 14(3) :353–368, 2011. 110

[91] H. U. Gerber and E. S. Shiu. Pricing lookback options and dynamic guarantees. *North American Actuarial Journal*, 7(1) :48–66, 2003. 50

[92] S. Graf, A. Kling, and J. Ruß. Risk analysis and valuation of life insurance contracts : Combining actuarial and financial approaches. *Insurance : Mathematics and Economics*, 49(1) :115–125, 2011. 51

[93] S. J. Grossman and J.-L. Vila. Optimal dynamic trading with leverage constraints. *The Journal of Financial and Quantitative Analysis*, 27(2) :151–168, 1992. 11

[94] H. Guennoun, A. Jacquier, and P. Roome. Asymptotic behaviour of the fractional Heston model. *Available at SSRN 2531468*, 2014. 19, 154

[95] P. S. Hagan, D. Kumar, A. S. Lesniewski, and D. E. Woodward. Managing smile risk. *Wilmott Magazine*, pages 84–108, 2002. 18, 129, 154

[96] M. Hardy. *Investment Guarantees : Modeling and Risk Management for Equity-Linked Life Insurance*. Wiley Finance, 2003. 14

[97] S. L. Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies*, 6(2) :327–343, 1993. 18, 129, 154

[98] D. Holz, A. Kling, and J. Ruß. Gmwb for life an analysis of lifelong withdrawal guarantees. *Zeitschrift für die gesamte Versicherungswissenschaft*, 101(3) :305–325, 2012. 50

[99] Y. Huang, P. Forsyth, and G. Labahn. Iterative methods for the solution of a singular control formulation of a GMWB pricing problem. *Numerische Mathematik*, 122 :133–167, 2012. 60

[100] G. Huberman and W. Stanzl. Price manipulation and quasi-arbitrage. *Econometrica*, 72(4) :1247–1275, 2004. 99

[101] J. Hull and A. White. The pricing of options on assets with stochastic volatilities. *The Journal of Finance*, 42(2) :281–300, 1987. 18, 129, 154

[102] J. Hull and A. White. Pricing interest-rate-derivative securities. *The Review of Financial Studies*, 3(4) :573–592, 1990. 6

[103] G. Iacurci. How insurers are losing when it comes to variable annuities. http://www.investmentnews.com/article/20160830/FREE/160839998/how-insurers-are-losing-when-it-comes-to-variable-annuities, August 30 2016. 4

[104] S. Jaimungal and D. Kinzebulatov. Optimal execution with a price limiter. *Risk*, pages 49–54, July 2014. 112, 114

[105] T. Jaisson and M. Rosenbaum. Rough fractional diffusions as scaling limits of nearly unstable heavy-tailed Hawkes processes. *The Annals of Applied Probability*, 26(5) :2860–2882, 2016. 19, 154

[106] C. Jiang, Y. Ma, and Y. An. The effectiveness of the VaR-based portfolio insurance strategy : An empirical analysis. *International Review of Financial Analysis*, 18 :185–197, 2009. 11, 84

[107] N. Ju and R. Zhong. An approximate formula for pricing American options. *Journal of Derivatives*, 72(3) :31–40, 1999. 35, 36, 37, 38

[108] D. B. Keim and A. Madhavan. Execution costs and investment performance : An empirical analysis of institutional equity trades. Working paper, School of Business Administration, University of Southern California, 1995. 107

[109] 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. 14

[110] 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. *European Actuarial Journal*, 4(2) :281–314, 2014. 8, 14

[111] A. Kolkiewicz and Y. Liu. Semi-static hedging for gmwb in variable annuities. *North American Actuarial Journal*, 16(1) :112–140, 2012. 14

[112] S. G. Kou. A jump-diffusion model for option pricing. *Management Science*, 48(8) :1086–1101, August 2002. 12, 80

[113] N. Kunitomo. Improving the parkinson method of estimating security price volatilities. *Journal of Business*, 65 :295–302, 1992. 132

[114] R. D. Lee and L. R. Carter. Modeling and forecasting u.s. mortality. *Journal of the American Statistical Association*, 87(419) :659–671, September 1992. 56

[115] H. E. Leland. Option pricing and replication with transaction costs. *The Journal of Finance*, 40(5) :1283–1301, 1985. 15, 98, 99, 100

[116] H. E. Leland and M. Rubinstein. The evolution of portfolio insurance. Published in Dynamic Hedging : A Guide to Portfolio Insurance, edited by Don Luskin (John Wiley and Sons, 1988), 1976. 76

[117] E. Lépinette and T. T. Quoc. Approximate hedging in a local volatility model with proportional transaction costs. *Applied Mathematical Finance*, 21(4) :313–341, 2014. 16, 98, 102, 123

[118] T. M. Li and R. Almgren. A fully-dynamic closed-form solution for δ-hedging with market impact. *To appear Operations Research*, 2013. 98

[119] T. M. Li and R. Almgren. Option hedging with smooth market impact. *Working paper*, October 2014. 98

[120] F. Longstaff and E. Schwartz. Valuing American options by simulation : A simple least-squares approach. *The Review of Financial Studies*, 14(1) :113–147, February 2001. 40

[121] J. Lorenz and R. Almgren. Mean-variance optimal adaptive execution. *Applied Mathematical Finance*, 18 :311–323, 2011. 100, 111

[122] K. Lott. *Ein Verfahren zur Replikation von Optionen unter Transaktionkosten in stetiger Zeit*. PhD thesis, Universität der Bundeswehr München. Institut für Mathematik und Datenverarbeitung, 1993. 102

[123] C. Marshall, M. Hardy, and D. Saunders. Measuring the effectiveness of static hedging strategies for a guaranteed minimum income benefit. *North American Actuarial Journal*, 16(2) :143–182, 2012. 14

[124] A. Medvedev and O. Scaillet. Approximation and calibration of short-term implied volatilities under jump-diffusion stochastic volatility. *Review of Financial Studies*, 20(2) :427–459, 2007. 21, 154, 158, 159

[125] R. C. Merton. Optimum consumption and portfolio rules in a continuous-time model. *Journal of Economic Theory*, 3 :373–413, 1971. 11, 76

[126] R. C. Merton. On the pricing of corporate debt : The risk structure of interest rates. *The Journal of Finance*, 29(2) :449–470, 1973. 98

[127] M. A. Milevsky and S. E. Posner. The titanic option : Valuation of the guaranteed minimum death benefit in variable annuities and mutual funds. *Journal of Risk and Insurance*, 68(1) :93–128, 2001. 50

[128] M. A. Milevsky and T. S. Salisbury. Financial valuation of guaranteed minimum withdrawal benefits. *Insurance : Mathematics and Economics*, 2005. 50

[129] M. A. Milevsky and T. S. Salisbury. Financial valuation of Guaranteed Minimum Withawal Benefits. *Insurance : Mathematics and Economics*, 38(1) :21–38, February 2006. 28, 31

[130] M. A. Milevsky and T. S. Salisbury. Financial valuation of guaranteed minimum withdrawal benefits. *Insurance : Mathematics and Economics*, 38(1) :21–38, 2006. 73

[131] T. Moenig and D. Bauer. Policyholder exercise behavior for variable annuities including guaranteed minimum withdrawal benefits. Working paper, 2011. 8

[132] J. Mühle-Karbe and M. Nutz. Small-time asymptotics of option prices and first absolute moments. *Journal of Applied Probability*, 48(4) :1003–1020, 2011. 19, 154, 155

[133] U. Nations. World population ageing report. Technical report, United Nations, 2015. 1

[134] A. Neuenkirch and T. Shalaiko. The order barrier for strong approximation of rough volatility models. *arXiv preprint arXiv1606.03854*, 2016. 19, 154

[135] C. J. Nuzman and V. H. Poor. Linear estimation of self-similar processes via Lamperti's transformation. *Journal of Applied Probability*, 37(2) :429–452, 2000. 20, 147, 150

[136] L. Ornstein and G. E. Uhlenbeck. On the theory of the Brownian motion. *Physical Review*, 36, 1930. 140

[137] J. F. B. P. d'Andria and L. Elie. Modèle analytique d'evaluation des options de remboursement anticipé. *Finance*, 12(2), 1991. 31

[138] D. Pain and J. Rand. Recent developments in portfolio insurance. *Bank of England Quarterly Bulletin*, Q1, 2008. 77

[139] M. Parkinson. The extreme value method for estimating the variance of the rate of return. *Journal of Business*, 53 :61–65, 1980. 19, 130, 131, 132

[140] A. F. Perold. Constant portfolio insurance. *Harvard Business School*, 1986. 76, 77

[141] A. F. Perold and R. Salomon. The right amount of assets under management. *Financial Analysts Journal*, 47 :31–39, 1991. 99

[142] J. Pézier. Rationalization of investment preference criteria. *ICMA Centre Discussion Papers in Finance DP2011-12*, 2011. 84

[143] G. Piscopo and S. Haberman. The valuation of guaranteed lifelong withdrawal benefit options in variable annuity contracts and the impact of mortality risk. *North American Actuarial Journal*, 15(1) :59–76, 2011. 8, 51

[144] J.-L. Prigent and F. Tahar. CPPI with cushion insurance. Available at SSRN eLibrary http ://ssrn.com/paper=675824, 2005. 77

[145] L. Rogers and S. Satchell. Estimating variance from high, low and closing prices. *Annals of Applied Probability*, 1 :504–512, 1991. 132

[146] J. M. Romero and J. Bautista. Exact solutions for optimal execution of portfolios transactions and the Riccati equation. eprint arXiv :1601.07961, January 2016. 116

[147] M. Rubinstein. Implied binomial trees. *Journal of Finance*, 49 :771–818, 1994. 117

[148] M. Rubinstein and H. E. Leland. Replicating options with positions in stock and cash. *Financial Analysts Journal*, 37(4) :63–72, 1981. 5

[149] P. Shah and D. Bertsimas. An analysis of the guaranteed withdrawal benefits for life option. Available at SSRN 1312727, 2008. 8, 50

[150] P. V. Shevchenko and X. Luo. A unified pricing of variable annuity guarantees under the optimal stochastic control framework. Available at SSRN 2773384, 2016. 51, 56

[151] J. Shu and J. E. Zhang. Testing range estimators of historical volatility. *Journal of Futures Markets*, 26(3), March 2006. 131

[152] P. Tankov. Pricing and hedging gap risk. *Journal of Computational Finance*, 13(3) :33–59, Spring 2010. 81, 86

[153] S. T. Tse, P. A. Forsyth, J. S. Kennedy, and H. Windcliff. Comparison between the mean-variance optimal and the mean-quadratic-variation optimal trading strategies. *Applied Mathematical Finance*, 20(5) :415–449, 2013. 111

[154] J. Wang and P. Forsyth. Maximal use of central differencing for Hamilton-Jacobi-Bellman PDEs in finance. *Journal on Numerical Analysis*, 46 :1580–1601, 2008. 125

[155] D. Yang and Q. Zhang. Drift-independent volatility estimation based on high, low, open, and close prices. *Journal of Business*, 73 :477–491, 2000. 132

[156] S. S. Yang and T.-S. Dai. A flexible tree for evaluating guaranteed minimum withdrawal benefits under deferred life annuity contracts with various provisions. *Insurance : Mathematics and Economics*, 52(2) :231–242, 2013. 51

[157] J. Yong and X. Y. Zhou. *Stochastic controls : Hamiltonian Systems and HJB Equations*. Springer, 1999. 112