



HAL
open science

Contributions to volatility modeling and risk management and to efficient numerical simulation in finance

Stefano de Marco

► **To cite this version:**

Stefano de Marco. Contributions to volatility modeling and risk management and to efficient numerical simulation in finance. Probability [math.PR]. Institut Polytechnique de Paris, 2022. tel-04554992

HAL Id: tel-04554992

<https://hal.science/tel-04554992v1>

Submitted on 22 Apr 2024

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

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



INSTITUT POLYTECHNIQUE DE PARIS

École doctorale de mathématiques Hadamard (ED 574)

CMAP-UMR 7641, École Polytechnique

Mémoire présenté pour l'obtention du

Diplôme d'Habilitation à Diriger les Recherches

Discipline : Mathématiques

par

Stefano DE MARCO

Contributions to volatility modeling and risk management
and to efficient numerical simulation in finance

CHRISTIAN BAYER (WIAS Berlin)

Rapporteurs : JEAN-PIERRE FOUQUE (University of California Santa Barbara)
PETER TANKOV (ENSAE)

Date de soutenance : 16 décembre 2022

Composition du jury :

BEATRICE ACCIAIO (ETH Zurich)

CHRISTIAN BAYER (WIAS Berlin)

JEAN-PIERRE FOUQUE (University of California Santa Barbara)

JIM GATHERAL (Baruch College, NYU)

CAROLINE HILLAIRET (ENSAE)

AHMED KEBAIER (Université d'Evry)

PETER TANKOV (ENSAE)

Remerciements

I would like to thank Christian Bayer, Jean-Pierre Fouque and Peter Tankov for accepting to be the reviewers for my habilitation manuscript. Their outstanding contributions in our research field have been important references and source of inspiration in my studies and in my research career ; I am honoured that my own work has received their evaluation.

I am grateful to Beatrice Acciaio, Jim Gatheral, Caroline Hillairet and Ahmed Kebaier for accepting to be part of my evaluation committee. Their presence and their appreciation are a great privilege to me.

I would not have been able to reach this point without the help, support and guidance of the leaders of the research teams in mathematical finance, numerical probability and statistical learning at Ecole Polytechnique, Emmanuel Gobet, Mathieu Rosenbaum and Nizar Touzi. Un grand merci à tous les trois pour leur aide, leurs conseils, pour leur rigueur scientifique qui a été un modèle, et en particulier à Emmanuel pour m'avoir tant appris sur les méthodes Monte Carlo, à Nizar pour m'avoir ouvert des portes sur des sujets que je ne connaissais pas, à Mathieu pour nos discussions autour de la volatilité stochastique (loin d'être achevées) devant un tableau, dans le Rer ou devant un verre (d'ailleurs Mathieu, je ne t'ai pas vu hier soir).

Merci à tous les collègues de l'équipe de mathématiques financières, Nicolas Baradel, Charles Bertucci, Fabrice Djete, et à Eduardo Abi-Jaber et Sergio Pulido, avec qui j'ai beaucoup échangé sur nos intérêts de recherche communs depuis leur arrivée au CMAP. Merci aux collègues de l'équipe stats apprentissage et simulation, Aymeric Dieuleveut, Remi Flamary, Marilou Gabrié, Josselin Garnier, Eric Moulines, Clément Rey, Erwan Scornet, des autres équipes de recherche au CMAP, Grégoire Allaire, Samuel Amstutz, Vincent Bansaye, (Giovanni ti ringrazio dopo), Quentin Cormier, Anne De Bouard, Maxime Breden, Stéphane Gaubert, Lucas Gerin, Carl Graham, Igor Kortchemski, Aline Lefebvre-Lepot, Cyril Marzouk, Marc Massot, Sylvie Méléard, Flore Nabet, Teddy Pichard, Gael Raoul, Milica Tomasevic, Cormac Walsh (au fait, as-tu éteint les sorbonnes ?), et à tous les collègues enseignants et chercheurs au CMAP, pour leur énergie et leur dynamisme qui contribuent à créer une si belle atmosphère au sein du laboratoire.

Une pensée particulière pour les collègues qui ont partagé avec moi pendant quelques années la lourde charge de la gestion du PA MAP, Xavier Allamigeon qui va rédiger son mémoire d'HDR en janvier, Erwan Le Pennec embarqué avec moi vers la création de nouvelles formations, et Karim Lounici. Au passage, je ne remercie pas la plateforme Synapses et les systèmes informatiques de l'Ecole qui nous ont fait perdre tant de temps, mais là c'est une autre histoire.

J'ai partagé de très bons moments, au CMAP et ailleurs, avec des amis qui continuent maintenant leur carrière dans d'autres établissements proches ou lointains, Ankush Agarwal, Julien Claisse, Michele Salvi. Merci les amis !

J'ai eu le plaisir de co-encadrer des brillants thésards, merci Florian pour tout le travail fait ensemble (on va bientôt finaliser le papier options américaines, t'inquiète), et j'ai le plaisir de continuer maintenant à travailler avec d'autres.

Merci aux présidents du Dep MAP de ces dernières années et à l'équipe admin du département, Nathalie Rodrigues et Stéphanie Clevenot et avant elles Leyla Marzuk et Nicoletta Bourgeois, pour leur aide indéfectible avec la planification et la mise en oeuvre des enseignements, ainsi qu'aux directeurs du CMAP et à l'équipe admin du labo, et notamment à Nasséra Naar et Alex Noiret, pour leur aide avec les missions, la gestion des financements, pour la prise en charge de mes demandes particulières, et pour leurs réponses patientes à mes questions.

Ces dernières années, j'ai eu le plaisir de travailler avec des collaborateurs exceptionnels, dans certains cas grâce aux Chaires de recherche du CMAP : j'ai énormément appris (et souvent rigolé) en travaillant avec

Pierre Henry-Labordère, en discutant avec Lorenzo Bergomi et en lisant son livre (une entreprise qui est loin d'être terminée).

J'ai eu la chance de travailler avec de brillants chercheurs qui étaient aussi des compatriotes, Giovanni Conforti et Paolo Pigato – j'espère que nos collaborations ne sont pas terminées.

Difficile de dire combien j'ai appris, sur la modélisation, l'application des mathématiques et la mise en oeuvre des méthodes numériques, et sur bien d'autre encore, en travaillant avec Zeliade Systems et son équipe de recherche quantitative, Pierre Cohort, Ismail Laachir, Claude Martini et maintenant Arianna Mingone, qui m'ont toujours transmis une grande énergie.

J'ai pu beaucoup m'enrichir, lors de plusieurs conférences et workshops, en discutant avec les meilleurs chercheurs dans mon domaine - Paul Gassiat, Stefan Gerhold, Julien Guyon, Blanka Horvath, Antoine Jacquier, Andrea Pallavicini, Adil Reghai, ainsi que Giorgia Callegaro et Martino Grasselli, que je côtoie depuis ma jeunesse scientifique en Italie. J'ai eu la grande chance, depuis le début de ma thèse de doctorat, de pouvoir profiter d'échanges avec toute la communauté de recherche en finance quantitative de la région parisienne, Bruno Bouchard, Jean François Chassagneux, Stéphane Crepey, Zorana Grbac, Gilles Pagès, Zhenjie Ren.

Avant d'arriver à l'Ecole Polytechnique, j'ai fait l'expérience d'une formation à la recherche par des maitres exceptionnels. Merci à Vlad Bally, à Peter Friz, mais aussi à Aurélien Alfonsi et aux autres enseignants du M2 Mathématiques et applications à Marne-la-Vallée.

Enfin je souhaiterais remercier mes parents, à qui je dois une bonne partie de ce que je suis (et sans doute une bonne partie mon intérêt pour les sciences) et mon frère, mon semblable et mon contrepoids.

Et merci à Lola*, qui a écouté patiemment mes récriminations quand je n'en pouvais plus du PA, qui aime la "smoothness of densities" et me pousse à ne plus m'occuper de finance quantitative, et au petit Saul, qui a déjà commencé à me dire "papa, tu me fatigues" quand je commence à lui expliquer un sujet - je me demande ce qu'il me dira quand j'essayerai de lui parler de l'application des maths - on verra bien.

List of articles

Only the underlined articles are presented in this habilitation thesis.

Published papers

- F. BOURGEY AND S. DE MARCO, Multilevel Monte Carlo simulation for VIX options in the rough Bergomi model, *Journal of Computational Finance*, 26 (2022), pp. 53–82.
- S. DE MARCO, On the harmonic mean representation of the implied volatility, *SIAM J. Financial Math.*, 12 (2021), pp. 551–565.
- F. BOURGEY, S. DE MARCO, E. GOBET, AND A. ZHOU, Multilevel Monte-Carlo methods and lower/upper bounds in Initial Margin computations, *Monte Carlo Methods and Applications*, 26 (2020), pp. 131–161.
- S. DE MARCO, A. AGARWAL, E. GOBET, J. LOPEZ-SALAS, F. NOUBIAGAIN, AND A. ZHOU, *Numerical approximations of McKean Anticipative Backward Stochastic Differential Equations arising in Initial Margin requirements*, *ESAIM: Proceedings and Surveys*, 65 (2019), pp. 1–26.
- A. AGARWAL, S. DE MARCO, E. GOBET, AND G. LIU, *Study of new rare event simulation schemes and their application to extreme scenario generation*, *Mathematics and Computer in Simulations*, (2018), pp. 89–98.
- S. DE MARCO AND P. FRIZ, Local volatility, conditioned diffusions, and Varadhan’s formula, *SIAM J. Financ. Math.*, 9 (2018), pp. 835–874.
- S. DE MARCO AND C. MARTINI, *Moment generating functions and Normalized implied volatilities: unification and extension via Fukasawa’s pricing formula*, *Quantitative Finance*, 18 (2018), pp. 609–622
- S. DE MARCO AND P. HENRY-LABORDÈRE, *Local volatility from American option*, *Risk magazine*, (2017).
- S. DE MARCO, C. HILLAIRET, AND A. JACQUIER, *Shapes of implied volatility with positive mass at zero*, *SIAM J. Financ. Math.*, 8 (2017), pp. 709–737.
- S. DE MARCO, A. JACQUIER, AND P. ROOME, *Two examples of non strictly convex large deviations*, *Electronic Communications in Probability*, 21 (2016), pp. 1–12.
- S. DE MARCO AND P. HENRY-LABORDÈRE, Linking vanillas and VIX options: a constrained martingale optimal transport problem, *SIAM J. Financial Math.*, 6 (2015), pp. 1171–1194.
- G. CONFORTI, J.-D. DEUSCHEL, AND S. DE MARCO, *On small-noise equations with degenerate limiting system arising from volatility models*, *Large Deviations and Asymptotic Methods in Finance*, Springer Proceedings in Mathematics and Statistics, Vol. 110 (2015).
- S. DE MARCO, P. FRIZ, AND S. GERHOLD, Rational Shapes of local volatility, *Risk magazine*, (2013), pp. 82–87.

- S. DE MARCO AND C. MARTINI, *The term structure of implied volatility in symmetric models with applications to Heston*, Int. J. Theor. Appl. Finance, 15 (2012).
- V. BALLY AND S. DE MARCO, *Some estimates in extended stochastic volatility models of Heston type*, Risk and Decision Analysis, 2 (2011), pp. 195–206.
- S. DE MARCO, *Smoothness and asymptotic estimates of densities for SDEs with locally smooth coefficients and applications to square root-type diffusions*, Ann. Appl. Probab., 21 (2011), pp. 1282–1321.

Submitted

- F. BOURGEY, S. DE MARCO, P. FRIZ, AND P. PIGATO, *Local volatility under rough volatility*, under minor revision for Mathematical Finance. <https://arxiv.org/abs/2204.02376>, 2022.
- F. BOURGEY, S. DE MARCO, AND E. GOBET, *Weak approximations and VIX option price expansions in forward variance curve models*. <https://arxiv.org/abs/2202.10413>, 2022.

Other works

- S. DE MARCO AND C. MARTINI, *Quasi-Explicit calibration of Gatheral's SVI model*, Zeliade Systems White Paper #5 (2009).
<https://zeliade.com/wp-content/uploads/whitepapers/zwp-0005-SVICalibration.pdf>

Ongoing work

- *Yet another analysis of the S&P500 at-the-money skew: crossover of different power-law behaviours*, with J. Delemotte
- *VIX in SSVI*, with A. Mingone and C. Martini
- *Asymptotics for American options in local volatility models*, with F. Bourgey

Abstract: In this habilitation thesis, we present a selection of some of our recent contributions to the field of volatility modeling and risk management and to the solution of numerical problems related to risk computations in finance. A class of numerical problems we consider is the approximate simulation of stochastic processes and of nested expectations by means of multilevel schemes, to which the first part of the manuscript is dedicated. In a second part, we consider a problem of robust risk management for volatility derivatives. We derive optimal super-replication strategies for VIX options based on Vanilla options on the underlying SP500 index and on VIX futures. In this part, we exploit tools from the theory of optimal transport with martingale constraints. The last part of the document is devoted to asymptotic expansion techniques for diffusions processes and other recent generations of financial models involving fractional processes (rough volatility models). The involved mathematical tools are pathwise large deviations and Malliavin calculus, the latter allowing to provide rigorous error estimates for the involved asymptotic expansions.

Keywords: Stochastic analysis, Malliavin calculus, Large deviations, Multilevel simulation, nested risks, robust hedging, volatility management, rough volatility, implied volatility surface, local volatility surface, VIX derivatives

Résumé : Dans ce rapport de sythèse, nous présentons une sélection de certaines de nos contributions récentes à des problèmes de modélisation et gestion des risques sur la volatilité des actifs, ainsi qu'au traitement de certains calculs de risques liés à des portefeuilles de dérivés. Une catégorie de problème numériques que nous traitons est la simulation approchée de processus stochastiques et de certaines espérance emboîtées à l'aide de la méthode multi-niveaux, à laquelle la première partie de ce manuscrit est consacrée. Dans une deuxième partie, nous considérons un problème de couverture robuste de dérivés de volatilité. Plus précisément, nous obtenons des stratégies optimales de sur-réplication pour les options sur l'indice VIX, construites à partir d'options Vanilles sur l'indice SP500 sous-jacent et de contrats futures VIX. Dans cette partie, nous exploitons des outils venant de la théorie du transport optimal avec contrainte de martingale. La dernière partie de ce document est dédiée à des problèmes de calcul asymptotique pour les processus de diffusion et autres classes récentes de modèles de prix faisant intervenir des processus fractionnaires (les modèles *rough volatility*). Les outils que nous employons ici viennent de la théorie de grandes déviations trajectorielles ainsi que du calcul de Malliavin, ce dernier nous permettant de formuler des estimations d'erreur pour les expansions asymptotiques.

Mots clés : Analyse stochastique, calcul de Malliavin, grandes déviations, simulation multi-niveaux, risques emboîtés, couverture robuste, risque de volatilité, volatilité *rough*, surface de volatilité implicite, surface de volatilité locale, dérivés sur le VIX

Table des matières

1	Introduction	1
1.1	Estimation of expectations in the presence of a bias : Monte Carlo and Multilevel Monte Carlo methods	1
1.1.1	An archetypal example : Euler scheme for SDEs	2
1.1.2	The Multilevel method	3
1.2	Forward variance modeling	5
1.2.1	A model class based on finite-dimensional Brownian motions	7
1.2.1.1	Instantaneous volatility and consistent modeling of S	8
1.2.1.2	An explicit class of forward variance models	9
1.2.2	On financial modeling with rough volatility	10
1.3	Elements of Martingale Optimal Transport	11
1.3.1	Classical optimal transport	11
1.3.2	Super-replication of exotic options and martingale optimal transport	12
I	Multilevel Monte Carlo methods for nested risk computations and option pricing	15
2	Multilevel Monte Carlo methods for nested risk computations	17
2.1	Nested expectations	17
2.1.1	Nested Monte Carlo estimator	18
2.2	Multilevel method	18
2.2.1	A further improvement : the antithetic Multilevel estimator	19
2.2.2	Irregular functions g	21
2.2.2.1	The case of step functions $g(x) = 1_{x>a}$	21
2.2.3	Our motivation : derivative pricing in the presence of Initial Margin valuation adjustments	22
2.3	Our assumptions and results for the nested expectation problem	24
2.3.1	Numerical experiments	26
3	An application to forward variance simulation in the rough Bergomi model	29
3.1	The computational problem	29
3.2	Discretization errors	31
3.3	Monte Carlo and Multilevel Monte Carlo simulation	33
3.3.1	Standard Monte-Carlo estimation	33
3.3.2	Multilevel scheme	34
3.3.3	Numerical experiments	35
3.3.4	The additional contribution of an efficient control variate	37

II	Pricing and risk management of VIX derivatives	39
4	Robust replication of VIX options : a constrained martingale optimal transport problem	41
4.1	Super-hedging of VIX options	41
4.1.1	Duality	43
4.2	An analytical upper-bound : extremal moment problem	44
4.2.1	Optimality	48
4.3	Numerical experiments	49
5	Weak approximations and VIX option price expansions in forward variance models	51
5.1	Our contribution in a nutshell	52
5.2	The main elements of our approximation procedure	53
5.2.1	Strong error estimates between VIX_T^2 and its proxy $VIX_{T,P}^2$	53
5.2.2	Option price expansion	54
5.3	Mixed (or “skewed”) exponential models	56
5.3.1	Numerical tests for option price formulas and implied volatilities	58
III	Asymptotic results for stochastic volatility	61
6	Asymptotics of the local volatility surface in stochastic volatility models	63
6.1	The local volatility surface	63
6.1.1	A related model-free result : the harmonic mean property of the implied volatility	65
6.2	Local volatility under rough volatility	65
6.2.1	Our toolbox : Large Deviations	67
6.2.2	Our main result : asymptotics of the Markovian projection in the large deviations regime	69
6.2.3	The consequence for the short-time local volatility skew	70
6.2.4	Numerical tests	72
	Bibliographie	79

Introduction

Contents

1.1 Estimation of expectations in the presence of a bias : Monte Carlo and Multilevel Monte Carlo methods	1
1.1.1 An archetypal example : Euler scheme for SDEs	2
1.1.2 The Multilevel method	3
1.2 Forward variance modeling	5
1.2.1 A model class based on finite-dimensional Brownian motions	7
1.2.1.1 Instantaneous volatility and consistent modeling of S	8
1.2.1.2 An explicit class of forward variance models	9
1.2.2 On financial modeling with rough volatility	10
1.3 Elements of Martingale Optimal Transport	11
1.3.1 Classical optimal transport	11
1.3.2 Super-replication of exotic options and martingale optimal transport	12

In this introductory chapter, we collect some basic results, present the most important objects and introduce the main problems that will be studied in the rest of the manuscript.

1.1 Estimation of expectations in the presence of a bias : Monte Carlo and Multilevel Monte Carlo methods

Suppose we want to estimate some quantity $z \in \mathbb{R}$ with a random estimator \hat{Z} . We have the classical bias-variance decomposition of the mean squared error

$$\text{MSE} = \mathbb{E} \left[(z - \hat{Z})^2 \right] = (\text{bias})^2 + \text{variance}$$

where

$$\text{bias} = \mathbb{E}[\hat{Z}] - z$$

and

$$\text{variance} = \text{Var}(\hat{Z}).$$

Unbiased estimation. A typical situation is the evaluation of expectations : we wish to evaluate $z = \mathbb{E}[Z]$ for a, say, square-integrable random variable Z . Of course, if the random variable Z can be simulated exactly, the empirical mean of independent and identically distributed (i.i.d.) samples $(Z_m)_{m=1, \dots, M}$,

$$\hat{Z}_M = \frac{1}{M} \sum_{m=1}^M Z_m,$$

offers an unbiased estimator of $\mathbb{E}[Z]$:

$$\text{MSE} = \text{Var}(\hat{Z}_M) = \frac{1}{M} \text{Var}(Z).$$

In order to achieve a target mean squared error of order $\text{MSE} = \mathcal{O}(\epsilon^2)$ for some given error tolerance $\epsilon > 0$, we therefore have to choose $M = \frac{\text{Var}(Z)}{\epsilon^2}$. The computational cost of the empirical mean \hat{Z}_M is linear in M (assuming each sample Z_m has $\mathcal{O}(1)$ cost), so that, eventually, the computational cost of the empirical mean is

$$\mathcal{C}_{\text{unbiased}} = \mathcal{O}(\epsilon^{-2}),$$

for a target mean squared error $\text{MSE} = \mathcal{O}(\epsilon^2)$. Any unbiased estimator with linear cost squares this property.

Of course, in several situations, the involved random variable Z cannot be simulated exactly. This is typically the case when non-linear functions of expectations are involved, or when the solution of a continuous-time stochastic differential equation is approximated via some discretization scheme.

Functions of expectations. Suppose that the aim is to evaluate a function of an expectation, say $g(\mathbb{E}[Z])$ for a given function g . The estimator obtained composing g with the empirical mean,

$$\hat{Z}_M = g\left(\frac{1}{M} \sum_{m=1}^M Z_m\right),$$

is not unbiased any more : it is classical that, if $g \in \mathcal{C}_b^2$, the estimator bias behaves as

$$\mathbb{E}\left[g\left(\frac{1}{M} \sum_{m=1}^M X_m\right)\right] - g(\mathbb{E}[X]) = \mathcal{O}\left(\frac{1}{M}\right) \quad \text{as } M \rightarrow \infty,$$

while the variance is still $\text{Var}\left(g\left(\frac{1}{M} \sum_{m=1}^M X_m\right)\right) = \mathcal{O}\left(\frac{1}{M}\right)$. Consequently, the MSE still behaves as

$$\text{MSE} = \mathcal{O}\left(\frac{1}{M}\right)$$

(in other words : the bias is asymptotically negligible in front of the statistical error), and we are back to the unbiased case : $\text{MSE} = \mathcal{O}(\epsilon^2)$ is achieved setting $M = \mathcal{O}(\epsilon^{-2})$, so that the computational cost of the estimator is still $\mathcal{C} = \mathcal{O}(M) = \mathcal{O}(\epsilon^{-2})$.

1.1.1 An archetypal example : Euler scheme for SDEs

Consider the Stochastic Differential Equation (SDE)

$$X_t = X_0 + \int_0^t b(s, X_s) ds + \sigma(s, X_s) dW_s, \quad t \leq T,$$

where $b(t, x)$ and $\sigma(t, x)$ are, say, Lipschitz functions of x uniformly over $t \in [0, T]$. The goal is to estimate $z = \mathbb{E}[g(X_T)]$ for a given function g ; when X_T cannot be simulated exactly, one can appeal to the Euler scheme with n time steps

$$\begin{cases} \hat{X}_{t_{k+1}}^n = \hat{X}_{t_k}^n + b(t_k, \hat{X}_{t_k}^n) \frac{T}{n} + \sigma(t_k, \hat{X}_{t_k}^n) (W_{t_{k+1}} - W_{t_k}) \\ \hat{X}_0^n = X_0, \end{cases}$$

where $t_k = k \frac{T}{N}$. The estimator based on M i.i.d. samples $(\hat{X}^{n,m})_{1 \leq m \leq M}$ of the Euler scheme with n steps is

$$\hat{Z}_{M,n} = \frac{1}{M} \sum_{m=1}^M g(\hat{X}_T^{n,m}).$$

If g is sufficiently smooth, the Euler scheme is known to have weak rate of convergence of order 1 : the estimator bias behaves as

$$\mathbb{E}[g(\hat{X}_T^n)] - \mathbb{E}[g(X_T)] = \mathcal{O}\left(\frac{1}{n}\right).$$

The variance of $\hat{Z}_{M,n}$ of course satisfies

$$\text{Var}\left(\frac{1}{M} \sum_{m=1}^M g(\hat{X}_T^{n,m})\right) = \frac{1}{M} \text{Var}\left(g(\hat{X}_T^n)\right) \leq \frac{C}{M}, \quad \forall n \geq 1.$$

Overall, we have

$$\text{MSE}(\hat{Z}_{M,n}) \leq C \left(\frac{1}{n^2} + \frac{1}{M} \right).$$

In order to achieve a target mean squared error $\text{MSE} = \mathcal{O}(\epsilon^2)$ for some given error tolerance $\epsilon > 0$, we have to take

$$n = \mathcal{O}(\epsilon^{-1}), \quad M = \mathcal{O}(\epsilon^{-2})$$

which yields the overall computational cost

$$C = Mn = \mathcal{O}(\epsilon^{-3}),$$

which is of course asymptotically much worse than the cost of order $\mathcal{O}(\epsilon^{-2})$ of an unbiased estimator. The natural question at this point is : can we do better ? In other terms, can one design an alternative estimator with a better computational complexity for the same target error level ?

1.1.2 The Multilevel method

We now aim at introducing the main ideas and notation for the so-called Multilevel method, introduced in the field of Monte Carlo simulation of stochastic processes by the seminal work of Giles [48], building on ideas of Heinrich [61, 62, 63]. In a nutshell, the principle of the Multilevel method is to combine estimators with different accuracy (different bias) and built with common random numbers, in such a way to profit from cancellations between their biases (in the spirit of Richardson-Romberg extrapolation) and, at the same time, to profit from variance reduction from the presence of control variates.

Let us present the construction of such an estimator more in details. In a general setting (encompassing the problem of nested expectations in the previous section, and many other problems related for example to the discretization of continuous time stochastic processes), assume we can exactly simulate approximations Z_l of a random variable Z with increasing accuracy (that is, decreasing bias)

$$\mathbb{E}[Z_l] - \mathbb{E}[Z] \rightarrow 0 \quad \text{as } l \rightarrow \infty$$

but increasing computational cost. Let us fix a maximum level of approximation $L \in \mathbb{N}$. The telescopic sum

$$\mathbb{E}[Z_L] = \mathbb{E}[Z_0] + \sum_{l=1}^L \mathbb{E}[Z_l - Z_{l-1}], \tag{1.1.1}$$

suggests to define the estimator

$$\widehat{Z}_{M_0, \dots, M_L} = \frac{1}{M_0} \sum_{m=1}^{M_0} Z_0^{(0,m)} + \sum_{l=1}^L \frac{1}{M_l} \sum_{m=1}^{M_l} \left(Z_l^{(l,m)} - Z_{l-1}^{(l,m)} \right) \quad (1.1.2)$$

where M_l is the number of samples used to approximate $\mathbb{E}[Z_l - Z_{l-1}]$ (with the convention $Z_{-1} = 0$). Since it combines estimators (the inner empirical means $\frac{1}{M_l} \sum_{m=1}^{M_l}$) associated to different levels of accuracy, $\widehat{Z}_{M_0, \dots, M_L}$ will be dubbed *multilevel* Monte Carlo estimator (MLMC, or simply ML, in short).

The random samples $Z_l^{(l,m)}$ and $Z_{l-1}^{(l,m)}$ are constructed according to the following recipe :

- ▶ For every level l , $(Z_l^{(l,m)})_{m \leq M_l}$ and $(Z_{l-1}^{(l,m)})_{m \leq M_l}$ are i.i.d. samples of Z_l and Z_{l-1} , respectively ;
- ▶ For every level l , $Z_l^{(l,m)}$ and $Z_{l-1}^{(l,m)}$ constructed with common random numbers, so that the variance at level l can be estimated via the strong error

$$\begin{aligned} V_l = \text{Var}(Z_l - Z_{l-1}) &\leq \|Z_l - Z_{l-1}\|_2^2 \mathbb{E}[(Z_l - Z_{l-1})^2] \\ &\leq 2 \mathbb{E}[(Z_l - Z)^2] + 2 \mathbb{E}[(Z_{l-1} - Z)^2]; \end{aligned}$$

- ▶ There is independence among levels : all the random samples used at level l are independent from the ones used at other levels.

The estimator (1.1.2) is parametrized by L and by the integers $(M_l)_{l=0, \dots, L}$ representing the simulation budget allocated at each level. We are free to choose the value of such hyper-parameters according to an optimality criterion, as we explain below.

Mean-squared error and complexity of the estimator. The computational cost of the Multilevel estimator $\widehat{Z}_{M_0, \dots, M_L}$ is

$$\mathcal{C} = \sum_{l=0}^L M_l \text{cost}(Z_l).$$

In light of (1.1.1), it is clear that $\mathbb{E}[\widehat{Z}_{M_0, \dots, M_L}] = \mathbb{E}[\mathbb{E}[Z_L]]$, so that $\text{bias}(\widehat{Z}_{M_0, \dots, M_L}) = \text{bias}(Z_L)$. By independence among the levels, we have

$$\text{MSE} = (\text{bias}(Z_L))^2 + \text{variance} = (\text{bias}(Z_L))^2 + \sum_{l=0}^L \frac{V_l}{M_l}.$$

We can now fix an target error tolerance ϵ and look for the estimator with the lowest computational cost among all those achieving the desired mean squared error. Doing so, we are left with the optimization problem

$$\min_{M_0, \dots, M_L} (\text{computational cost}) \quad \text{under the constraint} \quad \text{MSE} = \epsilon^2,$$

that is

$$\min_{M_0, \dots, M_L} \sum_{l=0}^L M_l \text{cost}(Z_l) \quad \text{under} \quad (\text{bias}(Z_L))^2 + \sum_{l=0}^L \frac{V_l}{M_l} = \epsilon^2. \quad (1.1.3)$$

The solution M_0^*, \dots, M_L^* of (1.1.3) and the final cost value $\sum_{l=0}^L M_l^* \text{cost}(Z_l)$ of course depend on the behaviour of $\text{cost}(Z_l)$, $\text{bias}(Z_L)$ and on the variance $V_l = \text{Var}(Z_l - Z_{l-1})$ for the specific case at hand.

The following theorem, usually referred to as ‘‘The Multilevel Monte Carlo Theorem’’, sums up the behavior of the resulting optimal ML estimator according to the available estimates for the bias, variance and cost at each level l .

Theorem 1.1 (Giles 2008 [48]). *Fix a parameter $h > 0$ (representing accuracy) and the sequence*

$$h_l = \frac{h}{2^l}, \quad l \geq 1.$$

Assume there exists constants c_1, c_2, c_3 and coefficients $\alpha, \beta, \gamma > 0$ such that

$$(i) \quad |\mathbb{E}[Z_l] - \mathbb{E}[Z]| \leq c_1 h_l^\alpha \quad (\text{bias estimate})$$

$$(ii) \quad \|Z_l - Z_{l-1}\|_2^2 \leq c_2 h_l^\beta \quad (\text{strong error estimate})$$

$$(iii) \quad \text{cost}(Z_l) \leq c_3 \left(\frac{1}{h_l}\right)^\gamma \quad (\text{cost estimate})$$

Then, there exist a number of levels L and sample sizes $(M_l)_{0 \leq l \leq L}$ such that the ML estimator achieves $\text{MSE} = \epsilon^2$ with computational cost

$$C = \begin{cases} \mathcal{O}\left(\epsilon^{-2-\frac{\gamma-\beta}{\alpha}}\right) & \text{if } \beta < \gamma \\ \mathcal{O}\left(\epsilon^{-2}(\log \epsilon)^2\right) & \text{if } \beta = \gamma \\ \mathcal{O}\left(\epsilon^{-2}\right) & \text{if } \beta > \gamma \end{cases}$$

Note that the strong error estimate in point (ii) allows to upper bound the variance at level l , via $V_l \leq \|Z_l - Z_{l-1}\|_2^2$. When applying Theorem 1.1 to the estimation problems considered in Chapters 2 and 3, we aim at keeping track of the values of the constants α, β and γ , which will be explicitly displayed next to our bias and variance estimates.

1.2 Forward variance modeling

We consider a market containing a (tradable or non-tradable) asset S , in which forward contracts – or future contracts, according to the type of market – and European options on S are traded for arbitrary maturities T . We denote F_t^T , $t \leq T$, the forward price of S quoted at time t for the maturity T .

- The *spot* log-contract on S , traded at time t for the maturity T , is the European option with payoff $-\frac{2}{T-t} \log \frac{S_T}{S_t}$.
- The *forward* log-contract on S on the future time interval $[T_2, T_1]$, $T_2 \geq T_1$, is the option with payoff $-\frac{2}{T_2-T_1} \log \frac{S_{T_2}}{S_{T_1}}$; such a contract can be traded at earlier times $t \leq T_1$.

Definition 1.1 (Log-contract forward variance). *The forward variance $V_t^{T_1, T_2}$ of the log-contract over the future interval $[T_1, T_2]$, observed at time $t \leq T_1$, is given by*

$$V_t^{T_1, T_2} = \text{price}_t \left(-\frac{2}{T_2 - T_1} \log \left(\frac{S_{T_2}}{S_{T_1}} \right) \right). \quad (1.2.1)$$

In (1.2.1), the notation $\text{price}_t(H)$ stands either for the observed market price of claim H , or for the price generated by a model – accordingly, we speak of market forward variance or model forward variance. For simplicity, in this manuscript we consider zero interest rates, and zero dividends and repo rates when the asset S is tradable. If this is not the case, S_{T_1} should be replaced inside (1.2.1) by the forward price $F_{T_1}^{T_2}$ observed at time T_1 for the maturity T_2 .

Due to the elementary identity $\log \frac{S_{T_2}}{S_{T_1}} = \log \frac{S_{T_2}}{S_t} - \log \frac{S_{T_1}}{S_t}$, we have, assuming linearity of the price operator price_t , the following relationship between forward variances

$$V_t^{T_1, T_2} = \frac{(T_2 - t)V_t^{t, T_2} - (T_1 - t)V_t^{t, T_1}}{T_2 - T_1}, \quad \forall t \leq T_1 \leq T_2.$$

Remark 1.1 (Forward variances are implied variances). *It is interesting to note that, in (1.2.1), we define a variance by means of a price. Is something missing in (1.2.1), in other words, are we missing a transformation that maps a price into a variance parameter? The answer is no, and this is clear once we have considered the Black-Scholes price of the log-contract : in a Black-Scholes model with volatility parameter σ , the log contract is worth*

$$\text{price}_t^{\text{BS}} \left(-\frac{2}{T_2 - T_1} \log \left(\frac{S_{T_2}}{S_{T_1}} \right) \right) = -\frac{2}{T_2 - T_1} \left(-\frac{1}{2} \sigma^2 (T_2 - T_1) \right) = \sigma^2.$$

This means that $\sqrt{V_t^{T_1, T_2}}$ in Definition 1.1 is precisely the unique volatility parameter σ has to be injected inside the Black-Scholes formula in order to retrieve the price of the log-contract on the right hand side of (1.2.1); in other words, $\sqrt{V_t^{T_1, T_2}}$ is the implied volatility (resp. $V_t^{T_1, T_2}$ the implied variance) of the forward log-contract over $[T_1, T_2]$, observed at time t .

Remark 1.2. *Forward variance can also be defined from Variance swaps quotes. In general, log-contract forward variances $V_t^{T_1, T_2}$ and Variance swap forward variances $\tilde{V}_t^{T_1, T_2}$ do not coincide – simply because they are implied variances of different contracts – even if their observed relative difference is typically small, see [13, Chapter 5].*

Being prices (or equivalently, implied variances, see Remark 1.1), forward variance are observable. An important example of a market quoting forward variances of log-contracts is the VIX market : when S is the SP500 index, the VIX index quoted by the Chicago Board of Options Exchange (CBOE) at a given (present or future) time T is

$$\text{VIX}_T := \sqrt{V_T^{T, T+\Delta}} = \sqrt{\text{price}_T \left(-\frac{2}{\Delta} \log \left(\frac{S_{T+\Delta}}{S_T} \right) \right)}, \quad (1.2.2)$$

where $\Delta = 30$ days is a fixed time horizon. Since the log-contract itself is not a traded object on the CBOE, its price is quoted by static replication of the payoff $\ln(S)$ with call and put options; the precise way the continuous static replication formula is discretized over the observed option strikes and maturities is described in the VIX White paper [26].

In practice, option prices are quoted only for a finite set of market maturities $(T_i)_{i=1, \dots, n}$. A stochastic dynamic model could target the underlying asset S and the set of discrete forward variances $(V_t^{T_i, T_{i+1}})_{i=1, \dots, n}$, resulting in a $n + 1$ -dimensional model. Instead of doing so, a more parsimonious approach, pioneered by Dupire [34] and then intensively studied by Buheler [24] and Bergomi [14, 13], is to model for instantaneous forward variances ξ_t^T , defined by

$$\xi_t^T = \frac{d}{dT} \left((T - t)V_t^{t, T} \right), \quad t \leq T, \quad (1.2.3)$$

assuming that, for every t , forward variances $V_t^{t, T}$ are differentiable with respect to their maturity parameter T . Once a model for the instantaneous – and therefore, strictly speaking, unobserved – object ξ_t^u has been

settled, the observed discrete forward variances are retrieved via integration over the parameter u ,

$$V_t^{t,T} = \frac{1}{T-t} \int_t^T \xi_t^u du \quad (1.2.4)$$

and

$$V_t^{T_1, T_2} = \frac{1}{T_2 - T_1} \int_{T_1}^{T_2} \xi_t^u du \quad (1.2.5)$$

We note in passing the analogy with forward rate modeling, where instantaneous forward rates $f(t, T)$ are defined from bond prices.

Remark 1.3 (Modeling approaches). *It can be instructive to situate the forward variance modeling considered in this section with respect to other existing modeling approaches.*

1. *Instantaneous volatility modeling : the target objects are the asset price S_t and its instantaneous volatility σ_t , typically modeled via a stochastic differential equation for the couple (S_t, σ_t) , a two-dimensional object. Classical stochastic volatility models such as the Heston and the SABR model follow this approach ; options on S have to be priced according to some numerical procedure and then implied volatilities of vanilla options – or forward variances – have to be extracted from the numerically computed prices.*
2. *Implied volatility surface modeling : the target objects of the stochastic modeling are the asset price S_t and the implied volatility surface of Vanillas. This approach is ambitious, for it requires to model the entire volatility surface, an infinite dimensional object parameterized by two variables (the option maturity T and the strike price K), which is known to be a difficult task, see [81] and [13, chapter 4]. The challenge is to build an explicit random surface satisfying at every point t in time the static no-arbitrage conditions for different values of T and K , while still satisfying the dynamic no-arbitrage conditions for every fixed (T, K) as t moves, and overall generating realistic joint dynamics for implied volatilities and the underlying asset S .*

Choosing to target instantaneous forward variances (and therefore integrated forward variances (1.2.5)), we are jointly modeling the asset S_t and the prices of a one-dimensional family of derivatives (here, specifically : log-contracts), as opposed to the two-dimensional implied volatility surface. This task is less ambitious than the full implied volatility surface in the modeling approach 2 above, while still keeping a considerable amount of flexibility and enhanced dynamical properties with respect to the simplified instantaneous volatility modeling.

Remark 1.4 (Forward variances are local martingales). *According to Definition 1.1, for any choice of T_1 and T_2 , the forward variance $(V_t^{T_1, T_2})_{t \leq T_1}$ is the price of an option with fixed payoff ; as a consequence, within a pricing model, forward variances (1.2.1) have to be modeled by local martingales under the pricing measure. This appealing feature of forward variances simplifies the modeling problem with respect to the modeling of implied volatilities of Vanilla options (which are not driftless) : we can freely specify the volatility of forward variances, and then simply have to set their drift to zero.*

1.2.1 A model class based on finite-dimensional Brownian motions

We want to set up a dynamic model for the random curve

$$T \mapsto \xi_t^T, \quad T \geq t,$$

together with a model for the asset price S . According to Remark 1.4, we look for models in which instantaneous forward variances have zero drift.

For every fixed T , we can ask that $(\xi_t^T)_{t \leq T}$ is the solution of the equation

$$\xi_t^T = \xi_0^T + \int_0^t a^T(s, \xi_s^T) K(T, s) \cdot dW_s, \quad t \leq T \quad (1.2.6)$$

where W a n -dimensional Brownian motion on a stochastic basis $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$, possibly with correlated components $\langle W_t^i, W_t^j \rangle = \rho_{ij} t$, and $K : [0, \infty)^2 \rightarrow \mathbb{R}^n$ is a deterministic kernel satisfying $K(T, s) = 0$ for $s > T$. The notation $K(T, s) \cdot dW_s$ stands for $\sum_{i=1}^n K_i(T, s) dW_s^i$. For every T , the function $a^T : [0, T] \times \mathbb{R} \rightarrow \mathbb{R}$ plays the role of a local volatility function for the forward variance ξ_t^T . Finally, $(\xi_0^T)_{T \geq 0}$ denotes the initial instantaneous forward variance curve – an observable market parameter (up to the computation of the derivative in (1.2.3)).

Note that (1.2.6) is a family of SDEs, indexed by the parameter T . For every fixed T , existence and uniqueness of solutions are standard issues and can be guaranteed assuming appropriate conditions on the function a^T and the kernel K . Being the solution to (1.2.6), every forward variance $(\xi_t^T)_{t \leq T}$ is a local martingale and a Markov process.

According to the choice of the kernel K (and of the function $a^T(\cdot)$), the model (1.2.6) can admit a Markovian representation of the forward variance curve, that is a representation of the form

$$\xi_t^T = f(t, T, X_t) \quad (1.2.7)$$

simultaneously for all $t \geq 0$ and all $T \geq t$, where $(X_t)_{t \geq 0}$ is some finite-dimensional Markov process. A representation such as (1.2.7) in general simplifies the simulation and computational issues related to the use of the model (1.2.6).

1.2.1.1 Instantaneous volatility and consistent modeling of S

Once the model (1.2.6) for forward variances has been fixed, we look for a joint model for the asset price S_t . A possible choice is to use the stochastic volatility model

$$dS_t = S_t \sqrt{\xi_t^t} dB_t \quad (1.2.8)$$

where B is a Brownian motion adapted to \mathcal{F}_t , possibly correlated with W . The diagonal process $\xi_t^t = \xi_t^T|_{t=T}$ has to be well defined in order for (1.2.8) to make sense.

Suppose we want to price a log-contract under the model (1.2.8). If we generate the price of derivatives with conditional expectations, by Itô's formula applied to $\log(S)$ we have

$$\begin{aligned} \text{price}_t^{\text{model}} \left(-\frac{2}{T_2 - T_1} \log \left(\frac{S_{T_2}}{S_{T_1}} \right) \right) &= \mathbb{E} \left[-\frac{2}{T_2 - T_1} \log \left(\frac{S_{T_2}}{S_{T_1}} \right) \middle| \mathcal{F}_t \right] \\ &= \frac{1}{T_2 - T_1} \int_{T_1}^{T_2} \mathbb{E} [\xi_u^u | \mathcal{F}_t] du = \frac{1}{T_2 - T_1} \int_{T_1}^{T_2} \xi_t^u du = V_t^{T_1, T_2}. \end{aligned} \quad (1.2.9)$$

Equation (1.2.9) shows that the prices of log-contracts in the asset price model (1.2.8) are precisely given by the forward variances generated by the model (1.2.6); in other words, the couple of equations (1.2.6)–(1.2.8) provides a consistent joint modeling of S_t and of all forward variances ξ_t^u .

The instantaneous variance in the asset price model (1.2.8),

$$V_t = \xi_t^t = \xi_0^t + \int_0^t a^t(s, \xi_s^t) K(t, s) \cdot dW_s, \quad t \geq 0,$$

is a stochastic process that integrates the moving kernel $K(t, \cdot)$ and the moving process $a^t(\cdot, \xi^t)$ against Brownian motion up to time t . We could refer to V_t as to a Volterra process due to the presence of the integration kernel K , even if it is worth to notice that we have not obtained a Volterra process in the terminology of Abi Jaber et al. [2], which would rather correspond to an equation of the form $V_t = \xi_0^t + \int_0^t a(s, V_s) K(t, s) \cdot dW_s$.

1.2.1.2 An explicit class of forward variance models

The model class (1.2.6) is very wide – in general, it is not obvious how one would specify a local volatility function $a^T(\cdot)$ for each instantaneous forward variance (even though some efficient parametric examples can be designed, such as the model family studied in section 5.3).

An effective instance of the model class (1.2.6) is obtained with the simple choice

$$a^T(t, x) = x.$$

Moreover, we assume $K(T, \cdot) \in L_{\text{loc}}^2(\mathbb{R}_+, \mathbb{R}^n)$ for every $T > 0$. In this setting, the explicit solution of (1.2.6) is given by

$$\xi_t^T = \xi_0^T \exp \left(\int_0^t K(T, s) \cdot dW_s - \frac{1}{2} \int_0^t K(T, s) \cdot \rho K(T, s) ds \right), \quad t \leq T, \quad (1.2.10)$$

where $\int_0^t K(T-s) \cdot \rho K(T-s) ds$ stands for $\sum_{i,j=1}^n \int_0^t K_i(T-s) \rho_{i,j} K_j(T-s) ds$.

It is clear that (1.2.10) defines a family of log-normal processes, indexed by the maturity parameter T . Though it might seem excessively simple, (1.2.10) provides a first effective solution to the forward variance modeling problem that still contains a good amount of flexibility via the choice of the kernel function K , the number n of Brownian factors and their correlations ρ_{ij} . From the point of view of simulation, an appealing feature of the model (1.2.10) is that only explicit functions of Gaussian random variables are involved (as a consequence, instantaneous forward variances ξ_t^T can be simulated exactly, even if integrated forward variances (1.2.4) and (1.2.5) cannot, as we will discuss in detail in Chapter 2).

In fact, the model family (1.2.10) is widespread and used in practice; it encompasses two of the main examples in the literature :

- The n -factors Bergomi's model [14], based on the exponential kernels

$$K_i(T, s) = \omega_i e^{-k_i(T-s)}, \quad \omega_i, k_i > 0 \quad i = 1, \dots, n. \quad (1.2.11)$$

- The so-called rough Bergomi model of Bayer, Friz and Gatheral [9], corresponding to $n = 1$ and to the power-law (or fractional) kernel

$$K(T, s) = \frac{\eta}{(T-s)^{\frac{1}{2}-H}}, \quad \eta > 0, \quad H \in (0, 1/2). \quad (1.2.12)$$

A n factor version of this model can of course be considered, setting $K_i(T, s) = \frac{\eta_i}{(T-s)^{\frac{1}{2}-H_i}}$, for possibly different parameters η_i and H_i ; the original rough Bergomi model designed [9] corresponds to (1.2.12).

In this setting, the instantaneous variance process induced by (1.2.10),

$$\begin{aligned} V_t = \xi_t^t &= \xi_0^t \exp \left(\int_0^t K(t, s) dW_s - \frac{1}{2} \int_0^t K(t, s)^2 ds \right) \\ &= \xi_0^t \exp \left(\eta \int_0^t (t-s)^{H-1/2} dW_s - \frac{\eta^2}{4H} t^{2H} \right), \end{aligned} \tag{1.2.13}$$

is an explicit function of the Gaussian Volterra process

$$\int_0^t K(t, s) dW_s = \int_0^t (t-s)^{H-1/2} dW_s,$$

which corresponds to the Riemann–Liouville Brownian motion. The process V admits a modification with trajectories that are β -Hölder for every $\beta < H < 1/2$, which explains the terminology “rough” for the resulting stochastic volatility model (1.2.8). Several authors have reported that a small value of H , namely $H \approx 0.1$, is appropriate to reconstruct the term structure of the at-the-money Vanilla implied volatility skew observed on the SP500 market, see (1.2.14) below, cf. Alos et al. [6], Fukasawa [41], Bergomi [13], and Bayer et al. [9].

1.2.2 On financial modeling with rough volatility

Without any ambition to exhaustively cover the numerous contributions to the field of rough volatility modeling and rough volatility estimation that have appeared since the original insights of Alos et al. [6] and Fukasawa [41] and the groundbreaking work of Gatheral, Jaisson and Rosenbaum [47], we would like to schematically sum up here some of the main properties, advantages and possible disadvantages of the rough volatility modeling framework.

A rough process such as (1.2.13) (possibly with other specifications of the kernel K , leading for example to Mandelbrot’s fractional Brownian motion $W_t^H = \int_0^t K(t, s) dW_s$) can be used as a model for the instantaneous realized volatility of an asset S , otherwise as a model for spot volatilities in an option pricing framework.

A rough volatility model for *realized* volatility

- has statistical evidence : the observed realized volatility of several stock indexes and of a number of individual stocks has been statistically estimated to be rough, see Gatheral et al. [47] and subsequent work by Fukasawa, Takabatake and Westphal [44] and Bolko et al. [17];
- has microstructural foundations, in the sense that it can be seen as the scaling limit of models for order flow dynamics at the order book level, see Jaisson and Rosenbaum [69], El Euch, Fukasawa and Rosenbaum [36], Jusselin and Rosenbaum [72];
- has interesting forecasting properties for future realized volatilities, see again Gatheral et al. [47].

On the other side, a rough volatility *pricing* model

- has an appealing calibration power to the implied volatility surface of large stock indexes (a static property), see again Bayer et al. [9];
- has a *parsimonious parameterization* when compared with other multi-factor stochastic volatility or forward variance models. As an example, the at-the-money (ATM) implied volatility skew

$\partial_k \sigma_{\text{BS}}(t, k)|_{k=0}$ of the asset price S , where t denotes the option maturity and $k = \log \frac{K}{S_0}$ the log-strike, behaves as a power-law for short maturities, and the power-law decay is controlled by the single parameter H :

$$\partial_k \sigma_{\text{BS}}(t, k)|_{k=0} \sim \text{const.} \frac{1}{t^{1/2-H}} \quad \text{as } t \rightarrow 0. \quad (1.2.14)$$

For comparison, the forward variance model (1.2.10) with exponential kernels (1.2.11) requires the combination of at least two kernels K_i (hence of two independent Brownian motions) in order to achieve a behavior analogous to (1.2.14) within a reasonable degree of accuracy (that is, from maturities of a few weeks up to maturities of a few years), see [14, 13].

- Has, generally speaking, the desirable dynamical properties (joint dynamics of the asset price and implied volatilities) of a stochastic volatility model;
- on the other hand, it comes with additional complexity and computational cost with respect to traditional Markovian models.

It is worth to mention that other types of rough volatility models have been suggested, that fall out of the model class (1.2.6). Instead of postulating a dynamic model for instantaneous forward variances ξ_t^T which then induces a dynamics on the instantaneous variance, it is possible to follow the (actually more restrictive) opposite path. Following this approach, one can start assuming that the instantaneous variance V is the solution of a Volterra stochastic differential equation (here in dimension one)

$$V_t = V_0 + \int_0^t b(V_s)K(t, s)ds + \int_0^t \sigma(s, V_s)K(t, s)dW_s, \quad (1.2.15)$$

and then define forward variances from $\xi_t^T = \mathbb{E}[V_T | \mathcal{F}_t]$. The family of equations (1.2.15) includes the rough Heston model of El Euch and Rosenbaum [37] and more general Affine Volterra models [2].

1.3 Elements of Martingale Optimal Transport

In Chapter 4, we will tackle a problem of super-replication of VIX option exploiting some methods from optimal transport theory. Let us recall here the essential elements and results that will be used therein.

1.3.1 Classical optimal transport

Let \mathcal{X} and \mathcal{Y} be polish spaces. We denote $\mathcal{P}_{\mathcal{X}}$ (resp. $\mathcal{P}_{\mathcal{Y}}$) the set of probability measures on \mathcal{X} (resp. \mathcal{Y}) and consider two measures $\mu \in \mathcal{P}_{\mathcal{X}}$ and $\nu \in \mathcal{P}_{\mathcal{Y}}$.

The mass-transport problem of Monge (1781) is stated as follows : for a given cost function $c(x, y)$, representing the cost to transport mass from the location $x \in \mathcal{X}$ to the location $y \in \mathcal{Y}$, we look for the transport map that minimizes the total transport cost of the mass distribution μ to the target distribution ν , that is

$$P_M := \inf \int_{\mathcal{X}} c(x, T(x))d\mu(x), \quad (1.3.1)$$

over measurable maps $T : \mathcal{X} \mapsto \mathcal{Y}$ such that $\mu \circ T^{-1} = \nu$.

Seen as a problem for the variable T , (1.3.1) is highly non linear, due to the presence of T inside the (in general non-linear) cost c and to the constraint $\mu \circ T^{-1} = \nu$. Nevertheless, it is well-known that (1.3.1) can

be converted into a linear problem, hence simplified, when moving to its relaxed Monge-Kantorovich version

$$P_{MK} := \inf \int_{\mathcal{X} \times \mathcal{Y}} c(x, y) d\mathbb{P}(x, y), \quad (1.3.2)$$

where now one is looking for a solution over the set of probability measures $\mathbb{P} \in \mathcal{P}_{\mathcal{X} \times \mathcal{Y}}$ with given marginals μ and ν , that is such that $\mathbb{P}(A \times \mathcal{Y}) = \mu(A)$, $\mathbb{P}(\mathcal{X} \times B) = \nu(B)$. The difference with (1.3.1) is that the Kantorovich formulation allows to potentially split mass from a location x to several locations y . The problem (1.3.2) for the variable \mathbb{P} has now both a linear objective and linear constraints.

Probabilistic formulation. Denoting (X, Y) the coordinate process on $(\mathcal{X}, \mathcal{Y})$ we can rewrite (1.3.2) as

$$P_{MK} = \sup_{\mathbb{P} \in \mathcal{P}(\mu, \nu)} \mathbb{E}^{\mathbb{P}}[c(X, Y)] \quad (1.3.3)$$

where we denote $\mathcal{P}(\mu, \nu) = \{\mathbb{P} \in \mathcal{P}_{\mathcal{X} \times \mathcal{Y}} : X \overset{\mathbb{P}}{\sim} \mu, Y \overset{\mathbb{P}}{\sim} \nu\}$ and we have considered a sup instead of an inf for reasons that will be clear later on. It is not difficult to see that, due to the marginal law constraints μ and ν , the set $\mathcal{P}(\mu, \nu)$ is tight and closed, hence weakly compact.

(1.3.2) – or (1.3.3) – is referred to as the primal problem of optimal transport. There is a dual problem that is naturally related to (1.3.3),

$$D_{MK} = \inf \left\{ \mu(\varphi) + \nu(\psi) : \varphi \in L^1(\mu), \psi \in L^1(\nu), \right. \\ \left. \varphi(x) + \psi(y) \geq c(x, y), \right\} \quad (1.3.4)$$

where $\mu(\varphi) = \int \varphi(x) \mu(dx)$ and $\nu(\psi) = \int \psi(y) \nu(dy)$. It is easy to see that weak duality holds : we have $\mu(\varphi) + \nu(\psi) = \mathbb{E}^{\mathbb{P}}[\varphi(X) + \psi(Y)] \geq \mathbb{E}^{\mathbb{P}}[c(X, Y)]$ for every admissible couple φ, ψ in (1.3.4) and every admissible \mathbb{P} in (1.3.3). Taking the inf over φ, ψ at the left hand side and the sup over \mathbb{P} at the right hand side, we obtain $D_{MK} \geq P_{MK}$.

Applying Rockafellar's convex duality, it is possible to see that (under some assumptions on the cost c) strong duality $D_{MK} = P_{MK}$ holds.

Theorem 1.2 (Kantorovich duality, see [83]). *Assume that the cost function c is upper semi-continuous and satisfies $c(x, y) \leq a(x) + b(y)$ for some $a \in L^1(\mu)$ and $b \in L^1(\nu)$. Then*

$$D_{MK} = P_{MK} = \mathbb{E}^{\mathbb{P}^*}[c(X, Y)] \quad (1.3.5)$$

for some $\mathbb{P}^* \in \mathcal{P}(\mu, \nu)$.

1.3.2 Super-replication of exotic options and martingale optimal transport

Now consider a discrete-time market containing a tradable asset

$$(S_0, S_1, S_2) \quad \text{at times} \quad t_0 < t_1 < t_2,$$

and assume that Vanilla options on S_1 and S_2 are liquidly traded. More specifically, we assume that call options $(S_1 - K_1)^+$ and $(S_2 - K_2)^+$ can be bought at time t_0 (today) at their market prices $C(t_1, K_1)$ and $C(t_2, K_2)$.

When S is a large-stock index, the prices of Vanillas are typically listed for a large number of strikes K . We can therefore reasonably assume that the prices $C(t_1, K)$ and $C(t_2, K)$ are given for all positive strikes

$K > 0$. We can always think that this is achieved thanks to an interpolation/extrapolation of the finite set of observed market prices with an arbitrage-free parametric (or non-parametric) option price surface (even if we should keep in mind that such a procedure introduces a user-based choice in the reference prices, which then cannot, *stricto sensu*, be considered “market prices” any more). It is well-known that the functions $C(t_1, \cdot)$ and $C(t_2, \cdot)$ uniquely identify two measures $\mu \in \mathcal{P}_{\mathbb{R}_+}$ and $\nu \in \mathcal{P}_{\mathbb{R}_+}$ via

$$C(t_1, K) = \int (s_1 - K_1)^+ \mu(ds_1), \quad \forall K > 0, \quad C(t_2, K) = \int (s_2 - K)^+ \nu(ds_2), \quad \forall K > 0.$$

In particular, μ and ν will have finite first moments, $\int x\mu(dx) + \int y\nu(dy) < \infty$.

By static replication, the price of any other European option $\varphi(S_1)$ or $\psi(S_2)$ with integrable payoff (whose terminal value depends only on S_1 , otherwise only on S_2) is given by

$$\text{price}(\varphi(S_1)) = \int \varphi(s_1)\mu(ds_1). \quad \text{price}(\psi(S_2)) = \int \psi(s_2)\nu(ds_2).$$

Now consider an exotic option with payoff $c(S_1, S_2)$ and maturity t_2 , for example a forward-start option $c(s_1, s_2) = (s_1 - s_2)^+$ or $c(s_1, s_2) = (\frac{s_1}{s_2} - K)^+$. We can set up a semi-static hedging strategy by statically holding options $\varphi(S_1)$ and $\psi(S_2)$ and by additionally delta-hedging with the underlying over the time interval $[t_1, t_2]$. The resulting super-replication price of the option $c(S_1, S_2)$ is then given by

$$D = \inf_{\varphi, \psi, h} \{\mu(\varphi) + \nu(\psi)\} \tag{1.3.6}$$

where now the inf is over

$$\varphi \in L^1(\mu), \psi \in L^1(\nu), h \in C_b(\mathbb{R}) : \varphi(s_1) + \psi(s_2) + h(s_1)(s_2 - s_1) \geq c(s_1, s_2). \tag{1.3.7}$$

Note that $\mu(\varphi) + \nu(\psi) = \int \varphi d\mu + \int \psi d\nu$ represents the initial wealth of the hedging strategy, required to purchase the options $\varphi(S_1)$ and $\psi(S_2)$ at time t_0 , and that the super-replication condition (1.3.7) takes into account the additional wealth coming from the delta-hedge position $h(s_1)$ that we hold between t_1 and t_2 .

We denote D the problem (1.3.6), for analogy with (1.3.4). Note nevertheless that (1.3.6) is actually the problem with start with in quantitative finance, rather than the primal problem of optimal transport (1.3.1).

The primal problem of martingale optimal transport. Denote (S_1, S_2) the coordinate process on $\mathcal{X} \times \mathcal{Y} = \mathbb{R}_+^2$. Exactly as in the standard optimal transport case, the super-hedging problem (1.3.6) has a dual counterpart :

$$P = \sup_{\mathbb{P} \in \mathcal{M}(\mu, \nu)} \mathbb{E}^{\mathbb{P}}[c(S_1, S_2)], \tag{1.3.8}$$

where

$$\mathcal{M}(\mu, \nu) = \{\mathbb{P} \in \mathcal{P}_{\mathbb{R}^2} : S_1 \stackrel{\mathbb{P}}{\sim} \mu, S_2 \stackrel{\mathbb{P}}{\sim} \nu, \mathbb{E}^{\mathbb{P}}[S_2|S_1] = S_1\} \subset \mathcal{P}(\mu, \nu)$$

is the set of martingale measures over the path space $\mathbb{R}_+ \times \mathbb{R}_+$ with given marginals μ, ν . Note that, while $\mathcal{P}(\mu, \nu)$ always contains at least the product measure $\mu \otimes \nu$, $\mathcal{M}(\mu, \nu)$ can be empty, but there is a useful characterization of the non-emptiness of $\mathcal{M}(\mu, \nu)$.

Theorem 1.3 (Strassen '65). *Let μ and ν be two measures on \mathbb{R} . Assume $\int |x|\mu(dx) + \int |y|\nu(dy) < \infty$, then $\mathcal{M}(\mu, \nu) \neq \emptyset$ if and only if $\mu \leq \nu$ for the convex order of probability measures, i.e.*

$$\mu(f) \leq \nu(f)$$

for all convex functions f .

The existence of a maximizer \mathbb{P}^* for (1.3.8) is relatively easy to prove. Assuming that the measures μ and ν in (1.3.8) have finite first moments and that the cost function c has sub-linear growth, it is not difficult to see that $\mathcal{M}(\mu, \nu)$ is a closed subset of $\mathcal{P}(\mu, \nu)$, so that $\mathcal{M}(\mu, \nu)$ is weakly compact, too. The existence of a maximizer \mathbb{P}^* then follows from the continuity of the map $\mathbb{P} \mapsto \mathbb{E}^{\mathbb{P}}[c(S_1, S_2)]$ over $\mathcal{M}(\mu, \nu)$.

More importantly, we have the martingale counterpart of the Kantorovich duality (1.3.5).

Theorem 1.4 (Duality for martingale optimal transport – Beiglböck, Henry-Labordère and Penkner [10]). *Assume the payoff c is upper semi-continuous and satisfies $c(s_1, s_2) \leq a(s_1) + b(s_2)$ for some $a \in L^1(\mu)$ and $b \in L^1(\nu)$. Then*

$$D = P = \mathbb{E}^{\mathbb{P}^*} [c(S_1, S_2)] \tag{1.3.9}$$

for some $\mathbb{P}^* \in \mathcal{M}(\mu, \nu)$.

This ground-breaking result has been followed by a large number of further contributions, studying the extension of the problem to the continuous-time framework or to more general super-hedging strategies. In Chapter 4 of this manuscript, we will consider the application of the tools outlined above to a problem of option hedging in the VIX market, which will lead us to define a martingale optimal transport problem similar to (1.3.8), but with an additional constraint on top of the martingale condition.

Première partie

**Multilevel Monte Carlo methods for
nested risk computations and option
pricing**

Multilevel Monte Carlo methods for nested risk computations

Contents

2.1	Nested expectations	17
2.1.1	Nested Monte Carlo estimator	18
2.2	Multilevel method	18
2.2.1	A further improvement : the antithetic Multilevel estimator	19
2.2.2	Irregular functions g	21
2.2.2.1	The case of step functions $g(x) = 1_{x>a}$	21
2.2.3	Our motivation : derivative pricing in the presence of Initial Margin valuation adjustments	22
2.3	Our assumptions and results for the nested expectation problem	24
2.3.1	Numerical experiments	26

In this chapter, we study the efficient implementation of some non-linear nested Monte Carlo problems, involving non-smooth functions of conditional expectations, that arise in the assessment of risks related to derivative portfolios. Our main motivating example comes from a problem of derivative pricing in the presence of Initial Margin valuation adjustments. The results presented in this chapter are taken from our article Bourgey, De Marco, Gobet and Zhou [21].

We return to the problem of efficient estimation of expectations introduced in section 1.1. In addition to the approximate simulation of SDEs via time discretization discussed in section 1.1.1, another important example in quantitative finance is the evaluation of expectations of functions of conditional expectations, that is the evaluation of

$$I = \mathbb{E}\left[g(\mathbb{E}[f(X, Y) | X])\right], \quad (2.0.1)$$

where $f : \mathbb{R}^d \times \mathbb{R}^{d'} \rightarrow \mathbb{R}$ and $g : \mathbb{R} \rightarrow \mathbb{R}$ are given functions, and X, Y are random variables. We will focus on the situation where X and Y are independent; several computational problems in finance can be cast under this form, see Examples 2.1 and 2.2 below.

Generally speaking, the evaluation of (2.0.1) is a two step problem : it requires the evaluation of (i) the inner conditional expectation $\mathbb{E}[f(X, Y) | X]$, and (ii) of the outer expectation $\mathbb{E}[g(\cdot)]$.

2.1 Nested expectations

Nested expectations of the form (2.0.1) are typical in pricing problems for American-type derivatives, but they also arise in a large variety of risk valuation problems (VaR or CVaR of a portfolio). In our work [21], the analysis of estimators for (2.0.1) was motivated by the computation of valuation adjustments to the price of a derivative in the presence of Initial Margin costs, as we explain in section 2.2.3.

Example 2.1 (Expected exposure of a derivative portfolio). *Consider a derivative portfolio with initial value f_0 and some time horizon T . Assume $\mathbb{E}[f(X, Y) | X]$ represents the future value of the portfolio, say at some future time $t \leq T$, conditional on the value of some market risk factors X at that moment. Here Y represents the additional noise that will be added to the risk factors between time t and T , for example under the forms of Brownian increments in a stochastic model –supposed to be independent from X . Then*

$$\mathbb{E}\left[\left(f_0 - \mathbb{E}[f(X, Y) | X]\right)^+\right]$$

represents the exposure of the trade at time t . Here $g(z) = (f_0 - z)^+$.

Example 2.2 (Tail distribution of a loss process). *Similar to Example 2.1, assume that $L(X) = \mathbb{E}[f(X, Y) | X]$ represents the loss of a derivative portfolio, conditional on the value of some market factor X . The aim is to evaluate the tail probability*

$$\mathbb{P}(L(X) \geq a) = \mathbb{P}(\mathbb{E}[f(X, Y) | X] \geq a) = \mathbb{E}\left[\mathbf{1}_{\mathbb{E}[f(X, Y) | X] \geq a}\right]$$

Here $g(z) = \mathbf{1}_{z \geq a}$ is an indicator function.

Other similar examples come from risk management problems in insurance and are related for example to stop-loss contracts, where $g(z) = \max(a, z)$.

2.1.1 Nested Monte Carlo estimator

It is rather natural to approximate (2.0.1) with the estimator

$$\hat{I}_{M,n} = \frac{1}{M} \sum_{m=1}^M g\left(\frac{1}{n} \sum_{j=1}^n f(X_m, Y_j^m)\right) \quad (2.1.1)$$

where $(X_m)_{m \geq 1}$ and $(Y_j^m)_{j, m \geq 1}$ are independent i.i.d. families having the distributions of X and Y respectively.

We are going to analyse the mean-squared error and the computation complexity of the estimator (2.1.1). If the function g is sufficiently smooth, namely if $g \in C_b^2$, and the random variable $f(X, Y)$ has finite moments (which will be a standing assumption in this chapter, unless otherwise stated), then we have the classical estimate

$$\text{bias} = \mathbb{E}\left[\hat{I}_{M,n}\right] - I = \mathcal{O}\left(\frac{1}{n}\right). \quad (2.1.2)$$

Since the variance of $\hat{I}_{M,n}$ is of order $\frac{1}{M}$, overall we get

$$\text{MSE} = \mathbb{E}\left[\left(\hat{I}_{M,n} - I\right)^2\right] \leq C \left(\frac{1}{n^2} + \frac{1}{M}\right),$$

for some constant C .

2.2 Multilevel method

The Multilevel method introduced in section 1.1.2 can be applied to the nested expectation problem (2.0.1). As done in (2.1.1), as an approximation of the inner conditional expectation it is natural to consider

$$Z_l = g\left(\frac{1}{n_l} \sum_{j=1}^{n_l} f(X, Y_j)\right),$$

where $n_l = \frac{1}{h_l} = 2^l$; in other words, we decide to double the number of points from a level to the following.

Consequently, in the Multilevel estimator (1.1.2) we can estimate $\mathbb{E}[Z_l - Z_{l-1}]$ by means of the difference

$$Z_l^{(l,m)} - Z_{l-1}^{(l,m)} = g\left(\frac{1}{n_l} \sum_{j=1}^{n_l} f\left(X_m^l, Y_j^{l,m}\right)\right) - g\left(\frac{1}{n_{l-1}} \sum_{j=1}^{n_{l-1}} f\left(X_m^l, Y_j^{l,m}\right)\right)$$

where $(X_m^l)_{l,m \in \mathbb{N}^*}$ and $(Y_j^{l,m})_{j,l,m \in \mathbb{N}^*}$ are two independent families of i.i.d. random variables having the distributions of X and Y respectively. Fully spelled out, the resulting Multilevel estimator reads

$$\begin{aligned} \hat{I}_{\mathbf{M}, \mathbf{n}} &= \frac{1}{M_0} \sum_{m=1}^{M_0} g\left(\frac{1}{n_0} \sum_{j=1}^{n_0} f\left(X_m^0, Y_j^{0,m}\right)\right) \\ &+ \sum_{l=1}^L \frac{1}{M_l} \sum_{m=1}^{M_l} \left\{ g\left(\frac{1}{n_l} \sum_{j=1}^{n_l} f\left(X_m^l, Y_j^{l,m}\right)\right) - g\left(\frac{1}{n_{l-1}} \sum_{j=1}^{n_{l-1}} f\left(X_m^l, Y_j^{l,m}\right)\right) \right\} \end{aligned} \quad (2.2.1)$$

where the number of outer and inner samples are now denoted by multi-indexes : $\mathbf{M} = (M_0, \dots, M_L)$, $\mathbf{n} = (n_0 < \dots < n_L)$. $\hat{I}_{\mathbf{M}, \mathbf{n}}$ is a combination of nested MC estimators of the form (2.1.1).

Complexity analysis (for smooth functions g). The cost for one simulation of $Z_l = g\left(\frac{1}{n_l} \sum_{j=1}^{n_l} f(X, Y_j)\right)$ is

$$\text{cost}(Z_l) = n_l \quad \rightsquigarrow \gamma = 1$$

We have already seen in (2.1.2) that, if $g \in C_b^2$, the bias of Z_l behaves as

$$\mathbb{E}[Z_l - Z] = \mathcal{O}\left(\frac{1}{n_l}\right) \quad \rightsquigarrow \alpha = 1$$

and the variance at level l as

$$V_l = \text{Var}(Z_l - Z_{l-1}) = \mathcal{O}\left(\frac{1}{n_l}\right) \quad \rightsquigarrow \beta = 1$$

An application of Theorem 1.1 yields a Multilevel estimator (2.2.1) achieving $\text{MSE} = \mathcal{O}(\epsilon^2)$ with computational cost

$$\mathcal{C}_{\text{ML}} = Mn = \mathcal{O}((\log \epsilon)^2 \epsilon^{-2}).$$

It can also be seen (by inspection of the solution of the optimization problem (1.1.3)) that such an estimator can be constructed setting $n_0 = \mathcal{O}(1)$, $L = \left\lceil \frac{-\log(n_0 \epsilon)}{\log(2)} \right\rceil$, $M_0 = \mathcal{O}(-\log(\epsilon) \epsilon^{-2})$ and $M_l = M_0 2^{-l}$ (note that, in this particular case, the product $M_l n_l$ is independent of l). We recognize a typical feature of Multilevel estimators : the optimal choice is to allocate a larger part M_l of the simulation effort on coarser levels (corresponding to small value of n_l) and progressively smaller simulation effort on finer levels.

2.2.1 A further improvement : the antithetic Multilevel estimator

In some situations, that include our nested expectation problem (2.0.1), the complexity of the Multilevel method can actually be nailed down to the asymptotically optimal complexity $\mathcal{O}(\epsilon^{-2})$ – a rather striking result.

This of course requires to modify the construction of the ML estimator. If we are able to find and simulate random variables $(Z_{l,l-1})_{l=1,;\dots,L}$ such that $\mathbb{E}[Z_{l,l-1}] = \mathbb{E}[Z_l - Z_{l-1}]$, it is of course still true that

$$\mathbb{E}[Z_L] = \mathbb{E}[Z_0] + \sum_{l=1}^L \mathbb{E}[Z_{l,l-1}].$$

On the other side, if $Z_{l,l-1}$ has smaller variance than $Z_l - Z_{l-1}$, the performance of the corresponding ML estimator will be improved.

For the nested expectation problem, such an approximation is offered by

$$Z_{l,l-1} = g\left(\frac{1}{n_l} \sum_{j=1}^{n_l} f(X, Y_j)\right) - \frac{1}{2} \left(g\left(\frac{1}{n_{l-1}} \sum_{j=1}^{n_{l-1}} f(X, Y_j)\right) + g\left(\frac{1}{n_{l-1}} \sum_{j=n_{l-1}+1}^{n_l} f(X, Y_j)\right) \right)$$

The resulting estimator, referred to as *antithetic* Multilevel estimator [49], is

$$\begin{aligned} \hat{I}_{\text{antith}}^{\text{ML}} = & \frac{1}{M_0} \sum_{m=1}^{M_0} g\left(\frac{1}{n_0} \sum_{j=1}^{n_0} f(X_m^0, Y_j^{0,m})\right) + \sum_{l=1}^L \frac{1}{M_l} \sum_{m=1}^{M_l} \left\{ g\left(\frac{1}{n_l} \sum_{j=1}^{n_l} f(X_m^l, Y_j^{l,m})\right) \right. \\ & \left. - \frac{1}{2} \left(g\left(\frac{1}{n_{l-1}} \sum_{j=1}^{n_{l-1}} f(X_m^l, Y_j^{l,m})\right) + g\left(\frac{1}{n_{l-1}} \sum_{j=n_{l-1}+1}^{n_l} f(X_m^l, Y_j^{l,m})\right) \right) \right\}, \quad (2.2.2) \end{aligned}$$

which is rather long to write, but still straightforward to simulate.

Complexity analysis (for smooth functions g). Still assuming $g \in C_b^2$, it is easy to see that the magnitude of the cost for constructing one simulation of $Z_{l,l-1}$ and the overall bias are unchanged : we still have

$$\text{cost}(Z_{l,l-1}) = n_l \quad \rightsquigarrow \gamma = 1$$

and, by construction,

$$\mathbb{E} \left[\hat{I}_{\text{antith}}^{\text{ML}} - Z \right] = \mathbb{E}[Z_L - Z - Z] = \mathcal{O}\left(\frac{1}{n_L}\right) \quad \rightsquigarrow \alpha = 1$$

With respect to the original ML estimator (2.2.1), the variance at level l is now improved from $\mathcal{O}(1/n_l)$ to

$$V_l = \text{Var}(Z_{l,l-1}) = \mathcal{O}\left(\frac{1}{n_l^2}\right) \quad \rightsquigarrow \beta = 2$$

as detailed for example in Giles [49].

According to Theorem 1.1, the antithetic ML estimator (2.2.2) now can achieve $\text{MSE} = \mathcal{O}(\epsilon^2)$ with computational cost

$$\mathcal{C}_{\text{antith}} = \mathcal{O}(\epsilon^{-2}).$$

Summing up, following the multivel construction it is possible (under some conditions on the function g and the random variable $f(X, Y)$, that we will precisely state in the following sections) to combine biased estimators in such a way to achieve the asymptotical complexity of an unbiased estimator.

2.2.2 Irregular functions g

In several practical applications, the function g is not smooth everywhere. In the examples 2.2 and 2.2, we have encountered functions of the form

$$g(x) = 1_{x>a}, \quad g(x) = \max(x, a).$$

Of course, the regularity of g impacts both the bias and the inner variance of Multilevel estimators, and therefore the optimal tuning of the parameters L and M_l . In order to estimate the bias and overall the performance of such estimators, there is a tradeoff between the smoothness of the function g and that of the probability distribution of the underlying random variables – the less regularity on g , the stronger the requirements we have to make on the underlying distribution.

2.2.2.1 The case of step functions $g(x) = 1_{x>a}$

In the existing literature, efforts have been made to evaluate the nested expectation I in (2.0.1) in the case of limited regularity, where g is a step function $g = 1_{[a, \infty)}$. This setting typically appears when evaluating tail distributions and quantiles, as in Example 2.2. To our knowledge, one of the earliest works dealing with this problem using nested Monte-Carlo methods is the seminal paper of Gordy and Juneja [53]. In [53], it is assumed that for $n \geq 1$, the couple of random variables $\left(\mathbb{E}[f(X, Y)|X], \sqrt{n} \left(\frac{1}{n} \sum_{j=1}^n f(X, Y_j) - \mathbb{E}[f(X, Y)|X]\right)\right)$ has a joint density g_n with respect to the two-dimensional Lebesgue measure, and that the partial derivatives $\partial_y g_n(y, \cdot)$ and $\partial_y^2 g_n(y, \cdot)$ admit finite moments up to order four, uniformly in n . Assumptions of this kind may look rather strong, and overall do not seem obvious to check : notably, a control on the moments of the joint law that is uniform over n does not seem easy to achieve. Under these conditions, the authors obtain the following behavior of the bias of $Z_l = g\left(\frac{1}{n_l} \sum_{j=1}^{n_l} f(X, Y_j)\right)$:

$$\mathbb{E}[Z_l] - \mathbb{E}[Z] = \mathcal{O}\left(\frac{1}{n_l}\right) \quad \rightsquigarrow \quad \alpha = 1.$$

Considering that the strong error behaves as

$$\|Z_l - Z_{l-1}\| = \mathcal{O}\left(\frac{1}{n_l^\beta}\right) \quad \text{with } \beta < 1/2,$$

we are in case $\beta < \gamma = 1$ of Theorem 1.1, so that a Multilevel estimator has at best complexity $\mathcal{O}(\epsilon^{-2-(1-\beta)})$ – which is worse than $\mathcal{O}(\epsilon^{-5/2})$ – for a target MSE of order $\mathcal{O}(\epsilon^2)$.

The analysis of the bias is pushed forward in the work of Giorgi, Lemaire and Pagès [51], where a higher order expansion is derived. In order to do so, the authors have to assume stronger regularity conditions on the densities of the underlying random variables : in [51], it is assumed that both couples of random variables $\left(\mathbb{E}[f(X, Y)|X], \frac{1}{n} \sum_{j=1}^n f(X, Y_j) - \mathbb{E}[f(X, Y)|X]\right)$ and $(\mathbb{E}[f(X, Y)|X], X)$ admit smooth densities on \mathbb{R}^2 for $n \geq 1$, and again that a certain number of their partial derivatives with respect to the first variable exist and are continuous¹. As a result of the bias expansion, the performance of the ML estimator can be improved applying the weighted Multilevel method of Lemaire and Pagès [77].

As a comparison with our setting, our Assumption 2.3 does not require existence of densities and their derivatives and therefore encompasses a more general framework than the one allowed in [53].

¹More precisely, denoting respectively p_1 and p_2 the two densities, in order to derive a bias expansion at order $R \in \mathbb{N}$, [51] assume existence of the partial derivatives $\partial_x^{(l)} p_1(x, y)$ for $l = 0, \dots, 2R + 1$, $\partial_x^{(l)} p_2(x, y)$ for $l = 0, \dots, 2R$, and that $\partial_x^{(2R+1)} p_1(x, y)$ is continuous.

2.2.3 Our motivation : derivative pricing in the presence of Initial Margin valuation adjustments

The linear risk-neutral pricing approach to the valuation of financial derivatives has undergone important modifications in the last years, under the influence of market regulators : for several type of trades, investment banks have to post collateral in an account supervised by a central counterparty (CCP, also called clearing house) in order to secure their positions. Every day, the CCP requires the involved banks to deposit a certain capital according to the risk exposure of their contracts. From the modeling point of view, taking into account this type of regulatory capitals in the valuation of a derivative trade gives rise to non-linear (backward stochastic- or partial-) differential equations.

One of these protection capitals is the Initial Margin (IM) deposit : in case of default of one of the CCP members, the aim of this capital is to cover potential losses experienced during the liquidation period of the defaulted member – concretely, the IM is materialized by the Value-at-risk or Conditional value-at-risk (CVaR) of the member's hedging portfolio over a time period Δ . To be more precise, let us consider the following mathematical setting : consider an option with payoff $\Phi(S_T)$ and assume the stochastic model $dS_t = \mu_t S_t dt + \sigma_t S_t dW_t$ for the underlying asset price, where $\sigma_t = \sigma(t, S_t)$. Taking into account the Initial Margin adjustment, the option delta-hedging portfolio has the dynamical formulation

$$dV_t^{\text{IM}} = r(V_t^{\text{IM}} - \delta_t^{\text{IM}} S_t) dt - R \text{CVaR}_\alpha(V_t - V_{t+\Delta} | \mathcal{F}_t) dt + \delta_t^{\text{IM}} dS_t, \quad (2.2.3)$$

where the Initial Margin cost is represented here by the conditional value-at-risk CVaR_α of the future portfolio increment over a time window Δ (typical reference values are $\Delta \approx 1$ day or $\Delta \approx 1$ week), r is a risk-free rate, assumed constant for simplicity, and R is the net interest rate of the account used to fund the IM cost. Note that equation (2.2.3) is a Backward Stochastic Differential Equation (BSDE) for the couple $(V_t^{\text{IM}}, \delta_t^{\text{IM}})$ containing a dependence with respect to the future law of the solution V_t^{IM} ; existence and uniqueness of solutions for such an equation are discussed in [3].

The simulation of the full equation (2.2.3) is not an easy task ; in view of numerical computations, since the time window Δ is small when expressed in annualized units, a possible approach is to apply an asymptotic expansion for the solution of the BSDE as Δ becomes small, as in Henry-Labordère [65, Section 4.2] and Agarwal et al. [3]. Let us briefly present the approximation procedure followed in [65] and [3]. When $\Delta = 0$ (that is : no IM correction), the option price at time t is given by the classical risk-neutral valuation price $\mathbb{E}[e^{-r(T-t)} \Phi(S_T) | S_t]$, with first derivative $\delta(t, S) = \partial_s \mathbb{E}[e^{-r(T-t)} \Phi(S_T) | S_t = s]$, where the expectation is computed under a probability measure such that $dS_t = r_t S_t dt + \sigma_t S_t dW_t$. Heuristically, for small values of Δ , the portfolio increment can be approximated by its delta component (otherwise said, by the increment of the stochastic integral part)

$$\begin{aligned} V_t^{\text{IM}} - V_{t+\Delta}^{\text{IM}} &\approx - \int_t^{t+\Delta} \delta^{\text{IM}}(u, S_u) \sigma_u S_u dW_u \approx -\delta^{\text{IM}}(t, S_t) \sigma_t S_t (W_{t+\Delta} - W_t) \\ &\stackrel{(d)}{\approx} -\delta^{\text{IM}}(t, S_t) \sigma_t S_t \times \sqrt{\Delta} G \end{aligned} \quad (2.2.4)$$

where $G \sim \mathcal{N}(0, 1)$. At the lower order in Δ , (2.2.4) suggests to approximate the IM correction in (2.2.3) with $\text{CVaR}_\alpha(V_t^{\text{IM}} - V_{t+\Delta}^{\text{IM}} | \mathcal{F}_t) \approx \sqrt{\Delta} C_\alpha |\delta^{\text{IM}}(t, S_t)| \sigma_t S_t$, where $C_\alpha := \text{CVaR}_\alpha(\mathcal{N}(0, 1)) = \frac{e^{-\frac{x^2}{2}}}{(1-\alpha)\sqrt{2\pi}} \Big|_{x=N^{-1}(\alpha)}$ is the CVaR of a standard Gaussian random variable. Such an approximation procedure leads to approximate equation (2.2.3) with the non-linear – but now standard – BSDE

$$dV_t = r(V_t - \delta_t S_t) dt - R \sqrt{\Delta} C_\alpha |\delta(t, S_t)| \sigma_t S_t dt + \delta(t, S_t) dS_t, \quad (2.2.5)$$

which can now be tackled with backward dynamical programming and Monte Carlo-based regressions. At the lowest order in Δ , the IM correction to the standard risk-neutral price of the trade is therefore given by

$$RC_\alpha \sqrt{\Delta} \mathbb{E} \left[\int_0^{T-\Delta} e^{-rt} \sigma_t S_t |\delta(t, S_t)| dt \right] =: RC_\alpha \sqrt{\Delta} (T - \Delta) \times I, \quad (2.2.6)$$

where $\delta(t, s) = \partial_s \mathbb{E} \left[\Phi(S_T) \mid S_t = s \right]$ is the reference standard risk-neutral portfolio delta (corresponding to $\Delta = 0$), and the integration upper bound indicates we assume that hedging operations are performed up to time $T - \Delta$. Note that the product of constants $RC_\alpha \sqrt{\Delta} (T - \Delta)$ is fixed once and for all; the computational problem boils down to an efficient evaluation of the expectation $I = \mathbb{E} \left[\frac{1}{T-\Delta} \int_0^{T-\Delta} e^{-rt} \sigma_t S_t |\delta(t, S_t)| dt \right]$.

To conclude this section, let us show that it is possible to cast the term I in (2.2.6) under the form of a nested expectation as (2.0.1). For simplicity, consider the Black-Scholes framework (2.3.4) where $\sigma_t \equiv \sigma$. Using the likelihood ratio method of Broadie and Glasserman [22], we can restore an expression of $\delta(t, S)$ in terms of an expectation

$$e^{(T-t)\delta(t, s)} = \partial_s \mathbb{E} \left[\Phi(S_T) \mid S_t = s \right] = \mathbb{E} \left[(\Phi(S_T) - \Phi(S_t)) H_{t,T} \mid S_t = s \right],$$

where the integration weight $H_{t,T}$, given in the general framework by the integration by parts formula of the Malliavin calculus, corresponds to $H_{t,T} = \frac{W_T - W_t}{T-t}$ in the Black-Scholes case. In the last expression, the conditionally centered term $-\Phi(S_t) H_{t,T}$ we have artificially introduced allows to reduce variance, playing the role of a control variate. Taking advantage of the independence between S_t and the future Brownian increments underpinning the integration weight $H_{t,T}$, it is easy to see that we can express $\mathbb{E}[|\delta(t, S_t)|]$ under the form (2.0.1). As an additional step in the evaluation of I in (2.2.6), instead of discretizing the time integral over $[0, T - \Delta]$ – which would produce a bias – we introduce an independent random variable with uniform distribution over $[0, T - \Delta]$, and set

$$I = \mathbb{E} \left[\frac{1}{T - \Delta} \int_0^{T-\Delta} \sigma S_t |\delta(t, S_t)| dt \right] = (T - \delta) \mathbb{E}[\sigma S_U |\delta(U, S_U)|].$$

Overall, the desired expectation I can be expressed under the form (2.0.1), using independent random variables $Y, Z \sim \mathcal{N}(0, 1)$ and $U \sim \mathcal{U}([0, T - \Delta])$, setting $X = (U, S_0 e^{(r - \frac{\sigma^2}{2})U + \sigma\sqrt{U}Z})$, $g(x) = |x|$ and defining the appropriate function f , see [21, Proposition 3.1] for details.

In the setting above, the function g in (2.0.1) is given by the absolute value function

$$g(x) = |x|.$$

More generally, we can consider a continuous and piecewise C_b^1 function (see our Assumption 2.1 below for precise conditions). With respect to previous works cited in section 2.2.2, we are in a situation of intermediate regularity of the function g – less than C^2 , but more regular than a step function – thus allowing us to drop restrictive assumptions on the distribution of the underlying random variables when studying the nested expectation (2.0.1). Postponing precise conditions to the next section, it will be sufficient for us to only exploit some (mild) regularity of the law of $\mathbb{E}[f(X, Y) \mid X]$ in the neighbourhood of the singularities of g , instead of considering the joint law of $\mathbb{E}[f(X, Y) \mid X]$ and its estimator. In this respect, assumption close to ours are considered in Giles and Haji-Ali [50], where g is a step function and the authors assume that in a neighbourhood of the discontinuity point of the step function, the random variable $\frac{\mathbb{E}[f(X, Y) \mid X]}{\sqrt{\text{Var}(f(X, Y) \mid X)}}$

has a bounded density; the resulting Multi-level estimator achieves mean-squared error ϵ^2 with complexity $\mathcal{O}(\epsilon^{-2}|\log \epsilon|^2)$. Once again, we would like to point out that such conditions on the underlying distributions are stronger than ours. In fact, our setting allows to treat the case of a butterfly option payoff (see Figure 2.1), for which the assumption of a bounded density seems difficult to check, while our milder Assumption 2.3 proves to be checkable in financial meaningful cases.

2.3 Our assumptions and results for the nested expectation problem

As pointed out above, the outer functions g we work with will have an intermediate regularity – not smooth, but more regular than an indicator function.

With reference to the nested expectation problem (2.0.1), our assumptions are the following :

Assumption 2.1. *g is continuous and piecewise \mathcal{C}^1 : precisely, there is a finite set of points*

$$D = \{-\infty = d_0 < d_1 < \dots < d_\theta < d_{\theta+1} = \infty\}$$

such that on every (d_i, d_{i+1}) , g is of class \mathcal{C}^1 , g' is bounded and Hölder continuous for some exponent $\eta \in (0, 1]$. g is continuous, piecewise \mathcal{C}_b^1 and g' is Lipschitz where it exists.

Assumption 2.2. *There exists $p > 2$ such that $\mathbb{E}[|f(X, Y)|^p] < \infty$.*

Assumption 2.3 (Small ball estimate around the singularities). *There exist positive constants ν , K and z_0 such that*

$$\mathbb{P}\left(\text{dist}(E_f(X), D) \leq z\right) \leq K z^\nu, \quad \forall z < z_0, \quad (2.3.1)$$

where $E_f(X) = \mathbb{E}[f(X, Y)|X]$ and $\text{dist}(y, D) := \min_{1 \leq i \leq \theta} |y - d_i|$.

Remark 2.1. *Note that, if the random variable $E_f(X) = \mathbb{E}[f(X, Y)|X]$ has a bounded density p , then Assumption 2.3 trivially holds true with $\nu = 1$, since in this case $\mathbb{P}(\text{dist}(E_f(X), D) \leq z) \leq \sum_{i=1}^{\theta} \mathbb{P}(|E_f(X) - d_i| \leq z) \leq 2\theta \|p\|_\infty z$. On the other hand, Assumption 2.3 is more general : it is stated in terms of the distribution of $E_f(X)$ and does not require existence or regularity of a density for $E_f(X)$.*

When $g(x) = |x|$ (the Initial Margin case presented in the previous), the singularity is at $x = 0$. Within a Black-Scholes model, for an option payoff such as a butterfly option – whose delta takes both positive and negative values (see Figure 2.1), we are able to show the following

Lemma 2.1. *For the problem of the Initial Margin correction (2.2.6) to the butterfly option within the model (2.3.4), assumptions 2.1, 2.2, 2.3 hold true with $\eta = 1$, any $p > 2$ and $\nu = \frac{1}{2} \left(1 \wedge \frac{\Delta}{(T-\Delta)(1+A)}\right)$ for any $A > 0$. Therefore, the small ball probability estimate*

$$\mathbb{P}\left(S_t |\delta(t, S_t)| \leq z\right) \leq K_\nu z^\nu$$

holds for the given value of $\nu < 1$, and Theorem 2.1 applies to this setting.

Turning to our main results, when the Assumptions 2.1–2.2–2.3 are in force, in [21] we prove the following bias and variance estimates.

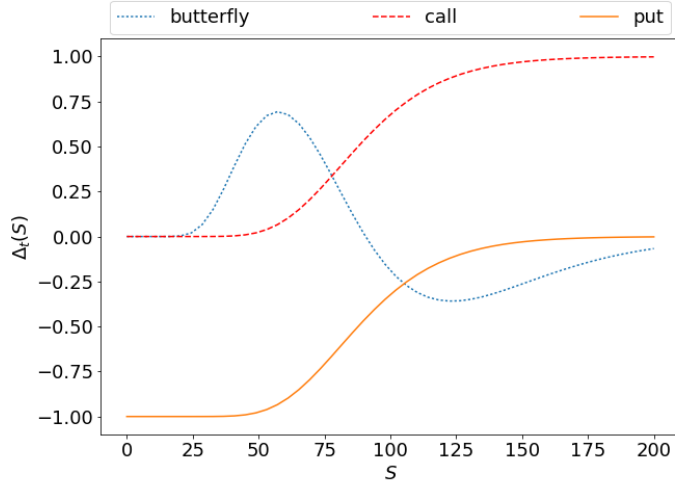


FIGURE 2.1 – Black-Scholes deltas for some standard options. The butterfly option payoff is $(x - (K + a))^+ + (x - (K - a))^+ - 2(x - K)^+$.

Proposition 2.1 (Bias). *The following estimate holds*

$$\left| \mathbb{E} \left[g \left(\frac{1}{n} \sum_{j=1}^n f(X, Y_j) \right) - g(\mathbb{E}[f(X, Y)|X]) \right] \right| \leq \frac{\kappa}{n^{\frac{1}{2} \left(1 + \frac{(p-1)\nu}{p+\nu} \wedge \eta \right)}} \quad \forall n, \quad (2.3.2)$$

for some positive constant κ .

With reference to Theorem 1.1, we have $\alpha = \frac{1}{2} \left(1 + \frac{(p-1)\nu}{p+\nu} \wedge \eta \right)$. Note that the smooth case with no singularities (that is, $g \in C_b^1$ with Lipschitz continuous derivative) corresponds to $\eta = 1$ and $\nu = \infty$. In this case, the exponent for n at the denominator of (2.3.2) becomes $\alpha = \frac{1}{2}(1 + (p-1) \wedge 1)$. If $p \geq 2$, this number is worth $\frac{1}{2} \cdot 2 = 1$. Therefore, in this case Proposition 2.1 provides the standard $\mathcal{O}\left(\frac{1}{n}\right)$ estimate for the bias in the presence of a smooth function g .

Proposition 2.2 (Inner variance for the antithetic ML estimator). *For every $l \geq 1$, the variance at level l is bounded by*

$$V_l = \text{Var}(Z_{l,l-1}) \leq \frac{\tilde{\kappa}}{n_l^{1 + \frac{(p-2)\nu}{2(p+\nu)} \wedge \eta}}, \quad \forall l \geq 1. \quad (2.3.3)$$

for some positive constant $\tilde{\kappa}$.

With reference to Theorem 1.1, we have $\beta = 1 + \frac{(p-2)\nu}{2(p+\nu)} \wedge \eta$. It is worth and important to notice that, under our standing assumptions, this number is always strictly greater than one. Once again, note that in the smooth case (that is : $g \in C_b^1$ with Lipschitz derivative, no singularities), we can take $\eta = 1$ and $\nu = \infty$. The exponent for n_l in (2.3.3) then becomes $\beta = 1 + \frac{p-2}{2} \wedge 1$. If we are allowed to take $p \rightarrow \infty$, this number tends to $1 + 1 = 2$, which corresponds to a $\mathcal{O}\left(\frac{1}{n_l^2}\right)$ estimate of the variance at level l .

We are in position to apply Theorem 1.1 and conclude on global convergence rates.

Theorem 2.1. *Let Assumptions 2.1, 2.2 and 2.3 hold true, and consider an error tolerance $\epsilon > 0$. There exist n_0, M_0, L such that the bound $\text{MSE} = \mathcal{O}(\epsilon^2)$ as $\epsilon \rightarrow 0$ is achieved with a computational complexity*

$\mathcal{O}(\epsilon^{-2})$, with the choice of parameters :

$$n_l = n_0 2^l, \quad M_l = M_0 2^{-\left(1 + \frac{(p-2)\nu}{4(p+\nu)} \wedge \frac{\eta}{2}\right)l}, \quad 0 \leq l \leq L.$$

Remark 2.2. Note also that the optimal number of Monte-Carlo samples at level l is of the form $M_l = M_0 2^{-(1+a)l}$ for a positive constant a , so that the computational cost at level l is proportional to $n_l M_l = n_0 M_0 2^{-al} = \mathcal{O}(\epsilon^{-2}) 2^{-al}$: as usual in the ML framework, the most expensive levels are the coarsest ones. As we have done for Propositions 2.1 and 2.2, it is instructive to consider the case corresponding to $\eta = 1$ in Assumption 2.1. If we can take p to be large in Assumption 2.2, M_l approaches $M_0 2^{-\left(1 + \frac{\nu}{4} \wedge \frac{1}{2}\right)l}$. When there are no singularities, then we can take $\nu \rightarrow \infty$ in the small ball estimate (2.3.1) and obtain $M_l = M_0 2^{-\frac{3}{2}l}$. When there is at least one singularity but we can still take $\nu = 1$, then we obtain $M_l = M_0 2^{-\frac{3}{4}l}$. Overall, Theorem 2.6 tells that, as soon as $\nu > 0$ and $p > 2$, we can make the choice $M_l = M_0 2^{-(1+a)l}$ for some positive a , which is still enough to achieve an overall cost of order $\mathcal{O}(\epsilon^{-2})$ for the ML estimator.

2.3.1 Numerical experiments

The Black-Scholes asset price model

$$S_t = S_0 e^{\left(r - \frac{\sigma^2}{2}\right)t + \sigma W_t}, \quad S_0 > 0, \quad (2.3.4)$$

provides a simple (yet meaningful) setting where we can evaluate explicit reference values (or unbiased estimates) for some of the involved nested expectations, which we can use to benchmark the accuracy of multi-level estimators.

We fix the model parameters in (2.3.4) to $r = 0.1$, $\sigma = 0.3$, and consider four different Portfolios $\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D}$, described below. To estimate a reference value for I in (2.2.6), we use an unbiased Monte-Carlo estimation of $\mathbb{E}[g(E_f(X))]$ (recall that $E_f(X) = \mathbb{E}[f(X, Y)|X]$ is explicit in the examples below) with $n_{\text{MC}} = 5 \times 10^7$ samples, and provide the associated 95%-confidence interval. We denote $\Phi(s, K, a) = (s - (K - a))_+ + (s + (K + a))_+ - 2(s - K)_+$ the butterfly option payoff with strikes $K - a$, K , and $K + a$.

Portfolio \mathcal{A} . We consider one butterfly option with payoff $\Phi(s, 100, 50)$, and set $S_0 = 90$. Here $I_{\mathcal{A}} = \mathbb{E}g(E_f^{\text{butterfly}}(X))$ is worth 10.720 ± 0.002 .

Portfolio \mathcal{B} . We choose Portfolio \mathcal{A} , but now with $S_0 = 30$, which implies $S_0 \ll K = 100$ and therefore the resulting samples will be far from the singularity of g with high probability (see Figure 2.2). Here $I_{\mathcal{B}} = 0.998 \pm 5 * 10^{-4}$.

Portfolio \mathcal{C} . We choose a more diversified portfolio, made of a linear combination of five different butterfly options. The final payoff is of the form

$$2\Phi(s, 10, 1) + 2\Phi(s, 20, 2) + 2\Phi(s, 40, 4) + \Phi(s, 50, 5) + 1.5\Phi(s, 80, 8).$$

In Figure 2.3 one can observe that the Portfolio delta, as a function of s , now has several zeros (playing the role of singularities for $g(z) = |z|$). We set $S_0 = 20$ (close to the singularities) and have $I_{\mathcal{C}} = 0.507 \pm 0.0002$ as a reference value.

Portfolio \mathcal{D} . We choose Portfolio \mathcal{C} , but now setting $S_0 = 100$ (further away from the singularities of the option's delta, see Figure 2.3 again). Here $I_{\mathcal{D}} = 1.263 \pm 0.0004$.

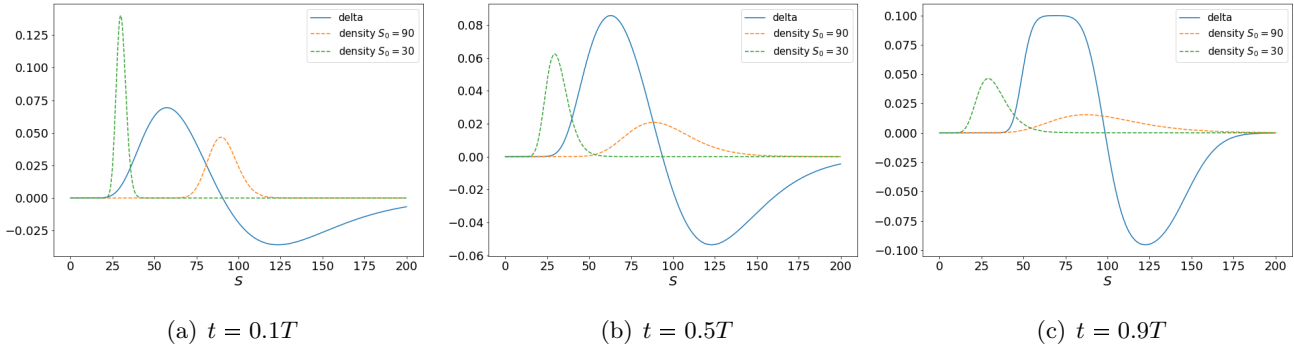


FIGURE 2.2 – Functions $s \rightarrow p_t(s)$ and $s \rightarrow 0.1 \times \delta_t(s)$ (we scale the delta function for visual reasons) for the Portfolios \mathcal{A}, \mathcal{B} and different values of t .

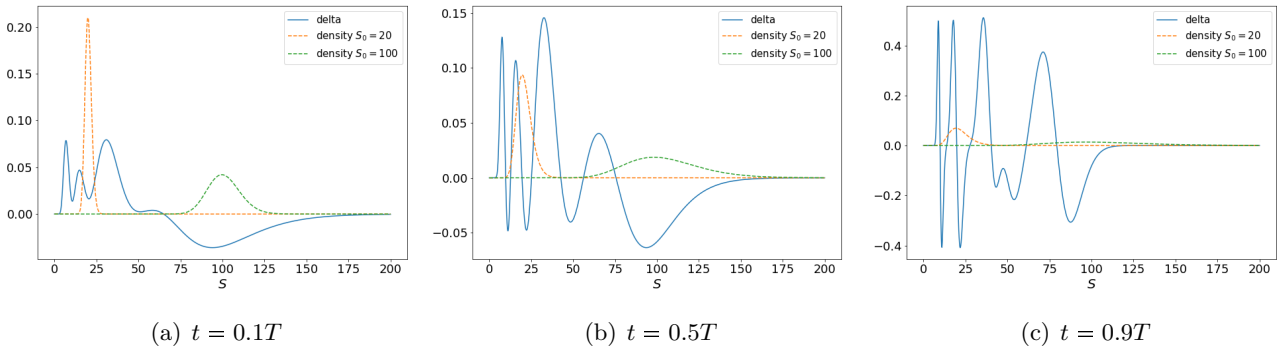


FIGURE 2.3 – Functions $s \rightarrow \delta_t(s)$ and $s \rightarrow p_t(s)$ for the Portfolios \mathcal{C}, \mathcal{D} and different values of t .

In Figures 2.2 and 2.3 we plot the Portfolio delta (as a function of s) for different times t , together with the probability density function of S_t , denoted p_t . Note that in Figure 2.2, we have rescaled the delta so to fit all the functions on the same graph. Observing the support of p_t , we see that in Portfolios \mathcal{A} and \mathcal{C} there is a high probability that $\delta_t(S_t)$ will change sign, as opposed to Portfolios \mathcal{B} and \mathcal{D} .

In Figure 2.4, we plot our estimates of the MSE's of the three different estimators : the plain Nested Monte Carlo estimator (2.1.1), the Multilevel estimator (2.2.1), and the enhanced antithetic Multilevel estimator (2.2.2), in terms of their respective computational costs. We observe that the antithetic ML estimator gives the best results in terms of size of the MSE for a fixed cost for the Portfolios \mathcal{A}, \mathcal{C} and \mathcal{D} , for which we have singularities in the delta. We retrieve a slope close to -1 for the ML estimator, whereas the other estimators show a smaller slope (in absolute value), which is in line with the theoretical results. On the other side, recall that Portfolio \mathcal{B} was constructed in such a way that the probability of visiting the zeros of the delta function is very small, so that the function g essentially does not introduce any bias. In this case, the three estimators provide similar results (and the antithetic estimator ML actually does not display the best results anymore) : the use of a multilevel estimator in this setting does not seem to provide concrete advantage with respect to a plain nested Monte Carlo procedure. In all the other situations, the antithetic MLMC estimator has the best performance.

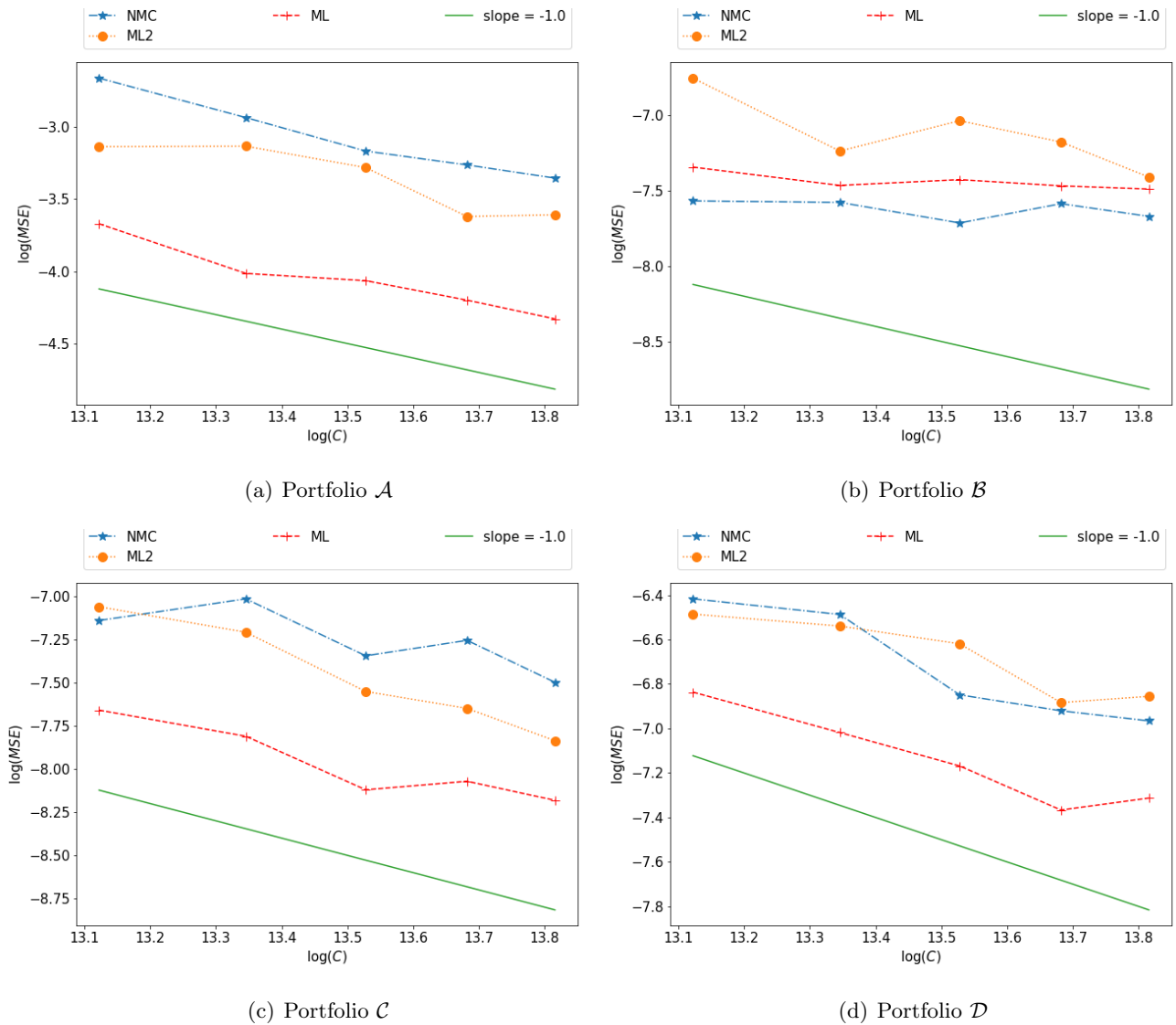


FIGURE 2.4 – $\log(\text{MSE})$ against \log of the computational cost for the Portfolios $\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D}$. ML stands for the antithetic Multi-level estimator (2.2.2), ML2 for the non-antithetic ML estimator (2.2.1), and NMC for the plain nested estimator(2.1.1).

An application to forward variance simulation in the rough Bergomi model

Contents

3.1	The computational problem	29
3.2	Discretization errors	31
3.3	Monte Carlo and Multilevel Monte Carlo simulation	33
3.3.1	Standard Monte-Carlo estimation	33
3.3.2	Multilevel scheme	34
3.3.3	Numerical experiments	35
3.3.4	The additional contribution of an efficient control variate	37

In this chapter, we consider the application of the Multilevel method to the simulation of forward variances (that is : the prices of log-contracts, see section 1.2) in a stochastic model within the family (1.2.10), the rough Bergomi model. The simulation of such objects requires the discretization of a stochastic process (the forward variance curve), hence introducing a bias. We will see that, by means of the multilevel methodology, the significant complexity of the basic Monte Carlo estimator (between $\mathcal{O}(\varepsilon^{-4})$ and $\mathcal{O}(\varepsilon^{-3})$ according to the involved discretization scheme) can be reduced to the optimal complexity $\mathcal{O}(\varepsilon^{-2})$ of an unbiased estimator with mean squared error of order $\mathcal{O}(\varepsilon^2)$. The results presented in this chapter are taken from our work Bourgey and De Marco [18].

3.1 The computational problem

The forward variance modeling framework (1.2.10) has an appealing simplicity : being explicit functions of a Gaussian process, instantaneous forward variances ξ_t^u can be simulated exactly. Nevertheless, this modeling approach still rises some non-trivial computational issues : recall that the instantaneous forward variance ξ_t^u is an unobservable object, while traded derivatives are written on integrated forward variances (1.2.4) or (1.2.5). More specifically, the squared VIX index at time T is

$$(\text{VIX}_T)^2 = \frac{1}{\Delta} \int_T^{T+\Delta} \xi_t^u \, du = \frac{1}{\Delta} \int_T^{T+\Delta} e^{X_T^u} \, du, \tag{3.1.1}$$

where we have defined

$$X_t^u := \log(\xi_t^u) = X_0^u - \frac{1}{2} \int_0^t K(u, s)^2 \, ds + \int_0^t K(u, s) \, dW_s. \tag{3.1.2}$$

In any forward variance model such as (1.2.10), then, the VIX is given by a continuous sum of log-normal random variables ; the resulting distribution of VIX_T is unknown (apart from the trivial case where $K \equiv 0$)

and the random variable VIX_T cannot be simulated exactly. Within this modeling framework, the price at time zero of an option on the VIX^2 with payoff φ is given by

$$\mathbb{E} [\varphi (VIX_T^2)] = \mathbb{E} \left[\varphi \left(\frac{1}{\Delta} \int_T^{T+\Delta} e^{X_T^u} du \right) \right]. \quad (3.1.3)$$

When $\varphi(x) = (\sqrt{x} - \kappa)_+$ (resp. $\varphi(x) = (\kappa - \sqrt{x})_+$), (3.1.3) provides the price of a VIX call option (resp. put option) with strike κ and maturity T .

The problem (3.1.3) is close in spirit to the valuation of Asian options with payoff $\varphi(A_T) = \varphi \left(\frac{1}{T} \int_0^T Y_t dt \right)$, where Y represents an asset price. There is a vast literature on the numerical methods that can be used to evaluate $\mathbb{E}[\varphi(A_T)]$ when Y is modeled by a diffusion process. Yet, there is a fundamental and structural difference with the pricing of a VIX option : while the Asian option payoff is based on the integral of a continuous-time Markov process with respect to its time parameter, there is, in general, no finite-dimensional Markovian structure underlying the problem (3.1.3), and this is precisely the case when K is a fractional kernel as in the rough Bergomi model (1.2.12). In particular then, numerical approaches based on PDEs, explored in the context of Asian options by [80, 89] and [33], are arguably out of the way (in a setting of finite dimensional PDEs at least).

Instead, in this chapter we are going to consider Monte Carlo-based pricing of VIX derivatives, and notably the application of the Multilevel Monte Carlo method presented in the previous chapter. In this respect, our contribution can be seen as an extension to the setting of rough fractional processes of the work of Ben Alaya and Kebaier [11], who studied the application of the Multilevel method to the problem of Asian option pricing in a geometric Brownian motion model.

From now on (in this chapter), we are going to focus on the rough Bergomi model (1.2.12).

Time discretization. Given a grid with $n \in \mathbb{N}^*$ time steps

$$\mathcal{G}_n := \{u_i := T + i h, \quad i = 0, \dots, n\}, \quad h := \frac{\Delta}{n}, \quad (3.1.4)$$

it is clear that the $n + 1$ -dimensional random vector $(X_T^{u_i})_{i=0, \dots, n}$ has Gaussian distribution $\mathcal{N}(\mu, C)$ with mean $\mu = (\mu(u_i))_{0 \leq i \leq n}$ and covariance matrix $C = (C(u_i, u_j))_{0 \leq i, j \leq n}$ given by

$$\mu(u_i) := \mathbb{E} [X_T^{u_i}] = X_0^{u_i} - \frac{1}{2} \int_0^T K(u_i, s)^2 ds = X_0^{u_i} - \frac{\eta^2}{4H} \left(u_i^{2H} - (u_i - T)^{2H} \right), \quad (3.1.5)$$

and

$$C(u_i, u_j) := \text{Cov} (X_T^{u_i}, X_T^{u_j}) = \int_0^T K(u_i, s) K(u_j, s) ds = \eta^2 \int_0^T (u_i - s)^{H-\frac{1}{2}} (u_j - s)^{H-\frac{1}{2}} ds. \quad (3.1.6)$$

The diagonal terms of the covariance matrix can be computed explicitly, $C(u_i, u_i) = \frac{\eta^2}{2H} \left(u_i^{2H} - (u_i - T)^{2H} \right)$, while the off-diagonal terms can be expressed in terms of hypergeometric functions¹, which can be efficiently evaluated offline using standard numerical libraries (we used the function `scipy.special.hyp2f1` from the `scipy` library [84] in our numerical tests).

¹More precisely, if $i < j$, we have

$$C(u_i, u_j) = \frac{\eta^2 (u_j - u_i)^{H-\frac{1}{2}}}{H + \frac{1}{2}} \left[u_i^{2H} {}_2F_1 \left(\frac{1}{2} - H, \frac{1}{2} + H; \frac{3}{2} + H; -\frac{u_i}{u_j - u_i} \right) - (u_i - T)^{2H} {}_2F_1 \left(\frac{1}{2} - H, \frac{1}{2} + H; \frac{3}{2} + H; -\frac{u_i - T}{u_j - u_i} \right) \right].$$

where ${}_2F_1(\cdot, \cdot; \cdot)$ is the hypergeometric function, see [79, Chapter 15].

An approximate simulation of the VIX random variable – and more generally of any integrated forward variance $V_t^{T_1, T_2} = \frac{1}{T_2 - T_1} \int_{T_1}^{T_2} e^{X_t^u} du$ – can be obtained via discretization of the time integral (3.1.1). We consider the following simulation schemes

Definition 3.1 (Discretization schemes for the VIX_T^2). *The right-point rectangle scheme based on the grid \mathcal{G}_n is given by*

$$\text{VIX}_T^{2, \mathcal{R}_n} := \frac{1}{n} \sum_{i=1}^n e^{X_T^{u_i}}, \quad (3.1.7)$$

and the trapezoidal scheme by

$$\text{VIX}_T^{2, \mathcal{T}_n} := \frac{1}{2n} \sum_{i=1}^n \left(e^{X_T^{u_i}} + e^{X_T^{u_{i-1}}} \right). \quad (3.1.8)$$

3.2 Discretization errors

In view of the application of the Multilevel method to the estimation of (3.1.3), we need to rely on asymptotic estimates of the strong and weak error associated to the discretization schemes in Definition 3.1.

Proposition 3.1 (L_p strong error; a slight extension of Proposition 2 in [67]). *Assume the initial forward variance curve $u \mapsto \xi_0^u = e^{X_0^u}$ is locally bounded. Then, for any $T > 0$ and $p > 0$,*

$$\left(\mathbb{E} \left[\left| \text{VIX}_T^2 - \text{VIX}_T^{2, \mathcal{R}_n} \right|^p \right] \right)^{\frac{1}{p}} = \mathcal{O} \left(\frac{1}{n} \right), \quad (3.2.1)$$

$$\left(\mathbb{E} \left[\left| \text{VIX}_T^2 - \text{VIX}_T^{2, \mathcal{T}_n} \right|^p \right] \right)^{\frac{1}{p}} = \mathcal{O} \left(\frac{1}{n^{1+H}} \right), \quad (3.2.2)$$

as $n \rightarrow \infty$.

We note that Proposition 3.1 is not limited to $H \in (0, 1/2)$ – it holds for any value of $H \in (0, 1)$ in the rough Bergomi model (1.2.12) – even if $H < 1/2$ corresponds to the more interesting situation in practice. Also note that, in general, the initial condition $(\xi_0^u)_{T \leq u \leq T+\Delta}$ for the instantaneous forward variance curve is an arbitrary function; in practical applications, ξ_0^u is bootstrapped from options data, or possibly from the VIX futures term-structure (according to the use the stochastic model is intended for). In any of these cases, the regularity of the curve $u \mapsto \xi_0^u$ is a user-based choice – typically, it is set to be either a continuous or a piece-wise constant function. In any case, we find it is more than enough to assume that the function ξ_0^u is locally bounded.

As a refinement of Proposition 3.1, we also provide the exact asymptotics of the L^2 strong error for the rectangle scheme, which shows that the strong rate of convergence is precisely $\frac{1}{n}$.

Theorem 3.1 (Exact asymptotics for the L^2 strong error, rectangle scheme). *Assume $H \in (0, \frac{1}{2})$ and that the initial instantaneous forward variance curve is kept constant over the VIX time window, that is $\xi_0^u = \xi_0$ for all $u \in [T, T + \Delta]$. Then, as $n \rightarrow \infty$,*

$$\left(\mathbb{E} \left[\left| \text{VIX}_T^2 - \text{VIX}_T^{2, \mathcal{R}_n} \right|^2 \right] \right)^{1/2} \sim \Lambda(\xi_0, T, \Delta, H) \frac{1}{n} \quad (3.2.3)$$

where

$$\Lambda(\xi_0, T, \Delta, H) := \frac{\xi_0}{2} \left(e^{\eta^2 \frac{T^{2H}}{2H}} + e^{\eta^2 \frac{(T+\Delta)^{2H} - \Delta^{2H}}{2H}} - 2e^{\eta^2 \int_0^T t^{H-\frac{1}{2}} (\Delta+t)^{H-\frac{1}{2}} dt} \right)^{1/2}.$$

Let us note that, in the proofs of Proposition 3.1 and Theorem 3.1 that we provide in [18], we exploit the explicit Gaussian character of the process $u \mapsto X_T^u$, which of course simplifies the analysis. On the other side, what is less straightforward is to handle the fractional kernel $\frac{1}{(u-s)^{1/2-H}}$ and its unbounded character when u approaches T from above and s approaches T from below.

Remark 3.1. *It is interesting to notice that the upper bound $\mathcal{O}\left(\frac{1}{n}\right)$ on the strong rate of convergence of the rectangle scheme does not depend on the fractional index H ; the dependence with respect to H is only seen at the level of a more accurate quadrature method such as the trapezoidal scheme in (3.2.2). For comparison, we note the strong discretization error for the spot asset price $dS_t = S_t \sqrt{V_t} dB_t$ is highly dependent on the value of H in a model with instantaneous variance V_t with Hölder regularity index H . Neuenkirch and Shalako [78] show that an Euler–Maruyama scheme for S based on n discretization points has a strong error of order $\mathcal{O}\left(\frac{1}{n^H}\right)$ when $V_t = e^{X_t}$ and X is a fractional Ornstein–Uhlenbeck process. The more satisfactory strong rate of convergence in Proposition 3.1 and Theorem 3.1 can be related to the higher regularity of the forward variance curve $u \mapsto \xi_T^u$ with respect to the low pathwise regularity of the instantaneous variance process V .*

Other classical estimates for comparison. It can be instructive to compare Proposition 3.1 and Theorem 3.1 with the behavior of the strong error in the discretization of the time integral of other processes in some – more standard – situations. Fix some time horizon T , and set $t_i = i\frac{T}{n}$ for $i \in \{1, \dots, n\}$.

- For the time integral of an Itô process $X_t = X_0 + \int_0^t b_s ds + \int_0^t a_s dW_s$, with, say, bounded coefficients a and b , it is a standard exercise of stochastic calculus to show that

$$\mathbb{E} \left[\left(\int_0^T X_t dt - \frac{T}{n} \sum_{i=1}^n X_{t_i} \right)^p \right]^{1/p} = \mathcal{O} \left(\frac{1}{n} \right).$$

In order to prove the estimate above, one can apply integration by parts on each sub-interval $\int_{t_i}^{t_{i+1}} (X_t - X_{t_i}) dt = - \int_{t_i}^{t_{i+1}} (t_{i+1} - t_i) dX_t$ and then use Itô’s isometry.

- For the Asian option underlying $\frac{1}{T} \int_0^T S_t dt$, when S_t is a log-normal asset price, Lapeyre and Temam [75] show that we still have

$$\mathbb{E} \left[\left(\frac{1}{T} \int_0^T S_t dt - \frac{1}{n} \sum_{i=1}^n S_{t_i} \right)^p \right]^{1/p} = \mathcal{O} \left(\frac{1}{n} \right).$$

Weak error. The discretization bias for the price of a VIX option $\mathbb{E}[\varphi(\text{VIX}_T^2)]$ with Lipschitz payoff φ can, of course, be upper bounded by the L^1 strong error multiplied by the Lipschitz constant of φ ; it is a straightforward consequence of Proposition 3.1 (or Theorem 3.1 for the rectangle scheme) that

$$\left| \mathbb{E} \left[\varphi(\text{VIX}_T^2) - \varphi(\text{VIX}_T^{2, \mathcal{R}_n}) \right] \right| = \mathcal{O} \left(\frac{1}{n} \right), \tag{3.2.4}$$

$$\left| \mathbb{E} \left[\varphi(\text{VIX}_T^2) - \varphi(\text{VIX}_T^{2, \mathcal{T}_n}) \right] \right| = \mathcal{O} \left(\frac{1}{n^{1+H}} \right). \tag{3.2.5}$$

In our numerical experiments in Figure 3.2 we can precisely observe the asymptotic behavior predicted by (3.2.4)-(3.2.5), in the case of VIX call options.

Remark 3.2. *A VIX call option with strike $\kappa > 0$ corresponds to the payoff $\varphi(x) = (\sqrt{x} - \kappa)_+$. In such a case, the function φ is Lipschitz, for it coincides with the integral of its bounded derivative φ' given by $\varphi'(x) = \frac{1}{2\sqrt{x}} \leq \frac{1}{2\kappa}$ for $x > \kappa^2$ and $\varphi'(x) = 0$ for $x < \kappa^2$.*

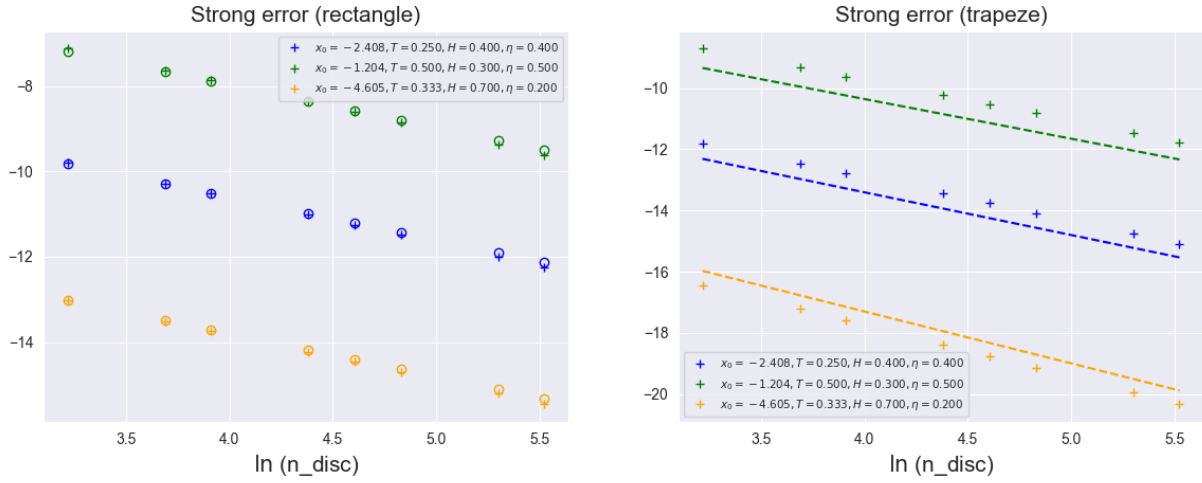


FIGURE 3.1 – Strong error of the discretization schemes for the VIX under the rough Bergomi model. *Left* : L^2 strong error, rectangle scheme VIX_T^{2,\mathcal{R}_n} , Monte Carlo estimate “+” and asymptotic expansion “o”, log-log scale. *Right* : L^2 strong error, trapezoidal scheme VIX_T^{2,\mathcal{T}_n} , Monte Carlo estimate “+” and straight lines with slope $-(1+H)$ for comparison.

Remark 3.3 (Adaptive grids). *Using a non-uniform discretization grid such as*

$$\left\{ u_i := T + \Delta \left(\frac{i}{n} \right)^a, \quad i = 0, \dots, n \right\}, \quad a > 0, \quad (3.2.6)$$

one can improve the asymptotic behavior of the weak error : according to [67, Corollary 1], the weak error of the trapezoidal scheme based on the grid (3.2.6) is $\mathcal{O}(n^{-2})$ instead of $\mathcal{O}(n^{-(1+H)})$ in (3.2.5), provided that $a > \frac{2}{H+1}$. It will be seen in section 3.3.2 that the optimal complexity $\mathcal{O}(\varepsilon^{-2})$ for an estimator with mean-squared error ε^2 can already be attained combining the multilevel method with a discretization scheme based on the uniform grid \mathcal{G}_n in (3.1.4). Appealing to the more complex mesh choice (3.2.6) is therefore not necessary in order to construct estimators of VIX options with optimal rate of convergence.

3.3 Monte Carlo and Multilevel Monte Carlo simulation

3.3.1 Standard Monte-Carlo estimation

Once the simulation scheme VIX_T^{2,\mathcal{D}_n} corresponding to either the rectangle VIX_T^{2,\mathcal{R}_n} or the trapezoidal scheme VIX_T^{2,\mathcal{T}_n} has been fixed, the plain Monte Carlo estimator of the expectation $\mathbb{E}[\varphi(VIX_T^2)]$ is

$$\hat{P}_{M,n} = \frac{1}{M} \sum_{m=1}^M \varphi(VIX_{T,m}^{2,\mathcal{D}_n}), \quad (3.3.1)$$

where \mathcal{D}_n stands for either \mathcal{R}_n (rectangle scheme) or \mathcal{T}_n (trapezoidal scheme), and $(VIX_{T,m}^{2,\mathcal{R}_n})_{1 \leq m \leq M}$ are i.i.d. independent copies of VIX_T^{2,\mathcal{D}_n} .

In the specific case of the rectangle scheme, according to Theorem 3.1, the estimator $\hat{P}_{M,n}$ has mean squared error that satisfies

$$\text{MSE} = \mathbb{E} \left[\left(\mathbb{E}[\varphi(VIX_T^2)] - \hat{P}_M^{\mathcal{R}_n} \right)^2 \right] = \text{bias}^2 + \text{variance} \leq C \left(\frac{1}{n^2} + \frac{1}{M} \right),$$

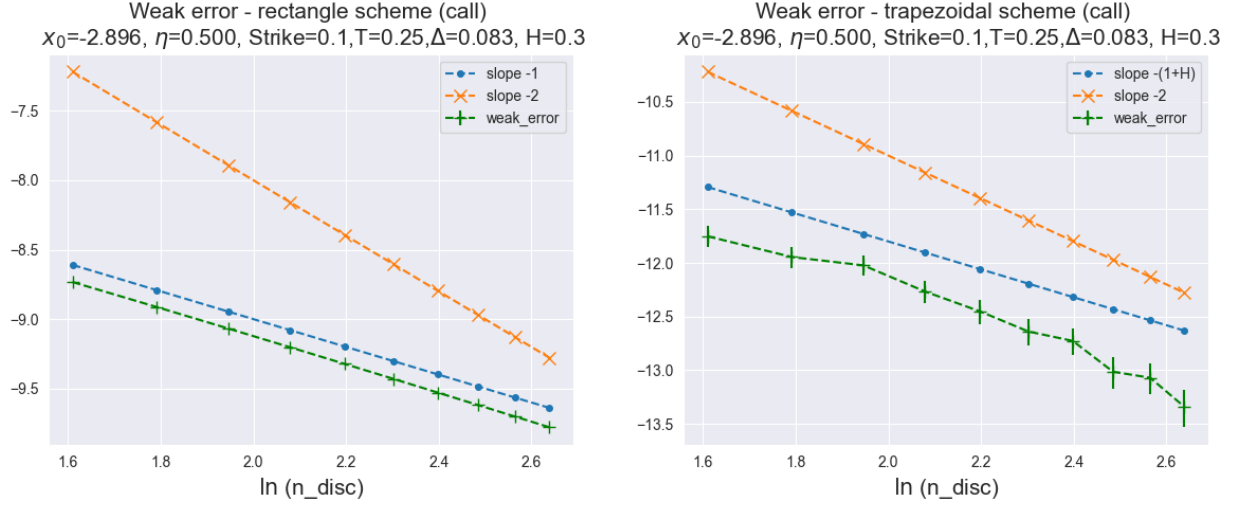


FIGURE 3.2 – Weak error of discretization schemes for the VIX under the rough Bergomi model. *Left* : weak error for a VIX call option, rectangle scheme VIX_T^{2,\mathcal{R}_n} , and straight lines with slopes equal to -1 and -2 for comparison. *Right* : weak error for a VIX call option, trapezoidal scheme VIX_T^{2,\mathcal{T}_n} , and straight lines with slopes $-(1+H)$ and -2 .

assuming that φ is Lipschitz so that (3.2.4) applies. If we want the mean-squared error to satisfy $MSE \leq \varepsilon^2$ for a given accuracy $\varepsilon > 0$, we have to set $M = \mathcal{O}(\varepsilon^{-2})$ and $n = \mathcal{O}(\varepsilon^{-1})$. The construction of VIX_T^{2,\mathcal{R}_n} requires to sample the $(n+1)$ -dimensional Gaussian vector $(X_T^{u_i})_{i=0,\dots,n}$ with given characteristics (3.1.5) and (3.1.6). We consider exact simulation of $(X_T^{u_i})_i$ based on the Cholesky decomposition $L_n L_n^T$ of the covariance matrix (3.1.6). We assume that, for the desired value of n , the required Cholesky matrix L_n has been computed once and for all as offline work, so that sampling a $(n+1)$ -dimensional Gaussian vector requires $\mathcal{O}(n^2)$ operations due to the matrix multiplication $L_n G$, with $G \sim \mathcal{N}(0, Id_n)$. The construction of VIX_T^{2,\mathcal{R}_n} in (3.1.7) from $(X_T^{u_i})_i$ requires an additional sum over n points, at a cost $\mathcal{O}(n)$. Overall, $\hat{P}_{M,n}$ achieves $MSE = \mathcal{O}(\varepsilon^2)$ with

$$\text{Cost}_{\text{MC}}^{\mathcal{R}_n} = (\mathcal{O}(n^2) + \mathcal{O}(n)) \times M = \mathcal{O}(\varepsilon^{-4}),$$

which shows that the plain Monte Carlo estimator $\hat{P}_{M,n}$ is very costly when compared to an unbiased estimator (with cost of order $\mathcal{O}(\varepsilon^{-2})$).

The same analysis for the trapezoidal scheme VIX_T^{2,\mathcal{T}_n} leads to

$$\text{Cost}_{\text{MC}}^{\mathcal{T}_n} = \mathcal{O}\left(\varepsilon^{-2(1+\frac{1}{1+H})}\right).$$

Note that the exponent $2\left(1 + \frac{1}{1+H}\right)$ takes values between -4 (when $H \rightarrow 0$) and -3 (when $H \rightarrow 1$).

3.3.2 Multilevel scheme

We can combine different estimators of the form (3.3.1) following the Multilevel method presented in section 1.1.2. Let $L \in \mathbb{N}^*$, and let $\mathbf{n} = (n_0, \dots, n_L)$ and $\mathbf{M} = (M_0, \dots, M_L)$ be integer multi-indexes representing respectively an increasing sequence of time steps and a sequence of Monte Carlo sample sizes. For the ease of notation, for all $\ell = 0, \dots, L$ we set

$$P_\ell^{\mathcal{D}} := \varphi\left(VIX_T^{2,\mathcal{D}_{n_\ell}}\right).$$

Scheme	Standard MC	Multilevel MC
Rectangle	$\mathcal{O}(\varepsilon^{-4})$	$\mathcal{O}(\ln^2(\varepsilon)\varepsilon^{-2})$
Trapezoidal	$\mathcal{O}\left(\varepsilon^{-2\left(1+\frac{1}{1+H}\right)}\right)$	$\mathcal{O}(\varepsilon^{-2})$

TABLE 3.1 – Summary of the different computational costs for the rectangle and trapezoidal schemes combined with standard and multilevel Monte Carlo. The target MSE is $\mathcal{O}(\varepsilon^2)$.

$P_\ell^{\mathcal{D}}$ is therefore the approximation of the VIX option payoff $\varphi(\text{VIX}_T^2)$ at level ℓ , based on either the rectangle scheme $\mathcal{D}_n = \mathcal{R}_n$ or the trapezoidal scheme $\mathcal{D}_n = \mathcal{T}_n$. Setting

$$n_\ell = n_0 2^\ell, \quad \ell = 0, \dots, L,$$

the construction of the $\text{VIX}_T^{2, \mathcal{D}_{n_\ell}}$ at different levels is straightforward : once we have generated the $(n_\ell + 1)$ -dimensional Gaussian sample $(X_T^{u_i})_{u_i \in \mathcal{G}_{n_\ell}}$ entering into $P_\ell^{\mathcal{D}}$ for some $\ell \geq 1$, we can extract a $(n_{\ell-1} + 1)$ -dimensional Gaussian sample that we use to construct $P_{\ell-1}^{\mathcal{D}}$, by selecting the $X_T^{u_{2k}}$ for $k = 1, \dots, n_0 2^{\ell-1}$.

The multilevel estimator of the VIX option price $\mathbb{E}[\varphi(\text{VIX}_T^2)]$ is therefore given by

$$\widehat{P}_{\mathbf{M}, \mathbf{n}}^{\mathcal{D}} := \frac{1}{M_0} \sum_{m=1}^{M_0} P_0^{\mathcal{D}, (0, m)} + \sum_{\ell=1}^L \frac{1}{M_\ell} \sum_{m=1}^{M_\ell} \left(P_\ell^{\mathcal{D}, (\ell, m)} - P_{\ell-1}^{\mathcal{D}, (\ell, m)} \right), \quad (3.3.2)$$

where, according to the construction detailed in section 1.1.2 : for every ℓ , the random variables $(P_\ell^{\mathcal{D}, (\ell, m)})_{1 \leq m \leq M_\ell}$ and $(P_{\ell-1}^{\mathcal{D}, (\ell, m)})_{1 \leq m \leq M_\ell}$ are independent copies of $P_\ell^{\mathcal{D}}$ and $P_{\ell-1}^{\mathcal{D}}$ respectively, and they are constructed using the same Gaussian samples, according to the procedure described above.

Exploiting the strong and weak error estimates from the previous section, and applying the general Theorem 1.1 on the multilevel method, we can prove that it is possible to construct an estimator (3.3.2) achieving $\text{MSE}_{\mathbf{M}, \mathbf{n}}^{\mathcal{D}} \leq \varepsilon^2$ with computational complexity $\mathcal{O}(\ln(\varepsilon)^2 \varepsilon^{-2})$ when \mathcal{D} is the right-point rectangle scheme and with the optimal complexity rate $\mathcal{O}(\varepsilon^{-2})$ when \mathcal{D} is the trapezoidal scheme.

Theorem 3.1 (Optimal Multilevel estimators for VIX options). *Suppose that the payoff function φ is Lipschitz, and consider $n_0 \in \mathbb{N}^*$. Then, for every tolerance $\varepsilon > 0$ there exist an initial number of samples $M_{0, \mathcal{R}}$ and a number of levels $L_{\mathcal{R}}$ (resp. $M_{0, \mathcal{T}}$ and $L_{\mathcal{T}}$) such that the rectangle (resp. trapezoidal) multilevel estimator $\widehat{P}_{\mathbf{M}, \mathbf{n}}^{\mathcal{R}}$ (resp. $\widehat{P}_{\mathbf{M}, \mathbf{n}}^{\mathcal{T}}$), defined in (3.3.2), has a mean-squared error satisfying $\text{MSE} \leq \varepsilon^2$ and a computational complexity $\mathcal{O}(\ln(\varepsilon)^2 \varepsilon^{-2})$ (resp. $\mathcal{O}(\varepsilon^{-2})$), obtained setting*

$$n_\ell = n_0 2^\ell, \quad M_{\ell, \mathcal{R}} = M_{0, \mathcal{R}} 2^{-2\ell}, \quad \forall \ell = 0, \dots, L_{\mathcal{R}}, \quad (3.3.3)$$

for the rectangle scheme, and respectively

$$n_\ell = n_0 2^\ell, \quad M_{\ell, \mathcal{T}} = M_{0, \mathcal{T}} 2^{-(2+H)\ell}, \quad \forall \ell = 0, \dots, L_{\mathcal{T}}, \quad (3.3.4)$$

for the trapezoidal scheme.

3.3.3 Numerical experiments

In order to compare the four different estimators, namely the plain MC and the MLMC estimators based on the rectangle and the trapezoidal schemes, we price an at-the-money VIX call option with maturity $T = 0.5$

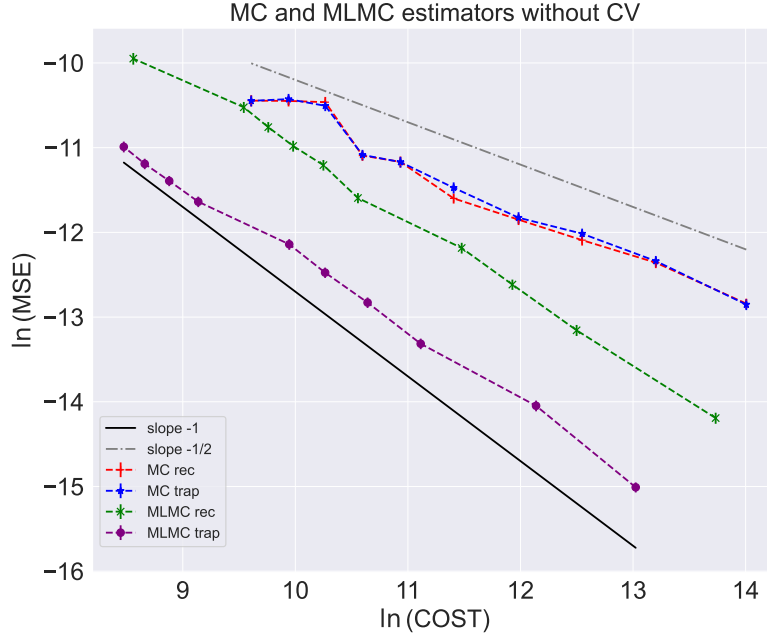


FIGURE 3.3 – Pricing of an ATM VIX call option with different estimators. The graph shows $\ln(\text{MSE})$ against $\ln(\text{Cost})$ for the Monte Carlo and the multilevel estimators based on the rectangle and trapezoidal schemes. The option maturity is $T = 0.5$ and the model parameters are : $H = 0.1$, $\eta = 0.5$, $X_0 = \ln(\xi_0) = \ln(0.235^2)$. The specification “without CV” in the title refers to the fact that the control variate presented in section 3.3.4 has not been used ; we display the performance of the estimators as defined in (3.3.2) and (3.3.1).

in the rough Bergomi model (1.2.12) with parameters $H = 0.1$, $\eta = 0.5$, $X_0 = \ln(\xi_0) = \ln(0.235^2)$. The reference price is computed with an intensive MC simulation over 500 discretization points and 2×10^7 i.i.d. samples. In [18], we also consider the case $\Delta = 1$ (as opposed to $\Delta = \frac{1}{12}$ for the VIX), which corresponds to a call option on the forward volatility $\sqrt{V_T^{T,T+1}}$ overlooking a one-year time window (the results being qualitatively analogous to the VIX case). We estimate the MSE for each method as $\frac{1}{N_{\text{MSE}}} \sum_{j=1}^{N_{\text{MSE}}} (\hat{p}_j - p)^2$ with $N_{\text{MSE}} = 10^3$, where p is the reference price and $(\hat{p}_j)_{1 \leq j \leq N_{\text{MSE}}}$ are N_{MSE} independent copies of either the multilevel or the plain MC estimators. In Figure 3.3, we display the behavior of the resulting MSE (along with the estimated 95%-confidence interval, materialized by the vertical error bars – which are rather small hence barely visible) against the computational cost in a log-log plot. We retrieve the expected asymptotic slopes : from our theoretical analysis, see Table 3.1, we expect a slope close to $-\frac{1}{2}$ for the plain MC estimator with the rectangle scheme, and a slope close to $-\frac{1+H}{2+H}$ for the MC estimator with the trapezoidal scheme. Since $\frac{1+H}{2+H} = \frac{1}{2} + \frac{H}{4} + \mathcal{O}(H^2)$, for small values of H the two slopes are very close, in line with what is observed in Figure 3.3 (recall we chose $H = 0.1$). For the multilevel method, we observe the expected slope close to -1 for both estimators (we actually expect slightly less than -1 for the multilevel method with a rectangle scheme, due to the logarithmic term in $\mathcal{O}(\varepsilon^{-2} \ln^2(\varepsilon))$). One can also see that, for a given cost, the estimated MSE is smaller for the multilevel estimators than for the plain MC estimators.

3.3.4 The additional contribution of an efficient control variate

Due to the structure of the problem, there is an efficient control variate technique (in the spirit of the control variate for Asian option pricing in [73]), already exploited by [67]. Noticing that the discretization schemes $\text{VIX}_T^{2,\mathcal{R}_n}$ and $\text{VIX}_T^{2,\mathcal{T}_n}$ in (3.1.7) and (3.1.8) are given by means of the exponentials $e^{X_T^{u_i}}$, it is reasonable to consider as a control variate the exponential of the mean

$$\overline{\text{VIX}}_T^2 = e^{\frac{1}{n} \sum_{i=1}^n X_T^{u_i}}. \quad (3.3.5)$$

Such a choice is appropriate notably because the parameter u_i is restricted to a small time window $\Delta \approx \frac{1}{12}$. The key ingredient is that $\overline{\text{VIX}}_T^2$ is lognormal : the Gaussian random variable $\frac{1}{n} \sum_{i=1}^n X_T^{u_i} \sim \mathcal{N}(\mu^n, (\sigma^n)^2)$ has explicit characteristics μ^n and $(\sigma^n)^2$ (see [18, section 2.2] for precise expressions). It is probably worth to point out that a quantitative analysis of the difference between the true VIX random variable VIX_T^2 and the control variate (3.3.5) will precisely be the object of a second study that we present in Chapter 5.

In practice, options on forward variance can be priced exploiting the additional contribution of the control variate (3.3.5). When considering the right-point rectangle scheme (3.1.7), we estimate option prices on the VIX^2 with the controlled estimator

$$\frac{1}{M} \sum_{m=1}^M \left[\varphi \left(\frac{1}{n} \sum_{i=1}^n e^{X_T^{u_i,m}} \right) - \varphi \left(e^{\frac{1}{n} \sum_{i=1}^n X_T^{u_i,m}} \right) \right] + \text{CV}_n, \quad (3.3.6)$$

where $(X_T^{u_1,m}, \dots, X_T^{u_n,m})_{1 \leq m \leq M}$ are M independent Monte Carlo samples of the Gaussian vector $(X_T^{u_1}, \dots, X_T^{u_n})$ and $\text{CV}_n := \mathbb{E} \left[\varphi \left(e^{\frac{1}{n} \sum_{i=1}^n X_T^{u_i}} \right) \right]$ is explicitly given by a Black-Scholes formula (see [18, section 2.2]). For the trapezoidal scheme (3.1.8), we use the same control variate.

In order to assess numerically the additional error reduction, we price the at-the-money call option considered in the previous section combining the MC and MLMC estimator with the control variate. The controlled MC estimator is given by (3.3.6) with $\varphi(x) = (\sqrt{x} - \kappa)_+$. In order to build an enhanced MLMC estimator, we inject the right-point rectangle control variate (3.3.5) at each level $0 \leq \ell \leq L$ of the multilevel scheme (3.3.2). The results are presented in Figure 3.4, where we plot the empirical means of the estimators over the N_{MSE} independent runs, along with their 95% confidence interval. First, comparing the left and the right graphs, we observe that the control variate significantly reduces the variance of each estimator, as expected. Then, one can see that the MC estimator with a rectangle scheme is still the worst estimator, while the MLMC estimator with a trapezoidal scheme still provides the best results – we also see that a standard MC estimator based on the trapezoidal scheme with the additional contribution of the control variate has a very satisfactory performance.

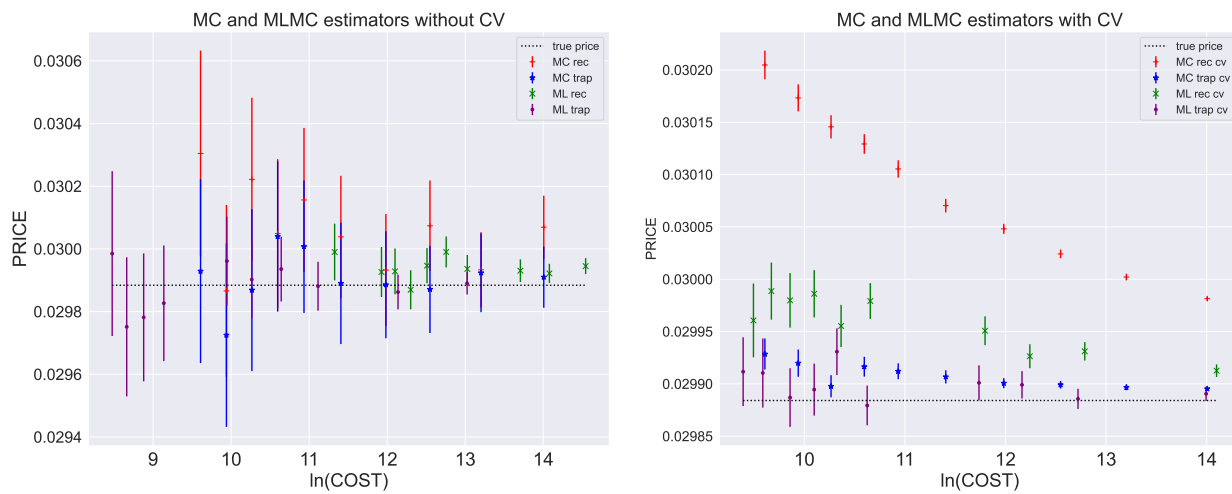


FIGURE 3.4 – Average prices of call options on the VIX against $\ln(\text{Cost})$ for the MC and MLMC estimators based on different discretization schemes, without any control variate (left) and with the introduction of the control variate (3.3.5) (right).

Deuxième partie

**Pricing and risk management of VIX
derivatives**

Robust replication of VIX options : a constrained martingale optimal transport problem

Contents

4.1 Super-hedging of VIX options	41
4.1.1 Duality	43
4.2 An analytical upper-bound : extremal moment problem	44
4.2.1 Optimality	48
4.3 Numerical experiments	49

In this chapter, we focus on the computation of model-independent price bounds and super-replication strategies for VIX options that are consistent with Vanillas on the underlying SP500 index and with the VIX future. After showing a duality result, we derive analytical (a priori non-optimal) bounds on the VIX option price and we characterize, in terms of an order condition between measures, the class of marginals distributions of the SP500 index for which these analytical bounds are actually optimal. The results presented in this chapter are taken from our work De Marco and Henry-Labordère [31] which, to our knowledge, provided one of the first mathematical results towards the robust super-replication of VIX options.

4.1 Super-hedging of VIX options

We consider the problem of hedging an option on the VIX index (1.2.2). Though we could consider a European option with arbitrary payoff, we will focus on a VIX call option $(\text{VIX}_{t_1} - K)^+$, where t_1 denotes the option maturity and K the option strike.

Hedging instruments. The hedging strategy will of course depend on the instruments we are allowed to trade on the VIX market and on the underlying SP500 market. In this chapter, we assume the hedging strategy can be constructed using the following operations :

- Trading at time $t_0 = 0$ in European options on the SP500 for maturities t_1 and $t_2 = t_1 + \Delta = t_1 + 30$ days, with arbitray payoffs $u_1(S_{t_1})$ and $u_2(S_{t_2})$. We assume these options can be traded on the SP500 market at time t_0 for their market prices

$$\text{price}^{\text{mkt}}(u_1(S_{t_1})) = \int u_1(s)\mu(ds), \quad \text{price}^{\text{mkt}}(u_2(S_{t_2})) = \int u_2(s)\nu(ds),$$

where the marginal laws μ and ν are identified from the prices of Vanilla options on the SP500 for maturities t_1 and t_2 .

- Trading at time $t_0 = 0$ in a future contract on VIX_{t_1} , paying at t_1 an amount equal to the value of VIX_{t_1} against its future price fixed at time t_0 , that we denote FVIX.
- Dynamic trading, between times t_1 and t_2 , in the underlying index S .
- Dynamic trading, between times t_1 and t_2 , of the forward log-contract introduced at the beginning of section 1.2 : a contract paying the amount $-\frac{2}{\Delta} \log \frac{S_{t_2}}{S_{t_1}}$ at maturity t_2 , where $\Delta = t_2 - t_1$.

Note that, by definition of the VIX (1.2.2), the forward log-contract will satisfy

$$\text{price}_{t_1}^{\text{mkt}} \left(-\frac{2}{\Delta} \log \frac{S_{t_2}}{S_{t_1}} \right) = \text{VIX}_{t_1}^2, \quad (4.1.1)$$

Our approach will be to look for a model-independent super-replication strategy for the VIX option. The dual version of the problem will be connected to a martingale optimal transport problem as introduced in section 1.3.2. When restricted to the space of measures $\mathbb{P} \in \mathbb{R}_+^2$ for the underlying SP500 index (S_{t_1}, S_{t_2}) , this new problem will correspond to the maximization of the expectation of a VIX option payoff with respect to a martingale measure \mathbb{P} with marginals μ, ν and with the additional constraint on the VIX future $\text{FVIX} = \mathbb{E}^{\mathbb{P}} \left[\sqrt{\mathbb{E}^{\mathbb{P}} \left[-\frac{2}{\Delta} \ln S_{t_2}/S_{t_1} | S_{t_1} \right]} \right]$. Note that this additional constraint, not present in the original martingale optimal transport, is non-linear with respect to the measure \mathbb{P} , making this optimal transport problem more involved. Our strategy will be to extend the state space by introducing an autonomous variable representing the value of VIX_{t_1} , which will be linked to (S_{t_1}, S_{t_2}) by a consistency condition (corresponding to the definition of the VIX (4.1.1)). As a result, we will be able to retrieve a linear problem, under the form of a martingale optimal transport problem with an additional (now linear) constraint.

From now on, $(S_{t_1}, S_{t_2}, \text{VIX}_{t_1})$ denotes the identity function on \mathbb{R}_+^3 . We denote $\mathcal{P}(X)$ the set of probability measures on a measurable space $(X, \beta(X))$. We assume that μ and ν are two probability measures on \mathbb{R}_+ having the same finite mean $\int_{\mathbb{R}_+} x\mu(dx) = \int_{\mathbb{R}_+} y\nu(dy) = S_0$ and satisfying the following condition

$$\log(S_1) \in L^q(\mu), \quad \log(S_2) \in L^q(\nu), \quad \text{for some } q > 1. \quad (4.1.2)$$

Exploiting the hedging instruments listed above, the robust super-hedging price of the VIX call option $(\text{VIX}_{t_1} - K)^+$ is defined by

$$\text{UB} := \inf_{u_1 \in L^1(\mu), u_2 \in L^1(\nu), \lambda \in \mathbb{R}, \Delta_S, \Delta_X} \mathbb{E}^{\mu}[u_1(S_{t_1})] + \mathbb{E}^{\nu}[u_2(S_{t_2})] + \lambda \text{FVIX} \quad (4.1.3)$$

over $u_1, u_2, \lambda, \Delta_S$ and Δ_X such that

$$u_1(s_1) + u_2(s_2) + \lambda\sqrt{x} + \Delta_S(s_1, x)(s_2 - s_1) + \Delta_X(s_1, x) \left(-\frac{2}{\Delta} \ln \left(\frac{s_2}{s_1} \right) - x \right) \geq (\sqrt{x} - K)^+ \quad \forall (s_1, s_2, \sqrt{x}) \in \mathbb{R}_+^3, \quad (4.1.4)$$

where the functions $\Delta_S, \Delta_X : \mathbb{R}_+ \times \mathbb{R}_+ \rightarrow \mathbb{R}$ are assumed to be bounded continuous functions on \mathbb{R}_+^2 . Note that (4.1.4) defines a linear infinite-dimensional programming problem. The variable x should be interpreted as the t_1 -value of a log-contract $-\frac{2}{\Delta} \ln \frac{s_2}{s_1}$, which corresponds to the square of the VIX index $\text{VIX}_{t_1}^2$.

Interpretation at t_1 . The super-replication strategy (4.1.4) provides the option hedger with the following recipe : at time t_0 , buy the options $u_1(S_{t_1})$ and $u_2(S_{t_2})$ and hold them statically until maturity, and enter into a position on λ VIX futures contracts with maturity t_1 . At time t_1 , introduce an additional delta hedge

position buying Δ_S shares of the underlying S , and buying Δ_X log-contracts $-\frac{2}{\Delta} \log \frac{S_{t_2}}{S_{t_1}}$ at their market price, to be held until time t_2 .

Since the VIX option expires at t_1 , it might seem strange to unwind the hedging position only at time t_2 . Nevertheless, the interpretation of (4.1.4) is the following : the wealth generated by the hedging portfolio on the left hand side of (4.1.4) at time t_1 will be sufficient to super-replicate the VIX option. Indeed, we have the following result : define

$$\text{UB}' := \inf_{u_1, u_2, \lambda} \mathbb{E}^\mu[u_1(S_{t_1})] + \mathbb{E}^\nu[u_2(S_{t_2})] + \lambda \text{VIX} \quad (4.1.5)$$

where the inf is taken over $u_1 \in L^1(\mu)$, $u_2 \in L^1(\nu)$ and $\lambda \in \mathbb{R}$ such that for all $s_1 \in \mathbb{R}_+$ and all $\text{vix}_{t_1} \in \mathbb{R}_+$,

$$u_1(s_1) + \mathbb{E}^\mathbb{P}[u_2(S_{t_2})] + \lambda \text{vix}_{t_1} \geq (\text{vix}_{t_1} - K)^+, \quad \forall \mathbb{P} \in \mathcal{P}_{s_1, \text{vix}_{t_1}} \quad (4.1.6)$$

where $\mathcal{P}_{s_1, \text{vix}_{t_1}}$ is the set of probability measures on \mathbb{R}_+ such that $\mathbb{E}^\mathbb{P}[S_{t_2}] = s_1$ and $\mathbb{E}^\mathbb{P}\left[-\frac{2}{\Delta} \ln\left(\frac{S_{t_2}}{s_1}\right)\right] = \text{vix}_{t_1}^2$. Condition (4.1.6) means that the VIX option $(\text{VIX}_{t_1} - K)^+$ can be super-replicated at t_1 by exercising at t_1 both the European option with payoff $u_1(s_1)$ and the VIX future, and selling the European option with payoff $u_2(S_{t_2})$ at the price $\mathbb{E}^\mathbb{P}[u_2(S_{t_2})|S_{t_1}, \text{VIX}_{t_1}]$ prevailing on the market at time t_1 , regardless of the pricing measure the market will be using at that moment.

Proposition 4.1 (Proposition 2.6 in [31]). *The two super-replication prices coincide : $\text{UB}' = \text{UB}$.*

4.1.1 Duality

We are going to state a duality result for the problem (4.1.3). The related measure set will be the set of all measures \mathbb{P} on the (pathspace) \mathbb{R}_+^3 whose first two components satisfy the martingale condition and have μ , ν , and such that $\mathbb{E}^\mathbb{P}[\text{VIX}_{t_1}] = \text{FVIX}$, that is

$$\begin{aligned} \mathcal{M}(\mu, \nu, \text{FVIX}) = \left\{ \mathbb{P} \in \mathcal{P}(\mathbb{R}_+^3) : S_{t_1} \stackrel{\mathbb{P}}{\sim} \mu, S_{t_2} \stackrel{\mathbb{P}}{\sim} \nu \right. \\ \mathbb{E}^\mathbb{P}[\text{VIX}_{t_1}] = \text{FVIX} \\ \mathbb{E}^\mathbb{P}[S_{t_2}|S_{t_1}, \text{VIX}_{t_1}] = S_{t_1} \\ \left. \mathbb{E}^\mathbb{P}\left[-\frac{2}{\Delta} \log \frac{S_{t_2}}{S_{t_1}} \middle| S_{t_1}, \text{VIX}_{t_1}\right] = \text{VIX}_{t_1}^2 \right\}. \end{aligned} \quad (4.1.7)$$

It is not difficult to see that the moment condition (4.1.2) implies that the set $\mathcal{M}(\mu, \nu, \text{FVIX})$ is tight and closed, hence weakly compact, see [31, Proposition 4.11] for details. We note that necessary conditions for non-emptiness of $\mathcal{M}(\mu, \nu, \text{FVIX})$ are the convex order condition $\mu \leq \nu$ and, by Jensen's inequality, the condition $\text{FVIX} \leq \sigma_{1,2}$, where

$$\sigma_{1,2}^2 = \text{price}_{t_0}^{\text{mkt}}\left(-\frac{2}{\Delta} \log \frac{S_{t_2}}{S_{t_1}}\right) = -\frac{2}{\Delta} (\mathbb{E}^\nu[\log(S_{t_2})] - \mathbb{E}^\mu[\log(S_{t_1})])$$

denotes the market price at time $t_0 = 0$ of the forward log-contract.

Theorem 4.1 (Duality for VIX options, Theorem 3.1 in [31]). *Let the VIX futures price FVIX be given. Assume that μ, ν are probability measures on \mathbb{R}_+ such that $\mathcal{M}(\mu, \nu, \text{FVIX})$ is non-empty. Then, the super-replication price in (4.1.3) satisfies*

$$\text{UB} = \sup_{\mathbb{P} \in \mathcal{M}(\mu, \nu, \text{FVIX})} \mathbb{E}^\mathbb{P}[(\text{VIX}_{t_1} - K)^+] = \mathbb{E}^{\mathbb{P}^*}[(\text{VIX}_{t_1} - K)^+] \quad (4.1.8)$$

for some $\mathbb{P}^* \in \mathcal{M}(\mu, \nu, \text{FVIX})$.

Theorem 4.1 tells that that the VIX option super-replication price is attained by a martingale measure, that is a model, calibrated to the t_1 and t_2 Vanilla smiles and to the VIX future. Note that, as a crucial step, by introducing a delta hedging on the forward log-contract, the problem (4.1.3) has been converted into a linear programming problem. We prove the duality result (4.1.8) mimicking the proof of the Kantorovich duality (1.3.9) in martingale optimal transport. Note that the existence of a maximizer \mathbb{P}^* follows from the continuity of the map $\mathbb{P} \mapsto \mathbb{E}^{\mathbb{P}}[(\text{VIX}_{t_1} - K)^+]$ over $\mathcal{M}(\mu, \nu, \text{FVIX})$ and the weak compactness of $\mathcal{M}(\mu, \nu, \text{FVIX})$.

Remark 4.1. *Guyon et al. [59] have shown a duality result for a structurally similar, even though simpler, problem : the super-replication of VIX futures given options on the SP500 with maturities t_1 and t_2 . With respect to our setting, VIX options are not involved in their work.*

Lower bound. Defining the option sub-replication price by

$$\text{LB} := \inf_{u_1 \in L^1(\mu), u_2 \in L^1(\nu), \lambda \in \mathbb{R}, \Delta_S, \Delta_X} \mathbb{E}^\mu[u_1(S_{t_1})] + \mathbb{E}^\nu[u_2(S_{t_2})] + \lambda \text{FVIX}$$

over $u_1, u_2, \lambda, \Delta_S$ and Δ_X such that

$$u_1(s_1) + u_2(s_2) + \lambda\sqrt{x} + \Delta_S(s_1, x)(s_2 - s_1) + \Delta_X(s_1, x) \left(-\frac{2}{\Delta} \ln \left(\frac{s_2}{s_1} \right) - x \right) \leq (\sqrt{x} - K)^+ \\ \forall (s_1, s_2, \sqrt{x}) \in \mathbb{R}_+^3,$$

we have the analogous duality result : $\text{LB} = \min_{\mathbb{P} \in \mathcal{M}(\mu, \nu, \text{FVIX})} \mathbb{E}^{\mathbb{P}}[(\text{VIX}_{t_1} - K)^+]$, see [31, Theorem 3.1].

4.2 An analytical upper-bound : extremal moment problem

For given data μ, ν and FVIX, the dual problem

$$\sup_{\mathbb{P} \in \mathcal{M}(\mu, \nu, \text{FVIX})} \mathbb{E}^{\mathbb{P}}[(\text{VIX}_{t_1} - K)^+] \quad (4.2.1)$$

and the original super-replication problem (4.1.3) can be approximated by finite dimensional problems and tackled numerically. In (4.2.1), the measures μ and ν can be approximated by measures μ_n, ν_n with finite support (on, say, n points), so that (4.2.1) is converted into a n^3 -dimensional problem (once the distribution of VIX_{t_1} has been discretized over n points, too) that can be tackled with standard methods such as the simplex algorithm, or possibly with an additional entropic regularization and the Sinkhorn algorithm. In its turn, (4.1.3) can be projected onto a finite-dimensional space of functions $u_1, u_2, \Delta_s, \Delta_X$ and the constraint (4.1.4) restricted to a finite set of points ; the resulting problem is again a finite-dimensional linear program, which can be tackled with standard algorithms, as we do in section 4.3.

In this section, we follow a different route : we evaluate an explicit – while a priori sub-optimal – upper bound $\overline{\text{UB}} \geq \text{UB}$. We will then be able to provide optimality conditions for $\overline{\text{UB}}$ in the next section.

Note that for every admissible $\mathbb{P} \in \mathcal{M}(\mu, \nu, \text{FVIX})$, one has

$$\mathbb{E}^{\mathbb{P}}[\text{VIX}_{t_1}] = \text{FVIX}, \quad \text{and} \quad \mathbb{E}^{\mathbb{P}}[\text{VIX}_{t_1}^2] = \sigma_{1,2}^2.$$

We can therefore relax (considerably) the problem (4.2.1), considering

$$\overline{\text{UB}} := \sup_{\mathbb{P} \in \mathcal{P}_{1,2}} \mathbb{E}^{\mathbb{P}}[(X - K)^+], \quad (4.2.2)$$

where X denotes the identity on \mathbb{R}_+ , and now the sup is taken over the set

$$\mathcal{P}_{1,2} := \{\mathbb{P} \in \mathcal{P}(\mathbb{R}_+) : \mathbb{E}^{\mathbb{P}}[X] = \text{FVIX}, \mathbb{E}^{\mathbb{P}}[X^2] = \sigma_{1,2}^2\}. \quad (4.2.3)$$

It is clear that (4.2.2) defines an upper bound $\overline{\text{UB}} \geq \text{UB}$.

The problem (4.2.2) a typical example of “extremal moment problem” : we are maximizing the expectation of a certain payoff under the constraint that some (here : the first and second) moments of the underlying distribution are fixed. This type of problem has been widely studied in the actuarial science literature ; for the case of piece-wise affine payoffs, see [25, 85, 86, 70] and the comprehensive review by Hürlimann [87]. Note that (4.2.2) is much easier to solve than (4.2.1). Let us explain how a solution can be constructed, for we believe the procedure is instructive and enlightening.

Since the objective function $\mathbb{P} \mapsto \mathbb{E}^{\mathbb{P}}[(X - K)^+]$ is linear (hence, convex), by Bauer’s maximum principle we can look for a solution to (4.2.2) within the extremal points of the convex set $\mathcal{P}_{1,2}$. Exploiting the following characterisation :

Theorem 4.2 (Winkler [88]). $\mathbb{P} \in \mathcal{P}_{1,2}$ is an extreme point of $\mathcal{P}_{1,2}$ if and only if $\#\text{supp}(\mathbb{P}) \leq 3$,

it will be enough to search for a solution $\overline{\mathbb{P}}$ with support on 2 or 3 points. It is not difficult to see that there is no maximizer $\overline{\mathbb{P}}$ charging three points ; a measure \mathbb{P} supported by two points will always perform better, due to the convexity of the payoff function $(x - K)^+$.

Now, by standard weak duality, we have

$$\overline{\text{UB}} \leq \inf\{a\sigma_{1,2}^2 + b\text{FVIX} + c : ax^2 + bx + c \geq (x - K)^+, \forall x \geq 0\}.$$

It is therefore sufficient to find a bi-atomic distribution $\overline{\mathbb{P}}(dx) = p\delta_{x_0}(dx) + (1 - p)\delta_{x_1}(dx) \in \mathcal{P}_{1,2}$ and constants $\bar{a}, \bar{b}, \bar{c}$ that attain the equality. To do so, we ask that $\inf_{x \in \mathbb{R}_+} \{ax^2 + bx + c - (x - K)^+\} = 0$ is attained at x_0, x_1 . This condition implies the following algebraic equations for (p, x_0, x_1, a, b, c) :

$$\begin{aligned} px_0 + (1 - p)x_1 &= m_1, & px_0^2 + (1 - p)x_1^2 &= m_2 \\ ax_0^2 + bx_0 + c &= (x_0 - K)^+, & ax_1^2 + bx_1 + c &= (x_1 - K)^+ \\ 2ax_0 + b &= 1_{x_0 \geq K} \text{ if } x_0 \neq 0, & 2ax_1 + b &= 1_{x_1 \geq K}, \end{aligned} \quad (4.2.4)$$

with $a > 0$. The following proposition reports the solution to (4.2.4) (which can be found for example in Jansen et al. [70]) ; the second part of the proposition translates the dual solution (a, b, c) in terms of the super-replicating strategy in (4.1.3).

Proposition 4.2 (Analytical upper bound). *The unique maximizer for the problem (4.2.2) is given by the bi-atomic measure*

$$\overline{\mathbb{P}}(dx) = p\delta_{x_0}(dx) + (1 - p)\delta_{x_1}(dx) \quad (4.2.5)$$

with

$$\begin{cases} x_0 = K - I; & x_1 = K + I; & p = \frac{K - \text{FVIX} + I}{2I(K)} & \text{if } K \geq K^*, \\ x_0 = 0; & x_1 = \frac{\sigma_{1,2}^2}{\text{FVIX}}; & p = \frac{\sigma_{1,2}^2 - \text{VIX}^2}{\sigma_{1,2}^2} & \text{if } K < K^*, \end{cases}$$

where $K^* = \frac{\sigma_{1,2}^2}{2\text{FVIX}}$ and $I = I(K) = \sqrt{\sigma_{1,2}^2 - \text{FVIX}^2 + (\text{FVIX} - K)^2}$. The value of the problem is then

$$\overline{\text{UB}}(\text{FVIX}, \sigma_{1,2}^2) = \mathbb{E}^{\overline{\mathbb{P}}}[(X - K)^+] = \begin{cases} \frac{1}{2}(\text{FVIX} - K + I) & K \geq K^* \\ \text{FVIX} - K \frac{\text{FVIX}^2}{\sigma_{1,2}^2} & K < K^*. \end{cases} \quad (4.2.6)$$

This bound is attained in (4.1.3) by the super-replication strategy

$$\bar{u}_1(s_1) = -\frac{2}{\Delta} \Delta_X \ln \frac{s_1}{S_0} + \nu, \quad \bar{u}_2(s_2) = \frac{2}{\Delta} \Delta_X \ln \frac{s_2}{S_0}, \quad \bar{\lambda}, \quad \bar{\Delta}_S(s_1, 0) = 0, \quad \bar{\Delta}_X(s_1, x) = \Delta_X, \quad (4.2.7)$$

where for $K \geq K^*$, we have $\Delta_X = -\frac{1}{4I}$, $\nu = \frac{-2KI + \sigma_{1,2}^2 - 2K(\text{FVIX} - K)}{4I}$, $\bar{\lambda} = \frac{1}{2} - \frac{K}{2I}$, and for $K \leq K^*$, $\Delta_X = -K \left(\frac{\text{FVIX}}{\sigma_{1,2}^2} \right)^2$, $\nu = 0$, $\bar{\lambda} = 1 - 2K \frac{\text{FVIX}}{\sigma_{1,2}^2}$.

Recall that the result above implies that the pathwise super-replication

$$\bar{u}_1(s_1) + \bar{u}_2(s_2) + \bar{\lambda} \sqrt{x} + \Delta_X \left(-\frac{2}{\Delta} \ln \left(\frac{s_2}{s_1} \right) - x \right) \geq (\sqrt{x} - K)^+ \quad (4.2.8)$$

holds for all $(s_1, s_2, \sqrt{x}) \in \mathbb{R}_+^3$. Note that the bound $\bar{\text{UB}}$ depends only on the market values FVIX and $\sigma_{1,2}$ and it is well-defined if and only if the condition $\text{FVIX} \leq \sigma_{1,2}$ holds.

Let us also note that the construction of the solution (4.2.8) has explicitly taken advantage of the fact that the log-contract defining the VIX value at t_1 can be written as the difference of two European options, i.e., $\ln s_2 - \ln s_1$. This would not be the case any more if we used the trapezoidal approximation of the log-contract applied by the CBOE to define the true market VIX [26] (even if we can consider that the log-contract still remains a reasonable approximation of the true VIX, due to the number of liquid strikes traded on the SP500).

Numerical values and comparison with market data. The analytical upper bound $\bar{\text{UB}}$ is straightforward to evaluate. In Figure 4.1, we have compared the value of $\bar{\text{UB}}$ (expressed in terms of VIX implied volatility) against market implied volatilities for VIX options with expiry $t_1 = 16$ Oct. 13 and pricing date $t_0 = 12$ Jul. 2013. Surprisingly enough, market prices of VIX options are above our analytical upper bound $\bar{\text{UB}} \geq \text{UB}$, for several values of the strike price K .

This surprising numerical result deserves some comments. The upper bound $\bar{\text{UB}}$ depends on the $t = 0$ value FVIX of the VIX future and on the value of the forward log-contract $\sigma_{1,2}$. While FVIX is quoted by the market, log-contracts are not directly traded, and they must be statically replicated using a strip of Vanilla options, which entails

- (i) a certain amount of uncertainty in the value of $\sigma_{1,2}$, which is sensitive to the interpolation and extrapolation method used for Vanilla option prices away from observed strikes (in other words : given the same finite set of option prices, different users will come up with different values of $\sigma_{1,2}$);
- (ii) the presence of option transaction costs, notably for low and high strikes.

In order to see the impact of the pricing uncertainty and transaction costs in the value of $\sigma_{1,2}$, we have added +0.5% and +1.0% to the reference value of the forward log-contract $\sigma_{1,2} = 18.15\%$ that was used in our tests. As the value of $\sigma_{1,2}$ increases, the arbitrage opportunities evaporate, as expected.

Remark 4.2 (On the joint SP500-VIX calibration problem). *Figure 4.1 seems to point to an unfeasibility of the joint calibration problem to SP500 and VIX options : if market prices for VIX options lie above the maximum price UB generated by admissible martingale models calibrated to the marginal laws of S_{t_1} and S_{t_2} and to the VIX future, it will be impossible to find a model that fits the market VIX option quotes (for otherwise this would contradict Theorem 4.1). But we should not forget a ‘‘compatibility issue’’ between the SP500 and the VIX option markets. It is well known that a forward contract on $\text{VIX}_{t_1}^2$ (which is not*

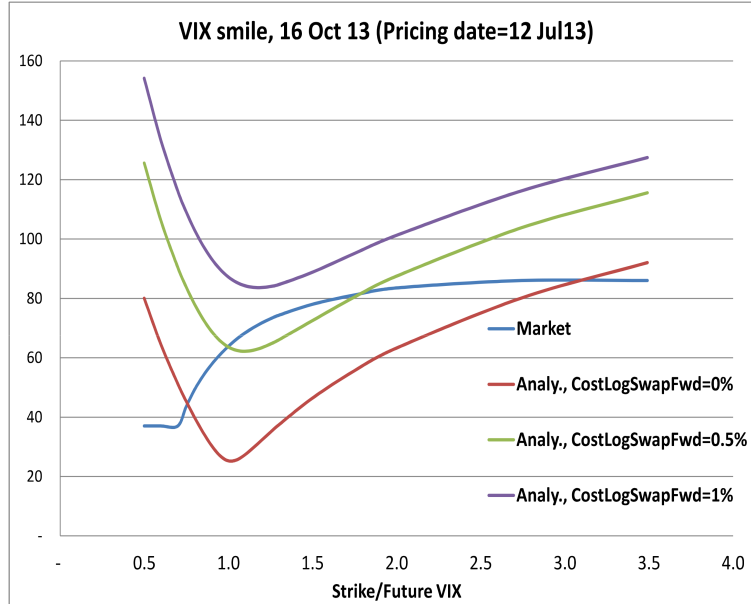


FIGURE 4.1 – Analytical upper bound \overline{UB} in Proposition 4.2 versus market values for VIX smiles observed on 12 Jul. 2013, for maturity $t_1 = 16$ Oct. 13. The (square-root of the) log-contract price is $\sigma_{1,2} = 18.15\%$, and the VIX futures quote is $FVIX = 18.05\%$.

traded on any of the two markets) can be synthesized in two different ways : from a strip of VIX options, by statical replication of the payoff x^2 , or by trading a forward log-contract $-\frac{2}{\Delta} \log \frac{S_{t_2}}{S_{t_1}}$, precisely as we did in the construction of the super-replication strategy (4.1.4). Let us denote $\text{price}^{\text{mkt}, \text{SP500}}(\text{VIX}_{t_1}^2)$ and $\text{price}^{\text{mkt}, \text{VIX}}(\text{VIX}_{t_1}^2)$ the prices of $\text{VIX}_{t_1}^2$ obtained according to the two different methods. It has been reported by several authors, see for example Bergomi [13, Chapter 7], that the two prices of $\text{VIX}_{t_1}^2$ do not always coincide (and their difference has sometimes been so large that arbitrage opportunities have appeared).

If the marginal laws μ, ν and the VIX option prices injected into a calibration procedure are such that the condition

$$\text{price}^{\text{mkt}, \text{SP500}}(\text{VIX}_{t_1}^2) = \text{price}^{\text{mkt}, \text{VIX}}(\text{VIX}_{t_1}^2) \quad (4.2.9)$$

is not satisfied, no model can possibly calibrate the given joint set of SP500 and VIX options. The condition (4.2.9) is a necessary condition for the existence of a model that achieves the joint calibration, and it should be ensured as a starting point, as explained in Guyon [57] – where a discrete-time model that does achieve the joint calibration is found. Our superhedging portfolio (4.1.4) is based on the construction of $\text{VIX}_{t_1}^2$ from options on the SP500, and therefore the data for our problem are the VIX future quote and the value of $\sigma_{1,2}^2 = \text{price}^{\text{mkt}, \text{SP500}}(\text{VIX}_{t_1}^2)$, regardless of the value of $\text{price}^{\text{mkt}, \text{VIX}}(\text{VIX}_{t_1}^2)$ obtained from VIX options.

In other words, at the time when the Figure 4.1 was generated, we did not check whether condition (4.2.9) was satisfied or not, that is to say, we did not check how much the value of $\sigma_{1,2} = 18.15\%$ we used in our formulas deviates from the price of $\text{VIX}_{t_1}^2$ that could be reconstructed from VIX options on the pricing date. It is likely that an important deviation from conditions (4.2.9) was experienced on that day.

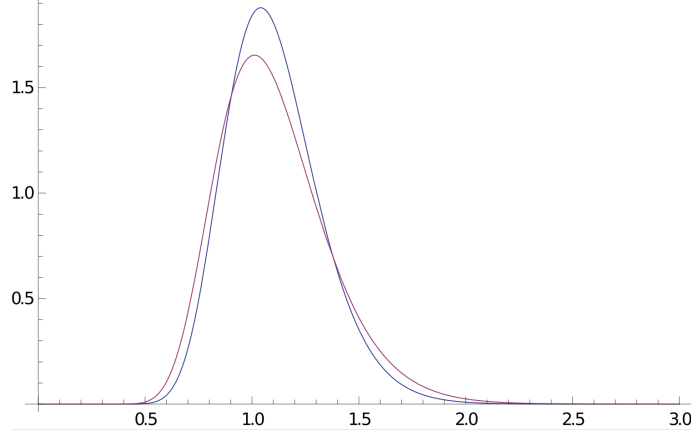


FIGURE 4.2 – Example of densities μ (blue curve) and ν (red curve) satisfying conditions (4.2.10) and (4.2.11) in Theorem 4.3 for the values of x_0, x_1 and p given in Proposition 4.2 when $\text{FVIX} = 0.3, \sigma_{1,2} = 0.4$ and $K = \text{FVIX}$ (corresponding values are $x_0 = 0.035, x_1 = 0.567$ and $p = 0.5$).

4.2.1 Optimality

It is natural to ask if there are situation in which the (a priori sub-optimal) analytical upper bound $\overline{\text{UB}}$ is actually optimal, that is, we have $\overline{\text{UB}} = \text{UB}$. It turns out that we are able to give conditions on the data (μ, ν, FVIX) that are equivalent to $\overline{\text{UB}} = \text{UB}$. The resulting condition on the marginals μ, ν is an order relationship.

Theorem 4.3 (Theorem 4.12 in [31]). *Let μ, ν be probability measures in $\mathcal{P}(\mathbb{R}_+)$ with same finite mean, satisfying condition (4.1.2) and such that $\sigma_{1,2}^2(\mu, \nu) > 0$. Let $\text{FVIX} \in (0, \sigma_{1,2}(\mu, \nu)]$. Denote $\text{UB}(\mu, \nu, \text{FVIX})$ the value of (4.1.3) and $\overline{\text{UB}}(\text{FVIX}, \sigma_{1,2}^2(\mu, \nu))$ the value of (4.2.2). The following are equivalent :*

(i) $\text{UB}(\mu, \nu, \text{FVIX}) = \overline{\text{UB}}(\text{FVIX}, \sigma_{1,2}^2(\mu, \nu))$

(ii) *There exist two couples of measures (μ_0, ν_0) and (μ_1, ν_1) on \mathbb{R}_+ such that*

$$\mu = p\mu_0 + (1 - p)\mu_1, \quad \nu = p\nu_0 + (1 - p)\nu_1, \tag{4.2.10}$$

and

$$\int f(z, \log(z)) \nu_i(dz) \geq \int f\left(y, \log(y) - \frac{\Delta}{2} x_i^2\right) \mu_i(dy), \quad i = 0, 1, \tag{4.2.11}$$

for every convex function $f : \mathbb{R}^2 \rightarrow \mathbb{R}$, where the coefficients p, x_0, x_1 have been given in Proposition 4.2.

Note that condition (4.2.11) entails that the measures (μ_i, ν_i) increase in the convex order (take $f(y, z) = \tilde{f}(y)$ with \tilde{f} convex in (4.2.11)), hence by (4.2.10) so do the measures (μ, ν) . Given two distributions μ and ν , it is arguably not straightforward how to check whether they satisfy (4.2.10)-(4.2.11), for (4.2.11) involves checking a condition for all convex functions on \mathbb{R}^2 . On the other hand, examples of a measures μ, ν satisfying conditions (4.2.10)-(4.2.11) can be explicitly constructed, as done in Example 4.14 in [31] and illustrated in Figure 4.2.

K/FVIX	$\text{UB}_{\text{num}} \times 100$	$\overline{\text{UB}} \times 100$
0.90	30.56	30.56
0.95	27.13	27.13
1.00	25.31	25.31
1.05	25.59	25.59
1.10	27.33	27.33
1.15	29.70	29.70

TABLE 4.1 – Numerical approximation of the optimal bound UB in (4.1.8) for VIX options (expiry 16 Oct. 13, pricing date = 12 Jul. 2013) and comparison with the analytical upper bound $\overline{\text{UB}}$. UB_{num} and $\overline{\text{UB}}$ are expressed in terms of VIX implied volatilities.

4.3 Numerical experiments

We have computed the optimal bound UB for VIX options with expiry 16 Oct. 13 (pricing date = 12 Jul. 2013) by numerically solving the linear program (4.1.3) with a simplex method. More precisely, the European payoffs u_1 and u_2 are decomposed over a basis of call options with payoffs $((s_j - K_i)^+)_{1 \leq i \leq n}$ and a log-contract, and we have used a fourth-order polynomial approximation in s_1 and x for the functions Δ_S and Δ_X .

The upper bound UB_{num} for VIX options with different strikes K is reported and compared with $\overline{\text{UB}}$ in Table 4.3, where both bounds are quoted in terms of VIX implied volatilities ($\times 100$). We see that $\overline{\text{UB}}$ and UB_{num} are indistinguishable, indicating that the bound $\overline{\text{UB}}$ seems optimal. We have checked that the numerical bound UB_{num} is not improved by increasing the range of liquid strikes and the rank of the polynomial approximations used for Δ_S and Δ_X .

This numerical result suggests that – on the pricing date considered here – the market marginals μ and ν are such that $\text{UB} = \overline{\text{UB}}$ (or, at least, such that $\overline{\text{UB}}$ and our numerical approximation of UB are indistinguishable), so that the optimality result proved in Theorem 4.3 seems to remain valid in this setting. Recall that the super-replication strategy (4.2.7) with price $\overline{\text{UB}}$ only contains semi-static positions in log-contracts and VIX futures, while the super-replicating portfolio (4.1.4) used to evaluate UB_{num} contains options $u_i(S_{t_i})$ with more general payoffs; the closeness of the values of $\overline{\text{UB}}$ and UB_{num} indicates that, maybe not surprisingly, log-contracts seem to be the most relevant Vanilla instruments involved in the super-hedging of VIX options.

Weak approximations and VIX option price expansions in forward variance models

Contents

5.1 Our contribution in a nutshell	52
5.2 The main elements of our approximation procedure	53
5.2.1 Strong error estimates between VIX_T^2 and its proxy $VIX_{T,P}^2$	53
5.2.2 Option price expansion	54
5.3 Mixed (or “skewed”) exponential models	56
5.3.1 Numerical tests for option price formulas and implied volatilities	58

This chapter can be seen as a companion of Chapter 3. As we did there, we consider the problem of VIX option pricing within a class of forward variance curve models. While Chapter 3 is centered on the efficient – and specifically, multilevel – simulation of the VIX random variable in such setting, the focus of the present chapter is on explicit approximation formulas for option prices, obtained via expansion procedures for expectations. The results presented in this chapter are taken from our work Bourgey, De Marco and Gobet [20].

The price of the VIX option (3.1.3) depends (among other model parameters) on the specification of the initial forward variance curve $u \mapsto \xi_0^u$ over the time window $(T, T + \Delta)$. In the present chapter, we are going to assume that

$$\xi_0^u = \xi_0 \text{ is constant over } u \in (T, T + \Delta).$$

This simple choice can usually be made in practice, for it is typically sufficient to consider variance curves that are piece-wise constants between VIX maturities. The case of a generic and non-constant curve ξ_0^u is considered in [20]. The restriction to a constant ξ_0 allows us to simplify the presentation of most of the results in this chapter. With reference to (1.2.10), let us set

$$Y_T^u := X_T^u - X_0^u = -\frac{1}{2} \int_0^T K^u(t)^2 dt + \int_0^T K^u(t) dW_t. \tag{5.0.1}$$

Notation for the present chapter. In this chapter, it will be convenient to slightly change the notation for the integration kernel with respect to section 1.2, and denote

$$K^u(t) = K(u, t) \quad \text{for } t \leq u.$$

Given the probability measure $\nu_0(du) = \frac{du}{\Delta}$ on the interval $[T, T + \Delta]$, we denote

$$\nu_0(f) = \int_T^{T+\Delta} f(u)\nu_0(du) = \frac{1}{\Delta} \int_T^{T+\Delta} f(u)du \quad (5.0.2)$$

the mean of an integrable functions f with respect to the measure ν_0 .

5.1 Our contribution in a nutshell

We go back to the VIX option pricing problem within the explicit class of forward variance models (1.2.10). We have seen in Chapter 3 that the expectation (3.1.3) can be approximated by coupling a discretization scheme with Monte Carlo or Multilevel Monte Carlo simulation. Asymptotic formulas for small VIX maturities T can also be investigated, as done in [4, 74]. In the present chapter, we explore an alternative approach based on analytical approximations of the form

$$\mathbb{E}[\varphi(\text{VIX}_T^2)] = \text{Main term} + \text{Correction terms} + \text{Error}. \quad (5.1.1)$$

Both the main and the correction terms will be easily computable using elementary log-normal distributions. In addition, we aim at providing error bounds in terms of the characteristics of the kernel K and the length of the time-window Δ , covering the case of non-smooth payoffs φ .

A proxy for the mean of exponentials. It has been noted by several authors, from the early works of Bergomi [14] to the more recent work of Bayer et al. [9], that the VIX in the exponential Gaussian models (1.2.10) is not far from a log-normal random variable (as a result, the VIX implied volatility smile generated by these models is unreasonably flat, see section 5.3 below for an improvement in this direction). More precisely, exploiting the classical idea of replacing the arithmetic mean of exponentials with their geometric mean, the law of

$$\text{VIX}_T^2 = \xi_0 \frac{1}{\Delta} \int_T^{T+\Delta} e^{Y_T^u} du \quad (5.1.2)$$

is arguably close to the law of

$$\text{VIX}_{T,P}^2 = \xi_0 e^{\frac{1}{\Delta} \int_T^{T+\Delta} Y_T^u du}, \quad (5.1.3)$$

where the subscript P stands for proxy. Note that $\text{VIX}_{T,P}^2$ has the appealing property of being a log-normal random variable,

$$\ln(\text{VIX}_{T,P}^2) \stackrel{d}{=} \mathcal{N}(\mu_P, \sigma_P^2)$$

where $\mu_P := -\frac{1}{2} \int_0^T \frac{1}{\Delta} \int_T^{T+\Delta} K^u(t)^2 du dt$ and $\sigma_P^2 := \int_0^T (\frac{1}{\Delta} \int_T^{T+\Delta} (K^u(t)) du)^2 dt$. Consequently, the approximate VIX option price $\mathbb{E}[\varphi(\text{VIX}_{T,P}^2)]$ is a Black-Scholes price. Such log-normal approximations of the VIX random variable have already been exploited by several authors, for example as an approximation of VIX futures in the rough Bergomi model by Jacquier, Martini and Muguruza [68], or as a way to generate a log-normal control variate for the simulation of the VIX in [67], as we also did in Chapter 3.

In [20], our contribution is precisely to further approximate the residual correction term

$$\mathbb{E}[\varphi(\text{VIX}_T^2)] - \mathbb{E}[\varphi(\text{VIX}_{T,P}^2)].$$

We compute an explicit approximation of this correction performing an expansion around the log-normal proxy; the details of the procedure are explained in section 5.2.2. The effectiveness of the resulting price

expansion hinges on the small value of the VIX time window Δ (recall that $\Delta = \frac{1}{12}$ for the VIX); in more precise terms, we provide a $\mathcal{O}(\Delta^\alpha)$ estimate of the error term of the expansion, for some explicit (model-dependent) α . We stress that our approach indeed does not rely on small-time asymptotics nor small-parameter (such as small volatility-of-volatility) asymptotics, and can therefore be applied to any option maturity and a wide range of parameter configurations.

Remark 5.1 (Evaluation of reference VIX option prices within the classical Bergomi model). *Let us recall that, for the standard Bergomi model, we have the Markovian representation of forward variances*

$$\text{VIX}_T^2 = \frac{1}{\Delta} \int_T^{T+\Delta} \xi_0^u f^u(T, X_T) du,$$

where

$$f^u(T, x) = \exp\left(\omega e^{-k(u-T)} x - \frac{1}{2} \omega^2 e^{-2k(u-T)} \text{Var}(X_T)\right)$$

and $(X_t)_{t \geq 0}$ is the Ornstein–Uhlenbeck process $X_t = -k \int_0^t X_s ds + W_s$, with $\text{Var}(X_T) = \frac{1-e^{-2kT}}{2k} \mathbf{1}_{k>0} + T \mathbf{1}_{k=0}$. Consequently, in order to obtain reference values for VIX option prices in the one-factor Bergomi model, instead of using a Monte Carlo estimator we can (and do) rely on a two-dimensional deterministic quadrature : we couple a Gauss–Legendre scheme for the integration with respect to the time parameter u , and a Gauss–Hermite scheme for the space dimension.

5.2 The main elements of our approximation procedure

Following Gobet and Miri [52], our approach is to introduce the interpolation

$$I(\varepsilon) := \xi_0 \int_T^{T+\Delta} e^{\nu_0(Y_T) + \varepsilon(Y_T^u - \nu_0(Y_T))} \frac{du}{\Delta}, \quad \varepsilon \in [0, 1], \quad (5.2.1)$$

which is such that $I(0) = \text{VIX}_{T,P}^2$ and $I(1) = \text{VIX}_T^2$. The function I allows us to write a representation of the difference $\text{VIX}_T^2 - \text{VIX}_{T,P}^2$. Under Assumptions 5.1 and 5.2 below, it is easy to see that the map $\varepsilon \mapsto I(\varepsilon)$ is smooth almost surely, with n th derivative given by

$$I^{(n)}(\varepsilon) = \xi_0 \int_T^{T+\Delta} (Y_T^u - \nu_0(Y_T))^n e^{\nu_0(Y_T) + \varepsilon(Y_T^u - \nu_0(Y_T))} \frac{du}{\Delta}. \quad (5.2.2)$$

Noticing that $I^{(1)}(0) = 0$, an application of Taylor’s theorem with integral remainder yields

$$\begin{aligned} \text{VIX}_T^2 - \text{VIX}_{T,P}^2 &= I(1) - I(0) \\ &= \int_0^1 (1-\varepsilon) I^{(2)}(\varepsilon) d\varepsilon = \frac{I^{(2)}(0)}{2} + \int_0^1 \frac{(1-\varepsilon)^2}{2} I^{(3)}(\varepsilon) d\varepsilon. \end{aligned} \quad (5.2.3)$$

The representations of $I(1) - I(0)$ in the second line of (5.2.3) will allow us to quantify the difference between VIX_T^2 and $\text{VIX}_{T,P}^2$.

5.2.1 Strong error estimates between VIX_T^2 and its proxy $\text{VIX}_{T,P}^2$

The first step in view of the derivation of option price expansions with quantitative error estimates is an estimate of the L^p norm of the difference $\text{VIX}_T^2 - \text{VIX}_{T,P}^2$.

Let us present the standing assumptions for this chapter.

Assumption 5.1. *The initial instantaneous forward variance curve $u \mapsto \xi_0^u$ is positive, bounded, and bounded away from zero.*

Assumption 5.2. *The kernel $K(\cdot)$ is such that $\int_0^T K^u(t)^2 dt < \infty$ for every $u \in [T, T + \bar{\Delta}]$, for some $\bar{\Delta} \geq 1$. Moreover, for any $p > 0$, there exists a positive constant C_p such that*

$$\frac{1}{\Delta} \int_T^{T+\Delta} e^{p \int_0^T K^u(t)^2 dt} du \leq C_p \quad (5.2.4)$$

for all $\Delta \leq \bar{\Delta}$.

Assumption 5.2 is a mild technical condition, which essentially means that the moments of ξ_T^u are integrable over $[T, T + \Delta]$. As a consequence of Assumption 5.2, all the moments of the random variable VIX_T^2 are also finite, by Jensen's inequality. It is easy to check that Assumption 5.2 is satisfied by the one-factor Bergomi model (1.2.11) and the rough Bergomi model (1.2.12).

In order to estimate the L^p norm of the difference $\text{VIX}_T^2 - \text{VIX}_{T,P}^2$, we need some additional upper bounds for the deterministic L^p norm of the difference between the diffusion coefficient $K^u(t)$ of the log-forward variance $\log(\xi_t^u)$ and its integral average $\nu_0(K(\cdot(t)))$ (respectively, for the difference between the drift coefficient $(K^u(t))^2$ and its integral average $\nu_0(K(\cdot(t))^2)$). We require that these deterministic L^p norms go to zero as Δ goes to zero with certain rates d_1 and d_2 , see conditions (5.2.5) and (5.2.6).

Assumption 5.3. *For any $p > 0$, there exist positive constants d_1, d_2, C such that*

$$\left(\int_T^{T+\Delta} \left| \int_0^T [K^u(t)^2 - \nu_0(K(\cdot(t))^2)] dt \right|^p \frac{du}{\Delta} \right)^{\frac{1}{p}} \leq C \Delta^{d_1}, \quad (5.2.5)$$

$$\left(\int_T^{T+\Delta} \left| \int_0^T [K^u(t) - \nu_0(K(\cdot(t)))]^2 dt \right|^p \frac{du}{\Delta} \right)^{\frac{1}{p}} \leq C \Delta^{d_2}, \quad (5.2.6)$$

Moreover,

$$\sup_{\Delta} |\mu_P| \leq C, \quad \frac{1}{C} \leq \inf_{\Delta} \sigma_P \leq \sup_{\Delta} \sigma_P \leq C. \quad (5.2.7)$$

Under Assumptions 5.1, 5.2, and 5.3, we are able to prove the following

Theorem 5.1. *For any $p \geq 1$ we have*

$$\|\text{VIX}_T^2 - \text{VIX}_{T,P}^2\|_p \leq c_p \Delta^{2d_1 \wedge d_2}, \quad (5.2.8)$$

for some constant c_p that does not depend on Δ .

5.2.2 Option price expansion

As discussed above, the leading order term in the approximation of the VIX option price $\mathbb{E}[\varphi(\text{VIX}_T^2)]$ will be given by the price on the proxy $\mathbb{E}[\varphi(\text{VIX}_{T,P}^2)]$. Assuming for a moment that the payoff function φ is smooth, a Taylor expansion around the point $\text{VIX}_{T,P}^2$ yields

$$\mathbb{E}[\varphi(\text{VIX}_T^2)] = \mathbb{E}[\varphi(\text{VIX}_{T,P}^2)] + \mathbb{E}\left[\varphi'(\text{VIX}_{T,P}^2) (\text{VIX}_T^2 - \text{VIX}_{T,P}^2)\right] + E_0 \quad (5.2.9)$$

where the remainder $E_0 = \int_0^1 (1 - \lambda) \mathbb{E}[\varphi''(\lambda \text{VIX}_T^2 + (1 - \lambda) \text{VIX}_{T,P}^2) (\text{VIX}_T^2 - \text{VIX}_{T,P}^2)^2] d\lambda$ will be treated as an error term. The difference $\text{VIX}_T^2 - \text{VIX}_{T,P}^2$ can be expanded using (5.2.3) : recalling from (5.2.2) the expression of the derivative $I^{(2)}$, we get

$$\begin{aligned} \text{VIX}_T^2 - \text{VIX}_{T,P}^2 &= \frac{1}{2} I^{(2)}(0) + \int_0^1 \frac{(1 - \varepsilon)^2}{2} I^{(3)}(\varepsilon) d\varepsilon \\ &= \frac{1}{2} \text{VIX}_{T,P}^2 \int_T^{T+\Delta} (Y_T^u - \nu_0(Y_T^u))^2 \frac{du}{\Delta} + \int_0^1 \frac{(1 - \varepsilon)^2}{2} I^{(3)}(\varepsilon) d\varepsilon, \end{aligned}$$

so that the second expectation on the right-hand side of (5.2.9) can eventually be written as

$$\mathbb{E}\left[\varphi'(\text{VIX}_{T,P}^2) (\text{VIX}_T^2 - \text{VIX}_{T,P}^2)\right] = \mathbb{E}\left[\varphi'(\text{VIX}_{T,P}^2) \frac{1}{2} \text{VIX}_{T,P}^2 \int_T^{T+\Delta} (Y_T^u - \nu_0(Y_T^u))^2 \frac{du}{\Delta}\right] + E_1 \quad (5.2.10)$$

where $E_1 = \mathbb{E}\left[\varphi'(\text{VIX}_{T,P}^2) \int_0^1 \frac{(1 - \varepsilon)^2}{2} I^{(3)}(\varepsilon) d\varepsilon\right]$ will be treated as a second error term.

Now, the random variable multiplying $\varphi'(\text{VIX}_{T,P}^2)$ inside (5.2.10) can be interpreted as the random weight appearing after the application of an integration-by-parts formula to higher-order derivatives of φ – which means that, in its turn, the expectation on the right-hand side of (5.2.10) can be rewritten in terms of a combination of derivatives of the form $\partial_\varepsilon^i \mathbb{E}[\varphi'(\text{VIX}_{T,P}^2 e^\varepsilon)]|_{\varepsilon=0}$. The important property of such higher-order derivatives of the expectation $\mathbb{E}[\varphi(\text{VIX}_{T,P}^2)]$ is to be explicit – they are Black-Scholes Greeks. The final expression of the expansion (5.2.9) will therefore contain a combination of a Black-Scholes price and some of its partial derivatives.

Leaving the details of the approach sketched above to [20], let us state here the final expression we obtain for the expansion (5.2.9) after the integration-by-parts procedure. We will make use of the following coefficients $(\gamma_i)_{i=1,2,3}$:

$$\begin{aligned} \gamma_1 &:= \frac{1}{8} \int_T^{T+\Delta} \left(\int_0^T [K^u(t)^2 - \nu_0(K^\cdot(t)^2)] dt \right)^2 \frac{du}{\Delta} + \frac{1}{2} \int_T^{T+\Delta} \left(\int_0^T [K^u(t) - \nu_0(K^\cdot(t))] dt \right)^2 \frac{du}{\Delta}, \\ \gamma_2 &:= -\frac{1}{2} \int_T^{T+\Delta} \left(\int_0^T \nu_0(K^\cdot(t)) [K^u(t) - \nu_0(K^\cdot(t))] dt \right) \left(\int_0^T [K^u(t)^2 - \nu_0(K^\cdot(t)^2)] dt \right) \frac{du}{\Delta}, \\ \gamma_3 &:= \frac{1}{2} \int_T^{T+\Delta} \left(\int_0^T \nu_0(K^\cdot(t)) [K^u(t) - \nu_0(K^\cdot(t))] dt \right)^2 \frac{du}{\Delta}. \end{aligned} \quad (5.2.11)$$

Theorem 5.2 (Option price approximation). *Let Assumptions 5.1, 5.2, and 5.3 be in force, and let $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ be a θ -Hölder continuous function for some $\theta \in (0, 1]$. The price of the VIX option with payoff $\varphi(\text{VIX}_T^2)$ satisfies*

$$\mathbb{E}[\varphi(\text{VIX}_T^2)] = \mathbb{E}[\varphi(\text{VIX}_{T,P}^2)] + \sum_{i=1}^3 \gamma_i \partial_\varepsilon^i \mathbb{E}[\varphi(\text{VIX}_{T,P}^2 e^\varepsilon)]|_{\varepsilon=0} + \mathcal{E}_\varphi, \quad (5.2.12)$$

where \mathcal{E}_φ is an error term such that $|\mathcal{E}_\varphi| \leq c \Delta^{3(d_1 \wedge \frac{d_2}{2})}$.

Note that VIX futures correspond to $\varphi(x) = \sqrt{x}$, VIX put options to $\varphi(x) = (\kappa - \sqrt{x})^+$ and call options to $\varphi(x) = (\sqrt{x} - \kappa)^+$. In all these cases, the function φ is $\frac{1}{2}$ -Hölder (and φ is even Lipschitz in the case of call options with strictly positive strike κ , see Remark 3.2).

Remark 5.2. *Although the payoff φ may fail to be smooth, condition (5.2.7) ensures that the lognormal proxy $\text{VIX}_{T,P}^2$ is non-degenerate, with the effect of regularizing the map $\varepsilon \mapsto \mathbb{E}[\varphi(\text{VIX}_{T,P}^2 e^\varepsilon)]$, so that the derivatives $\partial_\varepsilon^i \mathbb{E}[\varphi(\text{VIX}_{T,P}^2 e^\varepsilon)]$ are well-defined.*

Remark 5.3. *The adimensional coefficients $(\gamma_i)_{i \in \{1,2,3\}}$ are defined by deterministic integrals with respect to time variables. They depend on the option maturity T , on the time window Δ , and on the model parameters ξ_0^u and K , but not on the option payoff – which means that, in the case of call and put options, they can be evaluated once and for all strikes. Assuming (as we did in the present chapter) that the initial forward variance curve $u \mapsto \xi_0^u$ is constant over the VIX time window $(T, T + \Delta)$, the γ_i 's actually have analytical closed-form expressions in one-factor the Bergomi model (1.2.11), see [20, Proposition 7] for precise expressions. In the rough Bergomi model (1.2.12), the γ_i 's do not seem to admit a closed-form representation even when the initial variance curve ξ_0^u is constant, but in this case they can be written as separable functions of the remaining model parameters η and H with an explicit dependence with respect to η , see [20, Remark 12] for details.*

As a direct corollary of Theorem 5.2, we obtain explicit expansion formulas for the price of calls, puts, and futures on VIX_T , for which φ is $\frac{1}{2}$ -Hölder continuous and all the terms in (5.2.12) (apart from the error term) are explicit – precise expressions are given in [20, Corollary 6].

The assumptions of Theorem 5.2 are satisfied in our main examples. In our reference examples – the Bergomi model (1.2.11) and the rough Bergomi model (1.2.12) – we are able to check that the assumptions of Theorem 5.2 hold true, with a precise knowledge of the rates d_1 and d_2 appearing in Assumption 5.3.

Proposition 5.1. *In the one-factor Bergomi model (1.2.11) with parameter $k > 0$, Assumption 5.3 holds with $d_1 = 1$ and $d_2 = 2$. Consequently, Theorem 5.2 holds and the error term \mathcal{E}_φ in (5.2.12) is $\mathcal{O}(\Delta^3)$.*

Proposition 5.2. *Assume $H \in (0, 1) \setminus \{\frac{1}{2}\}$. In the rough Bergomi model, Assumption 5.3 holds with $d_1 = 1 \wedge 2H$ and $d_2 = 2H$. Consequently, Theorem 5.2 holds and the error term \mathcal{E}_φ in the expansion (5.2.12) is $\mathcal{O}(\Delta^{3H})$.*

Remark 5.4. *(Limiting case : constant kernel) When the kernel K is constant, which corresponds to $H = \frac{1}{2}$ in the rough Bergomi model and to $k = 0$ in the Bergomi model, we have $VIX_T^2 = VIX_{T,P}^2$. Correspondingly, in this case $\gamma_i = 0$ for every $i \in \{1, 2, 3\}$, and the expansion (5.2.12) holds with zero error term $\mathcal{E}_\varphi = 0$.*

The numerical tests provided in [20, section 2.5] show that the relative errors of the approximation formulas provided by Theorem 5.2 are extremely accurate – absolute relative errors for the rough Bergomi model with realistic parameters are less than 0.5% for futures contracts and less than 1.5% for at-the-money options, while for the standard Bergomi model (still with typical market-implied parameter values), relative errors are even smaller, less than (an impressive) $10^{-3}\%$ for the VIX futures, and less than 1% for at-the-money options. Instead of displaying these results here, we prefer to move to an enhanced version of the model class (1.2.10), for which we are also able to provide accurate approximation formulas.

5.3 Mixed (or “skewed”) exponential models

A class of models more general than (1.2.10), and capable of appropriately capturing the behavior of market VIX smiles, was introduced in Bergomi [15], who observed that a simple yet efficient way to twist the distribution of forward variances is to replace the exponential process $\xi_t^u = \xi_0^u e^{Y_t^u}$ stemming from (1.2.10) with a convex combination of two exponential functions. This observation yields the extended model family

$$\xi_t^u = \xi_0^u \left[\lambda e^{\alpha_1 Z_T^u - \frac{1}{2} \alpha_1^2 \text{Var}(Z_T^u)} + (1 - \lambda) e^{\alpha_2 Z_T^u - \frac{1}{2} \alpha_2^2 \text{Var}(Z_T^u)} \right] \quad (5.3.1)$$

where Z_T^u is the common Gaussian factor

$$Z_T^u = \int_0^T K^u(s) dW_s, \quad (5.3.2)$$

the positive parameters $\alpha_i > 0$ allow to tune the volatility of variances, and $\lambda \in [0, 1]$ plays the role of a mixing coefficient.

Remark 5.5. When $K^u(t) = e^{-k(u-t)}$, (5.3.1) corresponds to the mixed one-factor Bergomi model introduced in [15], and when $K^u(t) = (u-t)^{H-\frac{1}{2}}$, (5.3.1) yields the so-called mixed rough Bergomi model, introduced simultaneously by De Marco [32] and Guyon [56].

The squared VIX is of course still defined by integrated instantaneous forward variances, we see that VIX_T^2 is given by a convex combination of integral means of the form (5.1.2), that is

$$\text{VIX}_T^2 = \lambda \text{VIX}_{T,1}^2 + (1-\lambda) \text{VIX}_{T,2}^2 := \lambda \frac{1}{\Delta} \int_T^{T+\Delta} \xi_0^u e^{Y_{T,1}^u} du + (1-\lambda) \frac{1}{\Delta} \int_T^{T+\Delta} \xi_0^u e^{Y_{T,2}^u} du, \quad (5.3.3)$$

where

$$Y_{T,j}^u := \int_0^T K_j^u(t) dW_t - \frac{1}{2} \int_0^T K_j^u(t)^2 dt, \quad j \in \{1, 2\}, \quad (5.3.4)$$

and

$$K_j^u(t) = \alpha_j K^u(t).$$

Following the approach of section 5.2, we approximate each integral mean with a log-normal random variable, so that VIX_T^2 is eventually approximated by

$$\begin{aligned} \text{VIX}_{T,P}^2 &= \lambda \xi_0 e^{\frac{1}{\Delta} \int_T^{T+\Delta} Y_{T,1}^u du} + (1-\lambda) \xi_0 e^{\frac{1}{\Delta} \int_T^{T+\Delta} Y_{T,2}^u du} \\ &=: \lambda \text{VIX}_{T,P,1}^2 + (1-\lambda) \text{VIX}_{T,P,2}^2. \end{aligned}$$

The overall proxy $\text{VIX}_{T,P}^2$ is therefore a convex combination of correlated log-normal random variables. For $j \in \{1, 2\}$, we have

$$\ln(\text{VIX}_{T,P,j}^2) \stackrel{d}{=} \mathcal{N}(\mu_{P,j}, \sigma_{P,j}^2), \quad (5.3.5)$$

where $\mu_{P,j} := -\frac{1}{2} \int_0^T \nu_0(K_j(t)^2) dt$, $\sigma_{P,j}^2 := \int_0^T \nu_0(K_j(t)^2) dt$.

We have the following

Theorem 5.3. Let $\varphi \in \mathcal{C}_b^2$. In the mixed rough Bergomi model obtained setting $K^u(t) = (u-t)^{H-\frac{1}{2}}$ and $\alpha_j = \eta_j$ in (5.3.1)-(5.3.2), the price of the option with payoff $\varphi(\text{VIX}_T^2)$ is given by

$$\mathbb{E}[\varphi(\text{VIX}_T^2)] = \mathbb{E}[\varphi(\text{VIX}_{T,P}^2)] + \sum_{i=1}^3 \sum_{j=1}^2 \gamma_{i,j} P_{i,j} + \mathcal{E}_\varphi, \quad (5.3.6)$$

where \mathcal{E}_φ is an error term satisfying $|\mathcal{E}_\varphi| \leq c \Delta^{3(d_1 \wedge \frac{d_2}{2})}$ with d_1, d_2 given in (5.2.5)-(5.2.6), and

$$\mathbb{E}[\varphi(\text{VIX}_{T,P}^2)] = \mathbb{E}[\varphi(\nu(\xi_0) [\lambda e^{\mu_{P,1} + \sigma_{P,1} Z} + (1-\lambda) e^{\mu_{P,2} + \sigma_{P,2} Z}])], \quad (5.3.7)$$

$$P_{i,j} = \partial_\varepsilon^{i-1} \mathbb{E}[\Psi_j(\mu_{P,j} + \sigma_{P,j} Z + \varepsilon)]|_{\varepsilon=0}, \quad i \in \{1, 2, 3\}, \quad j \in \{1, 2\},$$

$$\Psi_1(x) = \partial_y \varphi \left(\nu(\xi_0) \left[\lambda e^{x+y} + (1-\lambda) e^{\frac{\eta_2}{2}(\eta_1 - \eta_2) \int_0^T \nu_0(K(t)^2) dt + \frac{\eta_2}{\eta_1} x} \right] \right) \Big|_{y=0}, \quad (5.3.8)$$

$$\Psi_2(x) = \partial_y \varphi \left(\nu(\xi_0) \left[\lambda e^{\frac{\eta_1}{2}(\eta_2 - \eta_1) \int_0^T \nu_0(K(t)^2) dt + \frac{\eta_1}{\eta_2} x} + (1-\lambda) e^{x+y} \right] \right) \Big|_{y=0}. \quad (5.3.9)$$

In (5.3.6), the coefficients $\gamma_{1,j}$, $\gamma_{2,j}$ and $\gamma_{3,j}$, $j \in \{1, 2\}$, are defined as in (5.2.11), replacing the kernel K with the corresponding K_j . A similar expansion holds for the mixed standard Bergomi model, taking $K^u(t) = e^{-k(u-t)}$ and replacing η_i with ω_i for $i \in \{1, 2\}$ in (5.3.8) and (5.3.9).

We note that the form of (5.3.7)-(5.3.8)-(5.3.9) is specific to the presence of a single Gaussian factor in (5.3.1) – in other words, specific to the mixed one-factor Bergomi model where $K_i^u(t) = \omega_i e^{-k(u-t)}$ (same value of k for the two kernels) and to the rough Bergomi model where $K_i^u = \eta_i(u-t)^{H-\frac{1}{2}}$ (same value of H). In these cases, the VIX proxy is a function of a single Gaussian random variable,

$$\text{VIX}_{T,P}^2 \stackrel{d}{=} \xi_0 \left[\lambda e^{\mu_{P,1} + \sigma_{P,1} Z} + (1 - \lambda) e^{\mu_{P,2} + \sigma_{P,2} Z} \right], \quad Z \stackrel{d}{=} \mathcal{N}(0, 1),$$

so that all the expressions in Theorem 5.3 can be evaluated with efficient one-dimensional Gaussian quadratures (we refer to [20, section 3.2] for more details).

5.3.1 Numerical tests for option price formulas and implied volatilities

Reference prices in the mixed rough Bergomi model are evaluated according to the Monte Carlo procedure described in Chapter 3 : we discretize the $\text{VIX}_{T,j}^2$ in (5.3.3) for $j \in \{1, 2\}$ with a rectangle scheme and simulate exactly the discretized variable for a large number $n = 300$ of discretization points. In the mixed standard Bergomi model, we exploit the Markovian representation in Remark 5.1 for each term $\text{VIX}_{T,j}^2$, $j \in \{1, 2\}$, and apply a two-dimensional deterministic quadrature with respect to the parameter u and to the space dimension.

We compare the resulting reference VIX implied volatilities with the approximate implied volatilities computed with our expansion in Theorem 5.3. To test the approximation formulas on different VIX smiles, we consider two different parameter scenarios in the rough and standard Bergomi models. We have considered options maturities equal to 1, 3, and 6 months.

Implied volatility for the mixed rough Bergomi model. Recall that the mixed rough Bergomi model is obtained by injecting $K^u = (u-t)^{H-\frac{1}{2}}$ and $\alpha_i = \eta_i$, $i \in \{1, 2\}$, in (5.3.1). We set $\xi_0 = 0.235^2$ and $H = 0.1$; the other model parameters can be found in Table 5.1. We evaluate the reference option prices with 10^6 Monte Carlo samples and 300 discretization points.

Scenario	1-month VIX futures	3-month VIX futures	6-month VIX futures
1	$0.218650 \pm 5 \times 10^{-6}$	$0.206308 \pm 5 \times 10^{-6}$	$0.196890 \pm 5 \times 10^{-6}$
2	$0.229001 \pm 3 \times 10^{-6}$	$0.224244 \pm 3 \times 10^{-6}$	$0.220472 \pm 3 \times 10^{-6}$

Scenario	η_1	η_2	λ
1	1.4	0.7	0.3
2	0.9	0	0.6

TABLE 5.1 – Term structure of VIX futures and model parameters for scenarios 1 and 2 in the mixed rough Bergomi model.

In Figure 5.1 we can see that, in the mixed model as well, our approximation formula proves to be very accurate – the reference and the approximate implied volatilities are barely distinguishable on the graphs. The related absolute relative errors are less than 1.6% for the parameter scenario 1 and less than 0.9% for scenario 2.

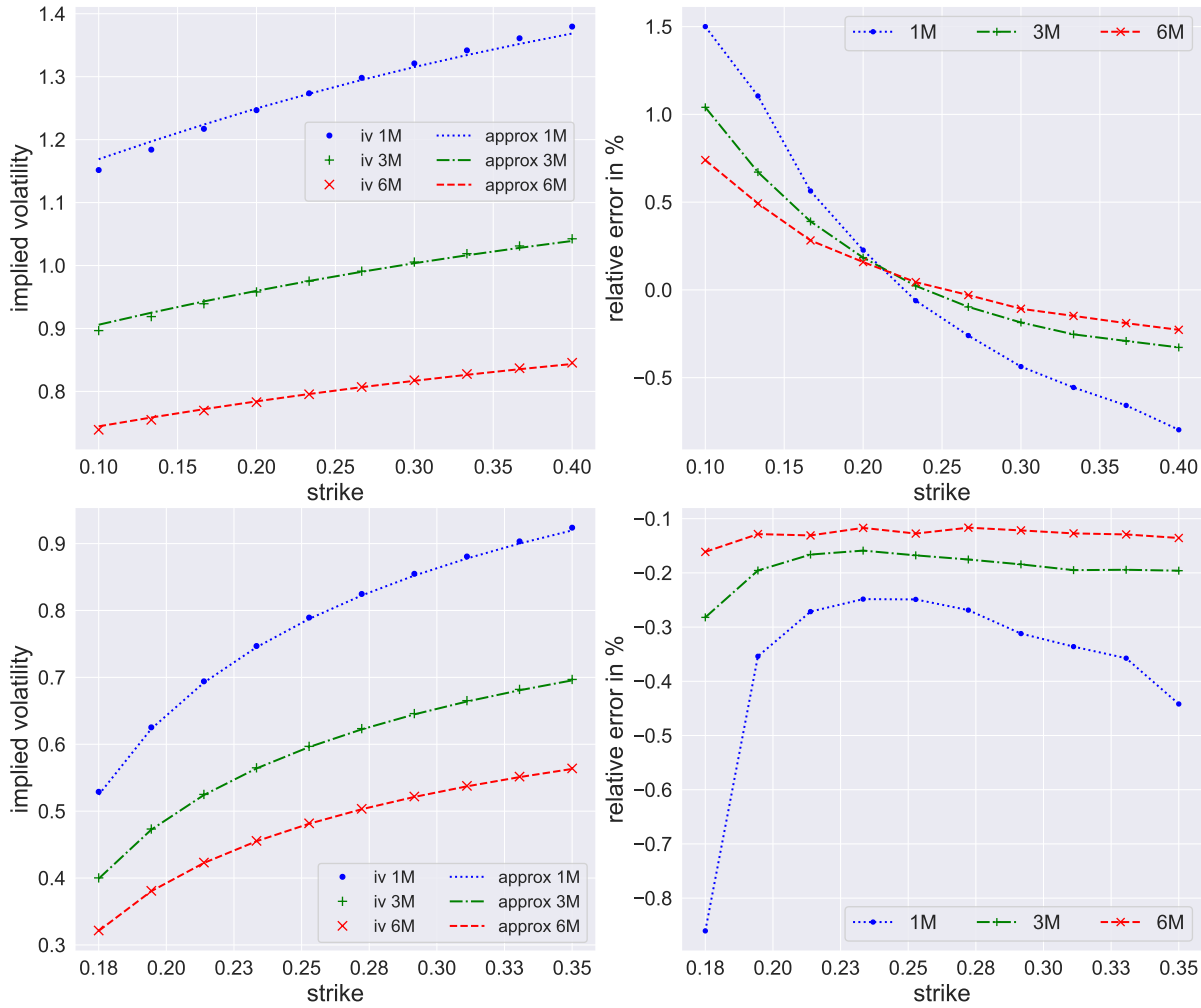


FIGURE 5.1 – VIX smiles in the mixed rough Bergomi model for $T = 1, 3, 6$ months (left), and corresponding percentage relative errors between the reference implied volatilities and their approximations from Theorem 5.3 (right), for parameter scenarios 1 (top figures) and 2 (bottom figures).

Implied volatility for the mixed standard Bergomi model. We perform a similar numerical analysis for the mixed one-factor standard Bergomi model, obtained setting $K^u(t) = e^{-k(u-t)}$ and $\alpha_i = \omega_i, i \in \{1, 2\}$, in (5.3.1). We set $\xi_0 = 0.2^2$ and $k = 1$. The other model parameters are given in Table 5.2.

Scenario	1-month VIX futures	3-month VIX futures	6-month VIX futures
3	0.172764	0.145976	0.130503
4	0.181527	0.165480	0.155141

Scenario	ω_1	ω_2	λ
3	0.5	6	0.3
4	10	2	0.2

TABLE 5.2 – Term structure of VIX futures and model parameters for scenarios 3 and 4 in the mixed standard Bergomi model.

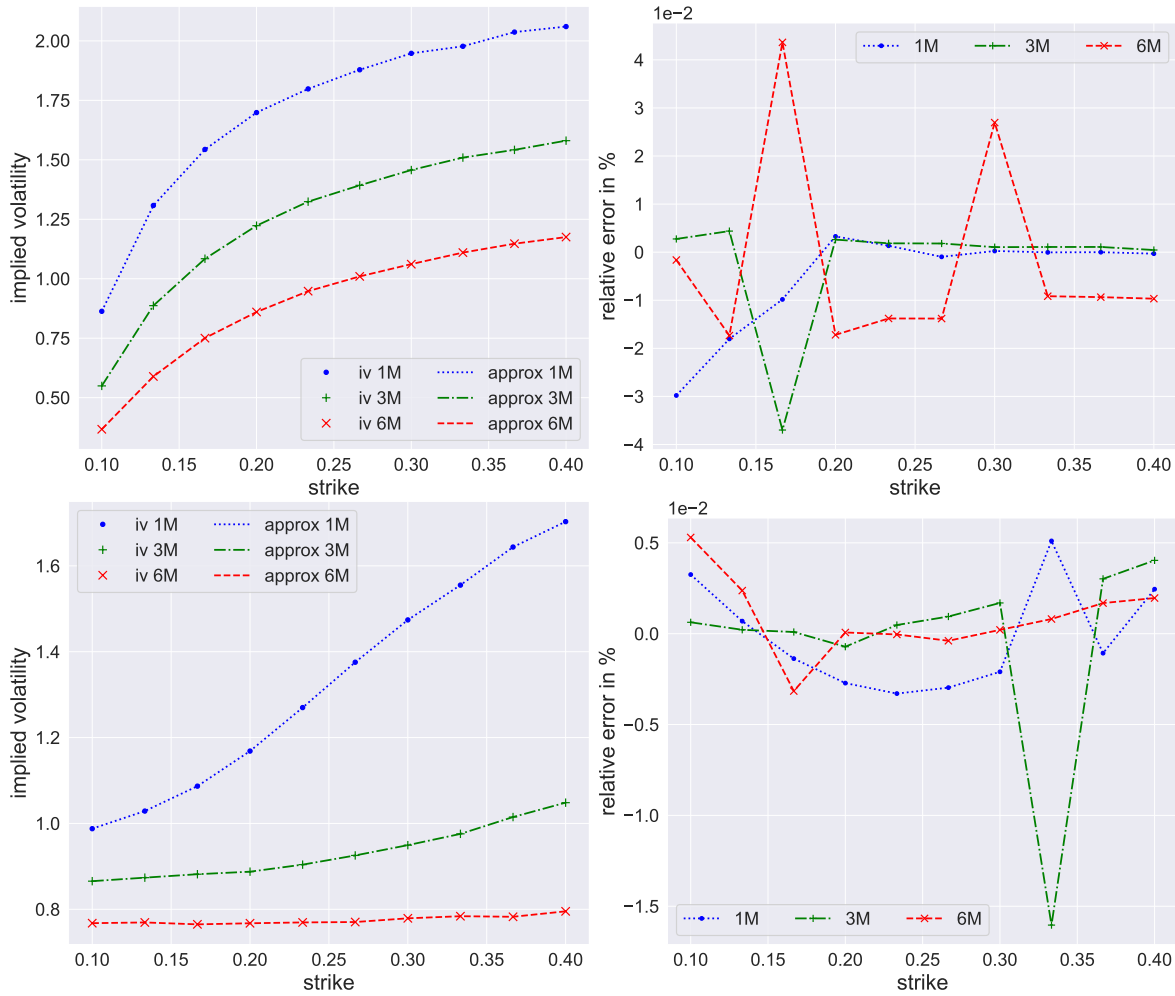


FIGURE 5.2 – VIX smiles in the mixed standard Bergomi model for $T = 1, 3, 6$ months (left figures), and relative errors between the reference implied volatilities and their approximations (right figures) for parameter scenarios 3 (top figures) and 4 (bottom figures).

As seen in Figure 5.2, the approximation formula from Theorem 5.3 turns out to be extremely accurate for the mixed Bergomi model, too (absolute relative errors are now smaller than $5 \times 10^{-2} \%$ for parameter scenario 3 and less than $2 \times 10^{-2} \%$ for scenario 4).

A calibration test of both models to VIX market data, performed in [20, section 3.3], illustrates the capability of the mixed exponential class (5.3.1) to appropriately fit the VIX implied volatility surface. The use of our pricing formula (5.3.6) of course offers a considerable gain in terms of execution time with respect to the reference pricing methods (in our tests, the full VIX implied volatility calibration with our expansion (5.3.6) was 3.5 times faster than the calibration based on two-dimensional quadrature for the mixed Bergomi model, and tens of times faster than the Monte Carlo-based calibration for the mixed rough Bergomi model).

Troisième partie

**Asymptotic results for stochastic
volatility**

Asymptotics of the local volatility surface in stochastic volatility models

Contents

6.1 The local volatility surface	63
6.1.1 A related model-free result : the harmonic mean property of the implied volatility . . .	65
6.2 Local volatility under rough volatility	65
6.2.1 Our toolbox : Large Deviations	67
6.2.2 Our main result : asymptotics of the Markovian projection in the large deviations regime	69
6.2.3 The consequence for the short-time local volatility skew	70
6.2.4 Numerical tests	72

In this chapter, we wish to understand the behavior of the local volatility surface generated by a – possibly rough – stochastic volatility model, at least asymptotically for short maturities and around the ATM point. In particular, we will examine the validity of traditional approximations linking the implied and the local volatility surfaces, such as the harmonic mean formula and the 1/2 skew rule, in the context of rough stochastic volatility. An application of our results will be the extrapolation of local volatility surfaces for very short maturities, as we discuss in section 6.2.3. The results presented in this chapter are taken from our works De Marco and Friz [29], De Marco, Friz and Gerhold [30], De Marco [28], and Bourgey, De Marco, Friz and Pigato [19].

6.1 The local volatility surface

Let $(t, K) \mapsto P(t, K)$ be an arbitrage-free put price surface on a single asset S , parametrized by the option maturity t and strike price K . We assume that P is smooth, namely $P \in C^{1,2}((0, \infty) \times (0, \infty))$. Let us recall Dupire’s formula [35] defining the the local volatility function σ_{loc} from the put price surface P (assuming zero interest, repository and dividend rates for the ease of notation) :

$$\sigma_{\text{loc}}(t, k)^2 = 2 \frac{\partial_t P(t, K)}{K^2 \partial_{KK} P(t, K)} \Big|_{K=S_0 e^k} \quad t > 0, k \in \mathbb{R}. \tag{6.1.1}$$

Note we define σ_{loc} as a function of time t and log-strike $k = \ln \frac{K}{S_0} \in \mathbb{R}$. The celebrated result by Dupire [35] states that the local volatility model $\bar{S}_t = S_0 + \int_0^t \bar{S}_u \sigma_{\text{loc}}\left(u, \log\left(\frac{\bar{S}_u}{S_0}\right)\right) dB_u$ admits a weak solution that reproduces the put option prices we started from, in the sense that $\mathbb{E}[(K - \bar{S}_t)^+] = P(t, K)$ for every t and K . Some regularity conditions are required for this construction to make sense; without ambitions to provide minimal requirements, let us mention that it will be enough that the function $\sigma_{\text{loc}}(t, k)$ is bounded

and Lipschitz with respect to k uniformly in t (see [66] for a more general construction, related to Kellerer's Theorem).

Let $\sigma_{\text{BS}}(t, k)$ denote the Black-Scholes implied volatility of the put price $P(t, K)|_{K=S_0 e^k}$. Injecting the identity $P(t, K) = P_{\text{BS}}(t, K, \sigma_{\text{BS}}(t, k))$ inside (6.1.1), we obtain Dupire's formula written in terms of the implied volatility σ_{BS} ,

$$\sigma_{\text{loc}}(t, k)^2 = \frac{\sigma_{\text{BS}}(t, k) + 2t \partial_t \sigma_{\text{BS}}(t, k)}{\left(t \partial_{kk} \sigma_{\text{BS}} - \frac{1}{4} t^2 \sigma_{\text{BS}} (\partial_k \sigma_{\text{BS}})^2 + \frac{1}{\sigma_{\text{BS}}} \left(1 - \frac{k \partial_k \sigma_{\text{BS}}}{\sigma_{\text{BS}}} \right)^2 \right) (t, k)}. \quad (6.1.2)$$

The computation leading from (6.1.1) to (6.1.2) is standard; details can be found for example in Lee [76].

Short-time limit and the harmonic mean formula. Formally taking the short maturity limit $t \rightarrow 0$ in (6.1.2), and assuming that the partial derivative $\partial_t \sigma_{\text{BS}}$, $\partial_k \sigma_{\text{BS}}$ and $\partial_{kk} \sigma_{\text{BS}}$ remain bounded as $t \rightarrow 0$, one obtains the equation

$$\sigma_{\text{loc}}(0, k)^2 = \frac{\sigma_{\text{BS}}(0, k)^2}{\left(1 - \frac{k \sigma'_{\text{BS}}(0, k)}{\sigma_{\text{BS}}(0, k)} \right)^2}, \quad (6.1.3)$$

relating the short-time limits $\sigma_{\text{loc}}(0, k) := \lim_{t \rightarrow 0} \sigma_{\text{loc}}(t, k)$ and $\sigma_{\text{BS}}(0, k) := \lim_{t \rightarrow 0} \sigma_{\text{BS}}(t, k)$ of the local and implied volatility surfaces, where we denote $\sigma'_{\text{BS}}(0, k) = \partial_k \sigma_{\text{BS}}(0, k)$. As already noted by Lee [76], if we think the local volatility σ_{loc} to be given, equation (6.1.3) can be seen as an ODE for the function $\sigma_{\text{BS}}(0, \cdot)$. It is a simple exercise to check that (6.1.3) is solved by the harmonic mean function

$$\sigma_{\text{BS}}(0, k) = H(0, k) := \frac{1}{\frac{1}{k} \int_0^k \frac{1}{\sigma_{\text{loc}}(0, y)} dy}, \quad k \in \mathbb{R},$$

which suggests that we can expect the harmonic mean $H(t, k) = \frac{1}{\frac{1}{k} \int_0^k \frac{1}{\sigma_{\text{loc}}(t, y)} dy}$ of the short-dated the local volatility $\sigma_{\text{loc}}(t, \cdot)$ to be a good approximation of the short-dated implied volatility $\sigma_{\text{BS}}(t, \cdot)$. A rigorous proof of the asymptotic statement

$$\sigma_{\text{BS}}(t, k) \underset{t \rightarrow 0}{\sim} H(t, k) \quad (6.1.4)$$

can be found in Berestycki et al. [12], or in Gatheral et al. [46] based on small-time heat kernel asymptotics.

The "1/2 skew" rule of thumb. Assuming the local volatility function is differentiable at $k = 0$, it is not difficult to see that, for fixed maturity t , the harmonic mean $H(t, k)$ satisfies the property

$$\partial_k H(t, k)|_{k=0} = \frac{1}{2} \partial_k \sigma_{\text{loc}}(t, k)|_{k=0}.$$

If we assume that the short-time approximation property $\sigma_{\text{BS}}(t, k) \approx H(t, k)$ from (6.1.4) also holds for the first derivative with respect to k , we obtain the so-called 1/2 skew rule

$$\partial_k \sigma_{\text{BS}}(t, k)|_{k=0} \approx \frac{1}{2} \partial_k \sigma_{\text{loc}}(t, k)|_{k=0} \quad \text{as } t \rightarrow 0, \quad (6.1.5)$$

in other words

$$\frac{\text{ATM implied vol skew}}{\text{ATM local vol skew}} \approx \frac{1}{2} \quad \text{as } t \rightarrow 0.$$

6.1.1 A related model-free result : the harmonic mean property of the implied volatility

With a view on the asymptotic approximation (6.1.4), it is worth to point out the following model-free result proved in [28] : any arbitrage-free implied volatility is in fact the harmonic mean of a positive function, for *any* fixed maturity (and not only in the small-maturity limit).

Theorem 6.1 (Theorem 2.4 in [28]). *Let σ_{BS} be an arbitrage-free implied volatility surface satisfying $\sigma_{\text{BS}}(t, k) > 0$ for every $t > 0$ and $k \in \mathbb{R}$, and $k \mapsto \sigma_{\text{BS}}(t, k) \in C^1(\mathbb{R})$. Then, there exists a unique strictly positive function $\mathfrak{s} : (0, \infty) \times \mathbb{R} \rightarrow \mathbb{R}$, with $k \mapsto \mathfrak{s}(t, k) \in C^0(\mathbb{R})$, such that $\sigma_{\text{BS}}(t, \cdot)$ is the harmonic mean of $\mathfrak{s}(t, \cdot)$:*

$$\sigma_{\text{BS}}(t, k) = \frac{1}{\frac{1}{k} \int_0^k \frac{1}{\mathfrak{s}(t, y)} dy} \quad \forall k \neq 0, \quad \forall t > 0, \quad (6.1.6)$$

and $\sigma_{\text{BS}}(t, 0) = \mathfrak{s}(t, 0)$.

The proof of Theorem 6.1 is based on Fukasawa's seminal result [40] about the strict monotonicity of the Black-Scholes maps $k \mapsto d_0(k) = -\frac{k}{\sqrt{t}\sigma_{\text{BS}}(t, k)} - \frac{\sqrt{t}\sigma_{\text{BS}}(t, k)}{2}$ and $k \mapsto d_1(k) = -\frac{k}{\sqrt{t}\sigma_{\text{BS}}(t, k)} + \frac{\sqrt{t}\sigma_{\text{BS}}(t, k)}{2}$, which holds for every t and for any arbitrage-free implied volatility σ_{BS} .

The link of the function \mathfrak{s} in the harmonic mean representation (6.1.6) with the local volatility σ_{loc} associated to σ_{BS} can be inspected starting from Dupire's formula (6.1.2). As expected, under appropriate regularity conditions on the behavior of σ_{BS} for small maturities (which are actually not met in the case of rough volatility models! See our results and discussion in the next section), the two functions \mathfrak{s} and σ_{loc} can be identified in the small time limit, but are different otherwise, see [28, Proposition 4.2 and Remark 4.3].

6.2 Local volatility under rough volatility

For reference, we recall the class of stochastic volatility models (1.2.8)-(1.2.10) induced by the forward variance modeling framework considered in section 1.2,

$$\begin{aligned} dS_t &= S_t \sqrt{V_t} \left(\rho dW_s + \sqrt{1 - \rho^2} d\bar{W}_s \right) \\ V_t &= \xi_0^t e^{\eta \int_0^t K(t, s) dW_s - \frac{\eta^2}{2} \int_0^t K(t, s)^2 ds}, \end{aligned} \quad (6.2.1)$$

where W, \bar{W} are two independent standard Brownian motions and ξ_0^t the deterministic initial forward variance curve (a market parameter). Note that, with respect to the notation in section 1.2, we have factored the dependence with respect to the vol-of-variance parameter η out of the kernel K . As already seen in previous chapters, our leading example is obtained when

$$K(t, s) = \frac{\sqrt{2H}}{(t - s)^{1/2-H}}, \quad H < 1/2.$$

The corresponding Gaussian process

$$\widehat{W}_t = (K * \dot{W})_t = \int_0^t K(t, s) dW_s \quad (6.2.2)$$

is the Riemann-Liouville Brownian motion, and in this case (6.2.1) corresponds to the rough Bergomi model of Bayer et al. [9].

In view of small-time asymptotic results, an important property of \widehat{W}_t that we will exploit is self-similarity

$$(\widehat{W}_{\varepsilon t}, t \geq 0) \stackrel{\text{law}}{=} \varepsilon^H (\widehat{W}_t, t \geq 0).$$

Our precise assumptions on the model coefficients. Our main result for this section (Theorem 6.2 below) holds for the class of models

$$\begin{aligned} dX_t &= -\frac{1}{2}V_t dt + \sqrt{V_t} \left(\rho dW_s + \sqrt{1-\rho^2} d\overline{W}_s \right) \\ V_t &= \sigma^2 \left(\int_0^t K(t,s) dW_s \right) = \sigma^2(\widehat{W}_t), \end{aligned} \tag{6.2.3}$$

where $X_t = \log \frac{S_t}{S_0}$ denotes the asset log-price and the smooth function $\sigma : \mathbb{R} \rightarrow \mathbb{R}$ satisfies the following

Assumption 6.1. *There exist $c_1, c_2, c_3, c_4 > 0$ such that*

(C1) $\sigma(x) \leq c_3 \exp(c_4|x|)$ (exponential upper bound)

(C2) $\sigma(x) \geq c_1 \exp(-c_2|x|)$ (exponential lower bound)

for all $x \in \mathbb{R}$.

Our choice of kernel is the Riemann-Liouville kernel $K(t,s) = \frac{\sqrt{2H}}{(t-s)^{1/2-H}}$ with $H \leq 1/2$, even if we suspect that our result could be extended to general self-similar Volterra processes.

With respect to the base model (6.2.1), we are able to handle more general functions than the exponential $\sigma^2(x) = e^x$, even if we do not include an explicit time dependence in the function σ – strictly speaking then, the model (6.2.1) where $V_t = \sigma^2(t, \widehat{W}_t) = \xi_0^t e^{\eta \widehat{W}_t - \frac{1}{2} \eta^2 t^{2H}}$ is not covered by our main Theorem 6.2, even if we can (and do) perform numerical tests for (6.2.1) indicating that Theorem 6.2 and its corollaries seem to hold in this slightly more general setting as well.

Local volatility as Markovian projection. It is well-known that, when option price surface P is generated by a stochastic volatility model such as (6.2.3), the local volatility function (6.1.1) has a representation as a conditional expectation [60, 23]

$$\sigma_{\text{loc}}(t, k)^2 = \mathbb{E}[V_t | X_t = k], \quad k \in \mathbb{R}, \tag{6.2.4}$$

which is usually referred to as “Markovian projection” of the stochastic volatility. Weak solvability and the marginal mimicking-property of the local volatility model with diffusion coefficient defined by (6.2.4) was proven in the seminal work of Gyöngy [60] under the assumptions that the original stochastic volatility model has bounded drift and diffusion coefficients and uniformly elliptic diffusion matrix (which is not the case for (6.2.3)), and then extended by Brunick and Shreve [23] who removed the boundedness and ellipticity assumptions in Gyöngy’s result.

Additional notation for this section. The theory of pathwise Large Deviations we are going to exploit in this section requires to introduce some additional notation. We denote AC the space of two-dimensional absolutely continuous paths $\left\{ (h, \bar{h}) : [0, 1] \rightarrow \mathbb{R}^2 : h_t = \int_0^t \dot{h}_s ds, \bar{h}_t = \int_0^t \dot{\bar{h}}_s ds \right\}$, and H^1 the Cameron–Martin space

$$H^1 = \left\{ (h, \bar{h}) \in AC : \dot{h}, \dot{\bar{h}} \in L^2(0, 1) \right\}$$

equipped with the Cameron–Martin norm $\|h, \bar{h}\|_{H^1}^2 = \int_0^1 (\dot{h}^2 + \dot{\bar{h}}^2) dt$. Using a notation analogous to (6.2.2), we denote

$$\hat{h}_t = (K^H * \dot{h})_t = \sqrt{2H} \int_0^t (t-s)^{H-1/2} dh_s$$

the “fractional” path associated to h .

6.2.1 Our toolbox : Large Deviations

We aim at short-time asymptotic results, which is why we target the family of time-changed processes $(X_{\varepsilon^{2t}})_{t \geq 0}$ as $\varepsilon \rightarrow 0$. Due to the time-scaling property of the Gaussian noise $\mathbf{W} = (W, \overline{W}, \widehat{W})$ driving the stochastic system (6.2.3), we have

$$(X_{\varepsilon^{2t}}, t \geq 0) \stackrel{d}{=} (X_t^\varepsilon, t \geq 0),$$

where the process X_t^ε now solves

$$X_t^\varepsilon = \int_0^t \sigma(\varepsilon^{2H} \widehat{W}_s) \varepsilon d\left(\rho dW_s + \sqrt{1 - \rho^2} d\overline{W}_s\right) - \frac{1}{2} \varepsilon^2 \int_0^t \sigma^2(\varepsilon^{2H} \widehat{W}_s) ds. \quad (6.2.5)$$

The different time scales ε for the Brownian component (W, \overline{W}) and ε^{2H} for the fractional component \widehat{W} both come into play in (6.2.5). In the framework of Large Deviations theory, we are able to transfer a large deviation principle from the Gaussian process $\varepsilon \mathbf{W}$ to the process X , but in the presence of a single scaling parameter $\hat{\varepsilon}$. It is still possible to apply this procedure to (6.2.5), but in order to do so, we have to rescale the space dimension as well, considering the new process

$$\tilde{X}_t^\varepsilon = \varepsilon^{2H-1} X_t^\varepsilon = \int_0^t \sigma(\hat{\varepsilon} \widehat{W}_s) \hat{\varepsilon} d\left(\rho dW_s + \sqrt{1 - \rho^2} d\overline{W}_s\right) - \frac{1}{2} \hat{\varepsilon}^{1 + \frac{1}{2H}} \int_0^t \sigma^2(\hat{\varepsilon} \widehat{W}_s) ds, \quad (6.2.6)$$

where now $\hat{\varepsilon} = \varepsilon^{2H}$.

It is well-known that a small noise Large Deviation Principle (LDP) holds for the family $(\tilde{X}_1^\varepsilon)_\varepsilon$ (hence for $\varepsilon^{2H-1} X_{\varepsilon^2}$) as $\varepsilon \rightarrow 0$, with speed $\hat{\varepsilon}^2 = \varepsilon^{4H}$ and rate function

$$\Lambda(x) = \inf \left\{ \frac{1}{2} \|h, \bar{h}\|_{H^1}^2 : \Phi_1(h, \bar{h}) = \int_0^1 \sigma(\hat{h}_t) d(\rho h_t + \sqrt{1 - \rho^2} \bar{h}_t) = x \right\}. \quad (6.2.7)$$

This result was first proved in the case where $V_t = \sigma(\widehat{W}_t)$ and σ is a function with linear growth in Forde and Zhang [38]. The more general exponential growth condition (C1) in Assumption 6.1 is no obstruction for an LDP to hold for the model (6.2.3), as was shown in Gulisashvili [54], where functions $\sigma(\cdot)$ with exponential growth are taken into account. Note that, in the deterministic map Φ_1 , the drift part of the stochastic equation (6.2.6) has vanished. This is typical phenomenon in small-time large deviations for SDEs : the time scale of the drift ($\hat{\varepsilon}^{1 + \frac{1}{2H}}$) is much faster than the time scale of the stochastic integral part ($\hat{\varepsilon}^{2H}$), with the consequence that the drift part of (6.2.6) is negligible in the limit.

It is clear that $\Lambda(x)|_{x=0} = 0$, for in this case the Cameron-Martin path (or optimal control) attaining the minimum in (6.2.7) is simply the null path $h = \bar{h} \equiv 0$. For other values of $x \neq 0$, as soon as $\Lambda(x) < \infty$, it follows from the lower semi-continuity of the Cameron-Martin norm $\|\cdot\|_{H^1}^2$, the compactness of its level sets and the continuity of the map Φ_1 that existence of a minimizer holds for (6.2.7) : we have

$$\Lambda(x) = \frac{1}{2} \|h^x, \bar{h}^x\|_{H^1}^2$$

for some element (h^x, \bar{h}^x) of H^1 . Moreover, for x small enough, such a minimizer is known to be unique [39] : we have $\Lambda(x) = \frac{1}{2} \|h^x, \bar{h}^x\|_{H^1}^2$ for a unique optimal control (h^x, \bar{h}^x) .

The time-space scaling described above naturally leads to consider asymptotics of distribution functions (for the original process X), of option prices and volatility surfaces along a curve in the maturity-log-strike

plane (t, k) , parametrized by H . Note indeed that, roughly speaking, fixing the value of \tilde{X}_1^ε in (6.2.6) amounts to

$$\tilde{X}_1^\varepsilon = x \longleftrightarrow X_{\varepsilon^2} = x \varepsilon^{1-2H} = x t^{1/2-H},$$

if we set $\varepsilon^2 = t$ in the last identity. More precisely, from the LDP (6.2.7), we have

$$-\varepsilon^{4H} \log \mathbb{P}(X_1^\varepsilon \geq x \varepsilon^{1-2H}) \rightarrow \Lambda(x), \quad \text{for } x \geq 0 \text{ as } \varepsilon \downarrow 0, \quad (6.2.8)$$

$$-\varepsilon^{4H} \log \mathbb{P}(X_1^\varepsilon \leq x \varepsilon^{1-2H}) \rightarrow \Lambda(x), \quad \text{for } x \leq 0 \text{ as } \varepsilon \downarrow 0, \quad (6.2.9)$$

and this small-noise LDP eventually translates to a short-time LDP for the process X_{ε^2} . In terms of call and put prices, the resulting short-time result reads as follows (see [38, Corollary 4.13])

$$-t^{2H} \log \mathbb{E}[(e^{X_t} - e^{x t^{1/2-H}})^+] \rightarrow \Lambda(x), \quad \text{for } x > 0 \text{ as } t \downarrow 0, \quad (6.2.10)$$

$$-t^{2H} \log \mathbb{E}[(e^{x t^{1/2-H}} - e^{X_t})^+] \rightarrow \Lambda(x), \quad \text{for } x < 0 \text{ as } t \downarrow 0. \quad (6.2.11)$$

Remark 6.1 (Precise conditions for option price asymptotics). *While the put price asymptotics (6.2.11) always holds, the unboundedness of the call option payoff requires some additional condition for (6.2.10) to hold : with reference to [39, Assumption A2], we will assume the following “1+ moment condition” whenever necessary :*

Assumption 6.2. *The process $S_t = e^{X_t}$ is a martingale, and there exist $p > 1$ and $t > 0$ such that $\mathbb{E}[S_t^p] < \infty$.*

It is known that such a condition on the moments of e^{X_t} is satisfied when σ has linear growth, cf. [38], while in the case $H = 1/2$, the same is true under much weaker assumptions (notably : $\sigma(\cdot)$ of exponential growth and $\rho < 0$ is enough, see [82, 71]). We expect similar results to hold for $H < 1/2$, but they have not been proved yet ; see the partial results available in Gassiat [45], where the martingale property of the rough Bergomi model is proved, and in Gulisashvili [55].

Let us also recall that the option price asymptotics (6.2.10) and (6.2.11) imply the following asymptotic formula for the Black–Scholes implied volatility

$$\sigma_{\text{BS}}^2(t, x t^{1/2-H}) \rightarrow \bar{\Sigma}^2(x) := \frac{x^2}{2\Lambda(x)} \quad \text{for } x \neq 0 \text{ as } t \downarrow 0, \quad (6.2.12)$$

see [38] and the related work [42]. Equation (6.2.12) has an immediate consequence for the estimation of the short-dated at-the-money implied volatility skew $\partial_k \sigma_{\text{BS}}(t, k)|_{k=0}$ (when it exists). Considering the following finite difference approximation of the ATM implied volatility skew (which is always well-defined)

$$\mathcal{S}_{\text{BS}}(t, x) := \frac{\sigma_{\text{BS}}(t, x t^{1/2-H}) - \sigma_{\text{BS}}(t, -x t^{1/2-H})}{2 x t^{1/2-H}},$$

then (6.2.12) entails

$$\mathcal{S}_{\text{BS}}(t, x) \sim \frac{\bar{\Sigma}(x) - \bar{\Sigma}(-x)}{2x} \frac{1}{t^{1/2-H}} \quad \text{as } t \downarrow 0, \quad (6.2.13)$$

An asymptotic expansion of the function $\bar{\Sigma}(x)$ as $x \rightarrow 0$, derived in [39], eventually allows to provide more explicit approximation formulas for the implied volatility surface $\sigma_{\text{BS}}(t, k)$ and its skew for small maturities t and around the at-the-money point $x = 0$.

6.2.2 Our main result : asymptotics of the Markovian projection in the large deviations regime

The tools presented in the previous section allow us to study the local volatility generated by the rough volatility model (6.2.3) at the large deviations regime,

$$\sigma_{\text{loc}}^2(t, k_t) = \mathbb{E}[V_t | X_t = k_t],$$

along the curve $(t, k_t) = (t, x t^{1/2-H})$ for some fixed value of x .

Theorem 6.2 (Theorem 3.3 in [19]). *Assume the growth conditions (C1) and (C2). Then, for $x \in \mathbb{R} \setminus \{0\}$ small enough,*

$$\sigma_{\text{loc}}^2(t, x t^{1/2-H}) = \mathbb{E}\left[\sigma^2(\widehat{W}_t) \middle| X_t = x t^{1/2-H}\right] \rightarrow \sigma^2(\hat{h}_1^x) \quad \text{as } t \downarrow 0, \quad (6.2.14)$$

where we recall that

$$\hat{h}_t^x = (K * \dot{h}^x)_t = \sqrt{2H} \int_0^t (t-s)^{H-1/2} d\dot{h}_s^x$$

is the fractional path coming from the minimizer (h^x, \bar{h}^x) of the rate function in (6.2.7).

Let us stress that Theorem 6.2 holds under the mild growth conditions of Assumption 6.1, while we do not require the 1+ moment condition in Assumption 6.2.

Numerical evaluation of the minimizing path h^x . It can be shown, see [38], that the rate function (6.2.7) satisfies

$$\Lambda(y) = \inf \left\{ \frac{(y - \rho G(h))^2}{2\bar{\rho}^2 F(h)} + \frac{1}{2} \langle \dot{h}, \dot{h} \rangle : \dot{h} \in L^2(0, 1) \right\},$$

with $F(h) = \langle \sigma^2(\hat{h}), 1 \rangle = \int_0^1 \sigma^2(\hat{h}_t) dt$ and $G(h) = \langle \sigma(\hat{h}), \dot{h} \rangle = \int_0^1 \sigma(\hat{h}_t) \dot{h}_t dt$. This alternative representation yields the rate function under the form of an unconstrained optimization problem (as opposed to the constrained optimization (6.2.7)), which can then be approximately solved by projection of the one-dimensional path \dot{h} over an orthonormal basis $\{\dot{e}_n\}_{n \geq 1}$ of L^2 , $\dot{h}_t = \sum_{n \geq 1} a_n \dot{e}_n(t)$. In practice, we truncate the sum at a certain order N and minimize over the coefficients $(a_n)_{1 \leq n \leq N}$; we obtain an approximation of the minimizer h^y and therefore of $\hat{h}_t^y = (K^H * \dot{h}^y)_t = \int_0^t K(t, s) \dot{h}_s ds$. In our numerical experiments in section 6.2.4, we choose the Fourier basis $\{\dot{e}_1(t) = 1, \dot{e}_{2n}(t) = \sqrt{2} \cos(2\pi n t), \dot{e}_{2n+1}(t) = \sqrt{2} \sin(2\pi n t), n \in \mathbb{N} \setminus \{0\}\}$, and observe that truncation of the sum at $N = 8$ provides a good accuracy. In the present context, this method is typically referred to as the Ritz method.

Some elements of the proof of Theorem 6.2. In order to prove Theorem 6.2, our starting point is a representation formula for the conditional expectation (6.2.4) that follows from the integration by parts formula of the Malliavin calculus,

$$\mathbb{E}[V_t | X_t = y] = \frac{\mathbb{E}[V_t \delta(X_t - y)]}{\mathbb{E}[\delta(X_t - y)]} = \frac{\mathbb{E}\left[\sigma(\widehat{W}_t)^2 1_{X_t \geq y} \int_0^t \frac{1}{\sqrt{1-\rho^2 \sigma(\widehat{W}_s)}} d\widehat{W}_s\right]}{\mathbb{E}\left[1_{X_t \geq y} \int_0^t \frac{1}{\sqrt{1-\rho^2 \sigma(\widehat{W}_s)}} d\widehat{W}_s\right]}. \quad (6.2.15)$$

When $y = x t^{1/2-H}$ and $x \neq 0$, we are able to study the asymptotics as $t \rightarrow 0$ of both the numerator and the denominator of the rightmost expression in (6.2.15). This follows from an application of the Laplace method on path space, building on ideas of Azencott [8] and Ben Arous [7], adapted to the rough volatility framework by Friz et al [39]. Very roughly speaking, when $t \rightarrow 0$ the process \widehat{W} inside the expectation concentrates around the path \hat{h}^x associated to the minimizer h^x of the rate function in (6.2.7). All the resulting common terms at the numerator and denominator of (6.2.15) simplify, apart from the single term $\sigma(\hat{h}^x)^2$ that only appears at the numerator.

6.2.3 The consequence for the short-time local volatility skew

The at-the-money local volatility skew is defined by the first derivative $\partial_k \sigma_{\text{loc}}(t, k)|_{k=0}$ (when it exists). Denote

$$\Sigma(x)^2 := \sigma^2(\hat{h}_1^x)$$

the short-time limit of the Markovian projection in (6.2.14). Considering the following finite difference approximation of the ATM local volatility skew,

$$\mathcal{S}_{\text{loc}}(t, x) := \frac{\sigma_{\text{loc}}(t, k_t) - \sigma_{\text{loc}}(t, -k_t)}{2k_t} = \frac{\sigma_{\text{loc}}(t, x t^{1/2-H}) - \sigma_{\text{loc}}(t, -x t^{1/2-H})}{2x t^{1/2-H}},$$

the asymptotic statement (6.2.14) immediately yields

$$\mathcal{S}_{\text{loc}}(t, x) \sim \frac{\Sigma(x) - \Sigma(-x)}{2x} \frac{1}{t^{1/2-H}} \quad (\text{local vol skew}) \quad (6.2.16)$$

as $t \rightarrow 0$, showing that, in addition to the ATM implied volatility skew in (6.2.13), the ATM *local* volatility skew also displays an explosive power-law behavior.

It is now just a matter of simple algebra to work out the asymptotic behavior of the ATM skews of the two surfaces. For reference and comparison with (6.2.16), recall from (6.2.13) that

$$\mathcal{S}_{\text{BS}}(t, x) \sim \frac{\bar{\Sigma}(x) - \bar{\Sigma}(-x)}{2x} \frac{1}{t^{1/2-H}} \quad (\text{implied vol skew})$$

Corollary 6.1 (The $H + 3/2$ skew rule). *Under assumptions (C1), (C2) and (C3),*

$$\lim_{t \rightarrow 0} \frac{\mathcal{S}_{\text{BS}}(t, x)}{\mathcal{S}_{\text{loc}}(t, x)} = \frac{\bar{\Sigma}(x) - \bar{\Sigma}(-x)}{\Sigma(x) - \Sigma(-x)}$$

if $\rho \neq 0$ (otherwise the implied and local volatility ATM skews are zero). Based on asymptotic expansions as $x \rightarrow 0$ of the function $\Sigma(x)$ derived in [19], we have

$$\lim_{x \rightarrow 0} \lim_{t \rightarrow 0} \frac{\mathcal{S}_{\text{BS}}(t, x)}{\mathcal{S}_{\text{loc}}(t, x)} = \frac{1}{H + 3/2}. \quad (6.2.17)$$

When $H = 1/2$, we are back to the classical $1/2$ skew rule presented in section 6.1. Otherwise, we dub the asymptotic result (6.2.17) the “ $H + 3/2$ skew rule”.

Remark 6.2. *Corollary 6.1 entails that the formal argument used to derive the $1/2$ rule (6.1.5) in section 6.1 does not hold anymore for the implied and local volatility surfaces generated by a rough stochastic volatility model. Notably, the boundedness of the partial derivatives $\partial_t \sigma_{\text{BS}}$, $\partial_k \sigma_{\text{BS}}$ and $\partial_{kk} \sigma_{\text{BS}}$ falls short – but in such a way that the limit of the ratio $\frac{\mathcal{S}_{\text{BS}}}{\mathcal{S}_{\text{loc}}}$ can still be identified and explicitly computed. As a consequence of (6.2.17), we also expect the short-time harmonic mean approximation (6.1.4) to break down in this context, see our related numerical tests in Figure 6.4*

Remark 6.3. *Fukasawa [43] provides asymptotic expansions for the implied volatility $\sigma_{\text{BS}}(t, x t^{1/2})$ generated by a class of rough stochastic volatility models encompassing rough Bergomi (note the different curve, $x t^{1/2}$ as opposed to $x t^{1/2-H}$, along which the volatility surface is evaluated, which hinges on the different asymptotic method applied in [43]). The asymptotic expansion for the implied volatility in [43, Theorem 2.1] is precisely expressed in terms of renormalized limit of the Markovian projection of the stochastic volatility model, see*

[43, Remark 2.1]. An adaptation of the arguments used in the proof of [43, Theorem 2.1] therefore allows to relate the implied volatility skew and the local volatility in the rough volatility model and to derive, at least formally, the “ $H+3/2$ ” skew rule with an independent argument¹.

An interesting connection with the “ $H+3/2$ ” rule can also be found in [43, Theorem 2.1], where the short-maturity limit of Skew Stickiness Ratio of Bergomi [16] is shown to satisfy

$$R = \lim_{\tau \rightarrow 0} \frac{\mathbb{E} \left[\log \frac{S_\tau}{S_0} (\sqrt{V_\tau} - \sqrt{V_0}) \right]}{\partial_k \sigma_{\text{BS}}(\tau, k)|_{k=0} \mathbb{E} \left[\left(\log \frac{S_\tau}{S_0} \right)^2 \right]} = H + \frac{3}{2}. \quad (6.2.18)$$

Though they represent limits of different objects, it would be interesting to explore more in details the possible connections between the limiting values in (6.2.18) and in (6.2.17).

Remark 6.4. Since the announcement of Theorem 6.2 and Corollary 6.1 at several conferences and meetings during 2021, other authors have explored the validity of the “ $H+3/2$ rule” in other rough or rough-like stochastic volatility models. In his Master thesis [27] (under the supervision of Riccardo Longoni, Andrea Pallavicini and Carlo Sgarra), Dall’Acqua provides numerical evidence for (6.2.17) in the framework of the lifted Heston model of Abi Jaber [1] (a Markovian approximation of rHeston), and provides a formal proof of the $H+3/2$ rule under the rough Heston model. In their recent preprint [5], Alos and co-authors proved again, and with a different proof, the asymptotics (6.2.17) in a class of rough volatility models under some regularity assumptions on the Malliavin derivatives of the involved processes, further providing an asymptotic rule for the at-the-money second derivative $\partial_{kk}(\cdot)|_{k=0}$ of the local and implied volatility functions.

Short-maturity extrapolation of local volatilities. Eventually, Theorem 6.2 provides us with an extrapolation recipe for local volatilities for very short maturities. Roughly speaking, fixed a (small) maturity t and a log-moneyness level k , formally plugging $y = \frac{k}{t^{1/2-H}}$ in the asymptotic statement

$$\sigma_{\text{loc}}^2(t, x t^{1/2-H}) \rightarrow \sigma^2(\hat{h}_1^x) \quad \text{as } t \rightarrow 0$$

we obtain the approximate expression

$$\sigma_{\text{loc}}^2(t, k) \approx \sigma^2(\hat{h}_1^x)|_{x=\frac{k}{t^{1/2-H}}} \quad \text{for small } t \text{ and small } k.$$

The limiting function $\sigma(\hat{h}_1^x)|_{y=\frac{k}{t^{1/2-H}}}$ can therefore be used in order to extrapolate a local volatility surface for very short maturities, in a way that is consistent with the behavior implied by a rough volatility model.

A specific context of application. The rough volatility model (6.2.3) can be enhanced with a leverage function $l(t, S)$

$$dS_t = S_t l(t, S_t) \sqrt{V_t} \left(\rho dW_s + \sqrt{1 - \rho^2} d\bar{W}_s \right).$$

yielding a Local-Stochastic Volatility model (LSV), whose calibration can be targeted with the particle method of Guyon and Henry-Labordère [58]. The LSV model calibrated to a given Dupire local volatility surface σ_{Dup} corresponds to

$$l(t, S_t) = \frac{\sigma_{\text{Dup}}(t, S_t)}{\sqrt{\mathbb{E}[V_t|S_t]}}.$$

In general, one wishes the leverage function $l(t, S)$ to be a small correction to the original stochastic volatility model (in other words : as close as possible to $l \equiv 1$). In practice, the local volatility σ_{Dup} coming from

¹Personal communication with Masaaki Fukasawa and Peter Friz.

market data has to be extrapolated for values of t smaller than the shorter observed maturity, and the choice of the extrapolation method is up to the user. If, for small t , the chosen extrapolation $\sigma_{\text{Dup}}(t, K)$ is qualitatively too different from the behavior of the conditional expectation $\mathbb{E}[V_t | S_t = K]$ in the rough volatility setting (for example : the behavior of the ATM skew of σ_{Dup} is far from the power law (6.2.16)), then the leverage function will have to compensate, deviating from the unit function. Under the pure rough volatility model ($l \equiv 1$), Theorem 6.2 and Corollary 6.1 describe the behavior of the markovian projection $\mathbb{E}[V_t | S_t]$ for small t : eventually, these statements give hints on how the market local volatility $\sigma_{\text{Dup}}(t, \cdot)$ should be extrapolated in order for $l(t, \cdot)$ not to deviate too much from the unit function.

Such a strategy for short-time local volatility extrapolation and the subsequent calibration of a LSV model with lifted Heston backbone is precisely explored in [27].

6.2.4 Numerical tests

We wish to estimate the conditional expectation (6.2.4) for a specific instance of the model (6.2.3), using Monte Carlo simulation. We consider the rough Bergomi model, for which the instantaneous variance process is given by

$$V_t = \xi_0 \exp\left(\eta \int_0^t \sqrt{2H}(t-s)^{H-1/2} dW_s - \frac{\eta^2}{2} t^{2H}\right), \quad (6.2.19)$$

where $\xi_0 = V_0$ is the spot variance ; since we focus on short-time asymptotics, only the short end $\xi_0 = \lim_{u \rightarrow 0} \xi_0^u$ of the variance curve will play a role in our estimations.

For a given time horizon $t > 0$ and a number N of time-steps, the random vector $(\log V_{t_k})_{1 \leq k \leq N}$, $t_k = k \frac{t}{N}$, has a multivariate Gaussian distribution with known mean and variance, and can therefore be simulated exactly via a Cholesky factorization of the covariance matrix. Of course, as already pointed out in Chapter 3, this method has a considerable complexity – cost $\mathcal{O}(N^3)$ for the Cholesky factorization and $\mathcal{O}(N^2)$ for the matrix multiplication required to get one sample of $(V_{t_k})_{0 \leq k \leq N}$ – but our focus here is on the accuracy of our estimations, rather than on their computational time. We construct approximate samples of the log-asset price $X_t = -\frac{1}{2} \int_0^t V_s ds + \int_0^t \sqrt{V_s} (\rho dW_s + \bar{\rho} d\bar{W}_s)$ using a forward Euler scheme on the same time-grid

$$X_t^N = -\frac{t}{2N} \sum_{k=0}^{N-1} V_{t_k} + \sum_{k=0}^{N-1} \sqrt{V_{t_k}} \left(\rho (W_{t_{k+1}} - W_{t_k}) + \bar{\rho} (\bar{W}_{t_{k+1}} - \bar{W}_{t_k}) \right).$$

Therefore, we obtain M i.i.d. approximate Monte Carlo samples $(X_t^{N,m}, V_t^m)_{1 \leq m \leq M}$ of the couple (X_T^N, V_t) , from which our estimators of the implied volatility and local volatility (6.2.4) are constructed, as detailed below. We estimate out-of-the-money put and call option prices by standard empirical means and evaluate the corresponding implied volatilities σ_{BS} by Newton's search.

The rough Bergomi model (6.2.19) parameters we used in our experiments are $S_0 = 1, \eta = 1.0, \rho = -0.7$, and $\xi_0 = 0.235^2$. We tested three different values of $H \in (0, 1/2]$, namely $H = 0.1, 0.3$, and 0.5 . We used $M = 1.5 \times 10^6$ Monte Carlo samples and $N = 500$ discretization points.

Estimator of the implied volatility skew. A representation of the first derivative $\partial_k \sigma_{\text{BS}}(t, k)$ can be obtained by differentiating the equation $\mathbb{E}[(S_0 e^{X_t} - S_0 e^k)^+] = C_{\text{BS}}(k, \sqrt{t} \sigma_{\text{BS}}(t, k))$ defining the implied volatility σ_{BS} with respect to the log-moneyness k . Taking the derivative with respect to k at both sides and using the expressions of the first-order Black–Scholes greeks $\partial_k C_{\text{BS}}(k, v)$ and $\partial_v C_{\text{BS}}(k, v)$, one gets

$$\partial_k \sigma_{\text{BS}}(t, k) = \frac{\Phi(d_2(k, v)) - \mathbb{P}(X_t \geq k)}{\sqrt{t} \phi(d_2(k, v))} \Big|_{v=\sqrt{t} \sigma_{\text{BS}}(t, k)},$$

where $d_2(k, v) = -\frac{k}{v} - \frac{v}{2}$, and ϕ (resp. Φ) denotes the standard Gaussian density (resp. cumulative distribution). The representation above of the implied volatility skew allows to avoid finite difference methods and only requires to estimate $\sigma_{\text{BS}}(t, k)$ and $\mathbb{P}(X_t \geq k)$, which we can do with the same Monte Carlo sample.

Estimator of the local volatility function. We have implemented and benchmarked two different estimators of the conditional expectation (6.2.4) : on the one side, a Nadaraya–Watson estimator with bandwidth δ ,

$$\sigma_{\text{loc}}^2(t, k) = \mathbb{E}[V_t | X_t = k] \approx \frac{\sum_{m=1}^M V_t^m K_\delta(X_t^{N,m} - k)}{\sum_{m=1}^M K_\delta(X_t^{N,m} - k)}, \quad (6.2.20)$$

with a Gaussian kernel $K_\delta(x) = \exp(-\delta x^2)$. On the other hand, it is a standard fact that, conditionally on $\mathcal{F}_t = \sigma(W_u : u \leq t)$, the instantaneous variance V_t is known, while the log-price X_t is normally distributed with mean $-\frac{1}{2} \int_0^t V_s ds + \rho \int_0^t \sqrt{V_s} dW_s$ and variance $(1 - \rho^2) \int_0^t V_s ds$. This property yields a representation of the Markovian projection $\sigma_{\text{loc}}(\cdot, \cdot)$ as the ratio of two expectations (already exploited, for example, in [64]) :

$$\sigma_{\text{loc}}^2(t, k) = \mathbb{E}[V_t | X_t = k] = \frac{\mathbb{E}[V_t \Pi_t(k)]}{\mathbb{E}[\Pi_t(k)]} \quad (6.2.21)$$

where

$$\Pi_t(k) = \frac{1}{\sqrt{\int_0^t V_s ds}} \exp\left(-\frac{1}{2(1-\rho^2) \int_0^t V_s ds} \left(k + \frac{1}{2} \int_0^t V_s ds - \rho \int_0^t \sqrt{V_s} dW_s\right)^2\right).$$

Estimator of the local volatility skew. Differentiating the right-hand side of (6.2.21) with respect to k , we obtain a representation of $\partial_k \sigma_{\text{loc}}(t, k)$:

$$\partial_k \sigma_{\text{loc}}(t, k) = \frac{\frac{\partial}{\partial k} \left(\frac{\mathbb{E}[V_t \Pi_t]}{\mathbb{E}[\Pi_t]} \right)}{2 \sigma_{\text{loc}}(t, k)} = \frac{\mathbb{E}[V_t \Pi_t] \mathbb{E} \left[\frac{U}{\int_0^t V_s ds} \Pi_t \right] - \mathbb{E} \left[\frac{U}{\int_0^t V_s ds} \Pi_t V_t \right] \mathbb{E}[\Pi_t]}{2(1-\rho^2) \mathbb{E}[V_t \Pi_t]^{1/2} \mathbb{E}[\Pi_t]^{3/2}}, \quad (6.2.22)$$

where Π_t is a shorthand for $\Pi_t(k)$, $U = U(k) = k + \frac{1}{2} \int_0^t V_s ds - \rho \int_0^t \sqrt{V_s} dW_s$, and $\frac{\partial \Pi_t}{\partial k} = -\frac{U}{(1-\rho^2) \int_0^t V_s ds} \Pi_t$. All the expectations appearing in (6.2.21) and (6.2.22) can be estimated based on the exact simulation of the discretized variance path $(V_{t_k})_{1 \leq k \leq N}$; we approximate the integrals $\int_0^t V_s ds$ and $\int_0^t \sqrt{V_s} dW_s$ using left-point Euler schemes. Note that the resulting non-parametric estimators based on (6.2.21) and (6.2.22) do not contain any kernel bandwidth or other hyper-parameters to be tuned, which is a clear advantage with respect to (6.2.20). We have nevertheless tested both estimators (6.2.20) and (6.2.21) for the regression function, and found perfect agreement between the two in our tests.

In Figure 6.1, we plot the term structure of the ATM implied and local volatility skews, for three different values of H and maturities up to $T = 0.5$ years. The power-law behavior observed for the local volatility skew in Figure 6.1 is consistent with Corollary 6.1. Figure 6.2 shows the ratio of the implied volatility ATM skew over the local volatility ATM skew, that is the ratio of the curves observed in Figure 6.1, for the different values of H : the numerical results are in very good agreement with the “ $H + 3/2$ rule” announced in Corollary 6.1. Additionally, we note that the ratio of the two skews appears to be rather stable : its value is almost constant for maturities up to $T = 0.5$ years, within our parameter setup.

An illustration of Theorem 6.2 is provided by Figure 6.3, where the function $y \mapsto \sigma_{\text{loc}}(t, y t^{1/2-H})$ is seen to approach its limit $\sigma(\hat{h}_1^y)$ when maturity decreases from $t = 0.5$ to $t = 0.05$. The residual error term $\sigma_{\text{loc}}(t, y t^{1/2-H}) - \sigma(\hat{h}_1^y)$ is seen to depend on H , lower values of H begin associated to higher errors.

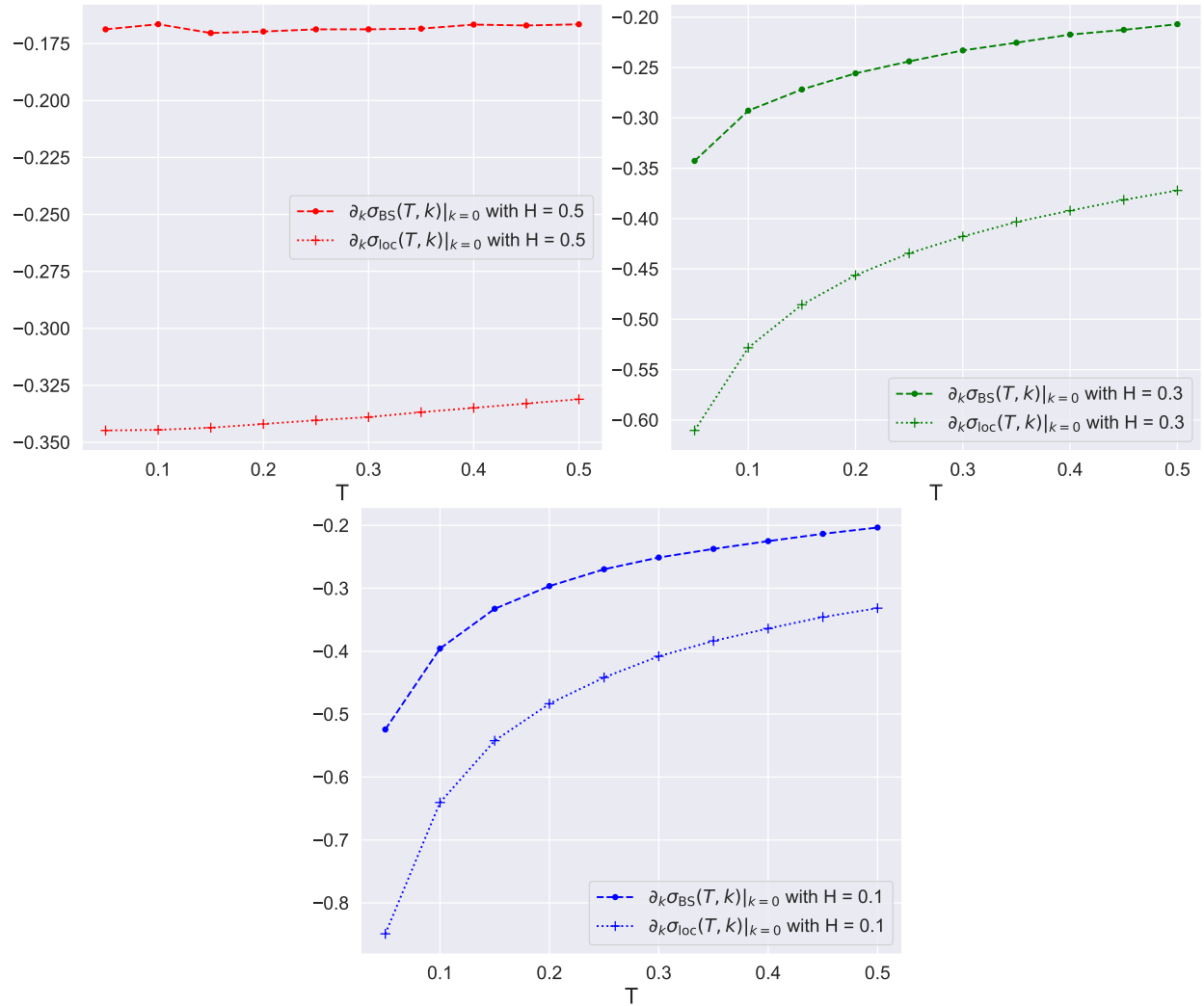


FIGURE 6.1 – At-the-money implied and local volatility skews in the rough Bergomi model (6.2.19) for $H = 0.5$ (red, top left figure), $H = 0.3$ (green, top right figure), and $H = 0.1$ (blue, bottom figure). The maturity T on the x -axis is expressed in years.

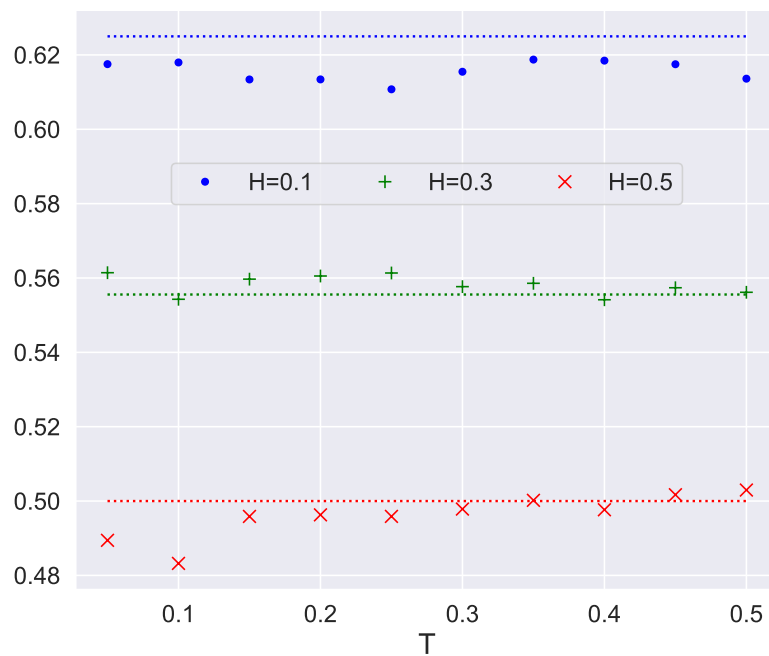


FIGURE 6.2 – Numerical evidence for the $\frac{1}{H+3/2}$ ratio rule stated in Corollary 6.1 : we plot the ratio of the at-the-money implied and local volatility skews $\frac{\partial_k \sigma_{BS}(T,k)|_{k=0}}{\partial_k \sigma_{loc}(T,k)|_{k=0}}$ for $H \in \{0.1, 0.3, 0.5\}$ against maturity T (in years). The dashed lines correspond to the constant values $\frac{1}{H+3/2}$ (blue for $H = 0.1$, green for $H = 0.3$, red for $H = 0.5$).

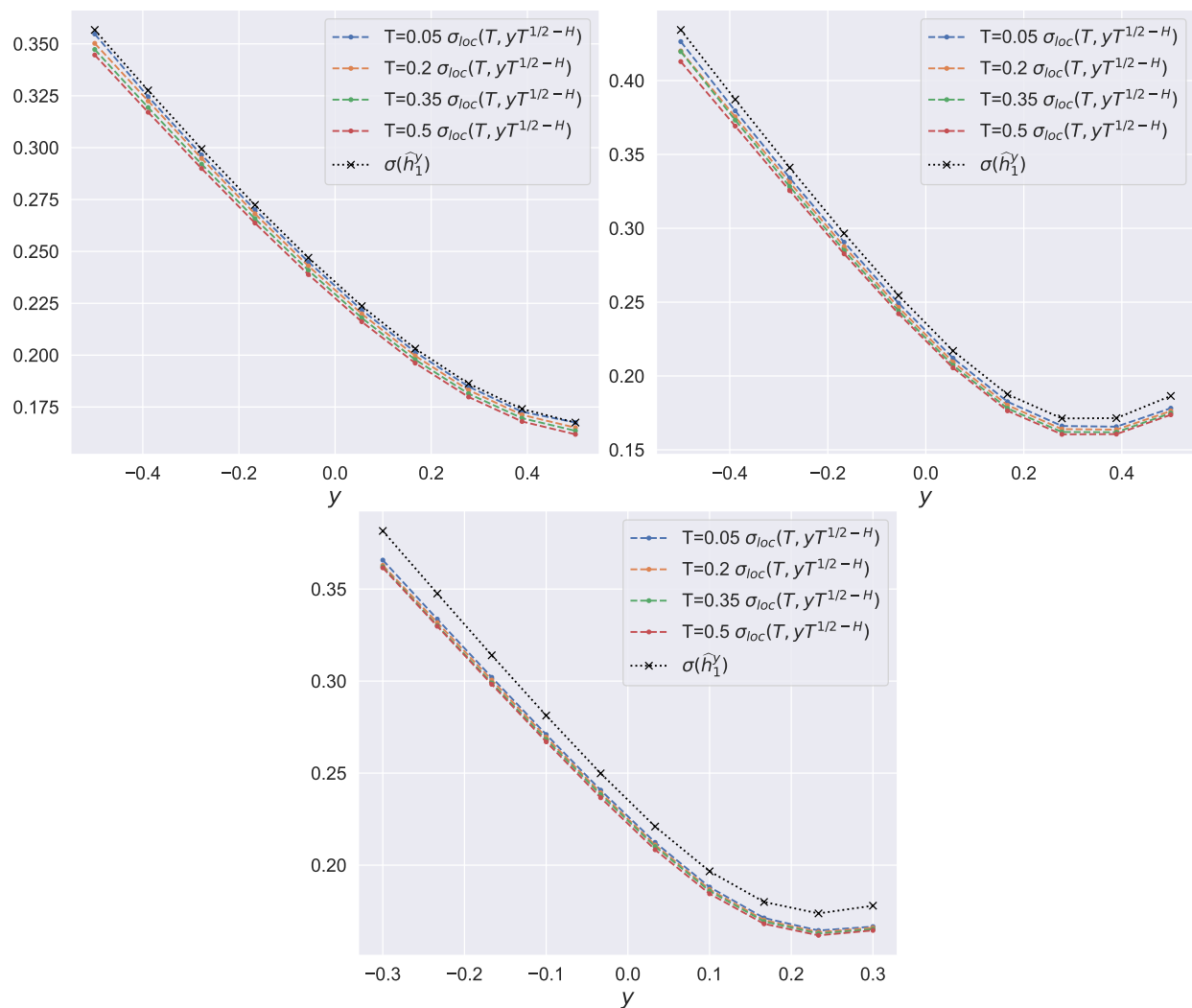


FIGURE 6.3 – Short-dated local volatility the rough Bergomi model (6.2.19) for $H = 0.5$ (top left figure), $H = 0.3$ (top right figure), and $H = 0.1$ (bottom figure). Recall that, according to Theorem 6.2, $\sigma_{loc}(T, yT^{1/2-H}) \rightarrow \sigma(\hat{h}_1^y)$ as $T \rightarrow 0$. The rate function minimizing path \hat{h}_t^y is evaluated using the Ritz projection method described in section 6.2.2 with $N = 8$ Fourier basis functions.

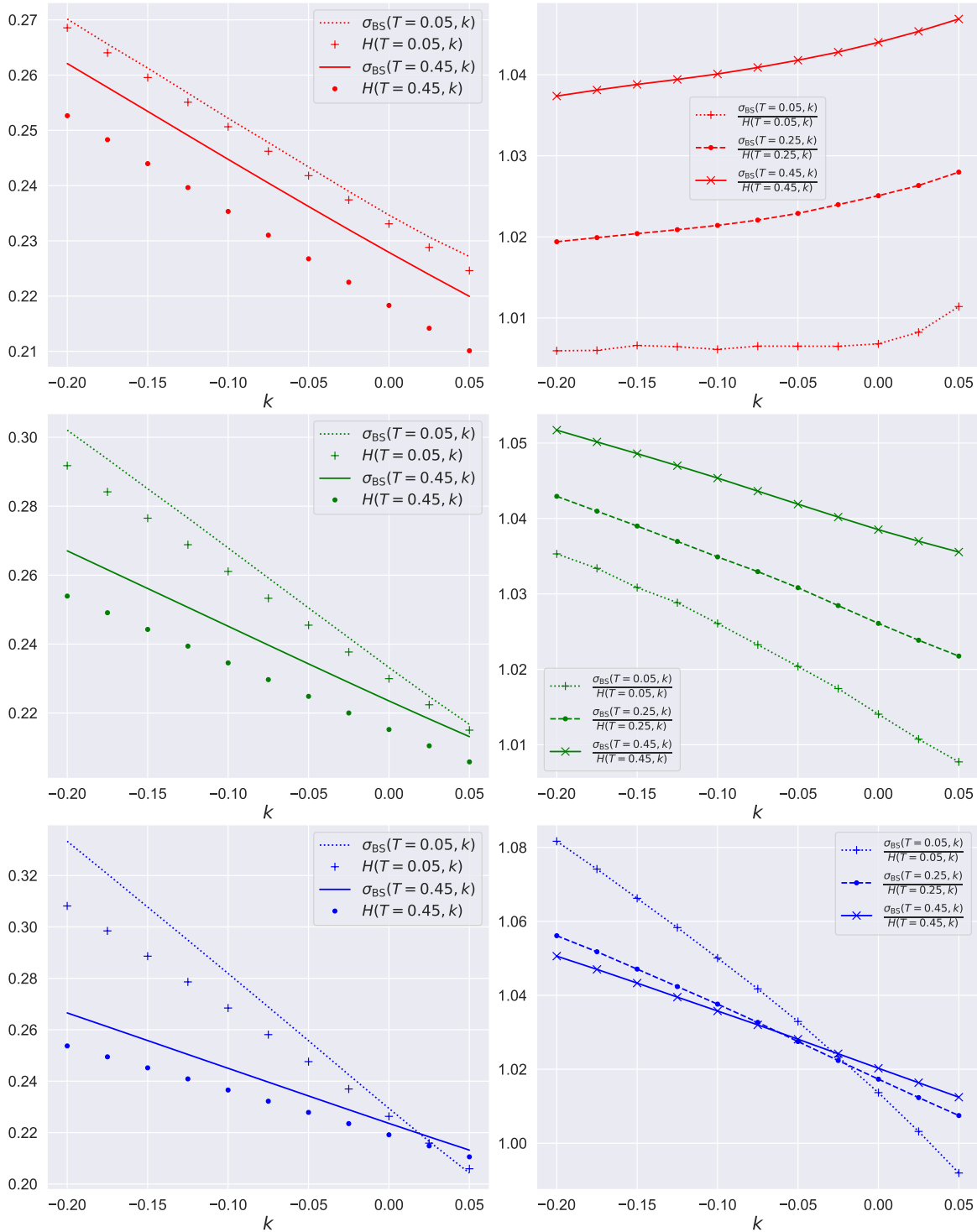


FIGURE 6.4 – Numerical evidence for the failure of the harmonic mean formula within the rough Bergomi model (6.2.19) (see Remark 6.2) : in the left figures, we compare the implied volatility $\sigma_{BS}(T, k)$ and the harmonic mean $H(T, k)$ of the local volatility defined in section 6.1, for two different maturities T and for $H = 0.5$ (red), $H = 0.3$ (green), and $H = 0.1$ (blue). In the right figures (same color conventions as the left figures), we plot the ratio $\frac{\sigma_{BS}(T, k)}{H(T, k)}$ of the two functions, expected to tend to 1 as $T \rightarrow 0$ when $H = 0.5$.

Bibliographie

- [1] Eduardo Abi Jaber. Lifting the heston model. *Quantitative Finance*, 19(12) :1995–2013, 2019. [71](#)
- [2] Eduardo Abi Jaber, Martin Larsson, and Sergio Pulido. Affine Volterra processes. *The Annals of Applied Probability*, 29(5) :3155 – 3200, 2019. [9](#), [11](#)
- [3] A. Agarwal, S. De Marco, E. Gobet, J. Lopez-Salas, F. Noubiagain, and A. Zhou. Numerical approximations of McKean anticipative backward stochastic differential equations arising in Initial Margin requirements. *ESAIM : Proceedings And Surveys*, 65 :1–26, Feb. 2019. [22](#)
- [4] Elisa Alòs, David Garc ıa-Lorite, and Aitor Muguruza Gonzalez. On Smile Properties of Volatility Derivatives : Understanding the VIX Skew. *SIAM Journal on Financial Mathematics*, 13(1) :32–69, 2022. [52](#)
- [5] Elisa Alos, David Garcia-Lorite, and Makar Pravosud. On the skew and curvature of implied and local volatilities. 2022. [71](#)
- [6] Elisa Alos, Jorge A Le on, and Josep Vives. On the short-time behavior of the implied volatility for jump-diffusion models with stochastic volatility. *Finance and Stochastics*, 11(4) :571–589, 2007. [10](#)
- [7] G erard Ben Arous. Methods de laplace et de la phase stationnaire sur l’espace de wiener. *Stochastics*, 25(3) :125–153, 1988. [69](#)
- [8] Robert Azencott. Petites perturbations al eatoires des systemes dynamiques : d eveloppements asymptotiques. *Bulletin des sciences math ematiques*, 109(3) :253–308, 1985. [69](#)
- [9] Christian Bayer, Peter Friz, and Jim Gatheral. Pricing under rough volatility. *Quantitative Finance*, 16(6) :887–904, 2016. [9](#), [10](#), [52](#), [65](#)
- [10] M. Beiglb ock, P. Henry-Labord ere, and F. Penkner. Model-independent bounds for option prices : A mass-transport approach. *Finance and Stochastics*, 17(3) :477–501, 2013. [14](#)
- [11] Mohamed Ben Alaya and Ahmed Kebaier. Multilevel Monte Carlo for Asian options and limit theorems. *Monte Carlo Methods Appl.*, 20(3) :181–194, 2014. [30](#)
- [12] H. Berestycki, J. Busca, and I. Florent. Asymptotics and calibration of local volatility models. *Quantitative Finance*, 2 :61–69, 2002. [64](#)
- [13] L. Bergomi. *Stochastic Volatility Modeling*. Chapman and Hall/CRC, 2016. [6](#), [7](#), [10](#), [11](#), [47](#)
- [14] Lorenzo Bergomi. Smile dynamics II. *Risk*, pages 67–73, 2005. [6](#), [9](#), [11](#), [52](#)
- [15] Lorenzo Bergomi. Smile dynamics III. *Risk*, pages 90–96, 2008. [56](#), [57](#)
- [16] Lorenzo Bergomi. Smile dynamics IV. *Risk*, pages 94–100, 2009. [71](#)
- [17] Anine E. Bolko, Kim Christensen, Mikko S. Pakkanen, and Bezirgen Veliyev. A gmm approach to estimate the roughness of stochastic volatility. arXiv, 2020. [10](#)

- [18] F. Bourgey and S. De Marco. Multilevel Monte Carlo simulation for VIX options in the rough Bergomi model. *Journal of Computational Finance*, 26(2) :53–82, 2022. 29, 32, 36, 37
- [19] F. Bourgey, S. De Marco, P. Friz, and P. Pigato. Local volatility under rough volatility. <https://arxiv.org/abs/2204.02376>, 2022. 63, 69, 70
- [20] F. Bourgey, S. De Marco, and E. Gobet. Weak approximations and VIX option price expansions in forward variance curve models. <https://arxiv.org/abs/2202.10413>, 2022. 51, 52, 55, 56, 58, 60
- [21] F. Bourgey, S. De Marco, E. Gobet, and A. Zhou. Multilevel Monte-Carlo methods and lower/upper bounds in Initial Margin computations. *Monte Carlo Methods and Applications*, 26(2) :131–161, 2020. 17, 23, 24
- [22] Mark Broadie and Paul Glasserman. Estimating security price derivatives using simulation. *Management Science*, 42(2) :269–285, 1996. 23
- [23] Gerard Brunick and Steven Shreve. Mimicking an Itô process by a solution of a stochastic differential equation. *The Annals of Applied Probability*, 23(4) :1584 – 1628, 2013. 66
- [24] Hans Buehler. Consistent variance curve models. *Finance Stoch.*, 10(2) :178–203, 2006. 6
- [25] H. Bühlmann, B. Gagliardi, H. Gerber, and E. Straub. Some inequalities for stop-loss premiums. *ASTIN Bull.*, 9(75) :75–83, 1977. 45
- [26] Chicago Board Options Exchange. The CBOE Volatility Index-VIX. www.cboe.com/micro/vix/vixwhite.pdf, pages 1–23, 2009. 6, 46
- [27] Enrico Dall’Acqua. A local-stochastic volatility model mimicking the rough Heston behaviour. Master thesis, Politecnico di Milano, 2012. 71, 72
- [28] S. De Marco. On the harmonic mean representation of the implied volatility. *SIAM J. Financial Math.*, 12(2) :551–565, 2021. 63, 65
- [29] S. De Marco and P. Friz. Local volatility, conditioned diffusions, and Varadhan’s formula. *SIAM J. Financ. Math.*, 9(2) :835–874, 2018. 63
- [30] S. De Marco, P. Friz, and S. Gerhold. Rational Shapes of local volatility. *Risk magazine*, pages 82–87, 2013. 63
- [31] S. De Marco and P. Henry-Labordère. Linking vanillas and VIX options : A constrained martingale optimal transport problem. *SIAM J. Financial Math.*, 6 :1171–1194, 2015. 41, 43, 44, 48
- [32] Stefano De Marco. Volatility derivatives in (rough) forward variance models, presentation at the Bachelier World Congress, Dublin, July 2018. 57
- [33] François Dubois and Tony Lelièvre. Efficient pricing of Asian options by the PDE approach. *Journal of Computational Finance*, 8(2) :55–64, 2004. 30
- [34] Bruno Dupire. Arbitrage pricing with stochastic volatility. <https://cims.nyu.edu/es-sid/ctf/stochvol.pdf>, 1993. 6

-
- [35] Bruno Dupire. Pricing with a smile. *Risk*, 7(1) :18–20, 1994. [63](#)
- [36] O. El Euch, M. Fukasawa, and M. Rosenbaum. The microstructural foundations of leverage effect and rough volatility. *Finance and Stochastics*, 22 :241–280, 2018. [10](#)
- [37] Omar El Euch and Mathieu Rosenbaum. The characteristic function of rough Heston models. *Mathematical Finance*, 29(1) :3–38, 2019. [11](#)
- [38] Martin Forde and Hongzhong Zhang. Asymptotics for rough stochastic volatility models. *SIAM Journal on Financial Mathematics*, 8(1) :114–145, 2017. [67](#), [68](#), [69](#)
- [39] Peter K Friz, Paul Gassiat, and Paolo Pigato. Precise asymptotics : Robust stochastic volatility models. *The Annals of Applied Probability*, 31(2) :896–940, 2021. [67](#), [68](#), [69](#)
- [40] M. Fukasawa. The normalizing transformation of the implied volatility smile. *Mathematical Finance*, 22(4) :753–762, 2012. [65](#)
- [41] Masaaki Fukasawa. Asymptotic analysis for stochastic volatility : martingale expansion. *Finance and Stochastics*, 15(4) :635–654, 2011. [10](#)
- [42] Masaaki Fukasawa. Short-time at-the-money skew and rough fractional volatility. *Quantitative Finance*, 17(2) :189–198, 2017. [68](#)
- [43] Masaaki Fukasawa. Volatility has to be rough. *Quantitative Finance*, 21(1) :1–8, 2021. [70](#), [71](#)
- [44] Masaaki Fukasawa, Tetsuya Takabatake, and Rebecca Westphal. Is Volatility Rough? *arXiv preprint arXiv :1905.04852*, 2019. [10](#)
- [45] Paul Gassiat. On the martingale property in the rough Bergomi model. *Electron. Commun. Probab.*, 24 :9 pp., 2019. [68](#)
- [46] J. Gatheral, E. P. Hsu, P. Laurence, C. Ouyang, and T.-H. Wang. Asymptotics of implied volatility in local volatility models. *Mathematical Finance*, 22(4) :591–620, 2012. [64](#)
- [47] Jim Gatheral, Thibault Jaisson, and Mathieu Rosenbaum. Volatility is rough. *Quantitative Finance*, 18(6) :933–949, 2018. [10](#)
- [48] Michael B Giles. Multilevel Monte Carlo path simulation. *Oper. Res.*, 56(3) :607–617, 2008. [3](#), [5](#)
- [49] Michael B Giles. Multilevel Monte Carlo methods. *Acta Numer.*, 24 :259–328, 2015. [20](#)
- [50] Michael B Giles and Abdul-Lateef Haji-Ali. Multilevel nested simulation for efficient risk estimation. *SIAM/ASA J. Uncertain. Quantif.*, 7(2) :497–525, 2019. [23](#)
- [51] Daphné Giorgi, Vincent Lemaire, and Gilles Pagès. Weak error for nested multilevel Monte Carlo. *Methodology and Computing in Applied Probability*, pages 1–24, 2020. [21](#)
- [52] Emmanuel Gobet and Mohammed Miri. Weak approximation of averaged diffusion processes. *Stochastic Process. Appl.*, 124(1) :475–504, 2014. [53](#)
- [53] M B Gordy and S Juneja. Nested Simulation in Portfolio Risk Measurement. *Management Science*, 56(10) :1833–1848, 2010. [21](#)

-
- [54] Archil Gulisashvili. Large deviation principle for volterra type fractional stochastic volatility models. *SIAM Journal on Financial Mathematics*, 9(3) :1102–1136, 2018. 67
- [55] Archil Gulisashvili. Gaussian stochastic volatility models : Scaling regimes, large deviations, and moment explosions. *Stochastic Processes and their Applications*, 130(6) :3648 – 3686, 2020. 68
- [56] Julien Guyon. On the joint calibration of SPX and VIX options, presentation at the Bachelier World Congress, Dublin, July 2018. 57
- [57] Julien Guyon. The joint S&P500/VIX smile calibration puzzle solved. *Risk*, 2020. 47
- [58] Julien Guyon and Pierre Henry-Labordère. Being particular about calibration. *Risk, January*, 2012. 71
- [59] Julien Guyon, Romain Menegaux, and 2017 Marcel Nutz. Bounds for VIX futures given S&P 500 smiles. *Finance and Stochastics*, 21(3) :593–630, 2017. 44
- [60] I. Gyongy. Mimicking the one-dimensional marginal distributions of processes having an Itô differential. *Probab. Th. Rel. Fields*, 71(4) :501–516, 1986. 66
- [61] Stefan Heinrich. Monte Carlo complexity of global solution of integral equations. *Journal of Complexity*, 14(2) :151–175, 1998. 3
- [62] Stefan Heinrich. The multilevel method of dependent tests. In *Advances in stochastic simulation methods*, pages 47–61. Springer, 2000. 3
- [63] Stefan Heinrich. Multilevel Monte Carlo methods. In *International Conference on Large-Scale Scientific Computing*, pages 58–67. Springer, 2001. 3
- [64] Pierre Henry-Labordère. Calibration of local stochastic volatility models to market smiles : A Monte-Carlo approach. *Risk Magazine, September*, 2009. 73
- [65] Pierre Henry-Labordère. Deep primal-dual algorithm for BSDEs : application of Machine Learning to CVA and IM. <https://ssrn.com/abstract=3071506>, 2017. 22
- [66] Francis Hirsch and Bernard Roynette. A new proof of Kellerer’s theorem. *ESAIM : Probability and Statistics*, 16 :48–60, 2012. 64
- [67] Blanka Horvath, Antoine Jacquier, and Peter Tankov. Volatility options in rough volatility models. *SIAM J. Financial Math.*, 11(2) :437–469, 2020. 31, 33, 37, 52
- [68] Antoine Jacquier, Claude Martini, and Aitor Muguruza. On VIX futures in the rough Bergomi model. *Quant. Finance*, 18(1) :45–61, 2018. 52
- [69] T. Jaisson and M. Rosenbaum. Rough fractional diffusions as scaling limits of nearly unstable heavy tailed Hawkes processes. *Ann. Appl. Probab.*, 26 :2860–2882, 2016. 10
- [70] K. Jansen, J. Haezendonck, and M. Goovaerts. Upper bounds on stop-loss premiums in case of known moments up to the fourth order. *Insurance : Mathematics and Economics*, 5 :315–334, 1986. 45

- [71] Benjamin Jourdain. Loss of martingality in asset price models with lognormal stochastic volatility. *preprint Cermics*, 267 :2004, 2004. 68
- [72] P. Jusselin and M. Rosenbaum. No-arbitrage implies power-law market impact and rough volatility. *Math. Finance*, 30 :1309–1336, 2019. 10
- [73] Angelien G Z Kemna and Antonius C F Vorst. A pricing method for options based on average asset values. *Journal of Banking & Finance*, 14(1) :113–129, 1990. 37
- [74] Chloe Lacombe, Aitor Muguruza, and Henry Stone. Asymptotics for volatility derivatives in multi-factor rough volatility models. *Mathematics and Financial Economics*, pages 1–33, 2021. 52
- [75] Bernard Lapeyre and Emmanuel Temam. Competitive Monte Carlo methods for the pricing of Asian options. *Journal of Computational Finance*, 5(1) :39–58, 2001. 32
- [76] Roger W. Lee. *Implied Volatility : Statics, Dynamics, and Probabilistic Interpretation*, pages 241–268. Springer US, Boston, MA, 2005. 64
- [77] Vincent Lemaire and Gilles Pagès. Multilevel Richardson–Romberg extrapolation. *Bernoulli*, 23(4A) :2643–2692, 2017. 21
- [78] Andreas Neuenkirch and Taras Shalaiko. The order barrier for strong approximation of rough volatility models. <https://arxiv.org/abs/1606.03854>, 2016. 32
- [79] Frank W J Olver, Daniel W Lozier, Ronald F Boisvert, and Charles W Clark, editors. *NIST handbook of mathematical functions*. U.S. Department of Commerce, National Institute of Standards and Technology, Washington, DC ; Cambridge University Press, Cambridge, 2010. 30
- [80] L Chris G Rogers and Zo Shi. The value of an asian option. *Journal of Applied Probability*, pages 1077–1088, 1995. 30
- [81] Martin Schweizer and Johannes Wissel. Arbitrage-free market models for option prices : the multistrike case. *Finance and Stochastics*, 12 :469–505, 2008. 7
- [82] Carlos A Sin. Complications with stochastic volatility models. *Advances in Applied Probability*, 30(1) :256–268, 1998. 68
- [83] C. Villani. *Topics in optimal transportation*. Graduate studies in Mathematics AMS, Vol. 58, 2003. 12
- [84] Pauli Virtanen, Ralf Gommers, Travis E Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, et al. SciPy 1.0 : fundamental algorithms for scientific computing in Python. *Nature methods*, 17(3) :261–272, 2020. 30
- [85] F. D. Vylder. Best upper bounds for integrals with respect to measures allowed to vary under conical and integral constraints. *Insurance : Mathematics and Economics*, 1 :109–130, 1982. 45
- [86] F. D. Vylder and M. Goovaerts. Upper and lower bounds on stop-loss premiums in case of known expectation and variance of the risk variable. *Mitt. der Ver schw Vers. Math*, pages 146–164, 1982. 45
- [87] F. D. Vylder and M. Goovaerts. Extremal moment methods and stochastic orders. *Boletín de la Asociación Matemática Venezolana*, XV :146–164, 2008. 45

- [88] G. Winkler. Extreme points of moment sets. *Mathematics of Operations Research*, 13 :581–587, 1988. [45](#)
- [89] J Zhang. A semi-analytical method for pricing and hedging continuously sampled arithmetic average rate options. *Journal of Computational Finance*, 5(1) :59–80, 2001. [30](#)