Advanced implied volatility modeling for risk management and central clearing

Arianna Mingone

INSTITUT POLYTECHNIQUE DE PARIS

NNT : 2023IPPAX088

Thèse de doctorat

Advanced implied volatility modeling for risk management and central clearing

Thèse de doctorat de l'Institut Polytechnique de Paris
préparée à École polytechnique

École doctorale n°574 Mathématiques Hadamard (EDMH)
Spécialité de doctorat: Mathématiques appliquées

Thèse présentée et soutenue à Palaiseau, le 21 novembre 2023, par

Arianna Mingone

Composition du Jury :

Emmanuel Gobet Professeur, Ecole Polytechnique (CMAP)	Président
Antoine Jacquier Reader, Imperial College London	Rapporteur
Mike Tehranchi Associate Professor, University of Cambridge	Rapporteur
Blanka Horvath Statut, University of Oxford	Examinatrice
Stefano De Marco Professeur associé, Ecole Polytechnique (CMAP)	Directeur de thèse
Claude Martini CEO, Zeliade Systems	Co-directeur de thèse
Jim Gatheral Professor, Baruch College	Invité
Vladimir Lucic Visiting Professor, Imperial College London	Invité

Remerciements

I would like to thank Antoine Jacquier, Vladimir Lucic and Mike Tehranchi for accepting to be the reviewers for this thesis. I also thank Antoine and Vladimir for the enlightening discussions we have had at Zeliade, and I would be glad to share our studies around a cup of coffee with Mike Tehranchi too.

I express my gratitude to Jim Gatheral and Emmanuel Gobet for being part of the jury, I am honoured by your presence.

Mes remerciements les plus chaleureux vont à Claude, qui m'a suivie et encouragée pendant ces trois années de thèse et travail chez Zeliade. Tu m'as toujours écoutée et aidée, en m'expliquant avec patience pour la cinquième fois les passages théoriques que je ratais. Nous avons eu des échanges d'idées pendant des heures, ensemble nous avons créé une équipe de recherche où chacun a son rôle fondamental et unique. Tu as accepté et compris mes faiblesses et nécessités et tu as su trouver un point de rencontre. Je t'estime énormément et je suis flattée de pouvoir travailler avec toi. Tu es le boss que tout le monde souhaite.

Chez Zeliade j'ai appris énormément de choses et pas seulement au niveau scientifique. Merci Pierre de m'avoir fait comprendre l'importance de l'application des mathématiques, avec ton habilité enviable de rester calme quand je panique; merci Ismail pour les confrontations techniques sur les mechanisms financiers et pour ton humilité sincère que j'admire; merci Fred pour ton enthousiasme et pour ton désir inné de partage; merci Niels pour l'air frais que tu as apporté à Zeliade, pour tes idées innovantes et pour tes projets engageants; merci Jean-Philippe de toujours résoudre les bêtises que je fais avec l'ordinateur.

Ringrazio calorosamente Stefano che, in parallelo al ruolo di direttore di tesi, mi ha sostenuta e consigliata in questo percorso di vita a Parigi. Grazie delle precise correzioni e dei fruttuosi scambi di idee, ma soprattutto grazie del supporto e del conforto che mi hai saputo dare quando ne avevo bisogno. Non dimenticheró mai il chilometro a piedi, caricato di barre e dischi di ghisa, che hai fatto per aiutarmi nel trasloco.

During these three years I had the opportunity to meet inspiring people with whom I had beneficial dialogues. A particular thank goes to Blanka Horvath, who has always shown a great respect and curiosity for me and my studies. Thank you for giving me the opportunity to give a talk in Munich and for asking for my advise in a selection process, your esteem is a privilege to me. A sincere thank goes to Rama Cont who invited me to present my works at the Oxford Mathematical and Computational Finance seminar, where I also met Samuel Cohen and Christoph Reisinger, with whom I had the chance to share our knowledge on common research fields. Thank you to Carlo Acerbi, Bruno Dupire, Mehdi El-Amrani, Masaaki Fukasawa, Bernard Gourion, Julien Guyon, Damien Lamberton, Mehdi Tomas who have shared their knowledge with me and suggested improvements to my articles.

Un ringraziamento ad Antonino Zanette, Marcellino Gaudenzi e Andrea Molent, che mi hanno accolta alla conferenza MathRiks a Udine e dove ho avuto l'opportunitá di

presentare i miei lavori.

Ringrazio la mamma e il papà che, nonostante non abbiano ancora capito di cosa mi occupi qui in Francia, sono fieri di me e raccontano a tutti i conoscenti ogni mio piccolo successo. Il vostro orgoglio mi lusinga e rende felice.

Ringrazio Anita, Veronica, Elisa ed Enrico che in questi anni mi hanno sempre sostenuta assicurandosi che la mia scorta di Disaronno non fosse mai a secco. La vostra presenza a questa difesa mi riempie di gioia, grazie davvero di esserci.

Infine, per ultimo perché il più importante, grazie Mansu di credere sempre in me, di sostenermi in ogni mia scelta, di sopportarmi durante le giornate ansiose, di accompagnarmi in tutti i miei progetti. Senza di te avrei comunque terminato la stesura di questa tesi -forse-, ma non l'avrei fatto con la stessa serenità d'animo che mi trasmetti ogni giorno.

Résumé

Dans la première partie de cette thèse, nous abordons la tâche non triviale de construire des surfaces de volatilité implicite sans arbitrage qui puissent être utilisées par les opérateurs de marché à des fins pratiques. Nous étudions en profondeur les contraintes d'arbitrage statique pour les portefeuilles d'options et nous les appliquons à des modèles de volatilité implicite connus. Tout d'abord, nous caractérisons complètement l'absence d'arbitrage Butterfly dans le modèle SVI de Gatheral, et nous étudions le cas de certains modèles sous-SVI à 3 paramètres, tels que le SVI symétrique, le SVI Vanishing Upward/Downward et le SSVI. Nous reconsidérons ensuite ce dernier modèle, étendu à plusieurs maturités, et nous combinons les conditions d'absence d'arbitrage Butterfly ainsi identifiées avec les conditions d'absence d'arbitrage Calendar Spread, déjà connues grâce à [Hendriks et Martini, The extended SSVI volatility surface, Journal of Comp Finance, 2019]. En conséquence, nous identifions un algorithme de calibration globale garantissant l'absence d'arbitrage pour le modèle eSSVI.

Dans un second temps, nous étudions la caractérisation d'une notion plus faible d'absence d'arbitrage Butterfly (que nous baptisons "weak no arbitrage condition"), c'est-à-dire les deux conditions de monotonie des fonctions d_1 et d_2 de la formule de Black-Scholes, identifiées par [Fukasawa, The normalizing transformation of the implied volatility smile, Math Fin, 2012]. Nous nous plaçons dans le cadre des smiles paramétrés en delta (suivant la convention typique sur les marchés de taux de change), et, comme résultat, nous caractérisons l'ensemble des smiles de volatilité satisfaisant cette condition faible de non arbitrage statique.

Enfin, en nous basant sur la remarque – simple mais, à notre connaissance, pas encore exploitée – que les options Call peuvent être vues comme des Calls écrits sur d'autres Calls, nous étudions les propriétés dynamiques de ces contrats.

Dans la deuxième partie, nous considérons le problème de la quantification du risque de contrepartie pour les portefeuilles d'options auquel les Chambres de compensation sont confrontées quotidiennement. Nous identifions une nouvelle formule model-free pour la VaR à court terme des portefeuilles d'options qui montre d'avoir des meilleurs performances que celles de l'approche classique de la Filtered Historical Simulation dans nos tests numériques. Enfin, nous considérons la notion d'Expected Shortfall, dont nous comparons différents types de mesures de backtesting.

Abstract

In the first part of this thesis we address the non trivial task of building arbitrage-free implied volatility surfaces which could be used by market operators for practical purposes. We study in depth static arbitrage constraints for option portfolios and apply them to notorious implied volatility models. We firstly fully characterize the absence of Butterfly arbitrage in the SVI model by Gatheral, and study the case of some 3-parameter sub-SVIs models, such as the Symmetric SVI, the Vanishing Upward/Downward SVI, and SSVI. We then reconsider the latter model, extended to multiple maturity slices, and combine the so identified conditions of no Butterfly arbitrage with the already known conditions of no Calendar Spread arbitrage by [Hendriks and Martini, The extended SSVI volatility surface, Journal of Comp Finance, 2019]. As a result, we identify a global calibration algorithm for the eSSVI model ensuring the absence of arbitrage.

Secondly, we study the characterization of a weaker notion of absence of Butterfly arbitrage (which we call "weak no arbitrage condition"), i.e. the two monotonicity requirements of the functions d_1 and d_2 in the Black-Scholes formula, identified by [Fukasawa, The normalizing transformation of the implied volatility smile, Math Fin, 2012]. We consider the framework of smiles parameterized in delta (following the typical convention on FX markets), and, as a result, we characterize the set of volatility smiles satisfying this weak condition of no static arbitrage.

Finally, based on the – simple but, to our knowledge, not yet exploited – remark that Call options can be seen as Calls written on other Calls, we study the dynamic properties of these contracts.

In the second part, we consider the problem of quantifying the counterparty risk for option portfolios that Central Clearing Counterparties face daily. We identify a new model-free formula for the short-term VaR of option portfolios which performs better than the classical approach of Filtered Historical Simulation in our numerical tests. Finally, we look at the notion of Expected Shortfall, and compare different types of backtesting measures.

List of papers being part of this thesis

- C. Martini and A. Mingone, *No arbitrage SVI*, SIAM Journal on Financial Mathematics, 13, 227-261, 2022.
- C. Martini and A. Mingone, *Explicit no arbitrage domain for sub-SVIs via reparametrization*, arXiv preprint https://arxiv.org/abs/2106.02418, 2021.
- C. Martini and A. Mingone, *Refined analysis of no Butterfly arbitrage domain for SSVI slices*, Journal of Computational Finance, 27(2), 1-32, 2023.
- A. Mingone, *No arbitrage global parametrization for the eSSVI volatility surface*, Quantitative Finance, 22(12), 2205-2217, 2022.
- A. Mingone, *Smiles in delta*, Quantitative Finance https://doi.org/10.1080/14697688.2023.2258932, 2023.
- C. Martini and A. Mingone, *Options are also options on options: how to smile with Black-Scholes*, arXiv preprint https://arxiv.org/abs/2308.04130, 2023.
- C. Martini and A. Mingone, *A closed form model-free approximation for the Initial Margin of option portfolios*, arXiv preprint https://arxiv.org/abs/2306.16346, 2023.
- Zeliade Systems, *Backtesting Expected Shortfall*, Zeliade Systems White Paper https://www.zeliade.com/wp-content/uploads/whitepapers/zwp-011-BacktestingExpectedShortfall.pdf, 2021.

Table of contents

Chapter 1

Introduction

The two general objectives of this thesis are:

- to study the problem of building implied volatility models that are simultaneously sufficiently handy and easy to apply in practice and compatible with the no-arbitrage constraints, especially considering the point of view of Central Clearing Counterparty houses (CCP);

- to directly address the problem of quantifying and backtesting the risk in option portfolios faced by CCPs.

These large institutions play the role of a central node in a financial network, providing the network participants protection against the so-called contagion risk, that is the propagation of the shock in the case of default of one of the members of the network. In order to guarantee such a protection, CCPs require initial margins to their Clearing Members, which are computed based on mathematical methods for risk management of financial portfolios and rely on the capability of generating future market scenarios that are sufficiently rich and diverse, while still being compatible with the fundamental requirement of absence of arbitrage opportunities on the market. Mathematically, this principle translates into the martingale property for the admissible stochastic models and it results into both static, that is to say at a given point in time, and dynamic, that is at different subsequent dates, no arbitrage conditions.

In particular for option portfolios, the surface of vanilla option prices on a financial asset, parameterized by the option maturity and the strike price, is one of the fundamental objects to work with. Equivalently, one can work with the implied volatility surface defined by such prices. Market operators on options hugely work with implied volatility since it allows to normalize prices among different underliers or different daily quotations on the same underlier. Implied volatility models have then the fundamental role to describe option prices by either a volatility smile (i.e. for a fixed maturity) or a whole surface indexed by the strike and the maturity. In practice, these models are used to both calibrate market prices and to interpolate/extrapolate prices in illiquid markets, with the additional benefit for parametric models of squeezing all the market

information into a few smile/surface parameters. In particular, implied volatility models are used with two possible main objectives:

- The calibrated surface can be used as input in the calibration of diffusion models such as Local Stochastic Volatility;
- The history of calibrated parameters is a way to encode the history of past option prices that demands only little storage capacity. The time series of calibrated parameters can then be used to inspect past option dynamics by looking at the variations of parameters, or to predict future moves or future price distributions via evolution models (starting from the basic Filtered Historical Simulation).

In order to accomplish such tasks, an implied volatility model has to satisfy the no arbitrage requirements.

1.1 Static Arbitrage-free representation and parametrization of implied volatility surfaces

When calibrating a surface of prices at a fixed point in time, the requirements which must be satisfied are of static type. Assuming there is a perfect market for the underlying asset and for the Call options, with short-selling allowed, and that there is no static buy-sell strategy involving the underlying asset and a finite set of Call options with a profit and loss which is strictly positive, the requirements of no static arbitrage can be classified into two categories: looking at the surface for a fixed slice, i.e., at a fixed maturity, prices must not present Butterfly arbitrages; looking at the relation among different slices, there must not be Calendar Spread arbitrages. The former arbitrage lives in the strike dimension; to be avoided, Call prices with fixed maturity T must be non-increasing and convex functions of the strike, bounded between the discounted (by the discount factor $D_0(T)$) Call payoff evaluated at the forward value $F_0(T)$, and the discounted forward. Observe that if the third property is satisfied, then the second one implies the first one because an increasing convex function cannot be bounded. Under the assumption of deterministic interest rates, the latter arbitrage arises when a normalized (by the discounted forward) Call with smaller maturity is quoted at a higher price than a normalized Call with larger maturity and same moneyness, so that the requirement is that normalized Call prices with fixed moneyness are a non-decreasing function of the time-to-maturity. We can sum up the requirements for no static arbitrage as in Table 1.1.

Among all implied volatility models, the most well-known and used ones are the Stochastic Volatility Inspired (SVI) model [33] for equity markets and the Stochastic Alpha Beta Rho (SABR) model [43] for rate markets. However, these two models do not automatically guarantee the absence of arbitrage and, so far, there were no explicit sufficient conditions on parameters for no arbitrage, even though the formulation of the models is very simple. The difficulty is hidden in the non-linear conditions for no arbitrage in terms of the implied volatility.

Type of arbitrage	Axis	Requirement on normalized Call prices
Butterfly	Fixed maturity	In $[(1-k)^+, 1]$ Non-increasing in moneyness Convex in moneyness
Calendar Spread	Fixed moneyness	Non-decreasing in time-to-maturity

Table 1.1: Requirements on normalized Call prices for no static arbitrage.

1.1.1 No arbitrage SVI

In Chapter 3 we study the SVI model, which is a model for the implied total variance (meaning, the square of the implied volatility times the time-to-maturity), and which was firstly proposed by Jim Gatheral in 2004:

$$\mathrm{SVI}(k) = a + b(\rho(k-m) + \sqrt{(k-m)^2 + \sigma^2})$$

where k is the log-forward moneyness, and (a, b, ρ, m, σ) parameters. This formula quickly became the benchmark at least on Equity markets, due to its ability to produce very good fits. Fabien Le Floch (head of research at Calypso) has a blog article on a situation where SVI *does not fit*, which is a good indicator of how rare such a situation is in practice. The practitioner literature on SVI and its variants is plentiful ([25], [45], [58], [11], [46]), and SVI is now part of every reference textbook on volatility models ([34], [39]).

In 2009, the whitepaper on the *Quasi-explicit calibration of Gatheral's SVI* ([26], also part of Stefano De Marco PHD thesis) proposed a simple trick to disambiguate the calibration of SVI, and became itself a reference calibration algorithm.

A remarkable fact is that, despite the simplicity of the formula, no Butterfly arbitrage conditions for a SVI smile remained up to now too intricate. So for instance in the algorithm [26] there is no guarantee that the calibrated parameter will be arbitrage-free. An interesting practical approach is provided in [29], where the no arbitrage constraints are expressed as a discretized set of Durrleman conditions and encoded as non-linear constraints in the optimizer; *stricto sensu* there also, there is no guarantee though that the calibrated parameter will be arbitrage-free. In this chapter, we solve this long-standing issue, fully characterizing the no Butterfly arbitrage for SVI smiles.

To do so, we deeply study the beautiful result by Fukasawa in [32], which states that the inverse functions of the $-d_1$ and $-d_2$ coefficients of the Black-Scholes formula should be increasing under no Butterfly arbitrage. In the following we will call these conditions the *Fukasawa weak conditions of no (Butterfly) arbitrage*.

Finally, it turns out that the fully explicit domain of the SVI parameters for which no Butterfly arbitrage holds is straightforward to code, resorting to root finding numerical routines (like the Brent algorithm) for the evaluation of the thresholds characterized in the computations. There are then 2 byproducts of this parametrization of the domain of high practical interest:

- a quick check routine that a given SVI parameter lies in the domain or not, which disentangles between 4 possible situations of arbitrage;
- a calibration algorithm, using any least-squares type objective function and a minimizer able to handle bounds.

Observe that the Calendar Spread arbitrage is not treated in this chapter, but it is reconsidered for a sub-SVI model in Chapter 5.

1.1.2 Explicit no arbitrage domain for sub-SVIs via reparametrization

From the SVI model, many sub-SVIs (meaning SVI with some frozen parameters, or re-parametrizations of SVI with less than 5 parameters) can be developed and studied. One among all, the *Surface SVI* (SSVI) model introduced in [36], with implied total variance of the form

$$\omega(k,T) = \frac{\theta(T)}{2}\big(1 + \rho\varphi(T)k + \sqrt{(\varphi(T)k+\rho)^2 + (1-\rho^2)}\big). \tag{1.1}$$

SVI is known to fit very well a large set of market data, and the challenging question of characterizing no Butterfly arbitrage in SVI is solved in Chapter 3 with a practical parametrization of the no arbitrage domain, leading to an efficient implementation of a calibration algorithm ensuring no Butterfly arbitrage. So why would one care about these sub-SVIs?

In fact, for several reasons. The first one is that SVI might be too rich in the sense that an excellent fit could also be achieved in most cases by sub-SVIs, with the additional benefit to stabilize the variation of the calibrated optimal parameter from one day to another, or between different maturity slices. A good theoretical reason to suspect this follows from considering SVI smiles with $\rho = 0$ and $m \neq 0$: indeed the correspondence with stochastic volatility models dictates that models with a zero correlation should yield symmetric smiles, which implies $m = 0$; in this sense SVI smiles with $\rho = 0$ and $m \neq 0$ should correspond to smiles which are not associated to stochastic volatility models, and are not very likely to be met in real-life market data. In this direction, one could also note that the result by Gatheral and Jacquier [35] that the Long Term Heston smile goes to SVI shows in fact that it goes to a sub-SVI, and in particular to an SSVI; indeed the SVI parameters are given by

$$a = \frac{\theta}{2}(1-\rho^2), \qquad b = \frac{\theta\varphi}{2}, \qquad \rho = \tilde{\rho}, \qquad m = -\frac{\rho}{\varphi}, \qquad \sigma = \frac{\sqrt{1-\rho^2}}{\varphi},$$

so that the long-term smile depends eventually only on the 3 parameters (θ, ρ, φ), with the constraint $\rho = 0 \implies m = 0$ enforced.

The second reason is that it is difficult to obtain no Calendar Spread arbitrage conditions on two SVI smiles attached to two different maturities, as discussed in [36]. This has been achieved for smiles corresponding to SSVI parameters, which are sub-SVI

ones with 3 parameters instead of 5, as shown in [44]. So in order to obtain tractable no arbitrage SVI surfaces, it may be required in practice to restrict the set of SVI parameters. Note that, in relation to the first point above, given the poor ability of SSVI to fit especially on the short term, the right balance between fitting ability and tractability might lie in-between SVI and SSVI, in some sub-SVI with 4 parameters.

Another hope is that there might be simplifications, due to the special structure of the different sub-SVIs under study, in the characterization of the no Butterfly arbitrage domain as described in Chapter 3, where it requires: the computation of the boundary of the *Fukasawa domain*, corresponding to the weak necessary condition of no arbitrage obtained by Fukasawa [32], via some root-finding algorithm; and two minimizations to get the lower bound of the no arbitrage domain for the SVI parameter σ.

In Chapter 4 we study in detail the following sub-SVIs:

1. The *Vanishing SVI* where $a = 0$ and $\rho = \pm 1$, with its two flavors *Vanishing upward* ($\rho = 1$), and *Vanishing downward* ($\rho = -1$); the second family may correspond to real-life smiles and is given by
$$w(k) = b\big(-(k-m) + \sqrt{(k-m)^2 + \sigma^2}\,\big);$$
2. The *Symmetric SVI* where $\rho = m = 0$, so that
$$w(k) = a + b\sqrt{k^2 + \sigma^2};$$
3. SSVI (slices) given by
$$w(k) = \frac{\theta}{2}\big(1 + \rho\varphi k + \sqrt{(\varphi k + \rho)^2 + 1 - \rho^2}\big).$$

For all those sub-SVIs, the objective is to find out an *explicit* parametrization for the Fukasawa domain and, when possible, of the full no arbitrage domain. Up to now and to the best of our knowledge, the only known volatility model (meaning: a formula for the implied volatility) with an explicit no arbitrage domain was the SSVI slice, with the (restrictive) conditions obtained by Gatheral and Jacquier [36], and extended to the characterization of the full no arbitrage domain in the decorrelated case in [41]. To this extent, the present work is a big leap forward, since we obtain 3 new families: the Vanishing Downward/Upward one, the Symmetric one, and the *correlated* SSVI, with explicit no arbitrage domains (a single-variable boundary function has to be computed numerically for SSVI).

The studied sub-SVIs are 3 parameters SVI; as discussed above, 3 parameters may be too little to produce a good fit on market data, so we would say that the interest here is essentially of academic nature: we hope that those families can help in the investigation of the theoretical properties of volatility smiles, or come as handy illustrations of those properties. Note though that the Vanishing Downward SVI could have a practical application to the case of decreasing market smiles which are often encountered for not-too-short maturities. Also, SSVI is used in practice; our investigation yields a parametrization of the SSVI slices satisfying the no Butterfly arbitrage property which is much more effective than the generic one presented in Chapter 3.

1.1.3 No arbitrage global parametrization for the eSSVI volatility surface

In Chapter 4 one of the studied sub-SVIs is SSVI, which has 3 parameters (θ, ρ, φ) for each slice $k \to \omega(k)$:

- θ is the At-The-Money (ATM) total implied variance;
- ρ is the correlation parameter, proportional to the slope of the smile at the ATM point;
- φ is proportional to the ATM curvature.

In [36], the authors formulated conditions for the surface to be arbitrage-free, in particular they found sufficient no Butterfly arbitrage conditions and, requiring the correlation parameter to be constant, no Calendar Spread arbitrage conditions. For what concerns Butterfly arbitrage, in Chapter 4 we already applied results obtained for SVI in Chapter 3 to SSVI and found the no arbitrage domain in terms of parameters, whose computation requires the minimization of a function depending on the two parameters θ and ρ. In order to simplify such domain and lighten algorithm computations, we reparametrized the no arbitrage domain, finally finding a domain that is a product of intervals for fixed ρ. The new domain only requires a root finding algorithm depending on the sole parameter ρ.

Even though these surfaces perform well and are of easy implementation, their calibration performances are not always satisfactory in practice because of the constraint of the constant correlation parameter. Hendriks and Martini [44] managed to extend the model by formulating no Calendar Spread conditions between two SSVI slices with different ρs and generalized the conditions to a continuous surface, giving birth to the *extended SSVI* (eSSVI). The quantitative research team at Zeliade Systems has then proposed a convenient way to robustly calibrate eSSVI surfaces in [20]. This calibration procedure can be summarized in a nutshell as follows:

1. Consider the SSVI model Equation (1.1) by Gatheral and Jacquier to model implied variance slices at fixed maturity;
2. Associate to each available option maturity on the market a slice (θ, ρ, φ) of SSVI parameters (not the same parameters for all slices), fulfilling the sufficient no Butterfly arbitrage condition obtained by Gatheral and Jacquier ([36] Theorem 4.2);
3. Calibrate the SSVI parameters of each slice sequentially with respect to option maturities, in a way that ensures the absence of Calendar Spread arbitrage as formulated by Hendriks and Martini ([44] Proposition 3.5);
4. Interpolate/extrapolate *linearly* the parameters; it can be proven that this eventually produces an arbitrage-free surface.

All in all, such a calibrated eSSVI is parametrized by a set of $3 \times N$ parameters where N is the set of maturities to which the model is calibrated, while SSVI would have $2 \times N + 1$ parameters.

The eSSVI calibration takes as underlying hypothesis the market availability of strikes and volatilities near the ATM point for each slice, since total variance is constrained to go through these points. The methodology works efficiently when market is very liquid, but when it is not the case, approximations of the ATM volatility could cause discrepancies between the calibrated model and real data. Furthermore, the arbitrage-free bounds are re-written based on a first order approximation, and they may not guarantee absence of arbitrage in case of illiquid markets.

These observations naturally bring the necessity of a more generic and *global* procedure, which is attained in Chapter 5 with a new calibration algorithm, based on a re-parametrization of the no-arbitrage domain of eSSVI as a product of intervals, suitable to be crunched by an optimization algorithm. Parameters calibration is no more performed sequentially slice by slice but globally on all slices. Hence, this is why we dub *Global eSSVI* the new model.

Figure 1.1 shows an example of the calibration performances of the Global eSSVI for the TA35 index computed by the Tel Aviv Stock Exchange. The global algorithm guarantees that, for every maturity, the eSSVI smile is Butterfly arbitrage-free and that there is no Calendar Spread arbitrage between different maturities. These arbitrage constraints together with the low market data liquidity entail that the first smile calibrates well essentially only the ATM point, while other maturities are overall well fitted.

1.1.4 Smiles in delta

So far, we have worked in the moneyness dimension for smiles, so that all no arbitrage conditions for the implied volatility have been stated using the moneyness notation. Indeed, the target SVI models are particularly used in equity markets, where all quotes are expressed via strike. However, this is not the case for FX markets, where OTC options are quoted in delta through the At-The-Money volatility, and risk reversal and strangle prices for different delta points. Furthermore, on Equity and Commodity markets, many trading firms analyze the risk of options portfolios at a given maturity on a grid of deltas instead of a grid of strikes. This fundamentally relates to the high-view of the risk as being driven at first order by a delta risk and a vega risk, which underpins in particular CME's SPAN methodology, which, although dated back to 1988, is still widely used in CCPs.

From this perspective, a first immediate question is whether the transformation of a smile $\hat{\sigma}$ in the strike space to a smile σ in the delta space is well-defined. In Chapter 6 we show that it is the case under the assumption that the map $k \to \delta(k) := N(d_1(k, \hat{\sigma}(k))$ is decreasing, where $d_{1,2} = -\frac{k}{\hat{\sigma}(k)\sqrt{T}} \pm \frac{\hat{\sigma}(k)\sqrt{T}}{2}$ denotes the classical quantity in the Black-Scholes formula, N the Gaussian cumulative density function, and k the log-forward moneyness. Thanks to a result of Fukasawa [32], we know that this property is fulfilled by smiles with no Butterfly arbitrage. Given the fact that a smile has no Butterfly

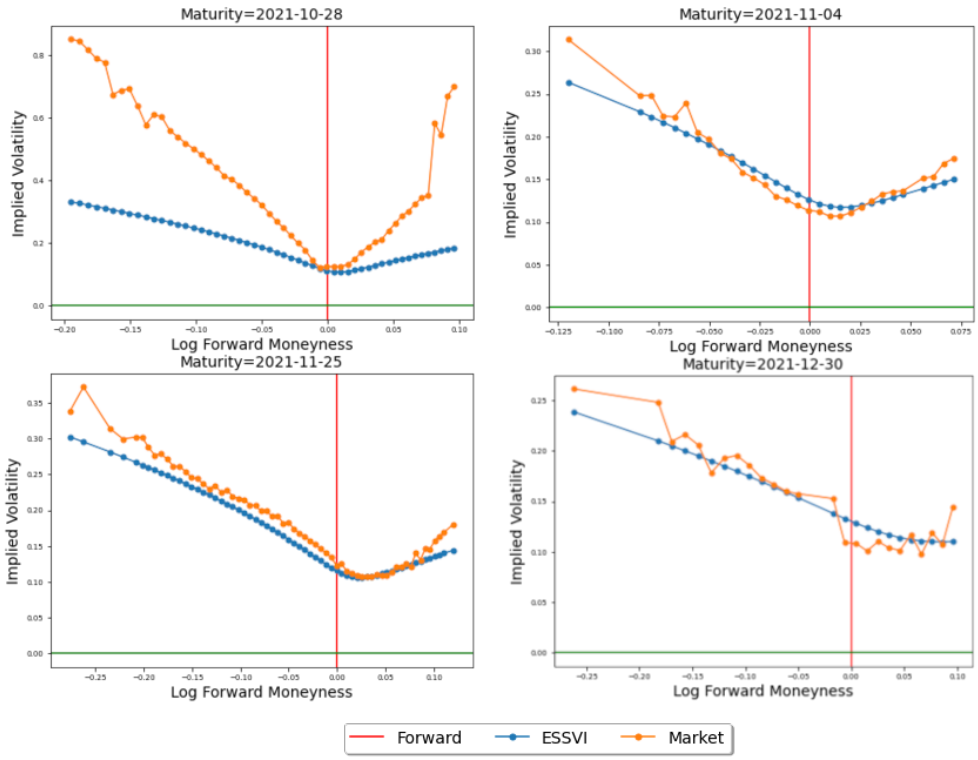

Figure 1.1: An example of Global eSSVI performances: calibrated implied volatility against real implied volatility for the Tel Aviv Stock Exchange index TA35 on the date 2021/10/26. Note the extremely short first maturity (2 days).

arbitrage if and only if the symmetrical smile (in log forward-moneyness) has no Butterfly arbitrage, such smiles will also have the property that the map $k \to \bar{\delta}(k) := N(d_2(k, \hat{\sigma}(k))$ is decreasing, since $d_1(k, \hat{\sigma}(-k))$ equals $-d_2(-k, \hat{\sigma}(-k))$.

The possibility of transforming a smile in the delta space into a smile in the strike space is also studied in this chapter, and a characterization of such smiles is obtained. This result sustains the widespread practice of the industry to calibrate smiles in the delta space to recover smiles in the strike space used to determine margins of options in strike. The chapter does not look at the question of reconstructing a smile in strike from the ATM, Butterfly and risk reversal market quotes in delta, which is sustained in [60]. The two main differences are that we work with full smiles in delta instead of the standard FX market quotes, and also that we account for no Butterfly arbitrage.

The second immediate question is how the no Butterfly arbitrage condition translates in the delta space. Surprisingly, this question is essentially an open one. Some results on the absence of arbitrage for implied volatility surfaces in the delta space can be found in [51]. We do not address the no Butterfly arbitrage in delta in this work, but

we consider instead the weaker condition obtained by Fukasawa that the two mappings $k \to d_{1,2}(k, \hat{\sigma}(k))$ are decreasing, and obtain an explicit parametrization of the smiles in delta fulfilling those conditions. This family can be useful in practice, since such smiles are expected to be not too far from fully (strongly) no arbitrage ones. Practitioners wishing to calibrate arbitrage-free smiles should then take into consideration such a parametrization since every arbitrage-free smile in delta can be represented through it.

From a theoretical view point, another take on our work is to consider the open question of parametrizing all the smiles with no Butterfly arbitrage. Such a parametrization would allow practitioners to calibrate implied volatility smiles with the certainty of fulfilling no Butterfly arbitrage and without range restrictions coming from the choice of a particular model (like Gatheral's SVI, see for example [34]). This question is probably a difficult one, and our work can be seen as a solution to the same question for a notion of weak arbitrage, through a move to the delta space. This could suggest that there is some hope to solve also the initial question in the delta space.

1.1.5 Options are also options on options: how to smile with Black-Scholes

Up to now, we have treated options as contracts written on an underlier, which could be an equity stock as in the SVI case, or a FX rate in the delta representation or a commodity, a rate, and so on. In Chapter 7 we rather exploit the remark that a (European-type) Call option with a strike $L > K$ on an underlier S can be also seen as a Call option with a strike $L-K$ written on the Call with strike K and the same maturity T; indeed when $L > K$ it holds for every value of the underlier $S(T)$ at maturity:

$$(S(T) - L)_+ = \big((S(T) - K)_+ - (L - K)\big)_+.$$

Under assumptions of perfect market (so that every asset has a single price, with no bid-asks) and no static arbitrage, this entails that the price (say, at time 0) of the two assets is the same. Denoting by $C(S, L)$ the price of standard Calls with strike L (the maturity T is the same for all the contracts), and $\hat{C}_K(\cdot, \star)$ the *relative* pricing function on the Call on Call contract, it means that the following equality holds:

$$C(S, L) = \hat{C}_K\big(C(S, K), L - K\big)$$

since the underlier price of the latter contract is the price of the Call with strike K.

This has no relation with the practice of the early days of equity-to-credit consisting in modeling the price of a stock as a Call option on the value of the underlying company, which led in turn to the fact that Calls on the stock can be valued with an option-on-option formula, as obtained by Merton in [55].

Here we study the relations between the price of options on options and the initial Call or Put prices at other strikes, finding that Calls are Calls on Calls, and Puts are Calls on Puts. Furthermore, the relative Call on Call pricing function leads to a natural transformation on the space of normalized (by the value of the underlier) Call prices as

functions of the moneyness, that we call the *Tehranchi space*, given by

$$\mathbb{T}_k c(x) := \frac{c\big(k + c(k)x\big)}{c(k)}$$

where $k = \frac{K}{S}$ and c is the normalized Call price. The derivative at zero with respect to the moneyness of $T_k c$ is in general strictly larger than -1, which corresponds to the fact that the underlier $C(S, K)$ vanishes with a positive probability at maturity. For this reason, we extend the family of transformations $\mathbb{T}_k$ to 2-parameter transformations, which allows to get a derivative at zero equal to -1.

These 2-parameter transformations allow to generate new close pricing formulas for options. Furthermore, the transformation that guarantees a derivative equal to -1 at the origin is somehow linked to the inversion of the volatility smile in the moneyness space. Indeed, applying twice this inversion in the case where the underlier may vanish at maturity provides a pricing function on an underlier which does not vanish at maturity and allows to define *lifted* Call on Call prices. We then study both Calls on Calls and lifted Calls on Calls, also looking at the properties of their implied volatility.

1.2 Risk metrics and margin computations for CCPs

The counterparty risk for exchange trades is generally mitigated thanks to CCPs, which take the role of the counterparty for each position: the CCP becomes the seller in front of the buyer and the buyer in front of the seller. In case of default of a clearing member, the CCP replaces the member until the defaulting member's positions are distributed among the surviving members through a liquidation of the portfolio performed through brokers and/or through an auction. The 2008 financial crisis entailed a strengthening of the regulations for CCPs, requiring solid risk frameworks in order to achieve the task of covering potential losses incurred by a default situation. As an example, the EMIR regulation lists the principles that a CCP must adopt to safely operate. In particular, in order to cover for the possible losses due to the liquidation of the defaulting member's portfolio, the CCP requires from its members to deposit collateral under the form of initial margin, of additional margins to cover liquidity and concentration or other specific risks, and of default fund contributions.

In this thesis, we focus on the initial margin, which is supposed to cover for the losses incurred in the case of liquidation of a given portfolio in normal market conditions. Article 41 of [22] requires CCPs to collect margins from the parties entering a transaction in a measure to be sufficient to cover the CCP potential exposures while liquidating the position. The margin must also be sufficient to cover at least 99.5% of these exposures in the case of Over-The-Counter (OTC) derivatives, and 99% for other financial products over the Margin Period Of Risk (MPOR), as recommended in article 24 of [23]. The MPOR is the required period for the portfolio liquidation, usually ranging between 2 days (for standard products) and 5 days (for OTC derivatives).

Therefore, CCPs face significant problems in calculating provisions: margins on member portfolios are computed based on the evaluation of the loss into which the CCP

could incur in case of the member's default, while liquidating its positions. The margin modeling is not universally specified among CCPs and there is not an industry standard, especially for complex products such as options. In this thesis we try to identify a simple and robust methodology which can be adopted by CCPs to margin option portfolios.

1.2.1 A closed form model-free approximation for the Initial Margin of option portfolios

Since the drafting of the EMIR regulation, CCPs have exploited different ways to compute margins for option portfolios. A first notorious methodology for complex portfolios is the SPAN algorithm of CME Group, which simulates joint risk scenarios for the underlier and the implied volatility and infers a conservative margin from these scenarios. However, this methodology has been overcome by more refined ones, which in most cases apply a Filtered Historical Simulation (FHS)-type algorithm [8] to selected risk factors in order to generate scenarios consistent with historical moves (examples are the SPAN2 by CME and the IRM2 by ICE). FHS is widely used among CCPs but its use on option markets is tricky and questionable. In particular, a straightforward use of FHS breaks down the structural relationships between risk factors, possibly generating highly implausible scenarios.

Different techniques rather than FHS for options margining have been studied in theory and eventually implemented. Some of these aim at reducing procyclicality, i.e. the excessive fluctuation of margins. As an example, the procyclicality control model by Wong and Zhang from OCC [67] relies on a scaling factor adjusting the dynamics of the ATM IV log returns to be higher during low-volatility periods and lower during high-volatility periods. More academic papers such as [24, 18] also look at the issue of computing option initial margins, additionally ensuring the absence of arbitrage for the generated scenarios. Indeed, in [24] the authors describe a generic algorithm which penalizes arbitrageable scenarios (in a static sense), which can be upgraded to any scenario generation algorithm already in production. In particular, the authors apply their methodology to Generative Adversarial Networks to simulate arbitrage-free implied volatility surfaces. In [18], an affine factor model for normalized Call option prices is defined and then calibrated, minimizing dynamic and static arbitrages. Scenarios are subsequently generated by neural SDEs which constrain the paths to live in the polytope defined by the no static arbitrage conditions. In this panorama, it is worth including the works by Bergeron at al. [10, 59] on the application of Variational Autoencoders used to reconstruct missing data on implied volatility surfaces (eventually with no arbitrage), and which can be tweaked to simulate scenarios based on historical movements.

In Chapter 8 we have two main objectives: the first one is to provide a practical and concrete panorama in options margining; the second, more ambitious, is to design a closed formula for the VaR of option portfolios, which is easy to understand and to implement. Specifically, we compute a short-term VaR formula which is model-free and coincides with an approximation of the VaR in Stochastic Volatility models and with

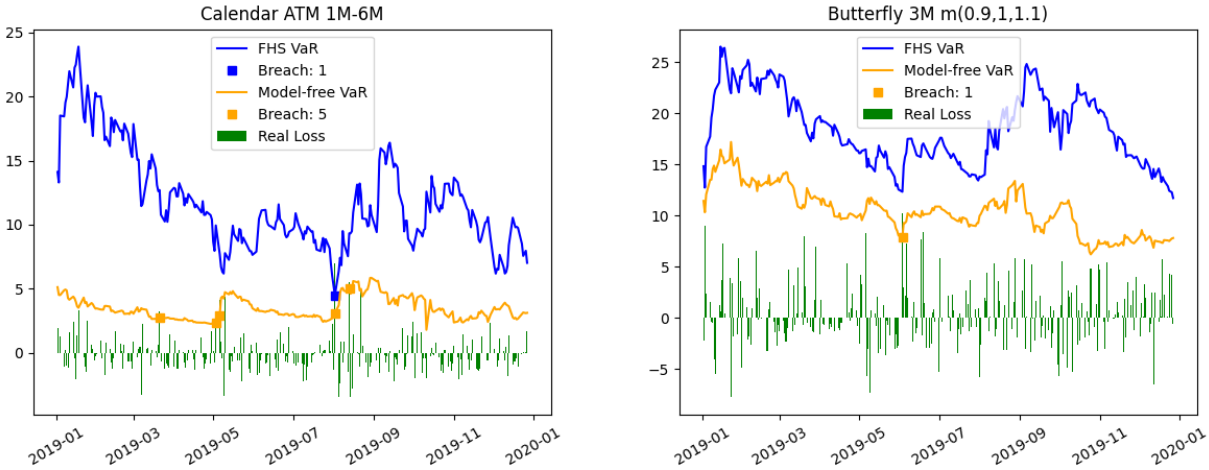

Figure 1.2: Margins obtained with the FHS algorithm (blue) and the short-term model-free VaR (orange), for a calendar spread ATM 1M-6M portfolio (left) and a Butterfly spread 3M with moneyness $(0.9, 1, 1.1)$ *portfolio (right).*

the exact VaR in the particular affine factor model for normalized Call prices proposed in [18] as the neural-SDE model. For the latter model, we show that it is actually possible to directly infer the VaR formula without any need of simulating scenarios, so that once the parameters of the model are calibrated, these can be plugged into a quasi-explicit formula to obtain the required margin. Also, considering the limit for small time steps, the formula becomes closed and it has the same form of our short-term model-free formula. Testing the short-term model-free formula, we obtain well-behaved margins which actually outperform the classic FHS ones in terms of regularity and adaptation to the market current behavior. For these reasons, we believe that the suggested short-term model-free formula could lay the foundations to a practical model-free formula for options margining.

Figure 1.2 compares margins obtained using FHS and margins from the short-term model-free formula. The tested portfolios are a calendar spread ATM 1M-6M and a Butterfly spread 3M with moneyness $(0.9, 1, 1.1)$, both on the S&P500, throughout year 2019. While the FHS VaR presents fewer breaches, it is much more irregular and it seems to move without a clear link to market patterns, compared to the short-term model-free VaR.

1.2.2 Backtesting Expected Shortfall

Article 49 of [23] is dedicated to the backtesting procedures that a CCP shall adopt in order to assess its margin coverage. Backtests are required to be performed daily and the coverage shall meet the levels identified by the EMIR regulation. For this reason, backtests need to be theoretically well-posed in order to output significant results.

CCPs compute margins via two possible metrics: a Value-at-Risk (VaR) or an Expected Shortfall (ES). While the former metric has several simple and robust statistical tests such as the Kupiec and Christoffersen tests, the latter metric lacks of a clear backtesting test, and this is one of the reasons why CCPs may prefer to use a VaR metric rather than an ES one. Several authors have addressed their studies to the ES backtesting, resulting in a lot of papers, each one claiming the efficacy of its test. In Chapter 9 we survey and compare the most common test statistics, in particular the Moldenhauer and Pitera test [57] and the Z_2, the Z_3 and the minimally biased tests by Acerbi and Szekely [1, 3].

Chapter 2

Introduction (français)

Les deux objectifs généraux de cette thèse sont les suivants:

- étudier le problème de la construction de modèles de volatilité implicite qui soient simultanément suffisamment maniables et faciles à appliquer en pratique et compatibles avec les contraintes de non-arbitrage, en particulier en considérant le point de vue des chambres de compensation (CCP);

- aborder directement le problème de quantification et de backtesting du risque dans les portefeuilles d'options auxquels sont confrontées les CCP.

Ces grandes institutions jouent le rôle d'un noeud central dans un réseau financier, fournissant aux participants du réseau une protection contre le risque de contagion, c'est-à-dire la propagation du choc en cas de défaut de l'un des membres du réseau. Pour garantir une telle protection, les CCP exigent des marges initiales à leurs membres de compensation, qui sont calculées sur la base de méthodes mathématiques de gestion des risques des portefeuilles financiers et reposent sur la capacité à générer des scénarios de marché futurs suffisamment riches et diversifiés, tout en étant compatibles avec l'exigence fondamentale d'absence d'opportunités d'arbitrage sur le marché. Mathématiquement, ce principe se traduit par la propriété de martingale pour les modèles stochastiques admissibles et se traduit par des conditions d'absence d'arbitrage à la fois statiques, c'est-à-dire à un moment donné, et dynamiques, c'est-à-dire à différentes dates ultérieures.

En particulier pour les portefeuilles d'options, la surface des prix d'options vanille sur un actif financier, paramétrée par la maturité de l'option et le prix d'exercice, est l'un des objets fondamentaux à traiter. De manière équivalente, on peut travailler avec la surface de volatilité implicite définie par ces prix. Les opérateurs de marché sur les options travaillent énormément avec la volatilité implicite car elle permet de normaliser les prix entre différents sous-jacents ou différentes cotations quotidiennes sur le même sous-jacent. Les modèles de volatilité implicite ont alors le rôle fondamental décrire les prix des options par un smile de volatilité (c'est-à-dire pour une maturité fixe) ou une surface entière indexée par le prix d'exercice et la maturité. En pratique, ces modèles sont utilisés à la fois pour calibrer les prix de marché et pour interpoler/extrapoler

les prix dans des marchés illiquides, avec l'avantage supplémentaire pour les modèles paramétriques de condenser toutes les informations de marché en quelques paramètres de smile/surface. En particulier, les modèles de volatilité implicite sont utilisés avec deux objectifs principaux possibles:

- La surface calibrée peut être utilisée en entrée dans la calibration de modèles de diffusion tels que la Volatilité Stochastique Locale;
- L'historique des paramètres calibrés est une manière d'encoder l'historique des prix d'options passés qui demande seulement peu de capacité de stockage. La série chronologique des paramètres calibrés peut alors être utilisée pour inspecter la dynamique des options passées en observant les variations des paramètres, ou pour prédire les mouvements futurs ou les distributions de prix futures via des modèles d'évolution (en commençant par la Simulation Historique Filtrée de base).

Pour accomplir de telles tâches, un modèle de volatilité implicite doit satisfaire aux exigences de non-arbitrage.

2.1 Représentation statique sans arbitrage et paramétrisation des surfaces de volatilité implicite

Lors de la calibration d'une surface de prix à un moment fixe, les exigences qui doivent être satisfaites sont de type statique. En supposant qu'il existe un marché parfait pour l'actif sous-jacent et pour les options d'achat, avec la possibilité de vente à découvert, et qu'il n'y a pas de stratégie d'achat-vente statique impliquant l'actif sous-jacent et un ensemble fini d'options d'achat avec un profit et une perte strictement positifs, les exigences de non-arbitrage statique peuvent être classées en deux catégories: en examinant la surface pour une tranche fixe, c'est-à-dire à une maturité fixe, les prix ne doivent pas présenter d'arbitrages de type Butterfly; en examinant la relation entre différentes tranches, il ne doit pas y avoir d'arbitrages de type Calendar Spread. L'arbitrage précédent vit dans la dimension du prix d'exercice; pour être évité, les prix des options d'achat avec une maturité fixe T doivent être des fonctions non croissantes et convexes du prix d'exercice, limitées entre la valeur actualisée (par le facteur d'actualisation $D_0(T)$) du payoff de l'option d'achat évaluée à la valeur forward $F_0(T)$, et le forward actualisé. Remarquez que si la troisième propriété est satisfaite, alors la deuxième implique la première car une fonction convexe croissante ne peut pas être bornée. Sous l'hypothèse de taux d'intérêt déterministes, l'arbitrage suivant se produit lorsqu'une option d'achat normalisée (par le forward actualisé) avec une maturité plus courte est cotée à un prix plus élevé qu'une option d'achat normalisée avec une maturité plus longue et la même moneyness, de sorte que l'exigence est que les prix des options d'achat normalisées avec une moneyness fixe soient une fonction non décroissante du temps jusqu'à l'échéance. Nous pouvons résumer les exigences pour l'absence d'arbitrage statique comme dans le Table 2.1.

Type d'arbitrage	Axe	Exigence sur les prix des options d'achat normalisées
Butterfly	Maturité fixe	Dans $[(1-k)^+, 1]$ Non croissant en moneyness Convexe en moneyness
Calendar Spread	Moneyness fixe	Non décroissant en temps jusqu'à l'échéance

Table 2.1: Exigences sur les prix des options d'achat normalisées pour l'absence d'arbitrage statique.

Parmi tous les modèles de volatilité implicite, les plus connus et les plus utilisés sont le modèle Stochastic Volatility Inspired (SVI) [33] pour les marchés d'actions et le modèle Stochastic Alpha Beta Rho (SABR) [43] pour les marchés de taux. Cependant, ces deux modèles ne garantissent pas automatiquement l'absence d'arbitrage et, jusqu'à présent, il n'y avait pas de conditions suffisantes explicites sur les paramètres pour qu'il n'y ait pas d'arbitrage, même si la formulation des modèles est très simple. La difficulté est cachée dans les conditions non linéaires pour l'absence d'arbitrage en termes de volatilité implicite.

2.1.1 No arbitrage SVI

Dans le Chapter 3, nous étudions le modèle SVI, qui est un modèle pour la variance totale implicite (c'est-à-dire, le carré de la volatilité implicite fois le temps jusqu'à l'échéance), et qui a été proposé pour la première fois par Jim Gatheral en 2004:

$$\mathrm{SVI}(k) = a + b(\rho(k-m) + \sqrt{(k-m)^2 + \sigma^2})$$

où k est la log-forward moneyness, et (a, b, ρ, m, σ) sont des paramètres. Cette formule est rapidement devenue la référence au moins sur les marchés actions, en raison de sa capacité à produire des ajustements très précis. Fabien Le Floch (responsable de la recherche chez Calypso) a un article de blog sur une situation où SVI *ne fit pas*, ce qui est un bon indicateur de la rareté d'une telle situation en pratique. La littérature pratique sur SVI et ses variantes est abondante ([25], [45], [58], [11], [46]), et SVI fait maintenant partie de tous les manuels de référence sur les modèles de volatilité ([34], [39]).

En 2009, le white paper sur la *Calibration quasi-explicite du SVI de Gatheral* ([26], également partie de la thèse de doctorat de Stefano De Marco) a proposé un simple astuce pour dissiper l'ambiguïté de la calibration de SVI, et est devenu lui-même un algorithme de calibration de référence.

Un fait remarquable est que, malgré la simplicité de la formule, la condition d'absence d'arbitrage Butterfly pour un smile SVI est restée jusqu'à présent trop complexe. Ainsi, par exemple, dans l'algorithme [26], il n'y a aucune garantie que le paramètre calibré sera sans arbitrage. Une approche pratique intéressante est fournie dans [29], où les contraintes de non-arbitrage sont exprimées sous forme d'un ensemble discret de conditions

de Durrleman et encodées comme contraintes non linéaires dans l'optimiseur; *stricto sensu* là aussi, il n'y a cependant aucune garantie que le paramètre calibré sera sans arbitrage. Dans ce chapitre, nous résolvons ce problème de longue date, caractérisant pleinement l'absence d'arbitrage Butterfly pour les smiles SVI.

Pour ce faire, nous étudions en profondeur le magnifique résultat de Fukasawa dans [32], qui stipule que les fonctions inverses des coefficients $-d_1$ et $-d_2$ de la formule de Black-Scholes doivent être croissantes sous l'absence d'arbitrage Butterfly. Dans ce qui suit, nous appellerons ces conditions les *conditions faibles de Fukasawa pour l'absence d'arbitrage (Butterfly)*.

Enfin, il s'avère que le domaine entièrement explicite des paramètres SVI pour lesquels il n'y a pas d'arbitrage Butterfly est simple à coder, en utilisant des routines numériques de recherche de racines (comme l'algorithme de Brent) pour l'évaluation des seuils caractérisés dans les calculs. Il y a alors 2 sous-produits de cette paramétrisation du domaine d'un intérêt pratique élevé:

- une routine de vérification rapide qu'un paramètre SVI donné se trouve dans le domaine ou non, ce qui dissocie entre 4 situations possibles d'arbitrage;
- un algorithme de calibration, utilisant une fonction objective de type moindres carrés et un minimiseur capable de gérer des bornes.

Remarquez que l'arbitrage Calendar Spread n'est pas traité dans ce chapitre, mais il est réexaminé pour un modèle sous-SVI dans le Chapter 5.

2.1.2 Explicit no arbitrage domain for sub-SVIs via reparametrization

Du modèle SVI, de nombreux sous-SVI (ce qui signifie SVI avec certains paramètres figés, ou reparamétrisations de SVI avec moins de 5 paramètres) peuvent être développés et étudiés. Parmi ceux-ci, le modèle *Surface SVI* (SSVI) introduit dans [36], avec une variance totale implicite de la forme

$$\omega(k,T) = \frac{\theta(T)}{2}\big(1 + \rho\varphi(T)k + \sqrt{(\varphi(T)k+\rho)^2 + (1-\rho^2)}\big). \tag{2.1}$$

On sait que SVI s'adapte très bien à un large éventail de données de marché, et la question difficile de caractériser l'absence d'arbitrage Butterfly dans SVI est résolue dans le Chapter 3 avec une paramétrisation pratique du domaine sans arbitrage, conduisant à une implémentation efficace d'un algorithme de calibration garantissant l'absence d'arbitrage Butterfly. Alors pourquoi s'intéresser à ces sous-SVI?

En fait, pour plusieurs raisons. La première est que SVI pourrait être trop riche dans le sens où un excellent ajustement pourrait également être obtenu dans la plupart des cas par des sous-SVI, avec l'avantage supplémentaire de stabiliser la variation du paramètre optimal calibré d'un jour à l'autre, ou entre différentes tranches de maturité. Une bonne raison théorique de soupçonner cela découle de la considération des smiles SVI avec $\rho = 0$ et $m \neq 0$: en effet, la correspondance avec les modèles de volatilité

stochastique indique que les modèles avec une corrélation nulle devraient produire des smiles symétriques, ce qui implique $m = 0$; dans ce sens, les smiles SVI avec $\rho = 0$ et $m \neq 0$ devraient correspondre à des smiles qui ne sont pas associés à des modèles de volatilité stochastique, et il est peu probable qu'ils se produisent dans des données de marché réelles. Dans cette direction, on pourrait également noter que le résultat de Gatheral et Jacquier [35] selon lequel le smile à long terme de Heston tend à SVI montre en fait qu'il tend à un sous-SVI, et en particulier à un SSVI; en effet, les paramètres SVI sont donnés par

$$a = \frac{\theta}{2}(1-\rho^2), \qquad b = \frac{\theta\varphi}{2}, \qquad \rho = \tilde{\rho}, \qquad m = -\frac{\rho}{\varphi}, \qquad \sigma = \frac{\sqrt{1-\rho^2}}{\varphi},$$

de sorte que le smile à long terme dépend finalement seulement des 3 paramètres (θ, ρ, φ), avec la contrainte $\rho = 0 \implies m = 0$ imposée.

La deuxième raison est qu'il est difficile d'obtenir des conditions d'absence d'arbitrage Calendar Spread sur deux smiles SVI attachés à deux maturités différentes, comme discuté dans [36]. Cela a été réalisé pour des smiles correspondant aux paramètres SSVI, qui sont des sous-SVI avec 3 paramètres au lieu de 5, comme le montre [44]. Ainsi, pour obtenir des surfaces SVI sans arbitrage gérables en pratique, il peut être nécessaire de restreindre l'ensemble des paramètres SVI. Noter que, par rapport au premier point ci-dessus, compte tenu de la faible capacité de SSVI à s'adapter notamment sur le court terme, le bon équilibre entre capacité d'ajustement et gérabilité pourrait se situer entre SVI et SSVI, dans un certain sous-SVI avec 4 paramètres.

Une autre espérance est qu'il puisse y avoir des simplifications, en raison de la structure spéciale des différents sous-SVI étudiés, dans la caractérisation du domaine sans arbitrage Butterfly comme décrit dans le Chapter 3, où cela nécessite: le calcul de la frontière du *domaine de Fukasawa*, correspondant à la condition nécessaire faible de non arbitrage obtenue par Fukasawa [32], via un algorithme de recherche de racines; et deux minimisations pour obtenir la borne inférieure du domaine sans arbitrage pour le paramètre SVI σ.

Dans le Chapter 4 nous étudions en détail les sous-SVI suivants:

1. Le *Vanishing SVI* où $a = 0$ et $\rho = \pm 1$, avec ses deux variantes *Vanishing upward* ($\rho = 1$), et *Vanishing downward* ($\rho = -1$); la deuxième famille pourrait correspondre à des smiles réels et est donnée par
$$w(k) = b\big(-(k-m) + \sqrt{(k-m)^2 + \sigma^2}\,\big);$$
2. Le *Symmetric SVI* où $\rho = m = 0$, donc
$$w(k) = a + b\sqrt{k^2 + \sigma^2};$$
3. SSVI (tranches) donné par
$$w(k) = \frac{\theta}{2}\big(1 + \rho\varphi k + \sqrt{(\varphi k + \rho)^2 + 1 - \rho^2}\big).$$

Pour tous ces sous-SVI, l'objectif est de trouver une paramétrisation *explicite* pour le domaine de Fukasawa et, si possible, du domaine sans arbitrage complet. Jusqu'à présent et à notre connaissance, le seul modèle de volatilité (c'est-à-dire une formule pour la volatilité implicite) avec un domaine sans arbitrage explicite était la tranche SSVI, avec les conditions (restrictives) obtenues par Gatheral et Jacquier [36], et étendues à la caractérisation du domaine sans arbitrage complet dans le cas décorrélé dans [41]. À cet égard, le présent travail est un grand pas en avant, car nous obtenons 3 nouvelles familles: celle du Vanishing Downward/Upward, celle du Symmetric, et le SSVI *correlé*, avec des domaines sans arbitrage explicites (une fonction de frontière à une seule variable doit être calculée numériquement pour le SSVI).

Les sous-SVI étudiés sont des SVI à 3 paramètres; comme discuté ci-dessus, 3 paramètres peuvent être trop peu pour produire un bon ajustement sur les données de marché, donc nous dirions que l'intérêt ici est essentiellement d'ordre académique: nous espérons que ces familles peuvent aider à l'étude des propriétés théoriques des smiles de volatilité, ou servir d'illustrations pratiques de ces propriétés. Remarquez cependant que le SVI Vanishing Downward pourrait avoir une application pratique dans le cas des smiles de marché décroissants qui sont souvent rencontrés pour des maturités pas trop courtes. De plus, le SSVI est utilisé en pratique; notre étude donne une paramétrisation des tranches SSVI satisfaisant à la propriété de non arbitrage Butterfly qui est bien plus efficace que celle présentée de manière générique dans le Chapter 3.

2.1.3 No arbitrage global parametrization for the eSSVI volatility surface

Dans le Chapter 4, l'un des sous-SVI étudiés est le SSVI, qui a 3 paramètres (θ, ρ, φ) pour chaque tranche $k \to \omega(k)$:

- θ est la variance totale implicite At-The-Money (ATM);
- ρ est le paramètre de corrélation, proportionnel à la pente du smile au point ATM;
- φ est proportionnel à la courbure ATM.

Dans [36], les auteurs ont formulé des conditions pour que la surface soit sans arbitrage, en particulier, ils ont trouvé des conditions suffisantes d'absence d'arbitrage Butterfly et, en exigeant que le paramètre de corrélation soit constant, des conditions d'absence d'arbitrage Calendar Spread. En ce qui concerne l'arbitrage Butterfly, dans le Chapter 4, nous avons déjà appliqué les résultats obtenus pour SVI dans le Chapter 3 à SSVI et trouvé le domaine sans arbitrage en termes de paramètres, dont le calcul nécessite la minimisation d'une fonction dépendant des deux paramètres θ et ρ. Afin de simplifier ce domaine et d'alléger les calculs d'algorithmes, nous avons reparamétré le domaine sans arbitrage, trouvant enfin un domaine qui est un produit d'intervalles pour ρ fixé. Le nouveau domaine ne nécessite qu'un algorithme de recherche de racines dépendant du seul paramètre ρ.

Bien que ces surfaces se comportent bien et soient faciles à implémenter, leurs performances de calibration ne sont pas toujours satisfaisantes en pratique en raison de la contrainte du paramètre de corrélation constant. Hendriks et Martini [44] ont réussi à étendre le modèle en formulant des conditions sans arbitrage Calendar Spread entre deux tranches SSVI avec des ρs différents et ont généralisé les conditions à une surface continue, donnant naissance au *SSVI étendu* (eSSVI). L'équipe de recherche quantitative de Zeliade Systems a ensuite proposé un moyen pratique de calibrer de manière robuste les surfaces eSSVI dans [20]. Cette procédure de calibration peut être résumée en quelques mots comme suit:

1. Considérer le modèle SSVI Equation (2.1) de Gatheral et Jacquier pour modéliser les tranches de variance implicite à maturité fixe;

2. Associer à chaque maturité d'option disponible sur le marché une tranche (θ, ρ, φ) de paramètres SSVI (pas les mêmes paramètres pour toutes les tranches), satisfaisant la condition suffisante d'absence d'arbitrage Butterfly obtenue par Gatheral et Jacquier ([36] Théorème 4.2);

3. Calibrer les paramètres SSVI de chaque tranche séquentiellement par rapport aux maturités des options, de manière à garantir l'absence d'arbitrage Calendar Spread comme formulé par Hendriks et Martini ([44] Proposition 3.5);

4. Interpoler/extrapoler *linéairement* les paramètres; il peut être prouvé que cela produit finalement une surface sans arbitrage.

Dans l'ensemble, un tel eSSVI calibré est paramétré par un total de $3 \times N$ paramètres où N est l'ensemble des maturités auxquelles le modèle est calibré, tandis que SSVI aurait $2 \times N + 1$ paramètres.

La calibration eSSVI part de l'hypothèse sous-jacente de la disponibilité sur le marché des strikes et des volatilités près du point ATM pour chaque tranche, puisque la variance totale est contrainte à passer par ces points. La méthodologie fonctionne efficacement lorsque le marché est très liquide, mais lorsque ce n'est pas le cas, des approximations de la volatilité ATM pourraient entraîner des écarts entre le modèle calibré et les données réelles. De plus, les bornes sans arbitrage sont réécrites sur la base d'une approximation du premier ordre, et elles peuvent ne pas garantir l'absence d'arbitrage en cas de marchés peu liquides.

Ces observations amènent naturellement la nécessité d'une procédure plus générique et *globale*, qui est atteinte dans le Chapter 5

2.1.4 Smiles in delta

Jusqu'à présent, nous avons travaillé dans la dimension de la moneyness pour les smiles, de sorte que toutes les conditions de non-arbitrage pour la volatilité implicite ont été énoncées en utilisant la notation de la moneyness. En effet, les modèles SVI cibles sont particulièrement utilisés sur les marchés des actions, où toutes les cotations sont

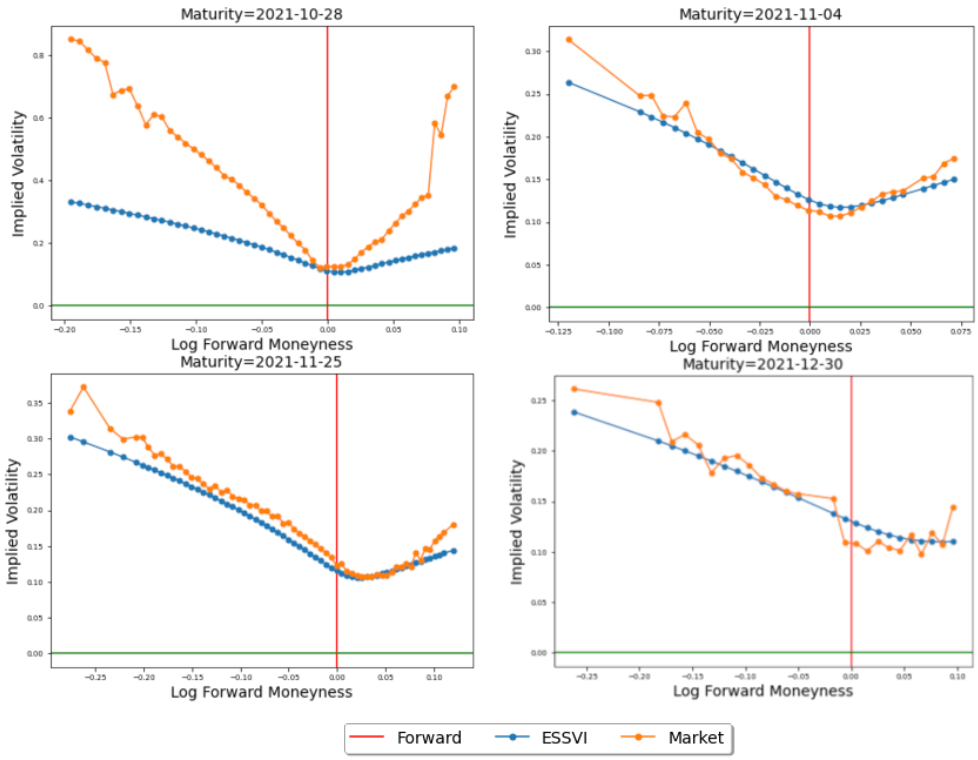

Figure 2.1: Un exemple des performances du Global eSSVI: volatilité implicite calibrée par rapport à la volatilité implicite réelle pour l'indice TA35 de la Bourse de Tel Aviv à la date du 26 octobre 2021. Notez la première échéance extrêmement courte (2 jours).

exprimées par strike. Cependant, ce n'est pas le cas pour les marchés des changes, où les options OTC sont cotées en delta à travers la volatilité At-The-Money, et les prix risk reversal et strangle pour différents points delta. De plus, sur les marchés des actions et des matières premières, de nombreuses sociétés de trading analysent le risque des portefeuilles d'options à une échéance donnée sur une grille de deltas plutôt que sur une grille de strikes. Cela est fondamentalement lié à la vision globale du risque comme étant principalement drivé par un risque delta et un risque vega, qui sous-tend en particulier la méthodologie SPAN de CME, qui, bien que datée de 1988, est encore largement utilisée parmi les CCP.

Dans cette perspective, une première question immédiate est de savoir si la transformation d'un smile $\hat{\sigma}$ dans l'espace des strikes en un smile σ dans l'espace des deltas est bien définie. Dans le chapitre Chapter 6, nous montrons que c'est le cas sous l'hypothèse que la fonction $k \to \delta(k) := N(d_1(k, \hat{\sigma}(k))$ est décroissante, où $d_{1,2} = -\frac{k}{\hat{\sigma}(k)\sqrt{T}} \pm \frac{\hat{\sigma}(k)\sqrt{T}}{2}$

désigne les quantités classiques dans la formule de Black-Scholes, N la fonction de densité cumulative gaussienne, et k la moneyness logarithmique forward. Grâce à un résultat de Fukasawa [32], nous savons que cette propriété est remplie par les smiles sans arbitrage Butterfly. Étant donné le fait qu'un smile n'a pas d'arbitrage Butterfly si et seulement si le smile symétrique (en moneyness forward logarithmique) n'a pas d'arbitrage Butterfly, de tels smiles auront également la propriété que la fonction $k \to \bar{\delta}(k) := N(d_2(k, \hat{\sigma}(k))$ est décroissante, puisque $d_1(k, \hat{\sigma}(-k))$ équivaut à $-d_2(-k, \hat{\sigma}(-k))$.

La possibilité de transformer un smile dans l'espace des deltas en un smile dans l'espace des strikes est également étudiée dans ce chapitre, et une caractérisation de tels smiles est obtenue. Ce résultat soutient la pratique généralisée de l'industrie consistant à calibrer des smiles dans l'espace des deltas pour récupérer des smiles dans l'espace des strikes utilisés pour déterminer les marges des options en strike. Le chapitre ne traite pas de la question de la reconstruction d'un smile en strike à partir des cotations ATM, des risk reversals et des strangles du marché en delta, qui est soutenue dans [60]. Les deux principales différences sont que nous travaillons avec des smiles complets en delta au lieu des cotations standard du marché des changes, et aussi que nous tenons compte de l'absence d'arbitrage Butterfly.

La deuxième question immédiate est de savoir comment la condition de non-arbitrage Butterfly se traduit dans l'espace des deltas. Étonnamment, cette question est essentiellement ouverte. Certains résultats sur l'absence d'arbitrage pour les surfaces de volatilité implicite dans l'espace des deltas peuvent être trouvés dans [51]. Nous n'abordons pas l'absence d'arbitrage Butterfly en delta dans ce travail, mais nous considérons plutôt la condition plus faible obtenue par Fukasawa selon laquelle les deux fonctions $k \to d_{1,2}(k, \hat{\sigma}(k))$ sont décroissantes, et obtenons une paramétrisation explicite des smiles en delta remplissant ces conditions. Cette famille peut être utile en pratique, car de tels smiles sont censés ne pas être très éloignés des smiles entièrement (fortement) sans arbitrage. Les praticiens souhaitant calibrer des smiles sans arbitrage devraient donc prendre en considération une telle paramétrisation puisque chaque smile sans arbitrage en delta peut être représenté à travers elle.

D'un point de vue théorique, une autre approche de notre travail consiste à considérer la question ouverte de la paramétrisation de tous les smiles sans arbitrage Butterfly. Une telle paramétrisation permettrait aux praticiens de calibrer des smiles de volatilité implicite avec la certitude de ne pas avoir d'arbitrage Butterfly et sans restrictions provenant du choix d'un modèle particulier (comme le SVI de Gatheral, voir par exemple [34]). Cette question est probablement difficile, et notre travail peut être vu comme une solution à la même question pour une notion d'arbitrage faible, à travers un passage à l'espace des deltas. Cela pourrait suggérer qu'il y a un espoir de résoudre également la question initiale dans l'espace des deltas.

2.1.5 Options are also options on options: how to smile with Black-Scholes

Jusqu'à présent, nous avons traité les options comme des contrats écrits sur un sous-jacent, qui pourrait être une action comme dans le cas du SVI, ou un taux de change dans la représentation delta ou une matière première, un taux d'intérêt, etc. Dans le chapitre Chapter 7, nous exploitons plutôt la remarque selon laquelle une option d'achat (de type européen) avec un prix d'exercice $L > K$ sur un sous-jacent S peut également être vue comme une option d'achat avec un prix d'exercice $L - K$ écrite sur l'option d'achat avec un prix d'exercice K et la même échéance T; en effet, lorsque $L > K$, cela est vrai pour chaque valeur du sous-jacent $S(T)$ à l'échéance:

$$(S(T) - L)_+ = \big((S(T) - K)_+ - (L - K)\big)_+.$$

Sous des hypothèses de marché parfait (de telle sorte que chaque actif a un prix unique, sans bid-ask) et d'absence d'arbitrage statique, cela entraîne que le prix (disons, au temps 0) des deux actifs est le même. En notant par $C(S, L)$ le prix des options d'achat standard avec un prix d'exercice L (l'échéance T est la même pour tous les contrats), et par $\hat{C}_K(\cdot, \star)$ la fonction *relative* de prix d'option d'achat sur option, cela signifie que l'égalité suivante est vérifiée:

$$C(S, L) = \hat{C}_K\big(C(S, K), L - K\big)$$

puisque le prix du sous-jacent du dernier contrat est le prix de l'option d'achat avec un prix d'exercice K.

Cela n'a aucune relation avec la pratique des premiers jours d'equity-to-credit consistant à modéliser le prix d'une action comme une option d'achat sur la valeur de l'entreprise sous-jacente, ce qui a conduit à son tour au fait que les options d'achat sur l'action peuvent être valorisées avec une formule d'option-sur-option, comme obtenue par Merton dans [55].

Ici, nous étudions les relations entre le prix des options sur options et les prix initiaux des options d'achat ou de vente à d'autres prix d'exercice, constatant que les options d'achat sont des options d'achat sur des options d'achat, et que les options de vente sont des options d'achat sur des options de vente. De plus, la fonction relative de prix des options d'achat sur options d'achat conduit à une transformation naturelle sur l'espace des prix des options d'achat normalisés (par la valeur du sous-jacent) en fonction de la moneyness, que nous appelons l'*espace de Tehranchi*, donnée par

$$\mathbb{T}_k c(x) := \frac{c\big(k + c(k)x\big)}{c(k)}$$

où $k = \frac{K}{S}$ et c est le prix normalisé de l'option d'achat. La dérivée en zéro par rapport à la moneyness de $T_k c$ est en général strictement supérieure à -1, ce qui correspond au fait que le sous-jacent $C(S, K)$ s'annule avec une probabilité positive à l'échéance. Pour cette raison, nous étendons la famille de transformations $\mathbb{T}_k$ à des transformations à 2 paramètres, ce qui permet d'obtenir une dérivée en zéro égale à -1.

Ces transformations à 2 paramètres permettent de générer de nouvelles formules de prix pour les options qui sont fermées. De plus, la transformation qui garantit une dérivée égale à -1 à l'origine est d'une certaine manière liée à l'inversion du smile de volatilité dans l'espace de la moneyness. En effet, en appliquant deux fois cette inversion dans le cas où le sous-jacent peut s'annuler à l'échéance, on obtient une fonction de prix sur un sous-jacent qui ne s'annule pas à l'échéance et permet de définir des prix d'achat sur options *élargis*. Nous étudions ensuite à la fois les options d'achat sur options d'achat et les options d'achat sur options d'achat élargies, en examinant également les propriétés de leur volatilité implicite.

2.2 Métriques de risque et calcul des marges pour les CCP

Le risque de contrepartie pour les échanges sur une bourse est généralement atténué grâce aux CCP, qui prennent le rôle de la contrepartie pour chaque position: la CCP devient le vendeur face à l'acheteur et l'acheteur face au vendeur. En cas de défaut d'un membre de compensation, la CCP remplace le membre défaillant jusqu'à ce que ses positions soient réparties entre les membres survivants grâce à une liquidation du portefeuille effectuée par l'intermédiaire de courtiers et/ou par le biais d'une enchère. La crise financière de 2008 a entraîné un renforcement de la réglementation des CCP, exigeant des cadres de risque solides afin d'accomplir la tâche de couvrir les pertes potentielles encourues en cas de situation de défaut. À titre d'exemple, le règlement EMIR énumère les principes qu'une CCP doit adopter pour fonctionner en toute sécurité. En particulier, afin de couvrir les éventuelles pertes dues à la liquidation du portefeuille du membre défaillant, la CCP exige de ses membres de déposer des garanties sous forme de marge initiale, de marges supplémentaires pour couvrir les risques de liquidité et de concentration ou d'autres risques spécifiques, ainsi que des contributions au fonds de défaut.

Dans cette thèse, nous nous concentrons sur la marge initiale, censée couvrir les pertes encourues en cas de liquidation d'un portefeuille donné dans des conditions de marché normales. L'article 41 de [22] exige que les CCP collectent des marges auprès des parties entrant dans une transaction dans une mesure suffisante pour couvrir les expositions potentielles de la CCP lors de la liquidation de la position. La marge doit également être suffisante pour couvrir au moins $99,5\%$ de ces expositions dans le cas des dérivés Over-The-Counter (OTC), et 99% pour d'autres produits financiers sur la période de risque de marge (MPOR), comme recommandé par l'article 24 de [23]. La MPOR est la période requise pour la liquidation du portefeuille, généralement comprise entre 2 jours (pour les produits standard) et 5 jours (pour les dérivés OTC).

Par conséquent, les CCP sont confrontées à des problèmes importants dans le calcul des provisions: les marges sur les portefeuilles des membres sont calculées sur la base de l'évaluation de la perte dans laquelle la CCP pourrait incurer en cas de défaut du membre, tout en liquidant ses positions. La modélisation des marges n'est pas spécifiée universellement entre les CCP et il n'existe pas de norme de l'industrie, en particulier pour les produits complexes tels que les options. Dans cette thèse, nous essayons d'identifier

une méthodologie simple et robuste qui peut être adoptée par les CCP pour marger les portefeuilles d'options.

2.2.1 A closed form model-free approximation for the Initial Margin of option portfolios

Depuis la rédaction du règlement EMIR, les CCP ont exploité différentes méthodes pour calculer les marges pour les portefeuilles d'options. Une première méthodologie notoire pour les portefeuilles complexes est l'algorithme SPAN du groupe CME, qui simule des scénarios de risque conjoints pour le sous-jacent et la volatilité implicite et en déduit une marge conservatrice à partir de ces scénarios. Cependant, cette méthodologie a été remplacée par des méthodes plus raffinées, qui dans la plupart des cas utilisent un algorithme de type Filtered Historical Simulation (FHS) [8] pour des facteurs de risque sélectionnés afin de générer des scénarios cohérents avec les mouvements historiques (des exemples sont le SPAN2 du CME et le IRM2 de l'ICE). FHS est largement utilisé parmi les CCP, mais son utilisation sur les marchés d'options est délicate et discutable. En particulier, une utilisation directe de FHS casse les relations structurelles entre les facteurs de risque, générant ainsi des scénarios hautement invraisemblables.

Différentes techniques plutôt que FHS pour le calcul des marges sur options ont été étudiées en théorie et finalement mises en oeuvre. Certaines visent à réduire la procyclicité, c'est-à-dire la fluctuation excessive des marges. Par exemple, le modèle de contrôle de la procyclicité de Wong et Zhang de l'OCC [67] repose sur un facteur d'échelle ajustant la dynamique des rendements logarithmiques de la volatilité implicite ATM pour être plus élevée pendant les périodes de faible volatilité et plus faible pendant les périodes de forte volatilité. Des articles plus académiques comme [24, 18] examinent également la question du calcul des marges initiales des options, garantissant également l'absence d'arbitrage pour les scénarios générés. En effet, dans [24], les auteurs décrivent un algorithme générique qui pénalise les scénarios arbitragés (d'un point de vue statique), qui peut être rajouté à n'importe quel algorithme de génération de scénarios déjà en production. En particulier, les auteurs appliquent leur méthodologie aux réseaux génératifs antagonistes pour simuler des surfaces de volatilité implicite sans arbitrage. Dans [18], un modèle de facteur affine pour les prix normalisés des options d'achat est défini puis calibré, minimisant les arbitrages dynamiques et statiques. Les scénarios sont ensuite générés par des SDE neuronaux qui contraignent les trajectoires à vivre dans le polytope défini par les conditions d'absence d'arbitrage statique. Dans ce panorama, il faut inclure les travaux de Bergeron et al. [10, 59] sur l'application des autoencodeurs variationnels utilisés pour reconstruire les données manquantes sur les surfaces de volatilité implicite (éventuellement sans arbitrage), et qui peuvent être ajustés pour simuler des scénarios basés sur les mouvements historiques.

Dans Chapter 8, nous avons deux objectifs principaux: le premier est de fournir un panorama pratique et concret du calcul des marges sur les options; le second, plus ambitieux, est de concevoir une formule fermée pour le VaR des portefeuilles d'options, qui soit facile à comprendre et à mettre en oeuvre. Plus précisément, nous calculons une

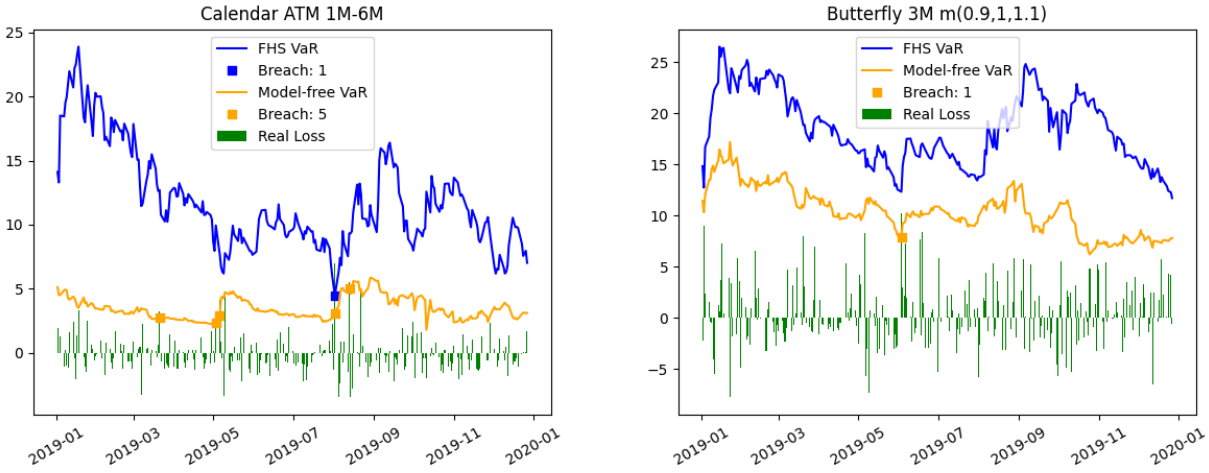

Figure 2.2: Marges obtenues avec l'algorithme FHS (bleu) et la VaR model-free pour les courts termes (orange), pour un portefeuille de calendar spread ATM 1M-6M (gauche) et un portefeuille de Butterfly spread 3M avec des moneyness $(0.9, 1, 1.1)$ *(droite).*

formule de VaR à court terme qui est libre de modèle et coïncide avec une approximation du VaR dans les modèles de volatilité stochastique et avec le VaR exact dans le modèle de facteur affine particulier pour les prix normalisés des options d'achat proposé dans [18] en tant que modèle de SDE neuronal. Pour ce dernier modèle, nous montrons qu'il est en fait possible d'inférer directement la formule de VaR sans avoir besoin de simuler des scénarios, de sorte que une fois que les paramètres du modèle sont calibrés, ceux-ci peuvent être introduits dans une formule quasi explicite pour obtenir la marge requise. De plus, en considérant la limite pour de petits laps de temps, la formule devient fermée et elle a la même forme que notre formule de modèle libre à court terme. En testant la formule modèl-free pour les courts termes, nous obtenons des marges qui se comportent bien et en fait mieux que celles du FHS classique, en termes de régularité et d'adaptation au comportement actuel du marché. Pour ces raisons, nous pensons que la formule model-free pour les courts termes suggérée pourrait poser les bases d'une formule model-free pratique pour le calcul des marges sur les options.

Figure 2.2 compare les marges obtenues en utilisant FHS et les marges de la formule model-free pour les courts termes. Les portefeuilles testés sont un calendar spread ATM 1M-6M et un Butterfly spread 3M avec des moneyness $(0.9, 1, 1.1)$, tous deux sur le S&P500, tout au long de l'année 2019. Bien que la VaR FHS présente moins de violations, elle est beaucoup plus irrégulière et semble évoluer sans lien clair avec les tendances du marché, par rapport à la VaR model-free pour les courts termes.

2.2.2 Backtesting Expected Shortfall

L'article 49 de [23] est consacré aux procédures de backtesting qu'une CCP doit adopter afin d'évaluer sa couverture de marge. Les backtests doivent être effectués quotidien-

nement et la couverture doit atteindre les niveaux identifiés par la réglementation EMIR. Pour cette raison, les backtests doivent être théoriquement bien posés afin de produire des résultats significatifs.

Les CCP calculent les marges via deux mesures possibles: un Value-at-Risk (VaR) ou un Expected Shortfall (ES). Alors que la première mesure dispose de plusieurs tests statistiques simples et robustes tels que les tests de Kupiec et Christoffersen, la seconde mesure manque d'un test de backtesting clair, et c'est l'une des raisons pour lesquelles les CCP peuvent préférer utiliser une mesure VaR plutôt qu'une mesure ES. Plusieurs auteurs ont consacré leurs études au backtesting ES, ce qui a donné lieu à de nombreux articles, chacun revendiquant l'efficacité de son test. Dans le Chapter 9, nous passons en revue et comparons les statistiques de test les plus courantes, en particulier le test de Moldenhauer et Pitera [57] ainsi que les tests Z_2, Z_3 et le test minimally biased par Acerbi et Szekely [1, 3].

Part I

Static Arbitrage-free representation and parametrization of implied volatility surfaces

Chapter 3

No arbitrage SVI

Abstract

We fully characterize the absence of Butterfly arbitrage in the SVI formula for implied total variance proposed by Gatheral in 2004. The main ingredient is an intermediate characterization of the necessary condition for no arbitrage obtained for any model by Fukasawa in 2012 that the inverse functions of the $-d_1$ and $-d_2$ of the Black-Scholes formula, viewed as functions of the log-forward moneyness, should be increasing. A natural rescaling of the SVI parameters and a meticulous analysis of the Durrleman condition allow then to obtain simple range conditions on the parameters. This leads to a straightforward implementation of a least-squares calibration algorithm on the no arbitrage domain, which yields an excellent fit on the market data we used for our tests, with the guarantee to yield smiles with no Butterfly arbitrage.

From:
C. Martini and **A. Mingone**, *No arbitrage SVI*, SIAM Journal on Financial Mathematics, 13, 227-261, 2022.

3.1 Structure of the chapter

The main objective of this chapter is to find a full characterization of no Butterfly arbitrage for Stochastic Volatility Inspired (SVI) smiles. SVI is a model for the implied total variance with formulation

$$\mathrm{SVI}(k) = a + b(\rho(k-m) + \sqrt{(k-m)^2 + \sigma^2})$$

where k is the log-forward moneyness, and (a, b, ρ, m, σ) parameters. We do not impose the condition $a \geq 0$, as is often done without justification; we work out the necessary and sufficient conditions in the full domain of the SVI parameters.

We start in Section 3.3 with a precise discussion of the meaning of no Butterfly arbitrage, which is based on [66] and on [62] for the corresponding statements in terms of volatility.

We proceed in Section 3.4 with a slight generalization of the beautiful result by Fukasawa in [32], which states that the inverse functions of the $-d_1$ and $-d_2$ coefficients of the Black-Scholes formula should be increasing under no Butterfly arbitrage. We need this generalization to handle all the configurations of SVI parameters. In this section we also clarify when and how Call and Put SVI option prices, given by the Black-Scholes formula with the SVI formula as argument, can be represented as expectations, using the results in [66].

The main ingredient (Section 3.5) is then to use a natural rescaling of SVI: we work with the parameters α, μ where $a = \sigma\alpha$ and $m = \sigma\mu$, and the dummy variable $l = \frac{k-m}{\sigma}$ instead of k. It turns out that the Fukasawa conditions for SVI in the new parameters do not involve σ. An interesting property of those conditions is that they provide the positivity of the first term of the Durrleman condition; based on the fact that in our case the complementing second term reads $\frac{1}{2\sigma}G_2(l)$ where G_2 does not depend on σ, ensuring the Durrleman condition yields a simple condition on σ. In Section 3.6 we give the full characterization of the Fukasawa conditions for SVI, and Section 3.7 finishes the work with the full characterization of no Butterfly arbitrage.

We provide in the last section (Section 3.8) numerical tests performed on data on index options purchased from the CBOE.

SVI models a volatility smile, not a volatility surface, so without ambiguity when we use the no arbitrage wording for SVI, we mean the absence of Butterfly arbitrage.

3.2 Domain of SVI parameters

The SVI model is defined when $a, m \in \mathbb{R}$, $b \geq 0$, $|\rho| \leq 1$, $\sigma \geq 0$. We recall that SVI is a convex function, with a minimum value given by $a + b\sigma\sqrt{1-\rho^2}$ (possibly attained at infinity if $|\rho| = 1$) and which goes to infinity as k goes to $\pm\infty$ (for $|\rho| < 1$). Since SVI models total variances, it is therefore required that $a + b\sigma\sqrt{1-\rho^2} \geq 0$.

If $\rho = -1$ the SVI smile decreases from ∞ to a, and if $\rho = +1$ the SVI smile increases from a to ∞.

3.3 The Durrleman condition and no arbitrage for SVI

Let S_0 denote the underlying asset value of standard Call options with a fixed maturity $t > 0$. Without loss of generality we assume that there is no interest rates nor dividend rates. In case of deterministic interest r and dividend rates δ, all the statements in this section still hold once S_0 is replaced by the Forward corresponding to the option maturity $F_t = S_0 \exp \int_0^t (r_s - \delta_s)ds$ and working with the numéraire of the option maturity.

3.3.1 Axiom of no Butterfly arbitrage

The condition of no Butterfly arbitrage is achieved when the Call price function with respect to the strike is (we follow the very careful treatment in [66]):

- convex;
- non-increasing;
- with value in the range $[(S_0 - K)^+, S_0]$.

These properties assume only that there is a perfect market for the underlying asset and for the Call options, with short-selling allowed, and that there is no static buy-sell strategy involving the underlying asset and a finite set of Call options with a Profit and Loss which is strictly positive.

We recall in particular that the *large moneyness behaviour* stating that the Call price function should go to zero at ∞ is an additional assumption, and does not strictly follow from the no arbitrage axiom.

In the case of a Call price function specified through an implied volatility: $C(K) = C_{\text{BS}}(k, \sqrt{w(k)})$ where w is the implied total variance $\sigma^2 T$ and $C_{\text{BS}}(k, a)$ is the Black-Scholes formula expressed as a function of the log-forward moneyness $k = \log \frac{K}{S_0}$ and the implied total volatility, the third property is automatically granted since the C_{BS} function is increasing with respect to its second argument and since the range bounds correspond to the limit when a goes to 0 and ∞.

Observe now that if the third property is satisfied, then the first one implies the second one because an increasing convex function cannot be bounded.

3.3.2 Smiles vanishing at some point

Can a volatility smile reach 0 at some (finite) point? Assume that it is the case, so $w(k_m) = 0$ at the log-forward moneyness k_m corresponding to some strike K_m. Then it means that the Call price with this strike is equal to its intrinsic value $(S_0 - K_m)^+$. If K_m lies on the right of S_0, the price is therefore 0, and by the property 2 above all the Call prices with $K > K_m$ will also be 0. If K_m lies on the left of S_0, the option price is $S_0 - K_m$; as the option price with a strike 0 is equal to $S_0 = S_0 - 0$, the convexity property implies that all the Call prices with $K < K_m$ are smaller than $S_0 - K$ which is the value of the chord between the points 0 and K_m. Since this value $S_0 - K$ is also lower bound for the Call prices, they are eventually equal to this value. So, in the implied volatility space, this means that $w = 0$ for $K \geq K_m$ in the first case, and $w = 0$ for $K \leq K_m$ in the second case.

This means that no arbitrage implies that smiles reaching 0 above (respectively below) the At-The-Money (forward) point will vanish above (respectively below) this point. In the case of SVI, smiles reach zero at most at a single strike, and only if $a + b\sigma\sqrt{1-\rho^2} = 0$ and $|\rho| < 1$, in which case they are strictly positive for other strike values, and there is a Butterfly arbitrage. So we can discard this case and assume $a + b\sigma\sqrt{1-\rho^2} > 0$ when $|\rho| < 1$.

3.3.3 No Butterfly arbitrage criterion for SVI

At this stage we know that SVI smiles with no Butterfly arbitrage are positive, and that the third property above is automatically satisfied. So there is no Butterfly arbitrage if and only if the first property holds. Now for positive smiles, as recalled in [36] after Lemma 2.2, with $w(k) = \mathrm{SVI}(k)$:

$$\begin{aligned} p(K) := \left.\frac{d^2 C_{\mathrm{BS}}}{dK^2}\right|_{K=S_0 e^k} &= \left.\frac{d^2 C_{\mathrm{BS}}(k, \sqrt{w(k)})}{dK^2}\right|_{K=S_0 e^k} \\ &= \frac{g(k)}{S_0 e^k \sqrt{2\pi w(k)}} \exp\left(-\frac{d_2(k, \sqrt{w(k)})^2}{2}\right) \end{aligned} \tag{3.1}$$

where d_2 is the standard coefficient of the Black-Scholes formula:

$$d_{1,2}(k, \sigma) = -\frac{k}{\sigma} \pm \frac{\sigma}{2}.$$

So convexity is equivalent to ask the function $g(k)$ ([36], equation 2.1)

$$g(k) := \left(1 - \frac{k\mathrm{SVI}'(k)}{2\mathrm{SVI}(k)}\right)^2 - \frac{\mathrm{SVI}'(k)^2}{4}\left(\frac{1}{\mathrm{SVI}(k)} + \frac{1}{4}\right) + \frac{\mathrm{SVI}''(k)}{2} \tag{3.2}$$

to be non-negative, which is usually called the *Durrleman condition* (cf. Theorem 2.9, condition (IV3) of [62]).

Note that the first derivative of the Call function with respect to the strike necessarily goes to 0 as K goes to ∞, and to a finite limit between -1 and 0 as K goes to 0, which means that the total mass of p is less than one, but not necessarily one, meaning that there could be a non-zero mass at zero. It will sum to one if and only if the limit is -1; in this case, p can be interpreted as a probability measure; the expectation of the underlying asset under this measure will be strictly less than the underlying asset value, unless the additional property that the Call price vanishes at infinity holds, in which case it will be exactly the underlying asset value (cf. Theorem 2.1.2 of [66]).

The above discussion can be translated in properties of the smile: we know from Theorem 2.9 in [62] that the large moneyness behaviour is one-to-one with the fact that $d_1(k)$ goes to $-\infty$ at infinity:

$$\lim_{k\to\infty} d_1(k, \sqrt{w(k)}) = -\infty.$$

The fact that there is no mass at zero, or, equivalently, that the derivative of the Call price with respect to the strike goes to -1 when the strike goes to zero, is equivalent to (cf. [31], Proposition 2.4):

$$\lim_{k\to-\infty} d_2(k, \sqrt{w(k)}) = +\infty.$$

In the case of SVI, the first condition translates to $b(1+\rho) < 2$ and the second one to $b(1-\rho) < 2$. In particular the following Lemma holds:

Lemma 3.1. *In SVI, the limit of* $d_1(k, \sqrt{w(k)})$ *for* k *going to* ∞*, is*

- $-\infty$ *if* $b(1+\rho) < 2$*;*
- 0 *if* $b(1+\rho) = 2$*;*
- ∞ *if* $b(1+\rho) > 2$.

The proof is simple and it is omitted. An important consequence to this result is that when $b(1+\rho) = 2$, the Call prices do not go to zero when the strike goes to infinity and so they are not given as the expectation of the payoff; we will come back to this situation in detail in Section 3.4. In such a case, this does not necessarily lead to an arbitrage and so the request $b(1+\rho) < 2$ is not a necessary condition fo the absence of arbitrage. We can summarize the previous discussion as follows:

Proposition 3.1 (No Butterfly arbitrage criterion for SVI)**.** *A necessary condition for no Butterfly arbitrage to hold in SVI is that* $SVI(k) > 0$ *for all* k*. Under this condition, there is no arbitrage in SVI if and only if the function* g *in Equation* (3.2) *is non-negative. In this case, the function* $p(K)$ *in Equation* (3.1) *where* $K = S_0 e^k$*, and* S_0 *is the underlying asset value, defines a positive density on* $\mathbb{R}_+$ *such that* $\int p(x)dx \leq 1$*.*

Moreover, the Call prices in SVI go to zero when the strike goes to infinity if and only if $b(1+\rho) < 2$*, and the derivative of the Call price (expressed in numéraire of the maturity) with respect to the strike goes to* -1 *if and only if* $b(1-\rho) < 2$*. In the first case* $\int xp(x)dx = S_0$ *and in the second case* $\int p(x)dx = 1$*.*

Note that the two conditions $b(1+\rho) = 2$ and $b(1-\rho) = 2$ can occur simultaneously if and only if $b = 2$ and $\rho = 0$.

3.3.4 Expectation-based representation of the Calls and Put prices in SVI

The issue of the representation of the Call and Put price functions by an expectation under our purely analytical assumptions has been settled by Tehranchi in [66]. Indeed re-starting from the assumption that the implied volatility function is such that the Call price function $K \to C(K)$ is convex, one can note that we are exactly in the situation of Theorem 2.1.2 in [66]. So there exists a non-negative random variable S_T such that $E[S_T] \leq S_0$ and $C(K) = S_0 - E[K \wedge S_T]$. What about the Put price? Synthesizing the strategy of selling a Call with strike K, buying the underlying and selling a quantity K of cash at time 0 yields a payoff $(X_T - K)^+ - X_T + K = (K - X_T)^+$, where X_T is the realized value of the underlying at maturity and the assumption of no arbitrage leads to $P(K) = C(K) - S_0 + K$ which is the Put-Call parity and gives the pricing formula for the Put $P(K) = K - E[K \wedge S_T]$. It is interesting to note that:

$$C(K) = S_0 - E[K \wedge S_T] = E[(S_T - K)^+] + S_0 - E[S_T],$$

whereas the usual expectation for the Put formula still holds:

$$P(K) = K - E[K \wedge S_T] = E[(K - S_T)^+].$$

Going back to SVI, when $b(1+\rho) < 2$ we are in the situation where $C(K) \to 0$ when $K \to \infty$, so that $E[S_T] = S_0$ in the above representation. In the case $b(1+\rho) = 2$ one has $\lim_{k\to\infty} d_1(k) = 0$, which plugged into the Black-Scholes formula gives $\lim_{K\to\infty} C(K) = \frac{S_0}{2}$; in turn this gives $E[S_T] = \frac{S_0}{2}$.

Observe also that since the equality $P_{\text{BS}}(k, \sigma(k)) = C_{\text{BS}}(k, \sigma(k)) - S_0 + K$ holds from the definition of the Black-Scholes functions C_{BS} and P_{BS}, it follows that $P(K) = P_{\text{BS}}(k, \sigma(k))$, so the Calls and Puts with the same strike have the same implied volatility. We can now state the following:

Proposition 3.2. *Let $C(K) := C_{BS}(k, \sqrt{SVI(k)})$ and $P(K) := P_{BS}(k, \sqrt{SVI(k)})$ be the Call and Put prices in SVI. Then there exists a positive random variable S_T such that:*

1. *$P(K) = K - E[K \wedge S_T] = E[(K - S_T)^+]$ and $C(K) = S_0 - E[K \wedge S_T] = E[(S_T - K)^+] + S_0 - E[S_T]$;*
2. *if $b(1+\rho) < 2$, $C(K) \to 0$ when $K \to \infty$, $E[S_T] = S_0$ and $C(K) = E[(S_T - K)^+]$;*
3. *if $b(1+\rho) = 2$, $C(K) \to \frac{S_0}{2}$ when $K \to \infty$, $E[S_T] = \frac{S_0}{2}$ and $C(K) = E[(S_T - K)^+] + \frac{S_0}{2}$.*

We turn now to the weak no Butterfly condition obtained by Fukasawa. Characterizing this intermediary condition will eventually lead us to our full characterization result.

3.4 Fukasawa necessary condition for no Butterfly arbitrage

We recall the beautiful model-free necessary no Butterfly arbitrage condition obtained by Fukasawa in [32]. Following Fukasawa, let us denote the Black-Scholes prices as $C_{\text{BS}}(k, \sigma) = S_0\Phi(d_1(k, \sigma)) - S_0 e^k \Phi(d_2(k, \sigma))$ for Calls and $P_{\text{BS}}(k, \sigma) = S_0 e^k \Phi(-d_2(k, \sigma)) - S_0\Phi(-d_1(k, \sigma))$ for Puts; the implied total volatility is $\sqrt{w(k)} = \sigma(k)$; for a given total implied volatility let us set

$$f_{1,2}(k) = -d_{1,2}(k, \sigma(k)).$$

Fukasawa proved in Theorem 2.8 of [32] that (under the hypothesis that option prices are given by the expectation of their payoff) if a total variance smile w, expressed as a function of the log-forward moneyness, has no Butterfly arbitrage, then the two functions f_1 and f_2 are necessarily strictly increasing with $f'_{1,2} > 0$.

3.4.1 A slight generalization of Fukasawa result

In the following, we generalize Fukasawa's result to the case where the only request on the Put prices is their convexity and differentiability, without requiring that they are given by the expectation of the payoff. Note that in the SVI case Proposition 3.2 (item 1) tells us that Put prices have actually this property. Nevertheless, we believe the section has its own interest. The proof is essentially Fukasawa's one.

Lemma 3.2. *Let Put prices be defined as the Black-Scholes Put prices with a total volatility $\sigma(k)$: $P(K) = P_{BS}(k, \sigma(k))$, where $K = S_0 e^k$. If the function P is convex and the total volatility is differentiable, then the functions $f_{1,2}$ are strictly increasing.*

Proof. Because the total volatility is differentiable, then also the Put prices are differentiable. Define $D_{\text{BS}}(K) := \frac{1}{K}\frac{\partial P_{\text{BS}}}{\partial k}(k,\sigma) \mid_{\sigma=\sigma(k)} = \Phi(f_2(k))$ and $D(K) := \frac{dP}{dK}(K)$. Note that here we do not use the equality $D(K) = E[I_{K>S_T}]$ whose proof requires that the Put prices are the expectation of their payoff. Since the Put prices are given by the Black-Scholes formula, they lie between $(K - S_0)^+$ and K, and since the function $K \to P(K)$ is convex, then its derivative lies necessarily between 0 and 1: $0 \leq D(K) \leq 1$. It holds

$$\begin{aligned} D(K) &= \frac{d}{dK} P_{\text{BS}}\big(\log(K/S_0), \sigma(\log(K/S_0))\big) \\ &= D_{\text{BS}}(K) + \frac{1}{K}\frac{\partial P_{\text{BS}}}{\partial \sigma}\big(\log(K/S_0), \sigma(\log(K/S_0))\big)\frac{d\sigma}{dk}\big(\log(K/S_0)\big) \\ &= D_{\text{BS}}(K) + \phi\big(f_2\big(\log(K/S_0)\big)\big)\frac{d\sigma}{dk}\big(\log(K/S_0)\big). \end{aligned} \tag{3.3}$$

We now check that

$$f_2(k)\frac{d\sigma}{dk}(k) < 1.$$

From the previous equations, $\frac{d\sigma}{dk}(k) = \frac{D(S_0e^k) - D_{\text{BS}}(S_0e^k)}{\phi(f_2(k))}$ and because of the bounds for $D(K)$, this quantity lies in $\left[-\frac{1-\Phi(-f_2(k))}{\phi(-f_2(k))}, \frac{1-\Phi(f_2(k))}{\phi(f_2(k))}\right]$. So when $f_2(k)$ is non-negative, $f_2(k)\frac{d\sigma}{dk}(k) \leq f_2(k)\frac{1-\Phi(f_2(k))}{\phi(f_2(k))}$. Otherwise, $f_2(k)\frac{d\sigma}{dk}(k) \leq -f_2(k)\frac{1-\Phi(-f_2(k))}{\phi(-f_2(k))}$. Both quantities are less than 1.

At this point, we verify that

$$f_1(k)\frac{d\sigma}{dk}(k) < 1. \tag{3.4}$$

It holds $KD(K) \geq P(K)$. Indeed, since P is convex, its tangent at K lies below the function P itself, so for any $x \geq 0$ one has $P(K) + (x - K)D(K) \leq P(x)$ and evaluating at $x = 0$ we obtain the target inequality since $P(0) = 0$. From this inequality, using Equation (3.3) and writing the explicit formula of $P(K)$, one gets $0 \leq S_0\Phi(f_1(k)) + S_0e^k\phi(f_2(k))\frac{d\sigma}{dk}(k) = S_0\Phi(f_1(k)) + S_0\phi(f_1(k))\frac{d\sigma}{dk}(k)$. If f_1 is non-positive, then $f_1(k)\frac{d\sigma}{dk}(k) \leq -f_1(k)\frac{1-\Phi(-f_1(k))}{\phi(-f_1(k))} < 1$. Otherwise, Equation (3.4) is automatically satisfied when $\frac{d\sigma}{dk}(k)$ is non-positive, while when it is positive, we notice that $f_1(k)\frac{d\sigma}{dk}(k) = f_2\frac{d\sigma}{dk}(k) - \sigma\frac{d\sigma}{dk}(k) < 1 - \sigma\frac{d\sigma}{dk}(k) < 1$.

Finally, we show that f_1 and f_2 are strictly increasing. Indeed, from their definition, $\frac{df_{1,2}}{dk}(k) = \frac{1}{\sigma(k)}\big(1 - \frac{d\sigma}{dk}(k)\big(\frac{k}{\sigma(k)} \pm \frac{\sigma(k)}{2}\big)\big) = \frac{1}{\sigma(k)}\big(1 - \frac{d\sigma}{dk}(k)f_{2,1}(k)\big) > 0$. □

What if we start from convex Call prices defined by the Black-Scholes Call prices instead of Put prices? In such case, as discussed in Section 3.3.4, one observes first that the Put prices come from the Black-Scholes Put prices and they are convex. Secondly,

looking at the Put-Call parity, one notices that the Call prices are convex iff the Put prices are convex. Applying the previous Lemma one finds again $f'_{1,2}(k) > 0$.

The last ingredient we will require is a natural change of parameters in SVI, that we describe in the next section altogether with the main argument of our full characterization.

3.5 Normalizing SVI

We now rescale SVI in the following way, which is natural:

$$\begin{aligned} SVI(k) &= \alpha\sigma + b\sigma\left(\rho\frac{k-m}{\sigma} + \sqrt{\left(\frac{k-m}{\sigma}\right)^2 + 1}\right) \\ &= \sigma N\left(\frac{k-m}{\sigma}\right) \end{aligned}$$

with $\alpha := a/\sigma$ and $N(l) := \alpha + b(\rho l + \sqrt{l^2+1})$. With this rewriting, the derivatives of the SVI model become

$$\begin{aligned} SVI'(k) &= N'\left(\frac{k-m}{\sigma}\right), \\ SVI''(k) &= \frac{1}{\sigma}N''\left(\frac{k-m}{\sigma}\right). \end{aligned}$$

Observe that the second derivative N'' is positive so N is strictly convex. Its only critical point is a minimum that we call $l^* = -\frac{\rho}{\sqrt{1-\rho^2}}$. We gather the important properties of N in the following:

Lemma 3.3 (Normalized SVI)**.** *Let $N(l) := \alpha + b(\rho l + \sqrt{l^2+1})$ where $a = \alpha\sigma$. Then N is strictly convex with a minimum at $l^* = -\frac{\rho}{\sqrt{1-\rho^2}}$, where $N(l^*) = \alpha + b\sqrt{1-\rho^2}$. Also:*

$$\begin{aligned} N'(l) &= b\left(\rho + \frac{l}{\sqrt{l^2+1}}\right), \\ N''(l) &= \frac{b}{(l^2+1)^{\frac{3}{2}}}. \end{aligned}$$

In particular as $l \to \pm\infty$:

$$N(l) \sim \alpha + b(\rho \pm 1)l, \qquad N'(l) \to b(\rho \pm 1), \qquad N''(l) \to 0,$$

and for every k

$$SVI(k) = \sigma N\left(\frac{k-m}{\sigma}\right).$$

In the above Lemma, note that the statement $N(l) \sim a' + b(\rho \pm 1)l$ covers the cases $b = 0$ and $b \neq 0$.

Hereafter we will also put $m = \mu\sigma$, so that $k = \sigma(l + \mu)$ and

$$SVI_{a,b,\rho,m,\sigma}(k) = \sigma N_{\alpha,b,\rho}\left(\frac{k}{\sigma} - \mu\right)$$

where the parameters have the following constraints:

$$b \geq 0, \qquad |\rho| \leq 1, \qquad \mu \in \mathbb{R}, \qquad \sigma \geq 0, \qquad \alpha + b\sqrt{1-\rho^2} \geq 0.$$

3.5.1 Expressing g with $f_{1,2}$ in rescaled parameters, and our main argument

There is a nice expression of g involving the functions $f_{1,2}$; indeed as shown e.g. in [27] (eq. A9) and [49] (eq. 14):

$$\left.\frac{\partial^2 C}{\partial K^2}\right|_{K=S_0 e^k} = \phi(f_2(k))\left(f_1'(k)f_2'(k)\sqrt{w(k)} + (\sqrt{w})''(k)\right)\frac{1}{S_0 e^k}$$

where ϕ is the standard Gaussian density. By identification this yields

$$g(k) = \left(f_1'(k)f_2'(k)\sqrt{w(k)} + (\sqrt{w})''(k)\right)\sqrt{w(k)}.$$

With our rescaled parameters, we have

$$g(k) = \left(1 - \frac{kN'\left(\frac{k}{\sigma} - \mu\right)}{2\sigma N\left(\frac{k}{\sigma} - \mu\right)}\right)^2 - \frac{N'\left(\frac{k}{\sigma} - \mu\right)^2}{4}\left(\frac{1}{\sigma N\left(\frac{k}{\sigma} - \mu\right)} + \frac{1}{4}\right) + \frac{N''\left(\frac{k}{\sigma} - \mu\right)}{2\sigma}$$

and writing $G(l) := g(\sigma(l + \mu))$ we find

$$G(l) = \left(1 - \frac{(l+\mu)N'(l)}{2N(l)}\right)^2 - \frac{N'(l)^2}{4}\left(\frac{1}{\sigma N(l)} + \frac{1}{4}\right) + \frac{N''(l)}{2\sigma}.$$

We can rewrite G as

$$\begin{aligned} G(l) &= \left(1 - N'(l)\left(\frac{(l+\mu)}{2N(l)} + \frac{1}{4}\right)\right)\left(1 - N'(l)\left(\frac{(l+\mu)}{2N(l)} - \frac{1}{4}\right)\right) + \frac{1}{2\sigma}\left(N''(l) - \frac{N'(l)^2}{2N(l)}\right) \\ &= G_1(l) + \frac{1}{2\sigma}G_2(l) \end{aligned}$$

where

$$\begin{aligned} G_1(l) &:= \left(1 - N'(l)\left(\frac{(l+\mu)}{2N(l)} + \frac{1}{4}\right)\right)\left(1 - N'(l)\left(\frac{(l+\mu)}{2N(l)} - \frac{1}{4}\right)\right), \\ G_2(l) &:= N''(l) - \frac{N'(l)^2}{2N(l)}. \end{aligned}$$

Call G_{1+} the first factor of G_1 and G_{1-} the second one. Then $f'_{1,2}(\sigma(l+\mu)) = f'_{1,2}(k) = \frac{1}{\sqrt{\sigma N\left(\frac{k}{\sigma}-\mu\right)}}\left(1 - N'\left(\frac{k}{\sigma}-\mu\right)\left(\frac{k}{2\sigma N\left(\frac{k}{\sigma}-\mu\right)} \pm \frac{1}{4}\right)\right) = \frac{G_{1\pm}(l)}{\sqrt{\sigma N(l)}}$ and the Fukasawa conditions correspond to $G_{1\pm} > 0$, which entails that $G_1 > 0$.

Completing the identification yields $G_1\left(\frac{k}{\sigma}-\mu\right) = f'_1(k)f'_2(k)w(k)$ and $\frac{1}{2\sigma}G_2\left(\frac{k}{\sigma}-\mu\right) = (\sqrt{w})''(k)\sqrt{w(k)}$.

It is now instrumental to observe that:

$$g(k) = G(l) = G_1(l) + \frac{1}{2\sigma}G_2(l)$$

where

- G_1 depends only on α, b, ρ, μ,
- G_2 depends only on α, b, ρ,

so that the dependency of G in σ is particularly simple; this is the main benefit of our rescaling of SVI.

Our main argument is now as follows: the Fukasawa conditions yield that it is necessary that $G_1 > 0$; for a given choice of b, ρ, α, this will give a condition on μ, which therefore characterizes the Fukasawa conditions in SVI. Given then a parameter μ satisfying this condition, the positivity of g (or G) can be casted as a simple condition on σ:

$$\sigma \geq \sup_l \frac{G_2(l)}{G_1(l)}$$

which yields the full characterization of no Butterfly arbitrage in SVI.

In Section 3.6 we investigate the conditions on G_1 related to the Fukasawa conditions, and in Section 3.7 this latter condition on σ.

3.5.2 Classifying the normalized SVI parameters

We will use the following notations to clarify the assumptions made on the SVI parameters:

- (A1) $\alpha + b\sqrt{1-\rho^2} > 0$ and $|\rho| < 1$,
- (A2) $\alpha \geq 0$ and $|\rho| = 1$,
- (B1) $b(1 \pm \rho) < 2$,
- (B2) $b(1+\rho) < 2$ and $b(1-\rho) = 2$,
- (B3) $b(1+\rho) = 2$ and $b(1-\rho) < 2$,
- (B4) $b(1+\rho) = 2$ and $b(1-\rho) = 2$, which is equivalent to $b = 2, \rho = 0$.

In the sequel, to avoid singularities in our computations, we will assume b positive since the case $b = 0$ is the Black-Scholes case, which is a trivial case of no arbitrage, and exclude the boundary cases $|\rho| = 1$, so work under assumption (A1). We revisit those boundary cases in Section 3.7.4 where we will assume (A2).

3.6 Investigating Fukasawa necessary no arbitrage conditions

3.6.1 Limits at infinity

We have the following:

Lemma 3.4 (Limits of G_1)**.**

$$\lim_{\pm\infty} G_1(l) = \Big(\frac{1}{2} - \frac{b(\rho \pm 1)}{4}\Big)\Big(\frac{1}{2} + \frac{b(\rho \pm 1)}{4}\Big).$$

In particular, $G_1(\infty) \geq 0$ *and* $G_1(-\infty) \geq 0$ *iff simultaneously* $b(1 \pm \rho) \leq 2$.

These conditions are conditions on the asymptotic slopes of the total variance smile, and are therefore related to the Roger Lee Moment formula [48]; this is a general fact for the Fukasawa conditions: [32] contains several asymptotic statement on f_1 and f_2 which are directly related to the asymptotic behaviour of $\frac{w(k)}{k}$.

3.6.2 The conditions as an interval for μ

Let us investigate the corresponding Fukasawa conditions of positivity of G_{1+} and G_{1-} in terms of SVI parameters. We start with the following:

Lemma 3.5. *Let*

$$L_\pm(l; \alpha, b, \rho) := 2N(l)\Big(\frac{1}{N'(l)} \mp \frac{1}{4}\Big) - l \tag{3.5}$$

where L_+ *is defined on* $]l^*, +\infty[$ *and* L_- *on* $]-\infty, l^*[$. *Then* $G_{1\pm} > 0$ *if and only if* $\sup_{l<l^*} L_-(l) < \inf_{l>l^*} L_+(l)$ *and*

$$\mu \in I_{\alpha,b,\rho} :=]\sup_{l<l^*} L_-(l), \inf_{l>l^*} L_+(l)[.$$

Proof. In order to have $G_{1\pm} > 0$, we need $\sup_{l<l^*} L_\pm(l) < \mu < \inf_{l>l^*} L_\pm(l)$. Indeed $G_{1\pm}(l) = 1 - N'(l)\Big(\frac{(l+\mu)}{2N(l)} \pm \frac{1}{4}\Big)$ so that $G_{1\pm}(l) > 0$ iff $1 \mp \frac{N'(l)}{4} > N'(l)\frac{(l+\mu)}{2N(l)}$. Since $L_+(l) < L_-(l)$ for every l, we obtain an interval for μ given by $\sup_{l<l^*} L_-(l) < \mu < \inf_{l>l^*} L_+(l)$. □

Remark 3.1. *In order to alleviate the notations, we will often suppress the list of parameters in* $L_\pm$*, or when we need it just denote the dependency in* α*,* (b, ρ) *being fixed.*

What are the basic properties of L_- and L_+?

Note that $L_-(l^{*-}) = -\infty$ and, under $b(1-\rho) < 2$, $L_-(-\infty) = -\infty$. It follows that l_- such that $L_-(l_-) = \sup_{l<l^*} L_-(l)$ lays in $]-\infty, l^*[$. Similarly, $L_+(l^{*+}) = +\infty$ and $L_+(+\infty) = +\infty$ when $b(1+\rho) < 2$, so l_+ such that $L_+(l_+) = \inf_{l>l^*} L_+(l)$ lays in $]l^*, +\infty[$. When $b(1-\rho) = 2$ then $L_-(-\infty) = -\frac{\alpha}{2}$ while when $b(1+\rho) = 2$ then $L_+(+\infty) = \frac{\alpha}{2}$. Indeed at infinity L_- behaves as $2\alpha\Big(\frac{1}{b(\rho-1)} + \frac{1}{4}\Big) + \frac{2+b(\rho-1)}{2}l$ while L_+ as

$2\alpha\left(\frac{1}{b(\rho+1)}-\frac{1}{4}\right)+\frac{2-b(\rho+1)}{2}l$. In these cases the supremum of L_- (or the infimum of L_+), could be reached at $-\infty$ (or $+\infty$).

Experiments show that not every choice of (α, b, ρ) leads to $L_-(l) < -\epsilon < 0$ for all $l < l^*$ and $L_+(l) > \epsilon > 0$ for all $l > l^*$, so the interval for μ could be empty: for example, for $\alpha = -0.8, b = 1$ and $\rho = 0.5$, we have $L_-(l_-) > L_+(l_+)$. This suggests that the situation is intricate; we show below that when $\alpha \geq 0$, the interval is non-empty.

The case $\alpha \geq 0$

In the case $\alpha \geq 0$, we can indeed demonstrate that the interval for μ is non-empty, with the following easy argument:

L_- is negative for $l < l^*$ iff $\frac{N}{2N'}(4+N')-l$ is negative. In this domain N' is negative, so the previous condition is equivalent to ask $N(4+N')-2lN' > 0$, or equivalently $2(N-lN')+N(2+N') > 0$. Let us consider the first term. We have $N-lN' = \alpha + b\sqrt{l^2+1}-\frac{bl^2}{\sqrt{l^2+1}}$ which is greater than 0 iff, multiplying by $\sqrt{l^2+1}$, also $\alpha\sqrt{l^2+1}+b > 0$ or equivalently $\alpha > -\frac{b}{\sqrt{l^2+1}}$. This holds for $\alpha \geq 0$ (note that the latter quantity reaches its maximum at $-\infty$ where it equals 0, so this proof cannot handle the case $\alpha < 0$).

We can now consider the second term. We want $2+N' > 0$. Since $N' > b(\rho-1)$, then $2+N' > 2+b(\rho-1) \geq 0$. So L_- is always strictly negative for $l < l^*$ and $\alpha \geq 0$.

Similarly, L_+ is positive for $l > l^*$ iff $2(N-lN')+N(2-N') > 0$. With the same arguments as before we obtain that L_+ is strictly positive for $l > l^*$ and $\alpha \geq 0$.

Under (B1), we showed $L_-(-\infty) = -\infty$ and $L_+(\infty) = \infty$, so the interval I is non-empty.

When $b(1-\rho) = 2$ or $b(1+\rho) = 2$ this result is still valid. Since in such cases $L_-(-\infty) = -\frac{\alpha}{2}$ and $L_+(+\infty) = \frac{\alpha}{2}$ respectively, then L_- is negative in $[-\infty, l^*[$ while L_+ is positive in $]l^*, +\infty]$ for $\alpha > 0$. Otherwise if $\alpha = 0$, $\sup_{l<l^*} L_-(l) = L_-(-\infty) = 0$ and $\inf_{l>l^*} L_+(l) = L_+(+\infty) = 0$ respectively.

We have proven the following:

Lemma 3.6 (SVI parameters fulfilling Fukasawa necessary no arbitrage conditions: case $\alpha \geq 0$)**.** *Assume (A1). For every (α, b, ρ) with $\alpha \geq 0$:*

- *under (B1), the interval $I_{\alpha,b,\rho}$ is non-empty and contains 0;*
- *under (B2),*
  - *if $\alpha > 0$, the interval $I_{\alpha,b,\rho}$ is non-empty and contains 0;*
  - *if $\alpha = 0$, the interval $I_{0,b,\rho}$ is non-empty and has 0 as left boundary;*
- *under (B3),*
  - *if $\alpha > 0$, the interval $I_{\alpha,b,\rho}$ is non-empty and contains 0;*
  - *if $\alpha = 0$, the interval $I_{0,b,\rho}$ is non-empty and has 0 as right boundary;*
- *under (B4),*

– *if* $\alpha > 0$, *the interval* $I_{\alpha,2,0}$ *is non-empty and contains* 0*;*

– *if* $\alpha = 0$, *the interval* $I_{0,2,0}$ *is empty.*

Computation of the interval for μ under (B1)

We tackle now the computation of the interval for μ in the general case where α is not necessarily positive, which is less straightforward. In this section we will assume (B1); we deal with the other cases in the dedicated Section 3.6.2.

We consider the function L_- for $l < l^*$ and L_+ for $l > l^*$. We have $L'_\pm(l) = 1 \mp \frac{N'}{2} - \frac{2NN''}{N'^2}$ and it follows that $L'_-(l_-) = L'_+(l_+) = 0$.

The corresponding equations in l are:

$$1 \mp \frac{b}{2}\left(\rho + \frac{l}{\sqrt{l^2+1}}\right) - \frac{2(\alpha + b(\rho l + \sqrt{l^2+1}))}{b\sqrt{l^2+1}(\rho\sqrt{l^2+1} + l)^2} = 0.$$

Actually, we don't need to solve these equations. Accordingly, we set:

$$g_{\pm(b,\rho)}(l) = \left(\rho\sqrt{l^2+1} + l\right)^2\left(\sqrt{l^2+1}\left(\frac{1}{2} \mp \frac{b\rho}{4}\right) \mp \frac{bl}{4}\right) - \left(\rho l + \sqrt{l^2+1}\right) \tag{3.6}$$

where $g_{+(b,\rho)}$ is defined on $[l^*, \infty[$ and $g_{-(b,\rho)}$ on $]-\infty, l^*]$. Then $L'_\pm(l) = 0$ iff $g_{\pm(b,\rho)}(l) = \frac{\alpha}{b}$.

The following technical result turns to be a key one:

Proposition 3.3. *Assume (A1) and (B1), and let* $g_{\pm(b,\rho)}$ *defined by Equation (3.6). Then* $g_{\pm(b,\rho)}(l^*) = -\sqrt{1-\rho^2}$, $g_{\pm(b,\rho)}(\pm\infty) = \infty$, *and* $g_{\pm(b,\rho)}$ *is either monotonous or with a single local minimum. Let* $s_\pm = l^*$ *if* $g_{\pm(b,\rho)}$ *is monotonous and* $s_\pm \neq l^*$ *such that* $g_{\pm(b,\rho)}(s_\pm) = -\sqrt{1-\rho^2}$ *otherwise. Then:*

- $L_-(x; bg_{-(b,\rho)}(x)) = \sup_{l<l^*} L_-(l; bg_{-(b,\rho)}(x))$ *for any* $x < s_-$, $L_-(x; bg_{-(b,\rho)}(x)) \to -\infty$ *when* $x \to -\infty$, *the function* $L_-(x; bg_{-(b,\rho)}(x))$ *is increasing and the function* $g_{-(b,\rho)}$ *is decreasing on* $]-\infty, s_-[$*;*

- $L_+(x; bg_{+(b,\rho)}(x)) = \inf_{l>l^*} L_+(l; bg_{+(b,\rho)}(x))$ *for any* $x > s_+$, $L_+(x; bg_{+(b,\rho)}(x)) \to +\infty$ *when* $x \to +\infty$, *and both the functions* $L_+(x; bg_{+(b,\rho)}(x))$ *and* $g_{+(b,\rho)}$ *are increasing on* $]s_+, \infty[$.

The proof is provided in Section 3.A. We display a typical plot of $g_{-(b,\rho)}$ and $g_{+(b,\rho)}$ in Figure 3.1.

This proposition has in turn two important corollaries:

Corollary 3.1. *Assume (A1) and (B1). There is a unique* (l_-, l_+) *such that* $l_- < l^* < l_+$ *and* $\alpha = bg_{-(b,\rho)}(l_-) = bg_{+(b,\rho)}(l_+)$. *The interval* $I_{\alpha,b,\rho}$ *is non-empty iff* $L_-(l_-; \alpha) < L_+(l_+; \alpha)$. *In this case the distance between* $L_+(l_+; \alpha)$ *and* $L_-(l_-; \alpha)$ *increases with* α.

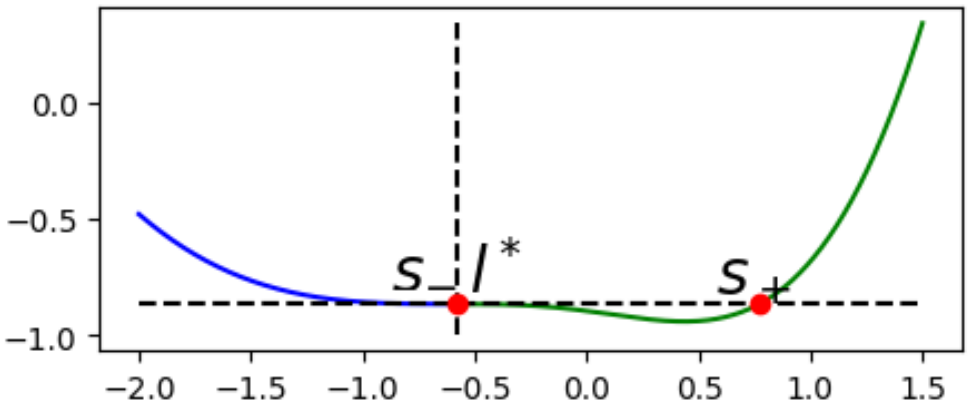

Figure 3.1: Typical plot of the functions $g_{\pm(b,\rho)}$ with $b = \frac{2}{3}$ and $\rho = \frac{1}{2}$. The horizontal dotted line corresponds to the level $-b\sqrt{1-\rho^2}$.

Proof. This follows directly from the previous analysis: the function $g_{+(b,\rho)}$ is increasing on $]s_+, \infty[$ and the function $g_{-(b,\rho)}$ is decreasing on $]-\infty, s_-[$ with $g_{\pm(b,\rho)}(\pm\infty) = \infty$. So the solutions $l_\pm = l_\pm(\alpha)$ of $bg_{\pm(b,\rho)}(l_\pm) = \alpha$ exist for $\alpha > -b\sqrt{1-\rho^2} = bg_\pm(s_\pm)$, and $l_\pm$ are monotone with respect to α. In turn, the function $L_+(l_+; bg_{+(b,\rho)}(l_+))$ increases and the function $L_-(l_-; bg_{-(b,\rho)}(l_-))$ decreases. Note that $l_- < s_-$ and $l_+ > s_+$ because $\alpha > -b\sqrt{1-\rho^2}$ from (A1). We can also use the fact that

$$\begin{aligned}\frac{d}{d\alpha}(L_+(l_+;\alpha) - L_-(l_-;\alpha)) &= L'_+(l_+)\frac{d}{d\alpha}l_+ - L'_-(l_-)\frac{d}{d\alpha}l_- + \partial_\alpha L_+(l_+;\alpha) - \partial_\alpha L_-(l_-;\alpha) \\ &= \partial_\alpha L_+(l_+;\alpha) - \partial_\alpha L_-(l_-;\alpha)\end{aligned}$$

where l_+ and l_- are functions of α given by $\alpha = bg_{+(b,\rho)}(l_+) = bg_{-(b,\rho)}(l_-)$. Now, the RHS is equal to $2\left(\frac{1}{N'(l_+)} - \frac{1}{N'(l_-)} - \frac{1}{2}\right)$ and since $\frac{1}{N'(l_+)} > \frac{1}{2}$ and $-\frac{1}{N'(l_-)} > \frac{1}{2}$, the previous quantity is greater than 1. □

Let $F(b,\rho)$ denote the unique value of α such that $L_+(l_+;\alpha) = L_-(l_-;\alpha)$ if there exists such a value for $\alpha > -b\sqrt{1-\rho^2}$, otherwise let $F(b,\rho) = -b\sqrt{1-\rho^2}$. Then $L_+(l_+;\alpha) > L_-(l_-;\alpha)$ if and only if $\alpha > F(b,\rho)$. In other words we define $F(b,\rho)$ as:

$$F(b,\rho) := \inf\{\alpha \mid L_+(l_+;\alpha) > L_-(l_-;\alpha)\} \vee -b\sqrt{1-\rho^2}$$

under the assumptions (A1) and (B1). We name F the Fukasawa threshold of SVI.

Figure 3.2 shows:

- in blue the function $l_- \to L_-(l_-; bg_{-(b,\rho)}(l_-))$ with $l_- < s_-$ where s_- is such that $g_{-(b,\rho)}(s_-) = -\sqrt{1-\rho^2}$;
- in green the corresponding value of $l_- \to L_+(l_+(bg_{-(b,\rho)}(l_-)); bg_{-(b,\rho)}(l_-))$.

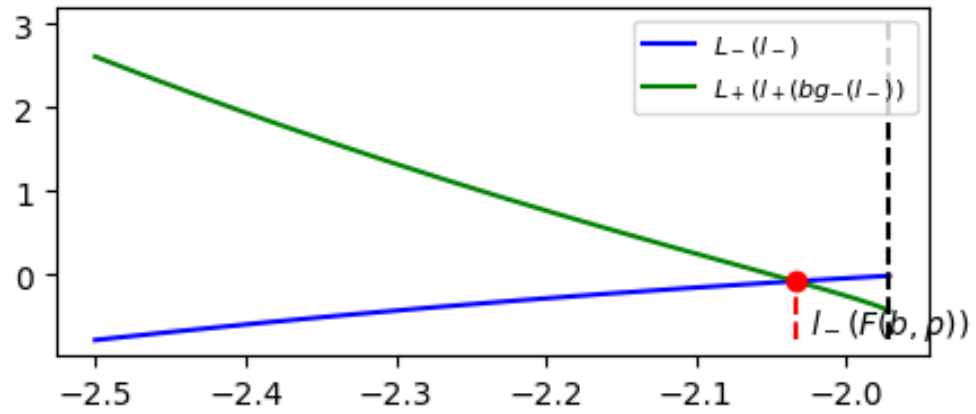

Figure 3.2: Plot of $L_-(l_-)$ and $L_+(l_+(bg_{-(b,\rho)}(l_-)))$ as functions of l_-, with $b = \frac{25}{21}$ and $\rho = \frac{2}{5}$. The vertical dotted line corresponds to the level $-b\sqrt{1-\rho^2}$.

The following corollary gives an easy criterion of existence of a Butterfly arbitrage:

Corollary 3.2. *Assume (A1) and (B1). If $\alpha \leq F(b,\rho)$ then for every choice of μ and σ, the SVI model does not satisfy the Fukasawa conditions.*

Study of the Fukasawa threshold under (B1)

In the previous section we showed that the difference $L_+(l_+; bg_{+(b,\rho)}(l_+)) - L_-(l_-; bg_{-(b,\rho)}(l_-))$ goes to infinity when increasing $\alpha = bg_{+(b,\rho)}(l_+) = bg_{-(b,\rho)}(l_-)$ to infinity, so there exists $\bar{\alpha}$ such that the interval for μ is non-empty; from the previous corollaries for each $\alpha > \bar{\alpha}$ the interval for μ is also non-empty. Decreasing α, we could bump into two situations:

- α reaches the value $F(b,\rho) > -b\sqrt{1-\rho^2}$ for which $L_+(l_+; F(b,\rho)) = L_-(l_-; F(b,\rho))$;
- α reaches the value $F(b,\rho) = -b\sqrt{1-\rho^2}$. In such case $l_\pm = s_\pm$.

Our simulations suggest that the first scenario always occurs.

Could we prove this? In this respect we can observe the following: it is equivalent to prove that $L_+(s_+; -b\sqrt{1-\rho^2}) < L_-(s_-; -b\sqrt{1-\rho^2})$.

If $s_+ = l^*$ then $L_+(s_+; -b\sqrt{1-\rho^2}) = -l^*$ and the function $L_+(l_+; bg_{+(b,\rho)}(l_+))$ is increasing. It follows that the function $L_-(l_-; bg_{-(b,\rho)}(l_-))$ cannot be increasing and $s_- < l^*$. We should show that $L_-(s_-; -b\sqrt{1-\rho^2}) > -l^*$.

When $s_- = l^*$ then $L_-(s_-; -b\sqrt{1-\rho^2}) = -l^*$ and the function $L_-(l_-; bg_{-(b,\rho)}(l_-))$ is increasing. Again, the function $L_+(l_+; bg_{+(b,\rho)}(l_+))$ cannot be increasing so $s_+ > l^*$. In this case we should prove that $L_+(s_+; -b\sqrt{1-\rho^2}) < -l^*$.

In the final case when both $g_{\pm(b,\rho)}$ have a minimum, it is enough to prove $L_-(s_-; -b\sqrt{1-\rho^2}) > -l^*$ and $L_+(s_+; -b\sqrt{1-\rho^2}) < -l^*$.

So to sum up, it would remain to prove that when $g_{-(b,\rho)}$ (or $g_{+(b,\rho)}$) has a minimum, $L_-(s_-; -b\sqrt{1-\rho^2}) > -l^*$ (or $L_+(s_+; -b\sqrt{1-\rho^2}) < -l^*$) to obtain the result in each case. We did not manage to conclude along those lines though.

Remark 3.2. *We don't know whether $F(b,\rho) > -b\sqrt{1-\rho^2}$ but we conjecture it. Indeed we prove in Section 3.B that there is a closed formula for $F(b,0)$ which satisfies $F(b,0) > -b$; the statement $F(b,\rho) > -b\sqrt{1-\rho^2}$ can be also assessed numerically.*

Symmetries

We can exploit the symmetry property of N with respect to ρ in order to restrict the required computations to the function L_- only.

Indeed $N(l;\alpha,b,\rho) = N(-l;\alpha,b,-\rho), N'(l;b,\rho) = -N'(-l;b,-\rho)$ and $N''(l;b) = N''(-l;b)$. This brings to the consideration that

$$L_-(l;\alpha,b,\rho) = -L_+(-l;\alpha,b,-\rho), \quad L_+(l;\alpha,b,\rho) = -L_-(-l;\alpha,b,-\rho),$$

so that

$$\begin{aligned}\inf_{l>l^*(\rho)} L_+(l;\alpha,b,\rho) &= -\sup_{l>l^*(\rho)} L_-(-l;\alpha,b,-\rho)\\ &= -\sup_{l<-l^*(\rho)} L_-(l;\alpha,b,-\rho)\\ &= -\sup_{l<l^*(-\rho)} L_-(l;\alpha,b,-\rho)\end{aligned}$$

so $L_+(l_+(\alpha,b,\rho);\alpha,b,\rho) = -L_-(l_-(\alpha,b,-\rho);\alpha,b,-\rho)$.

Note that $l_+(\alpha,b,\rho)$ is the unique $l > l^*(\rho)$ such that $L'_+(l;\alpha,b,\rho) = 0$ while $l_-(\alpha,b,-\rho)$ is the unique $l < l^*(-\rho)$ such that $L'_-(l;\alpha,b,-\rho) = 0$. Since $L'_+(l;\alpha,b,\rho) = L'_-(-l;\alpha,b,-\rho)$ and $-l_-(\alpha,b,-\rho) > -l^*(-\rho) = l^*(\rho)$, then $l_+(\alpha,b,\rho) = -l_-(\alpha,b,-\rho)$.

Lemma 3.7. *Assume (A1) and (B1). Then:*

- $L_+(l_+(\alpha,b,\rho);\alpha,b,\rho) = -L_-(l_-(\alpha,b,-\rho);\alpha,b,-\rho)$;
- $l_+(\alpha,b,\rho) = -l_-(\alpha,b,-\rho)$;
- $I_{\alpha,b,\rho} = \left]L_-(l_-(\alpha,b,\rho);\alpha,b,\rho), -L_-(l_-(\alpha,b,-\rho);\alpha,b,-\rho)\right[$.

From the above equations we also have $g_{+(b,\rho)}(l) = g_{-(b,-\rho)}(-l)$ so with easy arguments one gets $s_+(b,\rho) = -s_-(b,-\rho)$.

The cases (B2), (B3) and (B4)

Assume (B2) or (B4). Using the same definitions and following the proof of Proposition 3.3, we obtain that $g_{-(b,\rho)}(l)$ is increasing. Now since $g_{-(b,\rho)}$ is increasing on $]-\infty, l^*]$ and since $g_{-(b,\rho)}(l^*) = -\sqrt{1-\rho^2}$, it follows that there is no solution to the

equation $g_{-(b,\rho)}(l_-) = \frac{\alpha}{b}$. In this case so, the supremum of L_- is attained at $-\infty$ and it is $-\frac{\alpha}{2}$. Under (B3) or (B4), for symmetrical reasons L_+ attains its infimum $\frac{\alpha}{2}$ at ∞.

Under (B2), L_+ reaches its infimum in $]l^*, +\infty[$ while under (B3), L_- reaches its supremum in $]-\infty, l^*[$. Finally under (B4), $I_{\alpha,2,0} = \left]-\frac{\alpha}{2}, \frac{\alpha}{2}\right[$.

We can extend the definition of the Fukasawa threshold to the cases (B2), (B3) and (B4):

- under (B2), $F(b,\rho)$ denotes the unique value of α such that $L_-(l_-(\alpha,b,-\rho);\alpha,b,-\rho) = \frac{\alpha}{2}$ if there exists such a value for $\alpha > -b\sqrt{1-\rho^2}$, otherwise $F(b,\rho) = -b\sqrt{1-\rho^2}$: $F(b,\rho) := \inf\{\alpha \mid L_+(l_+;\alpha) > -\frac{\alpha}{2}\} \vee -b\sqrt{1-\rho^2}$;
- under (B3), $F(b,\rho)$ denotes the unique value of α such that $L_-(l_-(\alpha,b,\rho);\alpha,b,\rho) = \frac{\alpha}{2}$ if there exists such a value for $\alpha > -b\sqrt{1-\rho^2}$, otherwise $F(b,\rho) = -b\sqrt{1-\rho^2}$: $F(b,\rho) := \inf\{\alpha \mid \frac{\alpha}{2} > L_-(l_-;\alpha)\} \vee -b\sqrt{1-\rho^2}$;
- under (B4), $F(2,0) := 0$.

From Lemma 3.6, under cases (B2) and (B3) it holds $F(b,\rho) < 0$.

3.6.3 Conclusion

We can now state the full characterization of the Fukasawa necessary no arbitrage conditions for SVI:

Theorem 3.1 (SVI parameters $(\alpha,b,\rho,\mu,\sigma)$ fulfilling Fukasawa necessary no arbitrage conditions)**.** *Assume (A1). Then:*

- *under (B1), $F(b,\rho) < 0$ and the interval $I_{\alpha,b,\rho} = \left]L_-(l_-(\alpha,b,\rho);\alpha,b,\rho), -L_-(l_-(\alpha,b,-\rho);\alpha,b,-\rho)\right[$ is non-empty iff $\alpha > F(b,\rho)$;*
- *under (B2) (resp. (B3)), $F(b,\rho) < 0$ and the interval $I_{\alpha,b,\rho} = \left]-\frac{\alpha}{2}, -L_-(l_-(\alpha,b,-\rho);\alpha,b,-\rho)\right[$ (resp. $I_{\alpha,b,\rho} = \left]L_-(l_-(\alpha,b,\rho);\alpha,b,\rho), \frac{\alpha}{2}\right[$) is non-empty iff $\alpha > F(b,\rho)$;*
- *under (B4), the interval $I_{\alpha,2,0} = \left]-\frac{\alpha}{2}, \frac{\alpha}{2}\right[$ is non-empty iff $\alpha > F(2,0) = 0$.*

In every case, the Fukasawa conditions are satisfied iff $\mu \in I_{\alpha,b,\rho}$.

Except for $F(2,0)$, the result $F(b,\rho)$ negative holds even in the case $F(b,\rho) > -b\sqrt{1-\rho^2}$ because we have proven that for $\alpha \geq 0$ the interval for μ is always non-empty. In terms of the usual SVI parameters the conditions translate into $\frac{a}{\sigma} > F(b,\rho)$ and $\frac{m}{\sigma} \in I_{\frac{a}{\sigma},b,\rho}$.

Is the existence of the Fukasawa threshold surprising? We would say no: indeed the values of α too close to the lower bound $-b\sqrt{1-\rho^2}$ correspond to values of the smile too close to zero, and this will lead to an arbitrage as discussed in Section 3.3.2, so that one even expects that $F(b,\rho) > -b\sqrt{1-\rho^2}$.

The explanation of the range constraint for μ is less intuitive to us; we would say that it results from the geometrical constraint that the Fukasawa conditions impose on the shape of SVI, as follows from our computations.

3.6.4 Numerics

$F(b,\rho)$ at a fixed b We plot in Figure 3.3 the Fukasawa threshold at fixed $b = \frac{1}{2}$ as a function of ρ.

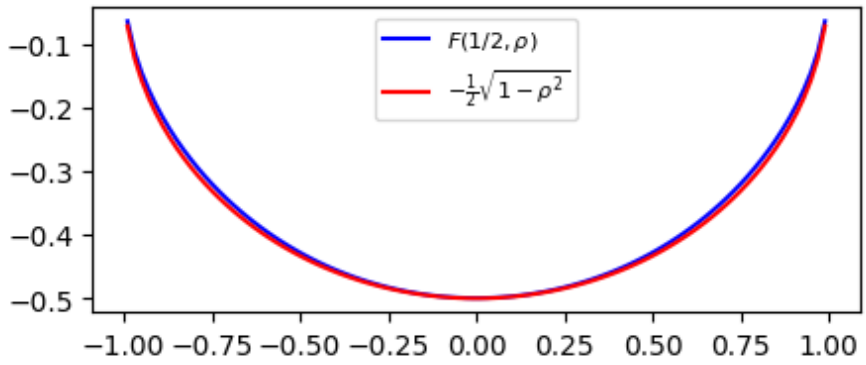

Figure 3.3: Plot of $F(b,\rho)$ as a function of ρ, with $b = \frac{1}{2}$.

The graph is symmetric with respect to ρ because $F(b,\rho)$ is the value of α such that $L_+(l_+(\alpha,b,\rho);\alpha,b,\rho) - L_-(l_-(\alpha,b,\rho);\alpha,b,\rho) = 0$, where $bg_{\pm(b,\rho)}(l_\pm(\alpha,b,\rho)) = \alpha$. But $L_+(l_+(\alpha,b,\rho);\alpha,b,\rho) = -L_-(l_-(\alpha,b,-\rho);\alpha,b,-\rho)$ so we look for α such that

$$L_-(l_-(\alpha,b,-\rho);\alpha,b,-\rho) + L_-(l_-(\alpha,b,\rho);\alpha,b,\rho) = 0$$

and this is symmetric with respect to ρ.

The red line is the level $\alpha = -b\sqrt{1-\rho^2}$ and it again supports our conjecture that $F(b,\rho) > -b\sqrt{1-\rho^2}$.

From the previous graph, it seems that $F(b,\rho)$ has monotonicity of the same sign as ρ.

$F(b,\rho)$ at fixed ρ as a function of b In Figure 3.4 we plot the Fukasawa threshold at fixed $\rho = \frac{1}{5}$ as a function of b.

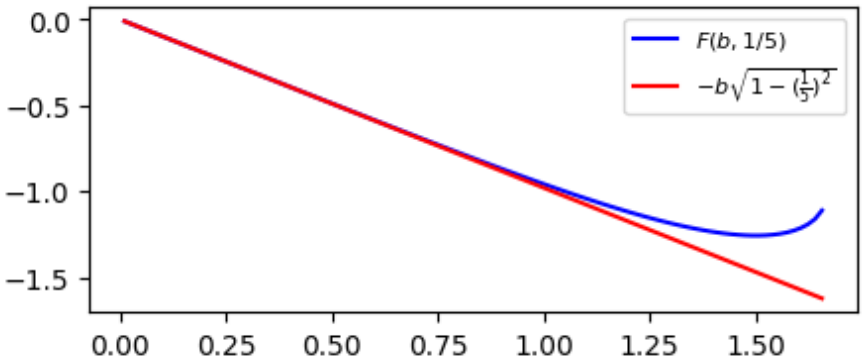

Figure 3.4: Plot of $F(b,\rho)$ as a function of b, with $b = \frac{1}{5}$.

$L_-(l_-(F(b,\rho),b,\rho);F(b,\rho),b,\rho)$ and $L_-(l_-(F(b,\rho),b,-\rho);F(b,\rho),b,-\rho)$ as functions of ρ The following Figure 3.5 shows in blue the function $L_-(l_-(F(b,\rho),b,\rho);F(b,\rho),b,\rho)$ (denoted for brevity as $L_-(F(b,\rho),\rho)$) with respect to ρ while in green the function $L_-(l_-(F(b,\rho),b,-\rho);F(b,\rho),b,-\rho)$ (or $L_-(F(b,\rho),-\rho)$) with respect to ρ. The fixed value for b is $\frac{3}{5}$.

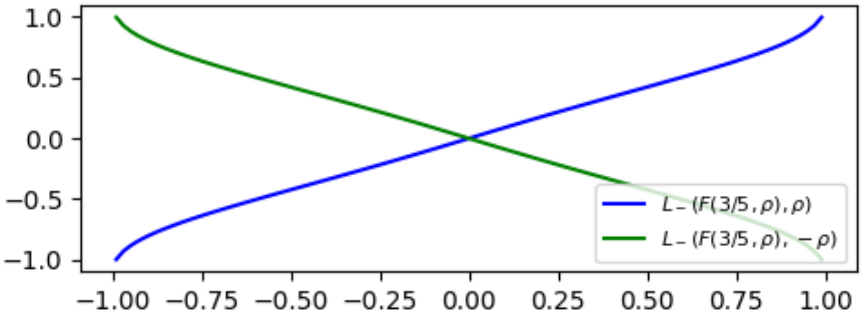

Figure 3.5: Plot of $L_-(l_-(F(b,\rho),\rho)$ and $L_-(F(b,\rho),-\rho)$ as functions of ρ, with $b = \frac{3}{5}$.

This graph also shows in blue the value of the two bounds for μ when they shrink to one point. Note that for $\rho = 0$ this is 0 for every b, while it depends on b for the other values of ρ.

The function $\rho \to L_-(l_-(F(b,\rho),b,\rho);F(b,\rho),b,\rho)$ is odd due to the symmetry of $\rho \to F(b,\rho)$. Furthermore, from the graph it seems that ρ and $L_-(l_-(F(b,\rho),b,\rho);F(b,\rho),b,\rho)$ have the same sign.

$L_-(l_-(F(b,\rho),b,\rho);F(b,\rho),b,\rho)$ as a function of b Figure 3.6 shows the function $L_-(l_-(F(b,\rho),b,\rho);F(b,\rho),b,\rho)$ (denoted as $L_-(F(b,\rho),\rho)$) with respect to b. Here we

fix $\rho = \frac{1}{2}$.

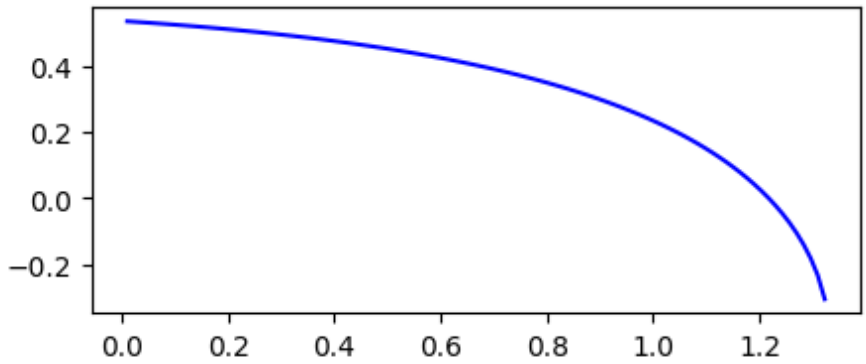

Figure 3.6: Plot of $L_-(l_-(F(b,\rho),\rho)$ as a function of b, with $\rho = \frac{1}{2}$.

3.6.5 Algorithm

We can *parametrize* the normalized SVI parameters satisfying the Fukasawa conditions as follows:

1. choose $\rho \in]-1,1[$ and b positive such that $b(1 \pm \rho) \leq 2$ by choosing $b' \in]0,1]$ and setting $b = b' \frac{2}{1+|\rho|}$;

2. compute numerically $F(b,\rho)$, and parametrize α by setting $\alpha = F(b,\rho) + u$ for positive u;

3. compute numerically (L_-, L_+) for this value of u, and parametrize μ by setting $\mu = \frac{(1+q)}{2} L_+ + \frac{(1-q)}{2} L_-$ for $q \in]-1,1[$.

The values in point 3 can be computed using the same functions employed to find $F(b,\rho)$, indeed it is sufficient to evaluate $L_-(l_-(\alpha,b,\rho);\alpha,b,\rho)$ and $-L_-(l_-(\alpha,b,-\rho);\alpha,b,-\rho)$.

If we are interested only by a test that a given parameter satisfies the Fukasawa conditions, we have the corresponding waterfall of failure possibilities that we define as follows:

1. $b(1-\rho) > 2$ or $b(1+\rho) > 2$: *failure of type 1*; otherwise:

2. $\alpha \leq F(b,\rho)$: *failure of type 2*; otherwise:

3. μ not in $I_{\alpha,b,\rho}$: *failure of type 3*.

Application to Axel Vogt parameters

The so-called *Axel Vogt example* (cf [36]) became the archetypal example of a smile with arbitrage. The SVI parameters are

$$(a, b, \rho, m, \sigma) = (-0.041, 0.1331, 0.3060, 0.3586, 0.4153),$$

and they are known to lead to a Butterfly arbitrage. Do they satisfy the Fukasawa conditions?

No, since the respective value for μ is 0.86347, while its arbitrage free interval is $]-0.72407, 0.82939[$.

The Fukasawa conditions are not satisfied because of μ. However $\alpha = -0.09872$ and $F(b, \rho) = -0.12663$, so $\alpha > F(b, \rho)$ and the interval for μ is non-empty. The problem here is due to μ, which is too large: we face a *failure of type 3*.

3.7 No arbitrage domain for SVI

3.7.1 Behaviour of the function G_2

Recall that the function G_2 is defined as

$$G_2(l) := N''(l) - \frac{N'(l)^2}{2N(l)} \tag{3.7}$$

and that it depends only on (α, b, ρ). As discussed in Section 3.5.1, G_2 is positively proportional to the second derivative of the *volatility* smile, meaning of $\sqrt{\text{SVI(k)}}$. Since the *variance* smile is convex and asymptotically linear on both sides, it is expected that G_2 will be asymptotically *negative*, while it is positive around the minimum of the smile. In particular it is expected that it will have zeros, on both sides of the minimum of the smile.

The zeros of G_2

In this section we prove the following:

Lemma 3.8 (Zeros of G_2)**.** *G_2 has exactly two zeros l_1, l_2 which satisfy $l_1 < l^* \wedge 0$ and $l_2 > l^* \vee 0$ such that $G_2(l) > 0 \iff l \in]l_1, l_2[$. Furthermore, $G_2(l) \to 0^-$ for $l \to \pm\infty$.*

Proof. For $l \to \pm\infty$ we have that the first addend behaves as bl^{-3} while the second as $-\frac{b(\rho\pm1)}{2}l^{-1}$, so G_2 behaves as $-\frac{b(\rho\pm1)}{2}l^{-1}$. This means that G_2 goes to 0^- as $l \to \pm\infty$. Since $G_2(l^*) = N''(l^*) > 0$ and G_2 is continuous, then there exists an interval $]l_1, l_2[$ containing l^* such that for every l in this interval, G_2 is positive. It follows that G_2 has at least two zeros. Deriving, we find the following interesting relationship between G_2' and G_2:

$$G_2'(l) = N'''(l) - \frac{N'(l)}{N(l)}G_2(l).$$

We will prove now that this relationship entails that the first zero of G_2 is negative. Indeed if $l_1 > 0$ is the first zero of G_2, since

$$N'''(l) = -\frac{3bl}{(l^2+1)^{\frac{5}{2}}}, \tag{3.8}$$

we have $G_2'(l_1) < 0$, which is not possible because $G_2(l)$ is negative for every $l < l_1$. If $l_1 = 0$, then $G_2'(0) = 0$ but 0 cannot be a point of local maximum for G_2, otherwise there would be a following zero $l_2 > 0$. In such case, $G_2'(l_2) < 0$ for Equation (3.8) but having G_2 so far negative, it should be increasing in l_2. Then 0 could at most be an inflection point. However,

$$G_2''(l) = N^{iv}(l) - \frac{N'(l)N'''(l)}{N(l)} + \left(2\frac{N'(l)^2}{N(l)^2} - \frac{N''(l)}{N(l)}\right)G_2(l)$$

so $G_2''(0) = N^{iv}(0) = -3b$, which is negative since $b > 0$. Therefore, the first zero l_1 of G_2 is necessarily negative. With similar arguments we obtain that the next zero l_2 must be non-negative. Suppose $l_2 = 0$. Then, as before, $G_2'(0) = 0$ and $G_2''(0) = -3b < 0$, so it would be a point of local maximum, which is not possible. Then l_2 must be positive.

Moreover, there cannot be other zeros for G_2. Indeed, suppose l_3 was the first zero after l_2. Then $l_3 > 0$ and from Equation (3.8) it should be $G_2'(l_3) < 0$ but this cannot be true since G_2 is negative in the left neighborhood of l_3.

This leads to the conclusion that G_2 has exactly two zeros, one positive and the other one negative. As a consequence, $G_2(0) = b\big(1 - \frac{b\rho^2}{2(\alpha+b)}\big) > 0$. This could have been obtained also from the fact that $\alpha + b\sqrt{1-\rho^2} \geq 0$ due to the positivity of N.

Then, we find that $G_2 > 0$ in $[l^*, 0]$ when $\rho \geq 0$ or in $[0, l^*]$ when $\rho < 0$. □

Substituting the explicit formulas for N, N' and N'' in Equation (3.7), we obtain

$$G_2(l) = \frac{b}{(l^2+1)^{\frac{3}{2}}} - \frac{b^2(\rho\sqrt{l^2+1}+l)^2}{2(l^2+1)(\alpha + b(\rho l + \sqrt{l^2+1}))}$$

which leads to the remark that $\frac{G_2(l)}{b} = \tilde{G}_{2,\frac{\alpha}{b},\rho}(l)$ where $\tilde{G}_{2,x,\rho}(l) := \frac{1}{(l^2+1)^{\frac{3}{2}}} - \frac{(\rho\sqrt{l^2+1}+l)^2}{2(l^2+1)(x+(\rho l+\sqrt{l^2+1}))}$, which reduces in general the study of G_2 to the study of a 2-parameters function.

In order to find the zeros of G_2 we should solve $2\frac{\alpha}{b} + b(2-l^2)\sqrt{l^2+1} - \rho^2(l^2+1)^{\frac{3}{2}} - 2\rho l^3 = 0$ or equivalently $2\frac{\alpha}{b} - 2\rho l^3 = ((\rho^2+1)l^2 + \rho^2 - 2)\sqrt{l^2+1}$.

Note that when $\rho = 0$ this equation is explicitly solvable.

Plot of a typical G_2 function

We plot in Figure 3.7 the function G_2 for the parameters $\alpha = \frac{1}{10}$, $b = \frac{1}{2}$, $\rho = -\frac{3}{10}$.

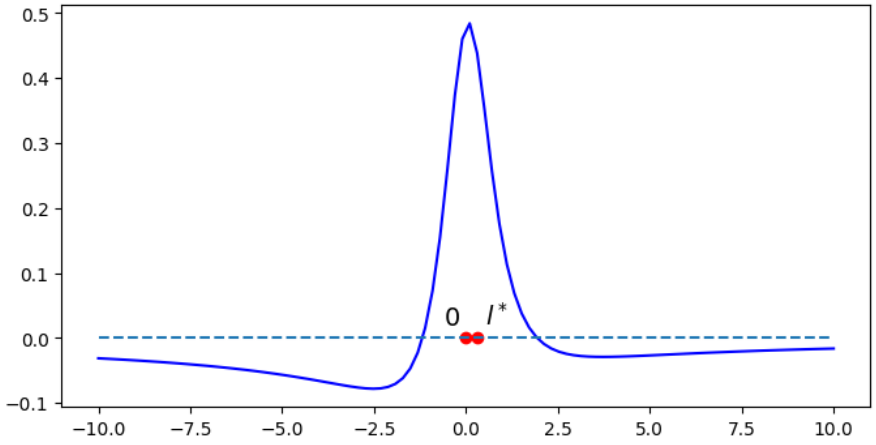

Figure 3.7: Plot of G_2 with $\alpha = \frac{1}{10}$, $b = \frac{1}{2}$ and $\rho = -\frac{3}{10}$.

3.7.2 The final condition on σ under (A1)

We recall that the non-negativity of the Durrleman condition in the case of SVI amounts to the non-negativity of the function

$$G(l) = G_1(l) + \frac{1}{2\sigma} G_2(l) \tag{3.9}$$

where G_1 and G_2 do not depend on σ.

We have proven that:

1. for every (α, b, ρ) with $b(1 \pm \rho) \leq 2$ and $\alpha > F(b, \rho)$, where $F(b, \rho) \leq 0$, there exists an interval for μ such that G_1 is positive on $\mathbb{R}$ (in fact each factor of G_1 is positive on $\mathbb{R}$). Moreover it is necessary that the conditions on (α, b, ρ) hold and that μ lies in this interval under no arbitrage.

2. for every (α, b, ρ) with $b(1 \pm \rho) \leq 2$ there exists an interval $]l_1, l_2[$ containing 0 and l^* such that $G_2(l) > 0$ iff $l \in]l_1, l_2[$.

We insist here again on the key property brought by the Fukasawa condition that it is *necessary* that G_1 is positive. This structures a lot the picture; previous to Fukasawa's observation, people investigating the positivity of G could not assume this. Another consequence is that under the Fukasawa conditions of Section 3.4, G is granted to be positive on $[l_1, l_2]$.

The last step is to exploit the fact that thanks to our re-parametrization, the dependency of G in σ is very simple. Let Π stand for a fixed set of parameters (α, b, ρ, μ) fulfilling the Fukasawa conditions. Then given the fact that $G_2(l) < 0$ for some l, it follows that if G is non-negative everywhere for (Π, σ), then G is also non-negative everywhere for every (Π, τ) with $\tau > \sigma$. As a consequence, there exists a function $\Pi \to \sigma^*(\Pi)$ such that G is non-negative everywhere for (Π, τ) iff $\tau \geq \sigma^*(\Pi)$.

The value of σ^* can be obtained asking the RHS of Equation (3.9) to be non-negative, which holds for $\sigma \geq \sup_l -\frac{G_2(l)}{2G_1(l)}$. Then

$$\sigma^*(\alpha, b, \rho, \mu) := \sup_{l<l_1 \vee l>l_2} -\frac{G_2(l)}{2G_1(l)}.$$

Since $G_2(l_1^-) = G_2(l_2^+) = 0^-$ and $G_2(\pm\infty) = 0^-$, the maximum of $-\frac{G_2(l)}{2G_1(l)}$ for $l < l_1 \vee l > l_2$ is reached for a finite real value in $]-\infty, l_1] \cup [l_2, +\infty[$.

We have therefore proven the following:

Theorem 3.2 (Necessary and sufficient no Butterfly arbitrage conditions for SVI under (A1))**.** *No Butterfly arbitrage in SVI entails that G_1 is positive, which requires $b(1 \pm \rho) \leq 2$. Under this condition:*

- *each of the factors of the function G_1 is positive on $\mathbb{R}$ if and only if $\alpha > F(b, \rho)$ and $\mu \in I_{\alpha,b,\rho}$;*
- *for such μ's, calling $l_1 < 0 < l_2$ the only zeros of G_2, the function G is positive in $]l_1, l_2[$ for every $\sigma \geq 0$ and the function G is non-negative on $\mathbb{R}$ if and only if $\sigma \geq \sigma^*(\alpha, b, \rho, \mu)$.*

Practical computation of σ^*

Computationally, it would be easier to implement an algorithm with bounded intervals for l. It is enough to substitute $h = \frac{1}{l}$ to obtain

$$\sigma^*(\alpha, b, \rho, \mu) := \sup_{\frac{1}{l_1}<h<\frac{1}{l_2}} -\frac{G_2(\frac{1}{h})}{2G_1(\frac{1}{h})}.$$

Under (A1) and (B1), for h which goes to $0^{\pm}$, the function G_2 goes to 0^- while G_1 is always positive under the Fukasawa conditions. So the function $f\left(\frac{1}{h}\right) = -\frac{G_2(\frac{1}{h})}{2G_1(\frac{1}{h})}$ goes to 0^+. This is a point of minimum for f in the interval $\left]\frac{1}{l_1}, \frac{1}{l_2}\right[$ because here the function is always positive.

To numerically compute σ^* we can use an algorithm which finds the maximum of f in $\left]\frac{1}{l_1}, 0\right[$ and in $\left]0, \frac{1}{l_2}\right[$ and then compares the two maxima.

It can be shown that $f'\left(\frac{1}{h}\right)$ goes to $\frac{4b(\rho-1)}{(2-b(\rho-1))(2+b(\rho-1))} < 0$ when h goes to 0^- while it goes to $\frac{4b(\rho+1)}{(2-b(\rho+1))(2+b(\rho+1))} > 0$ when h goes to 0^+. Furthermore, $f'\left(\frac{1}{l_1}\right) > 0$ and $f'\left(\frac{1}{l_2}\right) < 0$.

We plot in Figure 3.8 the function $f\left(\frac{1}{h}\right)$ with $b = \frac{1}{2}$, $\rho = -\frac{3}{10}$, $\alpha = \frac{1}{10}$ and $\mu = \frac{1}{10}$.

The function $f\left(\frac{1}{h}\right)$ seems to have always three extrema: two points of maximum (one in each interval $\left]\frac{1}{l_1}, 0\right[$ and $\left]0, \frac{1}{l_2}\right[$) and one point of minimum at 0. The sign of ρ does not imply in which of the two intervals the maximum lies.

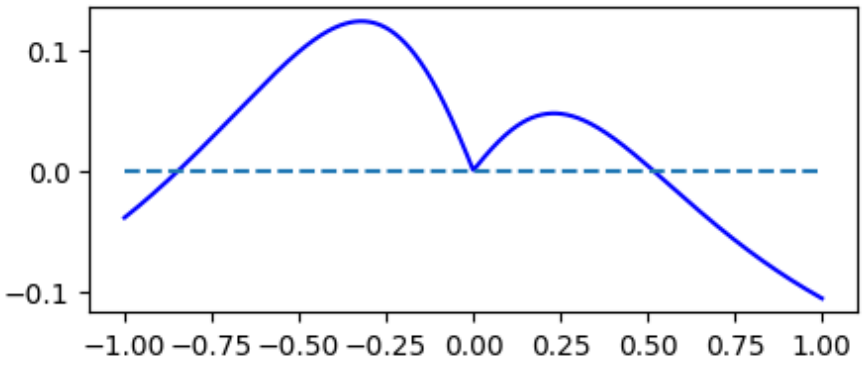

Figure 3.8: Plot of $f(\frac{1}{h})$ as a function of h, with $b = \frac{1}{2}$, $\rho = -\frac{3}{10}$, $\alpha = \frac{1}{10}$ and $\mu = \frac{1}{10}$.

For $\rho = 0$ and $\mu = 0$ the maxima have the same height, furthermore the two points of maximum are symmetrical with respect to 0, this last one is also the point of minimum. This follows from the fact that G_2 is symmetric for $\rho = 0$ and that when $\mu = 0$, also G_1 is symmetric.

Note that we have not proven that there is a single maximum on each side of 0. So a strict implementation should take into account the possibility that there are several ones, and use a global optimizer on each side. We strongly conjecture that there is in fact a single maximum on each side.

3.7.3 Algorithm under (A1)

We can now complete the algorithms stated for the Fukasawa conditions. For the parametrization of the no arbitrage domain, we just need to add the final step which specifies the range of σ:

1. choose $\rho \in]-1, 1[$ and b positive such that $b(1 \pm \rho) \leq 2$ by choosing $b' \in]0, 1]$ and setting $b = b' \frac{2}{1+|\rho|}$;

2. compute numerically $F(b, \rho)$, and parametrize α by setting $\alpha = F(b, \rho) + u$ for positive u;

3. compute numerically (L_-, L_+) for this value of u, and parametrize μ by setting $\mu = \frac{(1+q)}{2} L_+ + \frac{(1-q)}{2} L_-$ for $q \in]-1, 1[$;

4. compute numerically $\sigma^*(\alpha, b, \rho, \mu)$, and parametrize σ by setting $\sigma = \sigma^* + v$ where $v \geq 0$.

The main benefit of this parametrization is that it is eventually a simple product of intervals:

$$(\rho, b', u, q, v) \in]-1, 1[\times]0, 1]\times]0, \infty[\times]-1, 1[\times[0, \infty[$$

and this is perfectly suitable to feed optimization algorithms working with bounds, like the standard ones in the scipy.optimize scientific library.

A drawback to keep in mind is that sampling this product sub-space in a uniform way corresponds to a distorted sampling in the initial space.

There again, we can specify an algorithm which decides whether a SVI parameter lies or not in the no arbitrage domain:

1. $b(1-\rho) > 2$ or $b(1+\rho) > 2$: failure of type 1; otherwise:
2. $\alpha \leq F(b,\rho)$: failure of type 2; otherwise:
3. μ not in $I_{\alpha,b,\rho}$: failure of type 3; otherwise:
4. $\sigma < \sigma^*$: *failure of type 4.*

3.7.4 The monotonous case (A2)

In all the previous discussion, we have assumed $|\rho| < 1$ to avoid singular cases in our computations. What happens when $|\rho| = 1$? We discuss below the case $\rho = -1$, the case $\rho = 1$ follows by symmetry.

In this case the SVI smile is (convex) decreasing, and reaches its minimum α at infinity, so the domain of α is now $\alpha \geq 0$. Note that the boundary value 0 is allowed, unlike in the regular case, because the implied volatility does not vanish at any finite strike. The negative slope condition requires $b \leq 1$, and the positive (rightmost) one is automatically fulfilled.

Regarding the Fukasawa conditions, the proofs in Section 3.4 still hold with the convention that $l^* = +\infty$ so that N is decreasing. The interval for μ becomes $I_{\alpha,b,-1} =]L_-(l_-(\alpha,b,-1);\alpha,b,-1), +\infty[$, so exactly equal to $I_{\alpha,b,\rho}$ with the convention $L_-(l_-(\alpha,b,1);\alpha,b,1) = -\infty$. For $\alpha \geq 0$, we have $L_-(l) < 0$ for every l so $L_-(l_-(\alpha,b,-1);\alpha,b,-1) < 0$ also, and this interval always contains $[0,\infty[$. We can then extend the definition of the Fukasawa threshold to the case $\rho = -1$, putting $F(b,-1) = 0$. This implies that the interval for μ is non-degenerate even when $\alpha = F(b,-1) = 0$.

The function G_2 has only one negative zero l_1, above which it is always positive with $G_2(+\infty) = 0^+$ while $G_2(-\infty) = 0^-$. So $\sigma^* = \sup_{l<l_1} -\frac{G_2(l)}{2G_1(l)}$.

Theorem 3.3 (Necessary and sufficient no Butterfly arbitrage conditions for SVI, $\rho = -1$)**.** *No Butterfly arbitrage in SVI entails that G_1 is positive, which requires $b \leq 1$ and $\alpha \geq 0$. Under these conditions:*

- *each of the factors of the function G_1 is positive on $\mathbb{R}$ if and only if $\mu > L_-(l_-;\alpha,b,-1)$;*
- *for such μ's, calling $l_1 < 0$ the only zero of G_2, the function G is positive on $]l_1,\infty[$ for every $\sigma \geq 0$ and the function G is non-negative on $\mathbb{R}$ if and only if $\sigma \geq \sigma^*(\alpha,b,-1,\mu)$ where $\sigma^*(\alpha,b,-1,\mu) = \sup_{l<l_1} -\frac{G_2(l)}{2G_1(l)}$.*

Application: SVI decreasing to zero

Let us consider the case $\rho = -1$ and $a = 0$, so SVI is given by the formula $\text{SVI}(k) = b(-(k - \mu\sigma) + \sqrt{(k - \mu\sigma)^2 + \sigma^2})$ with $b \leq 1$.

Can we compute the lower bound for μ? Consider the equation $g_{-(b,-1)}(l) = 0$ or equivalently from Equation (3.6), $(-\sqrt{l^2+1} + l)\big(\sqrt{l^2+1}\big(\frac{1}{2} - \frac{b}{4}\big) + \frac{bl}{4}\big) + 1 = 0$. Simplifying, we obtain $2l(1-b)\sqrt{l^2+1} = 2(1-b)l^2 - (b+2)$ and squaring we find the two solutions $l = \pm\frac{b+2}{2\sqrt{3(1-b)}}$ when $b < 1$. The positive one does not solve the initial equation, so with the notations used in Section 3.4, we finally find $s_- = -\frac{b+2}{2\sqrt{3(1-b)}}$. If $b = 1$, then $s_- = +\infty$. Note that s_- corresponds to l_- when $\alpha = 0$, and we get that $L_-(l_-(0, b, -1); 0, b, -1) = -\sqrt{3(1-b)}$.

So for $\alpha = 0$:

- the Fukasawa conditions are satisfied if and only if $\mu > -\sqrt{3(1-b)}$;
- the unique zero of G_2 does not depend on b and is given by $l_1 = -\frac{1}{\sqrt{3}}$, and the parameters with no arbitrage are eventually given by $b \leq 1$, $\mu > -\sqrt{3(1-b)}$, $\sigma \geq \sigma^*(0, b, -1, \mu) = \sup_{l<-1/\sqrt{3}} -\frac{G_2(l)}{2G_1(l)}$.

3.8 Calibration experiments

Now that we have parametrized the no arbitrage domain, the design of a calibration algorithm is straightforward:

1. choose an objective function;
2. choose a starting point policy;
3. for the chosen starting points (possibly several of them), run a minimization algorithm of the objective function over the no arbitrage domain;
4. pick up the optimal parameters.

As objective function, we choose the classical least squares criterion, which takes as input the differences of the data and model total variances on the available set of log-forward moneyness. This will give equal weights to far-from-the-money points, where the precise value of the implied volatility, and so the accuracy of the calibration, matters less, and to close-to-the-money ones, which is not a desirable feature: it can be easily patched by adding weights given by the Vegas (computed once for all with the data points), so that the errors are more in line with losses, unit-wise. This would moreover stabilize the calibration from one day to another one, especially on illiquid markets, as discussed in detail in [58].

Now the big question for us is rather whether or not the no arbitrage constraint will deteriorate the quality of the fit, and we will also work on model generated data or on

index options data which are liquid ones, whence our choice of a standard non-weighted objective function.

Regarding the starting point policy, we are not big fans of *smart guess strategies* which try to compute the best starting point from the data. Such strategies can work brilliantly in many favorable situations, yet they might fail heavily on data with low quality (e.g. due to a dubious treatment by an internal department), or when faced with new market behavior and configurations. There is a clear risk of over-engineering here also. We would be more confident by using a *set* (with small cardinality) of starting points, possibly produced by a machine learning algorithm duly trained on the markets in scope. We implement a very basic version of this idea, which picks up uniformly generated points within the hyperrectangle of the no arbitrage domain, irrespective of the data.

The scipy function here used is the `least_squares` which lies in the optimize library. The method used is the `dogbox`, which handles bounds. The tolerances regarding the change of the cost function (`ftol`), the change of the independent variables (`xtol`) and the norm of the gradient (`gtol`) are all set at the Python numpy machine epsilon. The maximum number of function evaluations (`max_nfev`) is set at 1000.

Even though the arbitrage region does not impose an upper bound for α and σ, we choose arbitrary ones. In particular, we ask

$$\sigma \leq \max\left(\frac{|k_0|}{r}, \frac{|k_N|}{r}, 1.5\sigma^*\right)$$

with r as parameter to be chosen by the user (default value equal to 0.1). This bound is related to the fact that when $\frac{|k_i|}{\sigma}$ is below a threshold r, then the smile is almost flat and this causes uncertainty on the parameters to be chosen.

The upper bound for α is left to be chosen by the user. For the index option data we set $\alpha < 1$ since it is enough to achieve a very good fit, while for the model generated data, in order to have an almost perfect calibration, the upper bound actually depends on the α parameter used to generate data. We set in every case $\alpha < 3$, since we know a priori that all the data are generated with α lower than 3.

We provide below our calibration results on model generated data and then on market data.

3.8.1 On model data

To check the robustness of the algorithm we firstly run it on data generated by arbitrary SVI parameters with no arbitrage, and on the Axel Vogt parameters. We take a vector of 13 log-forward strikes taken from Table 3.2 of [29].

The parameters chosen for each of the graphs in Figure 3.9 are arbitrage-free. The red and the blue lines, which represent the total variances generated from the arbitrary parameters and the total variances obtained from the calibrated parameters respectively, overlap.

The fact that the fit is excellent can be seen by the Frobenius relative errors in Table 3.1.

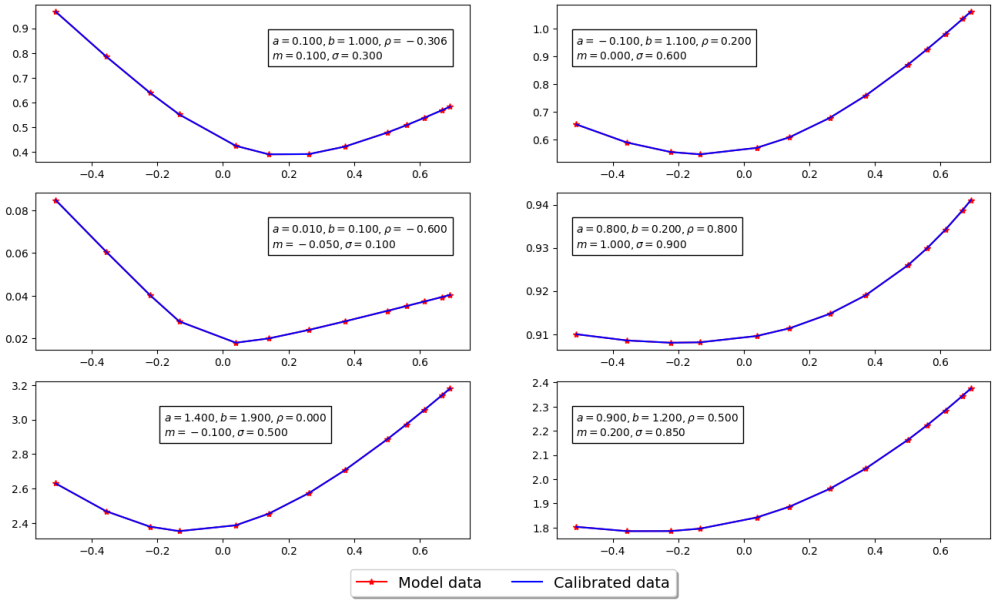

Figure 3.9: Model total variances generated by arbitrage-free parameters (in red) and calibrated total variances (in blue).

	a	b	ρ	m	σ	Relative error ($\times 10^{-16}$)
0	0.10	1.0	-0.306	0.10	0.30	1.48
1	-0.10	1.1	0.200	0.00	0.60	1.63
2	0.01	0.1	-0.600	-0.05	0.10	2.30
3	0.80	0.2	0.800	1.00	0.90	1.77
4	1.40	1.9	0.000	-0.10	0.50	2.35
5	0.90	1.2	0.500	0.20	0.85	2.25

Table 3.1: Frobenius relative errors for the total variances with arbitrage-free parameters calibrated on model total variances.

	a	b	ρ	m	σ	Relative error ($\times 10^{-14}$)
0	0.10	1.0	-0.306	0.10	0.30	0.10
1	-0.10	1.1	0.200	0.00	0.60	0.30
2	0.01	0.1	-0.600	-0.05	0.10	0.06
3	0.80	0.2	0.800	1.00	0.90	20.00
4	1.40	1.9	0.000	-0.10	0.50	0.10
5	0.90	1.2	0.500	0.20	0.85	3.00

Table 3.2: Frobenius relative errors for the parameters calibrated on model total variances.

	a	b	ρ	m	σ
Original	-0.041	0.1331	0.306	0.3586	0.4153
Calibrated	-0.0198444	0.102745	0.180754	0.266125	0.310459

Table 3.3: Axel Vogt parameters vs best fitting no arbitrage.

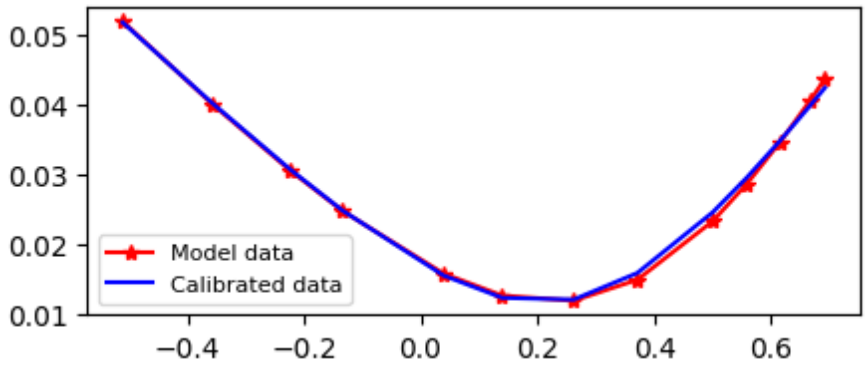

Figure 3.10: Total variances generated by the Axel Vogt parameters (in red) and total variances with arbitrage-free parameters (in blue).

Furthermore, also the Frobenius relative error on the parameters is low (Table 3.2). This means that the algorithm is robust and recovers the original data.

Axel Vogt parameters

For a matter of completeness we run our algorithm on the notorious Axel Vogt parameters, which lead to an arbitrage SVI. The original and the calibrated parameters are reported in Table 3.3 while the graphs of the original and arbitrage-free total variances are shown in Figure 3.10.

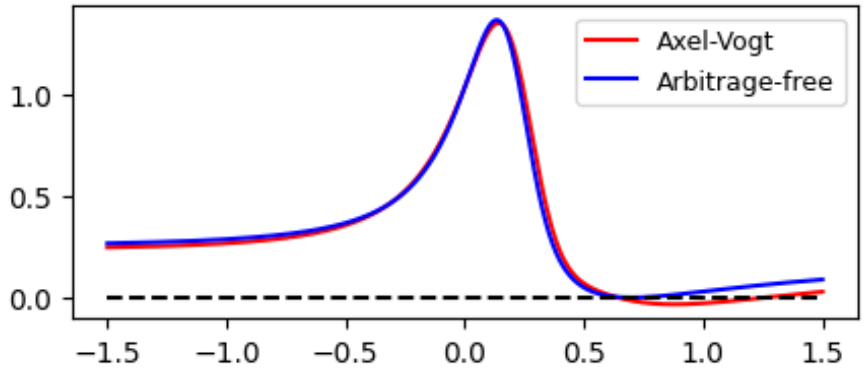

Figure 3.11: Plot of the functions g with the the Axel Vogt parameters (in red) and with the arbitrage-free parameters (in blue).

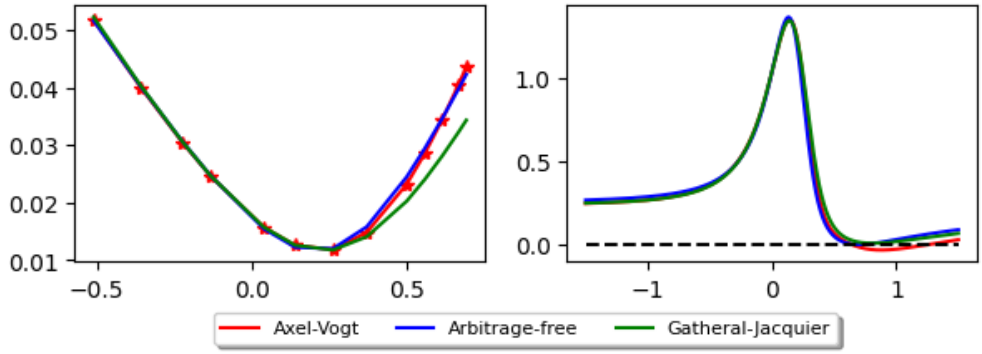

Figure 3.12: On the left, plot of the total variances generated by the Axel Vogt parameters (in red), the total variances with the arbitrage-free parameters (in blue) and the total variances with the Gatheral-Jacquier parameters (in green). On the right, plot of the function g with the the Axel Vogt parameters (in red), with the arbitrage-free parameters (in blue) and with the Gatheral-Jacquier parameters (in green).

Of course, the calibration is not perfect as in the previous case and the Frobenius error between the Axel Vogt total variances and the non arbitrage SVI corresponding total variances is 2.15%.

We compare the function g defined in Equation (3.2) with the original Axel Vogt parameters and the same function with the new arbitrage-free parameters in Figure 3.11.

From the plot it can be seen that the function g with the new arbitrage-free parameters can be very close to zero, but it is always positive.

In the following study, we compare the results obtained with the new arbitrage-free parameters and the ones with the parameters described in Example 5.1 of [36], which are also arbitrage free. Figure 3.12 shows that the fit of our new parameters is better than the one of Gatheral and Jacquier.

In Table 3.4 we compare the relative errors on the total variances for the two sets of

	a	b	ρ	m	σ	Relative error
Arbitrage-Free	-0.0198444	0.102745	0.180754	0.266125	0.310459	0.022
Gatheral-Jacquier	-0.0305199	0.102717	0.100718	0.272344	0.412398	0.133

Table 3.4: Frobenius relative errors for the total variances with arbitrage-free parameters vs Gatheral-Jacquier parameters.

arbitrage-free parameters.

3.8.2 On data from CBOE

We now turn to market data. We work with market data of good quality bought from the CBOE data store by Zeliade. They cover daily files for the DJX, SPX500 and NDX equity indices, with bid and ask prices.

To obtain implied total variances from the prices, we operate the classical treatment of inferring the discount factor and forward values at each option maturity by performing a linear regression of the (mid) Call minus Put prices with respect to the strike. Since the markets under study are very liquid, the fit is excellent and the residual error is extremely small.

Then, given the discount factor and forward values for each maturity, we are able (after working out the exact maturity of each contract from its code, if not provided explicitly) to compute the implied volatilities, for the bid and ask prices.

We feed the objective function with the implied volatility corresponding to the mid price, and plot below the implied volatilities for the calibrated model and the bid and ask market data. Results are reported in Figures 3.13 to 3.15.

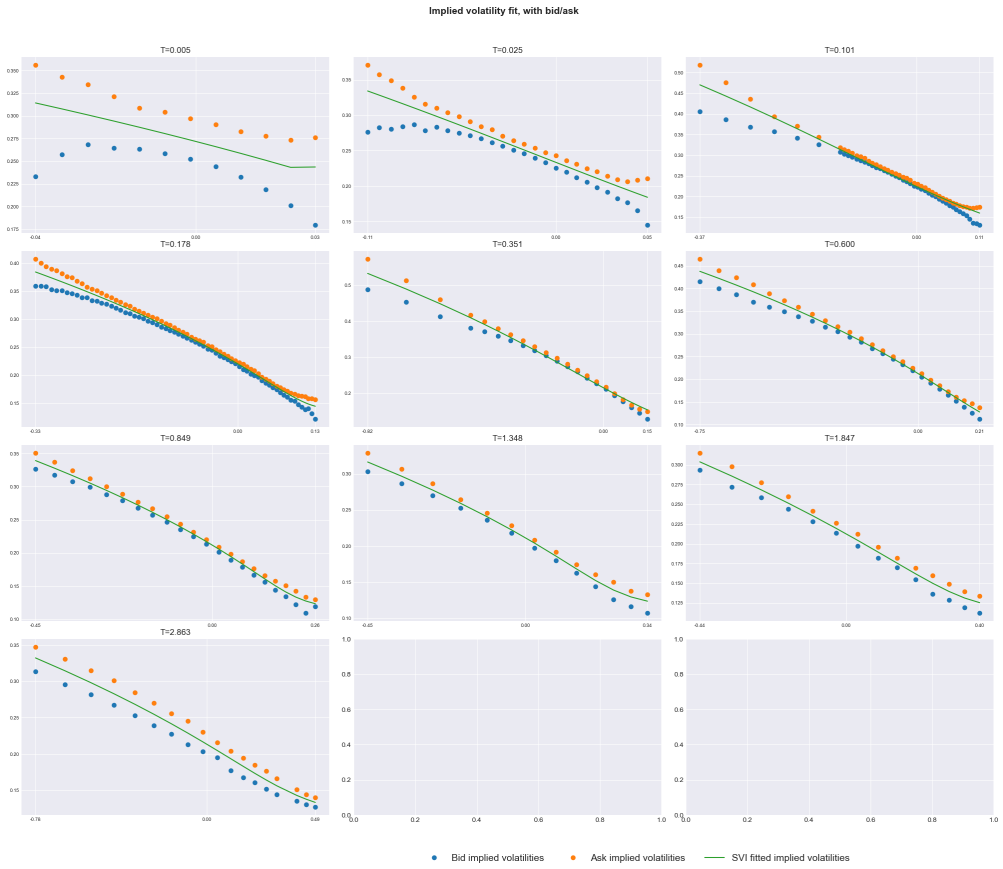

Figure 3.13: Fitted implied volatilities with arbitrage-free parameters (in green) and bid (in blue) and ask (in red) implied volatilities for the DJX index.

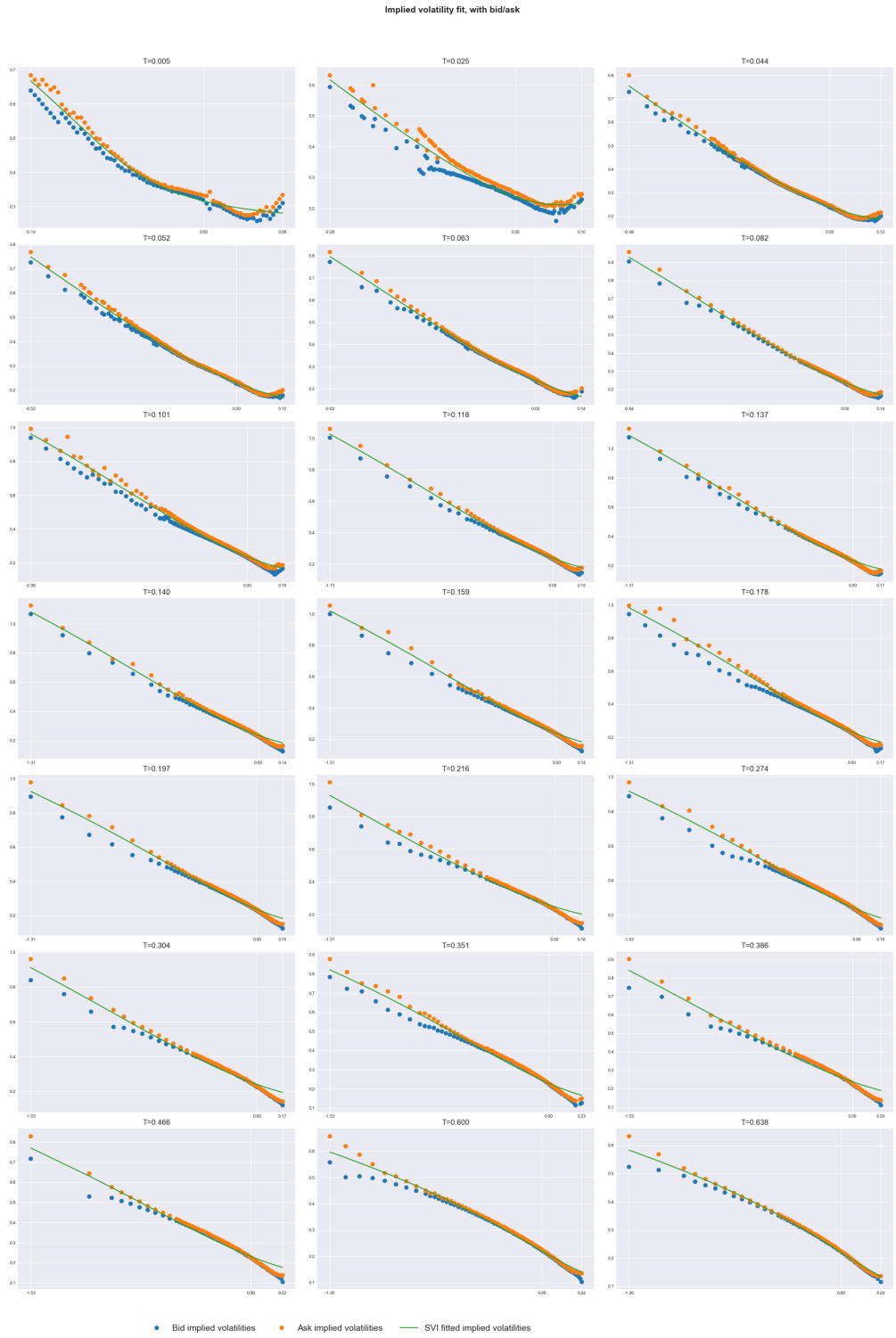

Figure 3.14: Fitted implied volatilities with arbitrage-free parameters (in green) and bid (in blue) and ask (in red) implied volatilities for the SPX500 index.

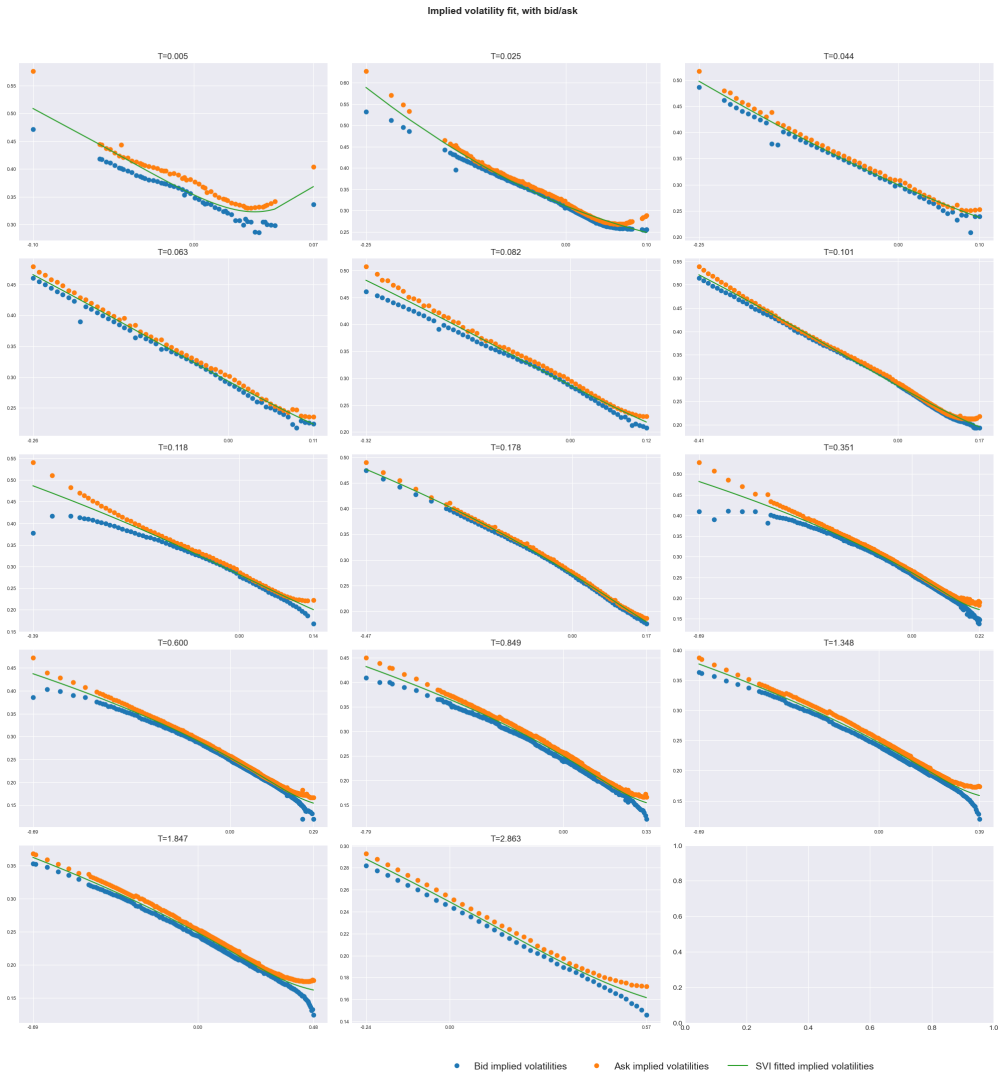

Figure 3.15: Fitted implied volatilities with arbitrage-free parameters (in green) and bid (in blue) and ask (in red) implied volatilities for the NDX index.

3.8.3 Discussion

From our experiments we draw several positive conclusions:

- the quality of fit is excellent, and there is no big loss resulting from the no arbitrage constraint;
- the implementation we have designed seems sufficiently robust in practice; of course such a statement should be re-assessed continuously;
- the payload of the root finding algorithms used to compute the Fukasawa threshold and the bounds for μ and σ is not an issue, the calibration is still reasonably fast

on a basic chip; the average for each maturity for the DJX data is 51.598 seconds, for the SPX data 36.490 seconds and for the NDX data 44.900 seconds.

Of course, there is room for improvement, at least at the level of the starting point strategy. One could also think of precomputing the numerical functions computed on the fly, or to design once for all explicit proxies for them, which would speed massively the execution of the algorithm.

3.9 Conclusion

Fukasawa's remark that the inverse of d_1 and d_2 functions of the Black-Scholes formula have to be increasing under no Butterfly arbitrage, paired with the natural rescaling of the SVI parameters which consists in scaling a and m by σ, allow us to fully describe the domain of no Butterfly arbitrage for SVI.

The no Butterfly arbitrage domain can be parametrized as an hyperrectangle, with 2 downstream algorithms of practical importance: one for checking that a SVI parameter lies or not in the no arbitrage domain, and the other one to effectively perform a calibration. Three functions have to be computed numerically by resorting to root-finding type algorithms; due to the fact that our careful mathematical analysis provided safe bracketing intervals for those functions, this can be achieved in a very quick manner. We provide calibration results on model and market data, the latter showing that there is no loss of fit quality due to imposing the no arbitrage constraint.

This analysis settles one important issue in the SVI saga. Other ones are still pending, like the study of sub-SVI parametrizations with 4 parameters instead of 5, in the spirit of SSVI (which has 3 parameters slice-wise), which could display more parameter stability than SVI and a better fit quality than SSVI; and also the question of the characterization of no Calendar Spread arbitrage for two SVI slices corresponding to different maturities.

3.A Proof of Proposition 3.3

Proof. Observe that at the point l^*, $\rho\sqrt{l^2+1}+l=0$ and also after computations, $\rho l+\sqrt{l^2+1}=\sqrt{1-\rho^2}$, so we have $g_{\pm(b,\rho)}(l^*)=-\sqrt{1-\rho^2}$. Furthermore, under (B1) it is easy to verify that $g_{\pm(b,\rho)}(\pm\infty)=\infty$.

We have $\frac{d}{d\alpha}L_\pm(l_\pm)=L'_\pm(l_\pm)\frac{d}{d\alpha}l_\pm+\partial_\alpha L_\pm(l_\pm)=\partial_\alpha L_\pm(l_\pm)$. Deriving Equation (3.5) with respect to α, we find $\partial_\alpha L_\pm(l_\pm)=2\Big(\frac{1}{N'(l_\pm)}\mp\frac{1}{4}\Big)$.

Since $N'(l)>0$ iff $l>l^*$ and $4\mp N'>0$, we have $\partial_\alpha L_-(l_-)<0$ and $\partial_\alpha L_+(l_+)>0$. So the function $\alpha\to L_-(l_-,\alpha)$ is decreasing while $\alpha\to L_+(l_+,\alpha)$ is increasing. It means that the bounds for μ are an increasing family of sets (possibly empty) parametrized by α. Consider the lower bound, so $l<l^*$. We can write the expression for $g_{-(b,\rho)}$ in another way. Note that

$$N(l)=\alpha+lN'(l)+N''(l)(l^2+1) \tag{3.10}$$

so we have

$$L'_-(l)=1+\frac{N'(l)}{2}-\frac{2N''(l)}{N'(l)^2}(\alpha+lN'(l)+N''(l)(l^2+1)).$$

Evaluating this in l_-, the LHS becomes 0 and we can isolate α, obtaining

$$g_{-(b,\rho)}(l)=\frac{1}{b}\left(\frac{N'(l)^2}{2N''(l)}\left(1+\frac{N'(l)}{2}\right)-lN'(l)-N''(l)(l^2+1)\right). \tag{3.11}$$

From this expression, we get the derivative of $g_{-(b,\rho)}$ such as $g'_{-(b,\rho)}(l)=\frac{N'(l)^2}{4b}\big(3-\frac{N'''(l)}{N''(l)^2}(N'(l)+2)\big)$, which is positive iff the second factor is positive. Substituting with the explicit expressions, we find that this holds iff $\frac{3}{b}\sqrt{l^2+1}\big(l(2+b\rho)+b\sqrt{l^2+1}\big)>0$ or equivalently $-l(2+b\rho)<b\sqrt{l^2+1}$.

Note that since $b(1-\rho)\le 2$, then $2+b\rho>0$. If $\rho<0$ then the equation is true for $l\ge 0$. For negative ls we can square, obtaining that it holds iff $(b^2(\rho^2-1)+4b\rho+4)l^2<b^2$. For $b(1-\rho)<2$, the coefficient of l^2 is positive, so the inequality holds iff $l>-\frac{b}{\sqrt{b^2(\rho^2-1)+4b\rho+4}}:=m_-$. Since ρ is negative, $m_-<l^*$. So in this case $g_{-(b,\rho)}(l)$ is increasing for $l>m_-$ and decreasing for $l<m_-$.

If $\rho\ge 0$, we proceed in a similar way taking the square and obtaining that, if $b(1-\rho)<2$, the inequality holds iff $l>-\frac{b}{\sqrt{b^2(\rho^2-1)+4b\rho+4}}:=m_-$. If $b\le\frac{2\rho}{1-\rho^2}$, then $m_-\ge l^*$ and g_- is always decreasing. Otherwise if $b>\frac{2\rho}{1-\rho^2}$, then $m_-<l^*$ and $g_{-(b,\rho)}$ is increasing for $l>m_-$ and decreasing for $l<m_-$. We can write α as a function of l_-, indeed $\alpha=bg_{-(b,\rho)}(l_-)$. This function has the same monotonicity as $g_{-(b,\rho)}$.

We obtain from the previous analysis that the function $x\to L_-(x;bg_{-(b,\rho)}(x))$ is:

- increasing for $x<m_-$ and decreasing for $x>m_-$ when $b>\frac{2\rho}{1-\rho^2}$;
- increasing for every $x<l^*$ when $b\le\frac{2\rho}{1-\rho^2}$.

Using Equation (3.10) and Equation (3.5) and substituting α with $bg_{-(b,\rho)}(x)$ considered as in Equation (3.11), we obtain

$$L_-(x; bg_{-(b,\rho)}(x)) = \frac{N'(x)^2}{N''(x)}\left(1 + \frac{N'(x)}{2}\right)\left(\frac{1}{N'(x)} + \frac{1}{4}\right) - x. \tag{3.12}$$

From here it can be seen that $L_-(x; bg_{-(b,\rho)}(x))$ goes to $-l^*$ when x goes to l^{*-}. Similarly, we can do all the equivalent computations for L_+. First, the function $g_{+(b,\rho)}$ can be re-written as

$$g_{+(b,\rho)}(l) = \frac{1}{b}\left(\frac{N'(l)^2}{2N''(l)}\left(1 - \frac{N'(l)}{2}\right) - lN'(l) - N''(l)(l^2+1)\right)$$

while

$$L_+(x; bg_{+(b,\rho)}(x)) = \frac{N'(x)^2}{N''(x)}\left(1 - \frac{N'(x)}{2}\right)\left(\frac{1}{N'(x)} - \frac{1}{4}\right) - x$$

and even in this case $L_+(x; bg_{+(b,\rho)}(x))$ goes to $-l^*$ when x goes to l^{*+}. We can study the monotonicity of $g_{+(b,\rho)}$, obtaining $g'_{+(b,\rho)}(l) = \frac{N'(l)^2}{4b}\left(-3 + \frac{N'''(l)}{N''(l)^2}(N'(l) - 2)\right)$.

Considering the second factor and substituting with the explicit expressions, the latter quantity is positive iff $-\frac{3}{b}\sqrt{l^2+1}\left(-l(2-b\rho) + b\sqrt{l^2+1}\right) > 0$ or equivalently $l(2-b\rho) > b\sqrt{l^2+1}$.

Here, since $b(1+\rho) \leq 2$, then $2 - b\rho > 0$. If $\rho > 0$ then the equation is false for $l \leq 0$. For positive ls we can square, obtaining that it holds iff $(b^2(\rho^2-1) - 4b\rho + 4)l^2 > b^2$. For $b(1+\rho) < 2$, the coefficient of l^2 is positive, so the inequality holds iff $l > \frac{b}{\sqrt{b^2(\rho^2-1)-4b\rho+4}} := m_+$. Since ρ is positive, $m_+ > l^*$. So in this case $g_{+(b,\rho)}(l)$ is increasing for $l > m_+$ and decreasing for $l < m_+$.

If $\rho \leq 0$, we proceed in a similar way taking the square and obtaining that, if $b(1+\rho) < 2$, the inequality holds iff $l > \frac{b}{\sqrt{b^2(\rho^2-1)-4b\rho+4}} := m_+$. If $b \leq -\frac{2\rho}{1-\rho^2}$, then $m_+ \leq l^*$ and $g_{+(b,\rho)}$ is always increasing. Otherwise if $b > -\frac{2\rho}{1-\rho^2}$, then $m_+ > l^*$ and $g_{+(b,\rho)}$ is increasing for $l > m_+$ and decreasing for $l < m_+$. Remember that the function $\alpha \to L_+(l_+, \alpha)$ is increasing. To recap, the function $x \to L_+(x; bg_{+(b,\rho)}(x))$ is:

- increasing for $x > m_+$ and decreasing for $x < m_+$ when $b > -\frac{2\rho}{1-\rho^2}$;
- increasing for every $x > l^*$ when $b \leq -\frac{2\rho}{1-\rho^2}$.

If $b \leq -\frac{2\rho}{1-\rho^2}$ then $\rho < 0$ and $b > \frac{2\rho}{1-\rho^2}$ while if $b \leq \frac{2\rho}{1-\rho^2}$ then $\rho > 0$ and $b > -\frac{2\rho}{1-\rho^2}$. This means that $L_+(x; bg_{+(b,\rho)}(x))$ and $L_-(x; bg_{-(b,\rho)}(x))$ cannot be both monotonous.

The last statement of the proposition is a direct consequence to the fact that $\frac{d}{dx}L_\pm(x; bg_{\pm(b,\rho)}(x)) = \partial_\alpha L_\pm(x; bg_{\pm(b,\rho)}(x))bg'_{\pm(b,\rho)}(x)$ where $\partial_\alpha L_-(x) < 0$ and $\partial_\alpha L_+(x) > 0$. □

3.B Computation of $F(b,0)$

In this appendix we compute $F(b,0)$ and prove that $F(b,0) > -b$.

With $\rho = 0$ we have $l^* = 0$ and

$$\begin{aligned} N &= \alpha + b\sqrt{l^2+1}, \\ N' &= \frac{bl}{\sqrt{l^2+1}}, \\ N'' &= \frac{b}{(l^2+1)^{\frac{3}{2}}}. \end{aligned}$$

Consider the particular case $b = 2$. Then we have already shown $F(2,0) = 0$, which is greater than -2.

Consider $b \neq 2$. Since $b > \frac{2\rho}{1-\rho^2} = 0$, then the function $l_- \to L_-(l_-; bg_{-(b,0)}(l_-))$ is increasing iff $l_- < m_-$ where $m_- = -\frac{b}{\sqrt{4-b^2}}$. Furthermore the interval for μ is $I_{\alpha,b,0} = \big]L_-(l_-(\alpha,b,0);\alpha,b,0), -L_-(l_-(\alpha,b,0);\alpha,b,0)\big[$ so it is symmetrical with respect to 0. The Fukasawa threshold $F(b,0)$ is then the solution to $L_-(l_-(F(b,0),b,0);F(b,0),b,0) = 0$.

From equation Equation (3.12) we obtain

$$L_-(l_-; bg_{-(b,0)}(l_-)) = b\frac{l_-^2}{2}\left(2\sqrt{l_-^2+1} + bl_-\right)\left(\frac{\sqrt{l_-^2+1}}{bl_-} + \frac{1}{4}\right) - l_-.$$

For $l_- < 0$, this expression is equal to 0 iff $(8+b^2)l = -6b\sqrt{l^2+1}$ and so iff l_- equals $l_-^* := -\frac{6b}{\sqrt{b^4-20b^2+64}}$. Then

$$F(b,0) = bg_{-(b,0)}\left(-\frac{6b}{\sqrt{b^4-20b^2+64}}\right)$$

where $g_{-(b,0)}(l) = \frac{l^2}{4}(2\sqrt{l^2+1} + bl) - \sqrt{l^2+1}$.

We now need to prove $g_{-(b,0)}(l_-^*) > -1$ or equivalently $l_-^* < s_-$. From the expression of $g_{-(b,0)}$, we immediately find that s_- satisfies $2(l^2-2)\sqrt{l^2+1} = -bl^3 - 4$, so we look for a negative root such that $\frac{-bl^3-4}{l^2-2} > 0$. This happens iff l lies outside the interval $\left[\left(-\frac{4}{b}\right)^{\frac{1}{3}}, -\sqrt{2}\right]$ if $b \leq \sqrt{2}$, or outside the interval $\left[-\sqrt{2}, \left(-\frac{4}{b}\right)^{\frac{1}{3}}\right]$ if $b > \sqrt{2}$. Squaring the previous equation and simplifying by l^3 we find $(4-b^2)l^3 - 12l - 8b = 0$. Call $P_b(l)$ the LHS.

At 0, this polynomial and its derivative are negative. Its local maximum is at $-\frac{2}{\sqrt{4-b^2}}$ and its value at this point is $\frac{16}{\sqrt{4-b^2}} - 8b$ which is always positive. So the polynomial has two negative roots and a positive one.

We can observe that $P_{\sqrt{2}}(-\sqrt{2}) = P_{\sqrt{2}}\left(\left(-\frac{4}{b}\right)^{\frac{1}{3}}\right) = 0$ with

- $P_b(-\sqrt{2}) = 2\sqrt{2}(b-\sqrt{2})^2 > 0$,
- $P_b\Big(\Big(-\frac{4}{b}\Big)^{\frac{1}{3}}\Big) = -\frac{4}{b}\Big(b^2 - 3(2b)^{\frac{2}{3}} + 4\Big) < 0$.

Then if $b < \sqrt{2}$ the root of interest s_- is the second negative root of the polynomial while if $b \geq \sqrt{2}$ it is the first negative root.

The value of the polynomial in l_-^* is

$$-\frac{8b((b^2-16)^2 - 36\sqrt{b^4-20b^2+64})}{(b^2-16)^2}$$

which is positive iff $b < \tilde{b}$ where $\sqrt{\frac{8}{5}} < \tilde{b} < \sqrt{2}$. The derivative of the polynomial evaluated in l_-^* is $\frac{24(5b^2-8)}{16-b^2}$, which is positive iff $b > \sqrt{\frac{8}{5}}$.

Then:

- if $b \leq \tilde{b}$ the polynomial is positive in l_-^* and s_- is its second root, so $l_-^* < s_-$;
- if $\tilde{b} < b < \sqrt{2}$ the polynomial is negative in l_-^* while its derivative is positive and s_- is its second root, so $l_-^* < s_-$;
- finally if $b \geq \sqrt{2}$ the polynomial is negative with a positive derivative in l_-^* so even if s_- is now its first root we have $l_-^* < s_-$.

Chapter 4

Explicit no arbitrage domain for sub-SVIs via reparametrization

Abstract

The no Butterfly arbitrage domain of Gatheral SVI 5-parameters formula for the volatility smile has been recently described. It requires in general a numerical minimization of 2 functions altogether with a few root finding procedures. We study here the case of some sub-SVIs (all with 3 parameters): the Symmetric SVI, the Vanishing Upward/Downward SVI, and SSVI, to which we apply the SVI results to provide the no Butterfly arbitrage domain. As side results: we prove that under simple requirements on parameters, SSVI slices always satisfy the weak conditions of no arbitrage by Fukasawa, i.e. the corresponding Black-Scholes functions d_1 and d_2 are always decreasing; and we find a simple subdomain of no arbitrage for the SSVI which we compare with the well-known Gather and Jacquier's subdomain. We simplify the so obtained no arbitrage domain into a parametrization with only one immediate numerical procedure required, proving an easy-to-implement calibration algorithm. Finally we show that the Long Term Heston SVI is in fact an SSVI, and characterize the horizon beyond which it is arbitrage-free.

From:
C. Martini and **A. Mingone**, *Explicit no arbitrage domain for sub-SVIs via reparametrization*, arXiv preprint https://arxiv.org/abs/2106.02418, 2021.
C. Martini and **A. Mingone**, *Refined analysis of no Butterfly arbitrage domain for SSVI slices*, Journal of Computational Finance, 27(2), 1-32, 2023.

4.1 Structure of the chapter

In this chapter we apply results in Chapter 3 and [53] to sub-SVI smiles with 3 parameters. We start with a look at general results on SVI regarding the no Butterfly arbitrage conditions in Section 4.2.1, where we remind theorems of [53] and illustrate a re-parametrization paradigm in the context of the bounds for σ.

This stage requires to compute the minimum of a function $\tilde{f}(l;\gamma,b,\rho,\mu)$, on an interval which depends on the parameters (γ,ρ); we know that $\tilde{f}$ goes to infinity at the bounds of this interval. Our strategy is then to study the critical points of $\tilde{f}$. It turns out that in full generality, the equation characterizing those critical points $\bar{l}$ reads

$$p(\bar{l};\gamma,\rho,\mu) = b^2 q(\bar{l};\gamma,\rho,\mu).$$

In some circumstances, we can then *use* $\bar{l}$ as a parameter, and obtain b^2 as $\frac{p(\bar{l};\gamma,\rho,\mu)}{q(\bar{l};\gamma,\rho,\mu)}$. In other circumstances, we use the fact that the equation $p - b^2 q = 0$ is, in full generality, *quadratic* in μ, and we use the same trick to back up μ once $\bar{l}$ is promoted to the status of a parameter.

Now this is only the easy part of the story, since the critical points of $\tilde{f}$ may correspond to other local minima than the absolute one, or, even worse, to local maxima. The hard part is to show that the chosen critical point $\bar{l}$, in a given domain, corresponds indeed to the global minimum of $\tilde{f}$; one way to prove this is that there is a *unique* solution to the critical points equation above. We manage to prove this unicity for the Symmetric and Vanishing case, and resort to a *numerical proof* for SSVI.

This leads to the following results:

- in Propositions 4.2 and 4.3 of Section 4.3, we obtain a fully explicit parametrization of the no arbitrage domain for the *Vanishing Upward* and the *Vanishing Downward* SVI
$$w(k) = b\big(-(k-m) + \sqrt{(k-m)^2+\sigma^2}\,\big);$$
- in Proposition 4.4 of Section 4.4, we find a parametrization of the no arbitrage domain for the *Extremal Decorrelated* SVI
$$w(k) = a + 2\sqrt{(k-m)^2+\sigma^2};$$
- in Proposition 4.5 of Section 4.5, we get the no arbitrage domain for the *Symmetric* SVI
$$w(k) = a + b\sqrt{k^2+\sigma^2};$$
- in Corollary 4.1 of Section 4.6, we find that SSVI slices
$$w(k) = \frac{\theta}{2}\big(1 + \rho\varphi k + \sqrt{(\varphi k+\rho)^2 + 1 - \rho^2}\big)$$
are always free of weak arbitrage, i.e. they always have decreasing d_1 and d_2 functions, and we derive the no arbitrage domain for SSVI in Proposition 4.6. The domain is quasi-explicit, in the sense that it requires the computation of a point of minimum depending on two parameters. With the aim of simplifying this domain, in Section 4.6.1 we find a new parametrization (Proposition 4.7) of free of arbitrage SSVI slices which allows for easier and quicker computations. In particular, the new domain is a product of intervals for fixed ρ and it only requires a root finding

depending on the sole parameter ρ. As a side result, we find a smart subdomain of no arbitrage and compare it in Section 4.6.1 with Gatheral and Jacquier sufficient conditions of no arbitrage for SSVI. In Section 4.6.2 we also re-visit the Long Term Heston SVI appoximation and show it is in fact of SSVI type; we prove in Proposition 4.9 that it is indeed free of Butterfly arbitrage as soon as T is larger than some fully explicit threshold, which completes Gatheral and Jacquier result.

4.2 Notations and preliminaries

As duly described in section 2.1 of [53], a smile is free of Butterfly arbitrage if the Call price function with respect to the strike is convex, non-increasing, and bounded between the discounted Call payoff evaluated at the forward value and the discounted forward. In the case of smiles defined via a model for the implied volatility, the third condition is always satisfied. As a consequence, the first condition implies the second, so that the only requirement to be satisfied for implied volatility models is the convexity of prices.

Furthermore, Fukasawa proved in [32] that any arbitrage-free volatility smile has decreasing functions $d_{1,2}(k) = -\frac{k}{\sqrt{\omega(k)}} \pm \frac{\sqrt{\omega(k)}}{2}$. We call this necessary condition *absence of weak arbitrage*. Smiles satisfying the absence of weak arbitrage have been characterized in Theorem 2.2 of [51], where Lucic finds a necessary and sufficient parametrization for smiles $\sigma(k)\sqrt{T} = \sqrt{\omega(k)}$ of the form

$$\sigma(k)\sqrt{T} = \begin{cases} -\sqrt{2k - \phi(k)} + \sqrt{-\phi(k)} & \text{if } k \leq k_*, \\ \sqrt{2k - \phi(k)} + \sqrt{-\phi(k)} & \text{if } k_* < k \leq k^*, \\ \sqrt{2k + \phi(k)} - \sqrt{\phi(k)} & \text{if } k > k^* \end{cases} \tag{4.1}$$

where $k_* < 0 < k^*$ and $\phi(k)$ is a continuous increasing function such that $\phi(k_*) = 2k_*$, $\phi(k^*) = 0$, and $\phi'(k) > 2$ for $k < k_*$, $\phi'(k) < 2$ for $k_* < k \leq k^*$.

With this in mind, in the next section we reconsider the conditions of no Butterfly arbitrage for the SVI model and in the following section we apply them to the sub-SVIs.

From now on, we denote with SVI the model for the implied total variance of the form

$$\mathrm{SVI}(k; a, b, m, \rho, \sigma) = a + b\big(\rho(k - m) + \sqrt{(k - m)^2 + \sigma^2}\big)$$

where k is the log-forward moneyness, $a,\ m \in \mathbb{R}$, $b \geq 0$, $\rho \in [-1, 1]$, $\sigma \geq 0$, $a + b\sigma\sqrt{1 - \rho^2} \geq 0$.

4.2.1 Necessary and sufficient no Butterfly arbitrage conditions for SVI

We set $\gamma = \frac{a}{b\sigma}$ and $\mu = \frac{m}{\sigma}$. Let us redefine the quantity

$$N(l; \gamma, \rho) := \gamma + \rho l + \sqrt{l^2 + 1}$$

such that $\text{SVI}(k) = b\sigma N\big(\frac{k-m}{\sigma}\big)$. The derivatives of N are

$$N'(l;\rho) = \rho + \frac{l}{\sqrt{l^2+1}}, \qquad N''(l) = \frac{1}{(l^2+1)^{\frac{3}{2}}}$$

so that N has a unique critical point which is a point of minimum equal to $l^* = -\frac{\rho}{\sqrt{1-\rho^2}}$. Since N is a rescaled total variance, it must be positive, so the constraint on the new variables are $\gamma \geq -\sqrt{1-\rho^2}$, $b \geq 0$, $\rho \in [-1,1]$, $\mu \in \mathbb{R}$, $\sigma \geq 0$. When $b = 0$, we recover the Black-Scholes case, which is free of Butterfly arbitrage under $\sigma > 0$. Then, when formulating theorems of non-arbitrage, we consider only the more difficult cases $b > 0$.

We define

$$\begin{aligned} h(l;\gamma,\rho,\mu) &:= 1 - N'(l;\rho)\frac{l+\mu}{2N(l;\gamma,\rho)}, \\ g(l;\rho) &:= \frac{N'(l;\rho)}{4}, \\ g_2(l;\gamma,\rho) &:= N''(l) - \frac{N'(l;\rho)^2}{2N(l;\gamma,\rho)}, \end{aligned} \tag{4.2}$$

and $G_1(l;\gamma,b,\rho,\mu) := G_{1+}(l;\gamma,b,\rho,\mu)G_{1-}(l;\gamma,b,\rho,\mu) := (h(l;\gamma,\rho,\mu) - bg(l))(h(l;\gamma,\rho,\mu) + bg(l;\rho))$.

The requirement that an SVI is (Butterfly) arbitrage-free is equivalent to the requirement that the function

$$G_1 + \frac{1}{2\sigma} b g_2$$

is non-negative. Fukasawa proved in [32] that the condition $G_{1+} > 0$ and $G_{1-} > 0$ are also necessary. Theorem 5.10 of [53], that we rewrite here, characterizes these conditions in the case of SVI. The statement requires the definition of the functions:

$$\begin{aligned} L_-(l;\gamma,b,\rho) &:= 2N(l)\big(\frac{1}{N'(l)} + \frac{b}{4}\big) - l, \\ g_{-(b,\rho)} &:= \big(\rho\sqrt{l^2+1} + l\big)^2\big(\sqrt{l^2+1}\big(\frac{1}{2} + \frac{b\rho}{4}\big) + \frac{bl}{4}\big) - \big(\rho l + \sqrt{l^2+1}\big). \end{aligned}$$

When $\gamma + \sqrt{1-\rho^2} > 0$ and $|\rho| < 1$, preliminary propositions show that under the case $b(1 \pm \rho) < 2$, there exist a unique $l_-(\gamma,b,\rho) < l^*$ and a unique $l_-(\gamma,b,-\rho) < l^*$ such that $g_{-(b,\rho)}(l_-(\gamma,b,\rho)) = \gamma$ and $g_{-(b,-\rho)}(l_-(\gamma,b,-\rho)) = \gamma$. In such way, the quantity

$$\tilde{F}(b,\rho) := \inf\big\{\gamma | - L_-(l_-(\gamma,b,-\rho);\gamma,b,-\rho) > L_-(l_-(\gamma,b,\rho);\gamma,b,\rho)\big\} \vee -\sqrt{1-\rho^2}$$

is well-defined and it is called the Fukasawa threshold. If instead $b(1-\rho) = 2$ (or $b(1+\rho) = 2$), there exists a unique $l_-(\gamma,b,-\rho) < l^*$ (resp. $l_-(\gamma,b,\rho)$) such that $g_{-(b,-\rho)}(l_-) = \gamma$ (resp. $g_{-(b,\rho)}(l_-) = \gamma$). When the former case arises while the latter does not, the Fukasawa threshold is defined as $\tilde{F}(b,\rho) := \inf\big\{\gamma | - L_-(l_-(\gamma,b,-\rho);\gamma,b,-\rho) > -\frac{b\gamma}{2}\big\} \vee -\sqrt{1-\rho^2}$. Vice versa, when it is the second case to be active while the first is not, the quantity is defined as $\tilde{F}(b,\rho) := \inf\big\{\gamma | \frac{b\gamma}{2} > L_-(l_-(\gamma,b,\rho);\gamma,b,\rho)\big\} \vee -\sqrt{1-\rho^2}$. Finally, when both cases are valid, so $b = 2$ and $\rho = 0$, the threshold is $\tilde{F}(2,0) := 0$. The aforementioned theorem is:

Theorem 4.1 (SVI parameters $(\gamma, b, \rho, \mu, \sigma)$ fulfilling Fukasawa necessary no arbitrage conditions)**.** *Assume $\gamma + \sqrt{1-\rho^2} > 0$ and $|\rho| < 1$. Then:*

- *if $b(1 \pm \rho) < 2$, $\tilde{F}(b, \rho) < 0$ and the interval $I_{\gamma,b,\rho} = \left]L_-(l_-(\gamma, b, \rho); \gamma, b, \rho), -L_-(l_-(\gamma, b, -\rho); \gamma, b, -\rho)\right[$ is non-empty iff $\gamma > \tilde{F}(b, \rho)$;*
- *if $b(1 - \rho) = 2$ (or $b(1 + \rho) = 2$) and $\rho \neq 0$, $\tilde{F}(b, \rho) < 0$ and the interval $I_{\gamma,b,\rho} = \left]-\frac{b\gamma}{2}, -L_-(l_-(\gamma, b, -\rho); \gamma, b, -\rho)\right[$ (resp. $I_{\gamma,b,\rho} = \left]L_-(l_-(\gamma, b, \rho); \gamma, b, \rho), \frac{b\gamma}{2}\right[$) is non-empty iff $\gamma > \tilde{F}(b, \rho)$;*
- *if $b = 2$ and $\rho = 0$, the interval $I_{\gamma,2,0} = \left]-\gamma, \gamma\right[$ is non-empty iff $\gamma > \tilde{F}(2, 0) = 0$.*

In every case, the Fukasawa conditions are satisfied iff $\mu \in I_{\gamma,b,\rho}$.

The final necessary and sufficient conditions for no Butterfly arbitrage in the case $\gamma + \sqrt{1-\rho^2} > 0$ and $|\rho| < 1$ are summed up in Theorem 6.2 of [53], in which σ^* is defined as

$$\sigma^*(\gamma, b, \rho, \mu) := \sup_{l < l_1 \vee l > l_2} -\frac{b g_2(l)}{2G_1(l)}$$

where $l_1 < 0 < l_2$ are the only zeros of g_2.

Theorem 4.2 (Necessary and sufficient no Butterfly arbitrage conditions for SVI, $\gamma + \sqrt{1-\rho^2} > 0$ and $|\rho| < 1$)**.** *No Butterfly arbitrage in SVI entails that G_1 is positive, which requires $b(1 \pm \rho) \leq 2$. Under this condition:*

- *each of the factors of the function G_1 is positive on $\mathbb{R}$ if and only if $\gamma > \tilde{F}(b, \rho)$ and $\mu \in I_{\gamma,b,\rho}$;*
- *for such μ's, calling $l_1 < 0 < l_2$ the only zeros of g_2, the function $G_1 + \frac{1}{2\sigma} b g_2$ is positive in $]l_1, l_2[$ for every $\sigma \geq 0$ and it is non-negative on $\mathbb{R}$ if and only if $\sigma \geq \sigma^*(\gamma, b, \rho, \mu)$.*

In the case of $|\rho| = 1$ and $\gamma \geq 0$, Theorem 6.3 of [53] holds:

Theorem 4.3 (Necessary and sufficient no Butterfly arbitrage conditions for SVI, $\rho = -1$)**.** *No Butterfly arbitrage in SVI entails that G_1 is positive, which requires $b \leq 1$ and $\gamma \geq 0$. Under these conditions:*

- *each of the factors of the function G_1 is positive on $\mathbb{R}$ if and only if $\mu > L_-(l_-; \gamma, b, -1)$;*
- *for such μ's, calling $l_1 < 0$ the only zero of g_2, the function $G_1 + \frac{1}{2\sigma} b g_2$ is positive on $]l_1, \infty[$ for every $\sigma \geq 0$ and it is non-negative on $\mathbb{R}$ if and only if $\sigma \geq \sigma^*(\gamma, b, -1, \mu)$ where $\sigma^*(\gamma, b, -1, \mu) := \sup_{l < l_1} -\frac{b g_2(l)}{2G_1(l)}$.*

From now on, we denote

$$f := -\frac{bg_2}{2G_1}, \qquad \tilde{f} := -\frac{G_1}{g_2}.$$

In this way, the value of the supremum of f is equal to b times half the reciprocal of the infimum of $\tilde{f}$, and the point at which the supremum of f is reached is exactly the point at which the infimum of $\tilde{f}$ is reached.

We present here some general observations that will be used in the next sections.

In order to calculate σ^*, one should find the supremum of the function f over a domain constituted of two open intervals: on the left of l_1, the first zero of g_2, and on the right of l_2, the second zero of g_2. From a computational point of view, one needs to perform two maximum searches and to compare the two found values to choose the highest. Sometimes this double search is not necessary, indeed for the sub-SVIs studied in this chapter, the signs of ρ and μ determine the interval where the global supremum of f lies. Since $l_1 < 0 < l_2$, the trick will be to compare the two quantities $f(l)$ and $f(-l)$, but first of all, it is necessary to study the two intervals of interest. In particular, the following Lemma holds:

Lemma 4.1. *If $\rho \geq 0$ and $l > 0$, then $g_2(l) \leq g_2(-l)$ and $l_2(\gamma, \rho) \leq -l_1(\gamma, \rho)$.*
If $\rho < 0$ and $l > 0$, then $g_2(l) > g_2(-l)$ and $l_1(\gamma, \rho) > -l_2(\gamma, \rho)$.

Proof. Fix $l > 0$, then for $\rho \geq 0$, it holds

$$\begin{aligned} g_2(-l) - g_2(l) &= \frac{l\rho\Big((l^2+1)(1-\rho^2) + 2\gamma\sqrt{l^2+1} + 1\Big)}{(l^2+1)N(l)N(-l)} \\ &\geq \frac{l\rho\Big(\sqrt{(l^2+1)(1-\rho^2)} - 1\Big)^2}{(l^2+1)N(l)N(-l)} \geq 0 \end{aligned}$$

since $\gamma \geq -\sqrt{1-\rho^2}$. So $g_2(l) \leq g_2(-l)$ and in particular $g_2(-l_1) \leq g_2(l_1) = 0$ so $l_2(\gamma, \rho) \leq -l_1(\gamma, \rho)$. These inequalities are strict for ρ strictly positive. Similarly for $\rho < 0$, we find $g_2(l) > g_2(-l)$ and $l_1(\gamma, \rho) > -l_2(\gamma, \rho)$. □

This Lemma has a direct consequence, which is that for $\rho \geq 0$, the supremum of f on the right of l_2 is higher than its supremum on the right of $-l_1$, since in the first case, f could attain it between l_2 and $-l_1$: $\sup_{l>l_2} f(l) \geq \sup_{l>-l_1} f(l)$. If we can show that for positive ls, $f(l) \geq f(-l)$, then the last term is greater than $\sup_{l>-l_1} f(-l)$. Making a change of variable, this quantity is equal to $\sup_{l<l_1} f(l)$ so this means that σ^* can be found as the supremum of f on the right of l_2.

The request $f(l) \geq f(-l)$ is satisfied if, for example, $G_1(l) \leq G_1(-l)$ because in the above Lemma we showed $g_2(l) \leq g_2(-l)$. The difference between $G_1(l)$ and $G_1(-l)$ can be written as $(h(l) - h(-l))(h(l) + h(-l)) - b^2(g(l) - g(-l))(g(l) + g(-l))$. The quantity $g(l) - g(-l)$ is equal to $\frac{l}{2\sqrt{l^2+1}}$, which is positive, while $g(l) + g(-l)$ is $\frac{\rho}{2}$, which again is non-negative for $\rho \geq 0$. Showing that the product with the h functions is non-positive would then be enough to reach the desired inequality.

Lemma 4.2. *Let $\rho = 0$. If $\mu \geq 0$, then $\sigma^*(\gamma, b, 0, \mu) = \sup_{l>l_2} f(l)$ while if $\mu < 0$, then $\sigma^*(\gamma, b, 0, \mu) = \sup_{l<l_1} f(l)$.*

Proof. In the case $\rho = 0$, it holds $l_2 = -l_1$ since g_2 is symmetric. It can be shown that in general

$$h(l) - h(-l) = -\frac{l}{\sqrt{l^2+1}} \frac{\rho\big(\gamma\sqrt{l^2+1}+1\big) + \mu\big(\gamma + (1-\rho^2)\sqrt{l^2+1}\big)}{N(l)N(-l)},$$
$$h(l) + h(-l) = \frac{\big(\gamma\sqrt{l^2+1}+1\big)\big(2\gamma + 2\sqrt{l^2+1} - \rho\mu\big) + l^2\big(\gamma + (1-\rho^2)\sqrt{l^2+1}\big)}{\sqrt{l^2+1}N(l)N(-l)}.$$

In particular for $\rho = 0$ and positive l, the sign of the former quantity is the sign of $-\mu$ while the numerator of the latter quantity is $(\gamma + \sqrt{l^2+1})(2\gamma\sqrt{l^2+1} + l^2 + 2)$, which is always positive for $\gamma > -1$. Then $G_1(l) - G_1(-l) = (h(l) - h(-l))(h(l) + h(-l))$ follows the sign of $-\mu$ and so does $f(-l) - f(l)$. If μ is non-negative, $f(l) \geq f(-l)$ so its supremum must be searched on the right of l_2. If μ is negative, the opposite inequality holds for f and its supremum lies on the left of l_1.

□

These Lemmas will be useful for the sub-SVIs studied in this chapter, however it must be noticed that the inequality $f(l) \geq f(-l)$ does not hold in general.

4.2.2 Smile inversion

Tehranchi proved in [66] that a curve of Call prices (with unit underlyer) parametrized by the strike κ is free of Butterfly arbitrage if and only if it is convex and satisfies $1 \geq C(\kappa) \geq (1-\kappa)^+$ for every $\kappa \geq 0$. Moreover, C has these properties if and only if $C^*(\kappa) := 1 - \kappa + \kappa C\big(\frac{1}{\kappa}\big)$ has them.

Assume now that in addition $1 > C(\kappa)$ for every κ. Then there is a unique total variance function $w(k)$ such that $C_{\mathrm{BS}}(\kappa, \sqrt{w(k)}) = C(\kappa)$ where C_{BS} is the (normalized) Black-Scholes Call price function and $k = \log\kappa$, and a unique $w^*(k)$ such that $C_{\mathrm{BS}}(\kappa, \sqrt{w^*(k)}) = C^*(\kappa)$. By the Put-Call parity for the Black-Scholes model it holds that

$$P_{\mathrm{BS}}(\kappa, \sqrt{w^*(k)}) = C_{\mathrm{BS}}(\kappa, \sqrt{w^*(k)}) + \kappa - 1 = \kappa C\Big(\frac{1}{\kappa}\Big).$$

Now the LHS is equal to $\kappa N\Big(\frac{k}{\sqrt{w^*(k)}} + \frac{\sqrt{w^*(k)}}{2}\Big) - N\Big(\frac{k}{\sqrt{w^*(k)}} - \frac{\sqrt{w^*(k)}}{2}\Big)$ and the RHS is equal to $\kappa\Big(N\Big(\frac{k}{\sqrt{w(-k)}} + \frac{\sqrt{w(-k)}}{2}\Big) - \frac{1}{\kappa}N\Big(\frac{k}{\sqrt{w(-k)}} - \frac{\sqrt{w(-k)}}{2}\Big)\Big)$. By the monotonicity of the function $u \to C_{\mathrm{BS}}(\kappa, u)$ it follows then that $w^*(k) = w(-k)$. Eventually, we can reword the symmetry of Tehranchi involution $C \to C^*$ at the implied volatility level: a smile w is free of Butterfly arbitrage if and only if the inverse smile $k \to w(-k)$ is.

Let us now relate the implied volatility of C with that of C^*.

In the case of SVI, this reduces to have

$$a + b\Big(\rho(k-m) + \sqrt{(k-m)^2+\sigma^2}\Big) = a^* + b^*\Big(\rho^*(-k-m^*) + \sqrt{(-k-m^*)^2+\sigma^{*2}}\Big) \quad (4.3)$$

for all k. We prove that this happens iff $a^* = a$, $b^* = b$, $\rho^* = -\rho$, $m^* = -m$ and $\sigma^* = \sigma$. Of course, a^*, b^*, ρ^*, m^* and σ^* do not depend on k^* so neither on k. Making the derivative with respect to k at the latter equality gives

$$b\left(\rho + \frac{k-m}{\sqrt{(k-m)^2+\sigma^2}}\right) = b^*\left(-\rho^* + \frac{k+m^*}{\sqrt{(k+m^*)^2+\sigma^{*2}}}\right). \tag{4.4}$$

Evaluating this for k going to $\pm\infty$, it must hold $b(\rho \pm 1) = b^*(-\rho^* \pm 1)$ or equivalently $b^* = b$ and $\rho^* = -\rho$. In the same equation, consider $k = m$, then it must hold

$$b\rho = b\left(\rho + \frac{m+m^*}{\sqrt{(m+m^*)^2+\sigma^{*2}}}\right)$$

which reduces to $m^* = -m$. At this point, Equation (4.4) becomes $\frac{k-m}{\sqrt{(k-m)^2+\sigma^2}} = \frac{k+m^*}{\sqrt{(k+m^*)^2+\sigma^{*2}}}$ so $\sigma^* = \sigma$. Finally, from Equation (4.3), it follows $a^* = a$.

Proposition 4.1 (Absence of Butterfly arbitrage for the inverse SVI)**.** *The SVI*(a, b, ρ, m, σ) *is Butterfly arbitrage-free iff the SVI*$(a, b, -\rho, -m, \sigma)$ *is Butterfly arbitrage-free.*

4.2.3 The b^* approach: reparametrizing from the critical point equation

We now explain how to properly switch the roles of the parameter b and of the critical point $\bar{l}$ depending on b of $\tilde{f}$. The key passage will be the exploitation of the characteristic equation of these critical points: $h(\bar{l})p(\bar{l}) - b^2 g(\bar{l})q(\bar{l}) = 0$. Three hypothesis must be verified:

1. the Fukasawa bound for γ must be a monotone function in b;
2. there is uniqueness for the critical points of $\tilde{f}$;
3. the functions p and q do not vanish at the same point.

In [53], the Fukasawa conditions read $\gamma > \tilde{F}(b, \rho)$. Firstly, suppose that for fixed ρ, **the function $b \to \tilde{F}(b, \rho)$ is monotone and surjective from $\left[0, \frac{2}{1+|\rho|}\right]$ to $[-1, 0]$.** Then, its inverse $\tilde{F}^{-1}(\cdot, \rho)$ is well defined on $[-1, 0]$ and we shall extend it to

$$\tilde{G}(\gamma, \rho) := \begin{cases} \tilde{F}^{-1}(\gamma, \rho) & \text{if } \gamma \in]-1, 0], \\ \frac{2}{1+|\rho|} & \text{if } \gamma > 0. \end{cases}$$

Secondly, suppose we have proven that **for $l > l_2$, the function $\tilde{f}$ has exactly one critical point $\bar{l}(\gamma, b, \rho, \mu)$ for $\rho \in [-1, 1], \gamma > -1, b \in [0, \tilde{G}(\gamma, \rho)], \mu \in]L_-, L_+[$.** This critical point must then be a local and global point of minimum. Note that we require the uniqueness also for $b = 0$ and $b = \tilde{G}(\gamma, \rho)$. This point satisfies $\tilde{f}'(\bar{l}) = 0$ or equivalently

$$h(\bar{l})p(\bar{l}) - b^2 g(\bar{l})q(\bar{l}) = 0$$

with $h(\bar{l}) = G_{1+}(l) + G_{1-}(l) > 0$, $g(l) > 0$ and $p(l) := h(l)g_2'(l) - 2h'(l)g_2(l)$ and $q(l) := g(l)g_2'(l) - 2g'(l)g_2(l)$.

In particular, either $q(\bar{l}) = p(\bar{l}) = 0$ or

$$b^2 = \frac{h(\bar{l})\big(h(\bar{l})g_2'(\bar{l}) - 2h'(\bar{l})g_2(\bar{l})\big)}{g(\bar{l})\big(g(\bar{l})g_2'(\bar{l}) - 2g'(\bar{l})g_2(\bar{l})\big)}$$

when $q(\bar{l}) \neq 0$.

It is natural to define on the set $B(\gamma,\rho,\mu) := \{l \in]l_2(\gamma,\rho), \infty[: p(l)q(l) > 0\}$ a positive function b^* by the formula

$$b^{*2}(l) := \frac{h(l)p(l)}{g(l)q(l)}. \tag{4.5}$$

Note that the function b^* might go on $B(\gamma,\rho,\mu)$ to levels not allowed for b; yet this function is continuous on its domain of definition and when $q(\bar{l}) \neq 0$ it holds that $b^2 = b^{*2}(\bar{l})$. As an immediate consequence, note that if $b_1 \neq b_2$ are such that $\bar{l}(\gamma,b_1,\rho,\mu), \bar{l}(\gamma,b_2,\rho,\mu) \in B(\gamma,\rho,\mu)$, then $\bar{l}(\gamma,b_1,\rho,\mu) \neq \bar{l}(\gamma,b_2,\rho,\mu)$. Note also that we don't know if B is one-piece, i.e. connected.

The question of interest is now the location of the set $Z(\gamma,\rho,\mu) := \{\bar{l}(\gamma,b,\rho,\mu), b \in]0,\tilde{G}(\gamma,\rho)[\}$ (note that we exclude 0, this will turn to be more convenient below) with respect to the sets $B(\gamma,\rho,\mu)$ and $\{p = q = 0\}$. The third and last hypothesis is that, **for the fixed parameters (γ,ρ,μ), the set $\{p = q = 0\}$ is empty**. Then $Z(\gamma,\rho,\mu)$ is contained in $B(\gamma,\rho,\mu)$. From the above remark, the function $\bar{l}(\gamma,\cdot,\rho,\mu)$ is injective; however at this stage we don't know whether it is continuous (which would imply it is either continuous increasing or continuous decreasing).

Remember that the proof of the uniqueness of the critical point of $\tilde{f}$ still holds for $b = \tilde{G}(\gamma,\rho)$. Then for $(l,b) \in]l_2,\infty] \times [0,\tilde{G}(\gamma,\rho)]$, the equation

$$h(l)p(l) - b^2 g(l)q(l) = 0$$

characterizes the points $\bar{l}(\gamma,b,\rho,\mu)$, since $\tilde{f}$ has a single local and global minimum. As discussed above, in the open set $B(\gamma,\rho,\mu)$, this equation defines a continuous function b^*; from the characterizing property we get that $Z(\gamma,\rho,\mu)$ eventually coincides with $b^{*-1}(]0,\tilde{G}(\gamma,\rho)[) \subseteq B(\gamma,\rho,\mu)$, so that in particular $Z(\gamma,\rho,\mu)$ is an open set. Furthermore, the function $b \to \bar{l}(\gamma,b,\rho,\mu)$ is the inverse of $b^* : Z(\gamma,\rho,\mu) \to]0,\tilde{G}(\gamma,\rho)[$.

It remains to prove that $Z(\gamma,\rho,\mu)$ is *one-piece*. Indeed, take a sequence $(b_n)_n$ such that $\bar{l}_n = \bar{l}(\gamma,b_n,\rho,\mu)$ goes to $\bar{l} \in \bar{Z}(\gamma,\rho,\mu)$. Since $(b_n)_n$ is a bounded sequence, it has a subsequence converging to a certain $b \in [0,\tilde{G}(\gamma,\rho)]$; call such subsequence as the original one, then the correlated subsequence of $\bar{l}_n$ still converges to $\bar{l}$. Since for every n we have $h(\bar{l}_n)p(\bar{l}_n) - b_n^2 g(\bar{l}_n)q(\bar{l}_n) = 0$ where all the functions are continuous, then taking the limit, it holds $h(\bar{l})p(\bar{l}) - b^2 g(\bar{l})q(\bar{l}) = 0$. From the characteristic equation above, this yields in turn that $\bar{l} = \bar{l}(\gamma,b,\rho,\mu)$. Also, b is unique because either $b = b^*(\bar{l})$ or $p(\bar{l}) = q(\bar{l}) = 0$, but the set $\{p = q = 0\}$ is empty. Then it must hold either that $\bar{l} \in Z(\gamma,\rho,\mu)$ or that $\bar{l}$ is the unique critical point $\bar{l}(\gamma,b,\rho,\mu)$ of $\tilde{f}$ when $b = 0$ or $b = \tilde{G}(\gamma,\rho)$. The boundary of $Z(\gamma,\rho,\mu)$ is the set $\{\bar{l}(\gamma,0,\rho,\mu), \bar{l}(\gamma,\tilde{G}(\gamma,\rho),\rho,\mu)\}$ and

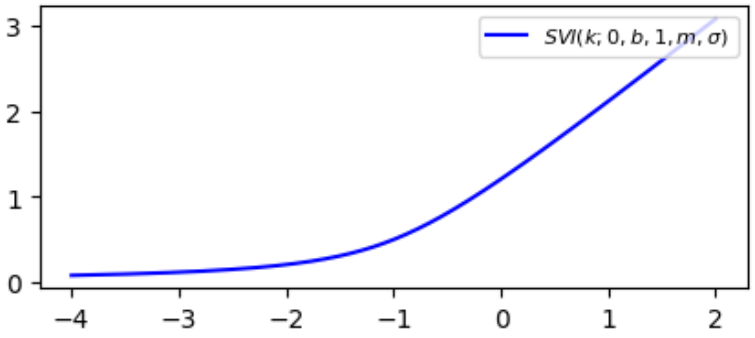

Figure 4.1: Vanishing Upward SVI with $b = \frac{1}{2}$, $m = -1$ and $\sigma = 1$.

because $Z(\gamma, \rho, \mu)$ is an open set, we get that $Z(\gamma, \rho, \mu) =]\bar{l}(\gamma, 0, \rho, \mu), \bar{l}(\gamma, \tilde{G}(\gamma, \rho), \rho, \mu)[$ or $Z(\gamma, \rho, \mu) =]\bar{l}(\gamma, \tilde{G}(\gamma, \rho), \rho, \mu), \bar{l}(\gamma, 0, \rho, \mu)[$.

As a consequence we also get that $\bar{l}(\gamma, \cdot, \rho, \mu)$ is either strictly increasing or strictly decreasing, and that $Z(\gamma, \rho, \mu)$ is a single connected component of $B(\gamma, \rho, \mu)$, with a boundary point $\bar{l}(\gamma, 0, \rho, \mu)$ which is the single zero of p, and the other boundary point lying within $B(\gamma, \rho, \mu)$.

4.3 Vanishing SVI

4.3.1 Vanishing (Upward) SVI

In this section, we work with the Vanishing Upward SVI and immediately recover the final results on the Vanishing Downward SVI in Section 4.3.4.

The Vanishing Upward SVI is the sub-SVI obtained by setting $\rho = 1$ and $a = 0$. The corresponding SVI formula becomes

$$\mathrm{SVI}(k; 0, b, 1, m, \sigma) = b(k - m + \sqrt{(k-m)^2 + \sigma^2}).$$

With our notations, $N(l) = l + \sqrt{l^2 + 1}$. Note that the Roger Lee conditions require $0 < b \leq 1$.

We plot in Figure 4.1 a Vanishing Upward SVI with $b = \frac{1}{2}$, $m = -1$ and $\sigma = 1$.

The wording *Vanishing* refers to the fact that the smile goes to 0 on the left, *upward* meaning it is increasing. The symmetric smile with $\rho = -1$ will be a *Vanishing Downward* one.

4.3.2 The Fukasawa conditions

Since $\gamma = 0$, the Fukasawa condition on γ is automatically satisfied. We cite here a result obtained in paragraph 5.3.1 of [53].

Lemma 4.3 (Fukasawa conditions for the Vanishing Upward SVI)**.** *A Vanishing Upward SVI with $0 < b \leq 1$ satisfies the Fukasawa conditions iff $\mu < \sqrt{3(1-b)}$.*

4.3.3 The condition on σ

Let us define the two following auxiliary functions:

$$\mu^*(x) := \frac{2(1-x)(2x^2-8x-1) + \sqrt{\begin{array}{r}4b^2x^6+8b^2x^5+8x^4(8-b^2)-4x^3(5b^2+32)\\ +x^2(96-b^2)+2x(5b^2-16)+4+3b^2\end{array}}}{2\sqrt{1-x^2}(2x^2-2x-1)}, \tag{4.6}$$

$$\sigma^*(x) := -\frac{4b\sqrt{1-x^2}(1-x-2x^2)}{4\big(2-x-\mu^*(x)\sqrt{1-x^2}\big)^2 - b^2(1+x)^2} \tag{4.7}$$

We show then the following:

Proposition 4.2 (Fully explicit no arbitrage domain for the Vanishing Upward SVI)**.** *A Vanishing Upward SVI with $b=1$ is arbitrage-free iff $\mu<0$ and $\sigma \geq -\frac{\mu}{2}$.*

A Vanishing Upward SVI with $0<b<1$ is arbitrage-free iff it can be parametrized as

$$SVI(k) = b\sigma\left(\frac{k}{\sigma} - \mu^*(x) + \sqrt{\left(\frac{k}{\sigma}-\mu^*(x)\right)^2+1}\right) \tag{4.8}$$

where $\frac{2+b}{4-b} < x < 1$ and $\sigma \geq \sigma^*(x)$.

Proof of Proposition 4.2

Note that

$$G_1(l) = \left(1 - \frac{(l+\mu)}{2\sqrt{l^2+1}}\right)^2 - \frac{b^2}{16}\left(1+\frac{l}{\sqrt{l^2+1}}\right)^2,$$
$$g_2(l) = \frac{1}{(l^2+1)^{\frac{3}{2}}} - \frac{\sqrt{l^2+1}+l}{2(l^2+1)}.$$

The g_2 function has only one positive zero. Indeed, its zeros solve $2 = l^2+1+l\sqrt{l^2+1}$, or yet $1-l^2 = l\sqrt{l^2+1}$. The only possible solution satisfies $l^2 \leq 1$, so $l_2 = \frac{1}{\sqrt{3}}$. In order to have no arbitrage, we then need $\sigma \geq \sigma^* = \sup_{l\in]l_2,+\infty[} -\frac{bg_2(l)}{2G_1(l)}$.

From now on, we operate a change of variable setting $x = \frac{l}{\sqrt{l^2+1}}$ so that the points $l = l_2$ and $l=\infty$ correspond to $x=\frac{1}{2}$ and $x=1$ respectively. Also, $\frac{1}{\sqrt{l^2+1}} = \sqrt{1-x^2}$.

We call J_1 and j_2 the functions G_1 and g_2 evaluated at x, so

$$J_1(x) = \left(1 - \frac{x}{2} - \frac{\mu}{2}\sqrt{1-x^2}\right)^2 - \frac{b^2}{16}(1+x)^2,$$
$$j_2(x) = \frac{\sqrt{1-x^2}}{2}(1-x-2x^2).$$

The first derivative of j_2 is $j_2'(x) = \frac{1}{2\sqrt{1-x^2}}(6x^3 + 2x^2 - 5x - 1)$ which is positive iff $x > \frac{2+\sqrt{10}}{6} := x_{m_2}$. Also, $j_2''(x) = \frac{-12x^4-2x^3+18x^2+3x-5}{2(1-x^2)^{\frac{3}{2}}}$ and the only inflection point of j_2 in $\left[\frac{1}{2}, 1\right]$ is $\frac{1}{2}$. Then, the function j_2 is null at $\frac{1}{2}$, decreases to its minimum $j_2(x_{m_2}) = -\frac{34\sqrt{2}-5\sqrt{5}}{108}$ and then increases to 0 for $x = 1$, furthermore it is always convex.

Let us turn to the study of J_1. We have $J_1(1) = \frac{1-b^2}{4}$ and we know that for the Fukasawa conditions, both J_{1+} and J_{1-} are positive, where $J_1 = J_{1+}J_{1-}$ and $J_{1\pm} = 1 - \frac{x}{2} - \frac{\mu}{2}\sqrt{1-x^2} \mp \frac{b}{4}(1+x)$. Considering the first derivatives, we know $J_1' = J_{1+}'J_{1-} + J_{1+}J_{1-}'$ where

$$J_{1\pm}'(x) = -\frac{1}{2} + \frac{\mu}{2}\frac{x}{\sqrt{1-x^2}} \mp \frac{b}{4}. \tag{4.9}$$

If $\mu \leq 0$, then $J_{1\pm}$ are decreasing and so is J_1. In particular, it attains its minimum at $x = 1$ and we get that $J = J_1 + \frac{1}{2\sigma}bj_2$ is always positive for

$$\sigma \geq -\frac{bj_2(x_{m_2})}{2J_1(1)} = \frac{(34\sqrt{2}-5\sqrt{5})b}{54(1-b^2)}$$

when $b < 1$. We have therefore the explicit no arbitrage sub-domain:

Lemma 4.4 (No arbitrage sub-domain for the Vanishing Upward SVI)**.** *A Vanishing Upward SVI with $0 < b < 1$, $\mu \leq 0$ and $\sigma \geq \frac{(34\sqrt{2}-5\sqrt{5})b}{54(1-b^2)}$ is arbitrage-free.*

This explicit sub-domain was obtained with not too much effort. Let us turn now to the more difficult task to obtain an explicit parametrization for the whole domain.

Uniqueness of the critical point of $\tilde{f}$ for the Vanishing SVI Let us study $\tilde{\phi}(x; b, \mu) = -\frac{J_1(x;b,\mu)}{j_2(x;b)}$. We want to prove that for each $b \in [0, 1]$ and $\mu \in]-\infty, \sqrt{3(1-b)}[$, there exists a unique value $x^* := x^*(b, \mu) \in \left[\frac{1}{2}, 1\right]$ such that $\inf_{x\in]\frac{1}{2},1[} \tilde{\phi}(x) = \tilde{\phi}(x^*)$.

The existence is obvious. The derivative of $\tilde{\phi}$ is

$$\begin{aligned}\tilde{\phi}'(x; b, \mu) = &\Big[4\mu^2(1-x^2)(2x^2-2x-1) - 8\mu(1-x)\sqrt{1-x^2}(2x^2-8x-1) + \\ &+ 2x^4(4-b^2) - 6x^3(b^2+12) + 3x^2(52-b^2) + 4x(b^2-22) + 3b^2\Big] \\ &/\Big[8(1-x^2)^{\frac{3}{2}}(1+x)(2x-1)^2\Big].\end{aligned} \tag{4.10}$$

The denominator of the above formula is always positive. Looking at the numerator, for $b = 1$ every coefficient of μ is negative when $x \in \left]\frac{1}{2}, 1\right[$ and $\mu < 0$, required from the Fukasawa condition. Then $\tilde{\phi}'(x; 1, \mu)$ is always negative and the inferior point of $\tilde{\phi}$ is reached at 1 for every $\mu < 0$, so that $x^*(1, \mu) = 1$. From now on we consider $b < 1$. Since for finite values of μ, $\tilde{\phi}(\frac{1}{2}; b, \mu) = \tilde{\phi}(1; b, \mu) = \infty$, the points which attain the infimum of $\tilde{\phi}$ are points of minimum belonging to $\left]\frac{1}{2}, 1\right[$ and such that $\tilde{\phi}'(x^*; b, \mu) = 0$.

What happens when $\mu = -\infty$? Dividing by μ^2 we still have $\frac{\tilde{\phi}'(x^*;b,\mu)}{\mu^2} = 0$, so making μ going to $-\infty$ and relying on the linearity of the limits, we obtain the equation $4(1-x^{*2})(2x^{*2}-2x^*-1) = 0$, whose only solution in $\left[\frac{1}{2}, 1\right]$ is 1, so $x^*(b, -\infty) = 1$.

Take now $\mu \leq 0$ and $|\nu| < -\mu$ (note in particular that $\mu < \nu$) and consider the quantity $\tilde{\phi}(x;b,\mu) - \tilde{\phi}(x;b,\nu)$. Unless a factor $b^{-1}(1+x)^{-1}(2x-1)^{-1}$, it is equal to $(\nu-\mu)\big(2-x-\frac{(\mu+\nu)}{2}\sqrt{1-x^2}\big)$ and, as a consequence, the quantity $\tilde{\phi}'(x;b,\mu) - \tilde{\phi}'(x;b,\nu)$ times $(1+x)(2x-1)$ equals

$$(\nu-\mu)\left(-1+\frac{(\mu+\nu)}{2}\frac{x}{\sqrt{1-x^2}}\right) + (4x+1)\big(\tilde{\phi}(x;b,\nu) - \tilde{\phi}(x;b,\mu)\big).$$

The function $\tilde{\phi}$ is decreasing in μ, indeed $\partial_\mu \tilde{\phi}(x) = -\frac{2-x-\mu\sqrt{1-x^2}}{b(1+x)(2x-1)}$ and $2-x-\mu\sqrt{1-x^2} > 2-x-\sqrt{3}\sqrt{1-x^2} \geq 0$. So given the fact that $\tilde{\phi}$ is decreasing in μ, we get in turn $\tilde{\phi}'(x;b,\mu) - \tilde{\phi}'(x;b,\nu) < 0$. This entails that if there is uniqueness, $x^*(b,\nu) < x^*(b,\mu)$ under those conditions.

We prove that indeed the uniqueness holds. We prove that if $\tilde{\phi}' \geq 0$, then $\tilde{\phi}'' > 0$, which means that once $\tilde{\phi}'$ becomes zero, then it will necessarily increase and it can never become 0 again (note that this property is slightly weaker than the convexity of $\tilde{\phi}$). We have $\tilde{\phi}' = \frac{J_1 j_2' - J_1' j_2}{j_2^2} \geq 0$ but the term $-J_1' j_2$ is negative, so j_2' must be positive. It holds $\tilde{\phi}'' = \frac{J_1 j_2'' - J_1'' j_2}{j_2^2} - 2\tilde{\phi}'\frac{j_2'}{j_2}$, where we have proven $-2\tilde{\phi}'\frac{j_2'}{j_2} \geq 0$. We now look at the quantity $n = J_1 j_2'' - J_1'' j_2$ and show that it is strictly positive to obtain the conclusion. Indeed, the denominator of n is $32\sqrt{1-x^2}(1-x)$ which is positive while its numerator is a quadratic function of μ with quadratic coefficient $-4(1-x^2)(2x-1)(4x^2-2x-3)$, linear coefficient $16\sqrt{1-x^2}(1-x)(2x-1)(2x^2-5x-4)$ and free term depending on b. The quadratic coefficient is positive, so n is a convex parabola as a function of μ. Furthermore, the linear coefficient, corresponding to $\partial_\mu n(x;\mu)|_{\mu=0}$ unless a positive factor, is negative, so $n(x;0) < n(x;\mu)$ for every $\mu < 0$. Then for negative μ's it is enough to prove that $n(x;0)$ is positive for every x. We already know that j_2'' is positive. In addition, $J_1'' = J_{1+}'' J_{1-} + 2J_{1+}' J_{1-}' + J_{1+} J_{1-}''$ where $J_{1\pm}''(x;0) = \frac{\mu}{2(1-x^2)^{\frac{3}{2}}}\Big|_{\mu=0} = 0$ and $J_{1\pm}'(x;0) < 0$. Then $J_1''(x;0) > 0$ and $n(x;0) = J_1(x;0) j_2''(x;0) - J_1''(x;0) j_2(x;0) > 0$.

For the final case $\mu > 0$, we need to introduce the μ^*-function.

The μ^* approach for the Vanishing SVI Assume $b < 1$ so that $J_1(1) > 0$ and let us make μ vary towards its upper bound defined by the Fukasawa condition. Then the function J_1 will eventually reach the x-axis level at a point $x_+^*(b)$. This point is necessarily a critical point of J_1 with μ set at the value of the upper bound. So $J_1(x_+^*(b)) = J_1'(x_+^*(b)) = 0$. Now observe that $\tilde{\phi}' = \frac{J_1 j_2' - J_1' j_2}{j_2^2}$ so that $\tilde{\phi}'(x_+^*(b); b, \sqrt{3(1-b)}) = 0$ also. In particular, if we know that $\tilde{\phi}$ (for this critical value of μ) has a single critical point, this must be $x_+^*(b)$.

We look for the solutions to $J_1(x) = 0$ when $\mu = \sqrt{3(1-b)}$. J_{1-} is always positive on $\left]\frac{1}{2}, 1\right[$ while $J_{1+}(x) = 1 - \frac{x}{2} - \frac{\mu}{2}\sqrt{1-x^2} - \frac{b}{4}(1+x)$ satisfies $\left(\left(1-\frac{b}{4}\right)x - \left(\frac{1}{2}+\frac{b}{4}\right)\right)^2 = 0$. This yields

$$x_+^*(b) = \frac{\frac{1}{2}+\frac{b}{4}}{1-\frac{b}{4}} = \frac{2+b}{4-b}$$

which is clearly larger than $\frac{1}{2}$ for positive b and smaller than 1 when $b < 1$.

As a sanity check we should verify that also $J_1'(x) = 0$ at this point. Indeed, $J_1'(x_+^*(b)) = J_{1+}'(x_+^*(b))J_{1-}(x_+^*(b)) + J_{1+}(x_+^*(b))J_{1-}'(x_+^*(b)) = J_{1+}'(x_+^*(b))J_{1-}(x_+^*(b))$. From Equation (4.9) we have that J_{1+}' is null for $\mu = \sqrt{3(1-b)}$ and $x = x_+^*(b)$.

In the limiting case $b = 1$ and $\mu = 0$, we have already seen that J_1 is decreasing and attains its minimum at 1. Furthermore, $\tilde{\phi}(1; 1, 0) = 0$, so also $\tilde{\phi}$ reaches its minimum at 1. We can then set $x^*(1) = 1$. The uniqueness of x^* follows from the fact that J_1 and so $\tilde{\phi}$ are positive in $\left[\frac{1}{2}, 1\right[$.

Take $x \in]x_+^*(b), 1[$. From Equation (4.10) we can see that $\tilde{\phi}'$ is a concave parabola as a function of μ. So in order to have a unique $\mu = \mu^*(x)$-solution to $\tilde{\phi}'(x; b, \mu) = 0$, it is enough to prove $\tilde{\phi}'(x; b, \sqrt{3(1-b)}) > 0$. Note that $x_+^*(b)$ is a zero and a point of minimum for $\tilde{\phi}(\cdot; b, \sqrt{3(1-b)})$, so in a right neighborhood of $x_+^*(b)$ we have $\tilde{\phi}'(x; b, \sqrt{3(1-b)}) > 0$. If for a certain x we rather have $\tilde{\phi}'(x; b, \sqrt{3(1-b)}) < 0$, then there exists a $x_-^*(b) > x_+^*(b)$ point of maximum such that $\tilde{\phi}'(x_-^*(b); b, \sqrt{3(1-b)}) = 0$ and $\tilde{\phi}''(x_-^*(b); b, \sqrt{3(1-b)}) < 0$. We show that this is not possible.

Let us look at J_1'. From Equation (4.9), we see that $J_{1+}'(x)$ is positive iff $\frac{x}{\sqrt{1-x^2}} > \frac{b+2}{2\mu}$ or $x > \frac{2+b}{\sqrt{4\mu^2+(2+b)^2}}$. For $\mu = \sqrt{3(1-b)}$ we obtain exactly that x has to be greater than $x_+^*(b)$. On the other side, $J_{1-}'(x)$ is positive iff $x > \frac{2-b}{\sqrt{4\mu^2+(2-b)^2}}$ which corresponds to $x > \frac{2-b}{\sqrt{b^2-16b+16}}$ for $\mu = \sqrt{3(1-b)}$. This quantity is inferior to $x_+^*(b)$ so also $J_{1-}'(x)$ is positive. Then $J_1'(x_-^*(b)) > 0$. Suppose $x_-^*(b) \geq x_{m_2}$, then $j_2'(x_-^*(b)) \geq 0$ and consequently $\tilde{\phi}'(x_-^*(b)) > 0$, which is not possible. So if it exists, $x_-^*(b)$ must be smaller than x_{m_2}.

We have $\tilde{\phi}'' = \frac{J_1 j_2'' - J_1'' j_2}{j_2^2} - 2\tilde{\phi}' \frac{j_2'}{j_2}$ so when evaluating in $x_-^*(b)$, the second term is null. The quantities $J_1(x_-^*(b))$, $-j_2(x_-^*(b))$ and $j_2''(x_-^*(b))$ are strictly positive. Also, $J_1'' = J_{1+}'' J_{1-} + 2J_{1+}' J_{1-}' + J_{1+} J_{1-}''$ and $J_{1\pm}''(x) = \frac{\mu}{2(1-x^2)^{\frac{3}{2}}}$ which are positive. Since also $J_{1\pm}$ and $J_{1\pm}'$ are positive in $x_-^*(b)$, then $J_1''(x_-^*(b)) > 0$. So $\tilde{\phi}''(x_-^*(b); b, \sqrt{3(1-b)})$ cannot be negative and this leads to a contradiction.

Consequently, for fixed $b \in]0, 1[$, there is a unique function $x \to \mu^*(x)$ such that $\tilde{\phi}'(x; b, \mu^*(x)) = 0$ is well defined for $x \in]x_+^*(b), 1[$ and takes values in $]-\infty, \sqrt{3(1-b)}[$. It is also continuous since it is defined as the first root of a second degree polynomial with continuously changing parameters. In particular, μ^* is defined as in Equation (4.6).

Note that the function μ^* cannot be extended to $b = 1$ since the inferior point of $\tilde{\phi}$ is reached at 1 for every $\mu < 0$. On the other hand, it can be extended to $b = 0$ since all the previous statements still hold.

We have already seen that for x in the preimage of $]-\infty, 0]$, the function $\mu^*(x)$ is injective. We show that this still holds in the preimage of $]0, \sqrt{3(1-b)}[$. If this is not the case, there exists a critical point $\hat{x}$ such that $\mu^{*\prime}(\hat{x}) = 0$. Taking the derivative with respect to x at the members of $\tilde{\phi}'(x; \mu^*(x)) = 0$, we obtain $\tilde{\phi}''(x; \mu^*(x)) + \partial_\mu \tilde{\phi}'(x; \mu^*(x))\mu^{*\prime}(x) = 0$, so at $\hat{x}$ it holds simultaneously $\tilde{\phi}'(\hat{x}; \mu^*(\hat{x})) = 0$ and $\tilde{\phi}''(\hat{x}; \mu^*(\hat{x})) = 0$. We want to prove that this is not possible.

At $\hat{x}$, we have $\tilde{\phi}''\,|_{\mu=\mu^*(\hat{x})} = \frac{J_1 j_2'' - J_1'' j_2}{j_2^2}\,|_{\mu=\mu^*(\hat{x})}$. Since we have already proven that $j_2'' > 0$ on $\left]\frac{1}{2}, 1\right[$, it is enough to prove that $J_1'' > 0$. We consider $\mu \geq 0$. It can be shown that $J_1''(x) = \frac{(4-b^2-4\mu^2)(1-x^2)^2+4\mu\sqrt{1-x^2}(2x^3-3x+2)}{8(1-x^2)^2}$ and $J_1'''(x) = \frac{3\mu(2x-1)}{2(1-x^2)^{\frac{5}{2}}}$, which is positive iff $x > \frac{1}{2}$. So J_1'' is increasing. We have $J_1''\left(\frac{1}{2}\right) = -\frac{b^2}{8} - \frac{\mu^2}{2} + \frac{\mu}{\sqrt{3}} + \frac{1}{2}$ which is positive iff $-4\mu^2 + \frac{8}{\sqrt{3}}\mu + 4 - b^2 > 0$; this concave parabola is positive at zero and its positive root is $\frac{2+\sqrt{16-3b^2}}{2\sqrt{3}}$. This quantity is greater than $\sqrt{3(1-b)}$ so J_1'' is always positive for $\mu \in [0, \sqrt{3(1-b)}[$.

The proof that the function $\mu^*(x)$ is injective corresponds to the proof of the uniqueness of $x^*(b,\mu)$ in the set $]x_+^*(b), 1[$. In order to obtain the uniqueness in the whole $\left]\frac{1}{2}, 1\right[$, we prove now that $\tilde{\phi}$ is strictly decreasing in $\left]\frac{1}{2}, x_+^*(b)\right[$, so that any critical point x^* cannot live in this set. For the continuity of μ^* and since $\mu^*(1) = -\infty$, for every $\mu \leq 0$ there exists $x > x_+^*(b)$ such that $\mu^*(x) = \mu$. Furthermore, for any fixed x, we have proven that there is at most one possible $\mu \leq 0$ such that $\tilde{\phi}'(x; b, \mu) = 0$. So if $x < x_+^*(b)$, there is no $\mu \leq 0$ satisfying the latter equation.

Consider now $\mu > 0$. Firstly, we prove that $\tilde{\phi}'$ is increasing with respect to μ. From Equation (4.10) it is evident that it holds true iff $\mu < \frac{(1-x)(2x^2-8x-1)}{\sqrt{1-x^2}(2x^2-2x-1)}$. The right hand side is a decreasing function for $x \in \left]\frac{1}{2}, 1\right[$ so it is enough to check the inequality in $x_+^*(b)$. Here the condition becomes $\mu < \frac{b^2-8}{b^2+4b-8}\sqrt{3(1-b)}$ and this is greater than $\sqrt{3(1-b)}$, so the inequality holds true.

So in order to prove that $\tilde{\phi}'(x; b, \mu)$ is negative, it is enough to prove that $\tilde{\phi}'(x; b, \sqrt{3(1-b)})$ is non-positive. Suppose now that there exists $x < x_+^*(b)$ such that $\tilde{\phi}'(x; b, \sqrt{3(1-b)}) > 0$. Since $\tilde{\phi}'(\frac{1}{2}; b, \sqrt{3(1-b)})$ is negative, there is an intermediate point at which $\tilde{\phi}'$ is null and then becomes positive up to x. Also $\tilde{\phi}'(x_+^*(b); b, \sqrt{3(1-b)})$ is null so there is $x_-^* < x_+^*(b)$ such that $\tilde{\phi}'(x_-^*; b, \sqrt{3(1-b)}) > 0$ and $\tilde{\phi}''(x_-^*; b, \sqrt{3(1-b)}) = 0$. However it holds that $\tilde{\phi}''(x_-^*; b, \sqrt{3(1-b)}) = \frac{J_1(x_-^*) j_2''(x_-^*) - J_1''(x_-^*) j_2(x_-^*)}{j_2(x_-^*)^2} - 2\tilde{\phi}'(x_-^*)\frac{j_2'(x_-^*)}{j_2(x_-^*)}$. We have already proven that J_1'' is positive for all $x \in \left]\frac{1}{2}, 1\right[$ so the first term is positive in $x_-^* < x_+^*(b)$. Then, it must be $\tilde{\phi}'(x_-^*)\frac{j_2'(x_-^*)}{j_2(x_-^*)} > 0$ or equivalently $j_2'(x_-^*) < 0$. Note that $J_1'(x_-^*)$ is negative because J_1'' is positive in $\left]\frac{1}{2}, 1\right[$ and for $\mu = \sqrt{3(1-b)}$ the function J_1 is null at $x_+^*(b)$. Then $\tilde{\phi}'(x_-^*) = \frac{J_1(x_-^*) j_2'(x_-^*) - J_1'(x_-^*) j_2(x_-^*)}{j_2(x_-^*)^2}$ is negative, which is a contradiction.

To sum up, we have shown that for fixed $b \in [0, 1[$ and $\mu < \sqrt{3(1-b)}$, there is only one x such that $\tilde{\phi}'(x; b, \mu) = 0$ and this x lives in $]x_+^*(b), 1[$. Furthermore, for fixed $b \in [0, 1[$ and $x \in]x_+^*(b), 1[$, there is only one μ in $]-\infty, \sqrt{3(1-b)}[$ such that $\tilde{\phi}'(x; b, \mu) = 0$. If $b = 1$ and $\mu < 0$, the function $\tilde{\phi}(x; 1, \mu)$ is decreasing and reaches its infimum $-\frac{\mu}{2}$ at $x = 1$.

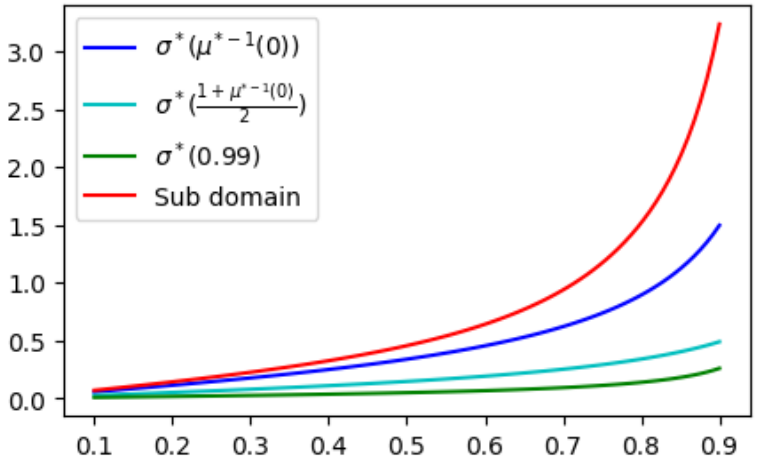

Figure 4.2: Comparison between the Vanishing Upward SVI arbitrage-free sub-domain and domain as functions of b.

Numerical illustration of Proposition 4.2.

Figure 4.2 compares the no arbitrage sub-domain for σ reported in Lemma 4.4 with the full no arbitrage domain found in Proposition 4.2, as functions of b. The red line represents the sub-domain. The blue line corresponds to the full domain for x^* equal to the zero of the μ^*-function. The reason why we take this point, which depends on b, is that the sub-domain is defined for $\mu \leq 0$, so x^* must be greater than the zero of $\mu^*(\cdot)$. Then, the green line shows the full domain for $x^* = 0.99$ (for the chosen values of b, 0.99 is always greater than the zero of $\mu^*(\cdot)$) and the light blue line is the full domain for an intermediate point between the blue and the green ones. As expected, the red line is above all the other lines.

4.3.4 Vanishing (Downward) SVI

The Vanishing Downward SVI is the sub-SVI obtained by setting $\rho = -1$ and $a = 0$. The corresponding SVI formula becomes

$$\mathrm{SVI}(k; 0, b, -1, m, \sigma) = b(-k + m + \sqrt{(k-m)^2 + \sigma^2}).$$

We plot in Figure 4.3 a Vanishing Downward SVI with $b = \frac{1}{2}$, $m = 1$ and $\sigma = 1$.

From Proposition 4.1, we know that the Vanishing Downward SVI is arbitrage-free iff $\mathrm{SVI}(k; 0, b, 1, -m, \sigma)$ is arbitrage-free, and this corresponds to a Vanishing Upward SVI. This means that the previous results still hold for the Vanishing Downward SVI setting $m(\rho = -1) = -m(\rho = 1)$. In particular, we redefine the quantity σ^* as

$$\sigma^*(x) := -\frac{4b\sqrt{1-x^2}(1-x-2x^2)}{4\big(2-x+\mu^*(x)\sqrt{1-x^2}\big)^2 - b^2(1+x)^2}$$

and Proposition 4.2 becomes:

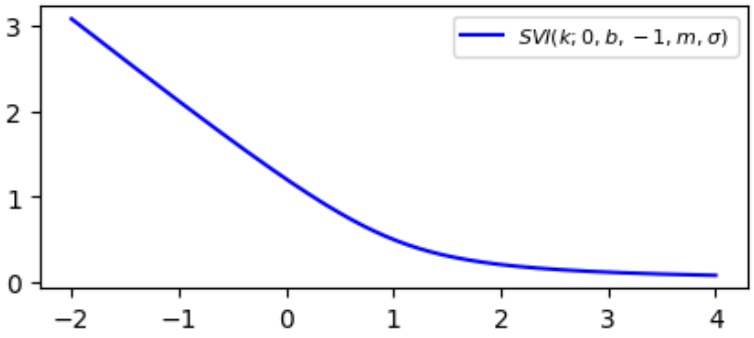

Figure 4.3: Vanishing Downward SVI with $b = \frac{1}{2}$, $m = 1$ and $\sigma = 1$.

Proposition 4.3 (Fully explicit no arbitrage domain for the Vanishing Downward SVI)**.** *A Vanishing Downward SVI with $b = 1$ is arbitrage-free iff $\mu > 0$ and $\sigma \geq \frac{\mu}{2}$.*

A Vanishing Downward SVI with $0 < b < 1$ is arbitrage-free iff it can be parametrized as

$$SVI(k) = b\sigma\left(-\frac{k}{\sigma} - \mu^*(x) + \sqrt{\left(\frac{k}{\sigma} + \mu^*(x)\right)^2 + 1}\right) \tag{4.11}$$

where $\frac{2+b}{4-b} < x < 1$ and $\sigma \geq \sigma^*(x)$.

4.4 Extremal Decorrelated SVI

The coefficient ρ in SVI should correspond to the leverage factor in stochastic volatility models, i.e. the stock/vol returns correlation. In particular, when it is zero, the volatility is independent from the stock which leads to symmetric smiles (see e.g. [61, 65]). In terms of SVI parameters this means that m should also be zero.

This is not automatically enforced in SVI, where ρ and m are distinct parameters. Therefore we call:

- *Decorrelated SVI* sets of SVI parameters where $\rho = 0$;
- *Extremal Decorrelated SVI* a Decorrelated SVI with $b = 2$;
- *Symmetric SVI* a Decorrelated SVI with $m = 0$.

The two latter families intersect into an SVI with $b = 2$ and $\rho = m = 0$.

The *Extremal Decorrelated* SVI is the sub-SVI obtained by setting $b = 2$ and $\rho = 0$. The corresponding SVI formula becomes

$$\text{SVI}(k; a, 2, 0, m, \sigma) = a + 2\sqrt{(k-m)^2 + \sigma^2}.$$

With our notations, $N(l) = \gamma + \sqrt{l^2 + 1}$.

We plot in Figure 4.4 an Extremal Decorrelated SVI with $a = 8$, $m = 2$ and $\sigma = 2$.

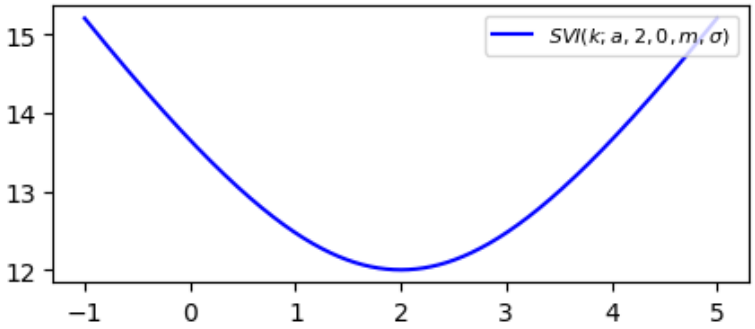

Figure 4.4: Extremal Decorrelated SVI with $a = 8$, $m = 2$ and $\sigma = 2$.

4.4.1 The Fukasawa conditions

The Roger Lee conditions on b and ρ are satisfied. In subsection 5.2.5 of [53] we prove that the Fukasawa conditions are satisfied iff $\gamma > \tilde{F}(2,0) = 0$ and $\mu \in I_{\gamma,2,0} =]-\gamma, \gamma[$. Setting $\mu = q\gamma$ we obtain the following.

Lemma 4.5 (Fukasawa conditions for the Extremal Decorrelated SVI)**.** *An Extremal Decorrelated SVI satisfies the Fukasawa conditions iff it can be parametrized by*

$$SVI(k) = 2\sigma\left(\gamma + \sqrt{\left(\frac{k}{\sigma} - q\gamma\right)^2 + 1}\right) \tag{4.12}$$

with $\gamma > 0$, $q \in]-1, 1[$ and $\sigma > 0$.

4.4.2 The condition on σ

We will prove the following characterization of the no Butterfly arbitrage domain for the Extremal Decorrelated SVI:

Proposition 4.4 (Fully explicit no arbitrage domain for the Extremal Decorrelated SVI)**.** *An Extremal Decorrelated SVI is arbitrage-free iff it can be parametrized as*

$$SVI(k) = 2\sigma\left(\gamma + \sqrt{\left(\frac{k}{\sigma} - q\gamma\right)^2 + 1}\right) \tag{4.13}$$

with $\gamma > 0$, $q \in]-1, 1[$ and $\sigma \geq \frac{1}{\gamma(1-|q|)}$.

Proof of Proposition 4.4

Setting $\mu = q\gamma$ with $q \in]-1, 1[$, the function G_1 is

$$G_1(l) = \frac{-2q\gamma l^3 + l^2((3+q^2)\gamma^2 + 3) - 4q\gamma l + 4(1+\gamma^2) + 2\gamma\sqrt{l^2+1}(l^2 - 2q\gamma l + 4)}{4(l^2+1)(\gamma + \sqrt{l^2+1})^2}$$

and since $\rho = 0$, $g_2(l) = -\frac{(l^2-2)\sqrt{l^2+1}-2\gamma}{2(l^2+1)^{\frac{3}{2}}(\gamma+\sqrt{l^2+1})}$ is symmetric, so $l_1 = -l_2$. In particular, working with $l \geq 0$, to find its roots it is sufficient to solve a third degree equation in $\sqrt{l^2+1}$.

We operate the substitution $y = \sqrt{l^2+1}$ so that

$$G_1(l(y)) = \frac{2\gamma y^3 + ((3+q^2)\gamma^2+3)y^2 + 6\gamma y + 1 + (1-q^2)\gamma^2 - 2q\gamma l(y)(y^2+2\gamma y+1)}{4y^2(y+\gamma)^2},$$

$$g_2(l(y)) = -\frac{y^3-3y-2\gamma}{2y^3(y+\gamma)}.$$

From Lemma 4.2, it follows that the function $f = -\frac{bg_2}{2G_1}$ is such that for $l > l_2$ and q negative, $f(l) < f(-l)$ while for q positive, $f(l) > f(-l)$ and for $q = 0$ it is symmetric. So if we want to find the supremum of f in $]-\infty, l_1[\cup]l_2, \infty[$, it is enough to look at $]-\infty, l_1[$ for $q < 0$ and $]l_2, \infty[$ for $q \geq 0$. Note also that $f(l;\gamma,q) = f(-l;\gamma,-q)$.

Suppose $q \geq 0$ and $l > l_2$. The function f has its limit at ∞ equal to $\frac{1}{\gamma(1-q)}$. We want to prove that this is also the supremum of f. This would be true iff $f(l) < \frac{1}{\gamma(1-q)}$ for every l or equivalently iff $-g_2(l)\gamma(1-q) < G_1(l)$. Using the change of variable, this reads also

$$\frac{y^3-3y-2\gamma}{y^3(y+\gamma)}\gamma(1-q) < \Big[2\gamma y^3 + ((3+q^2)\gamma^2+3)y^2 + 6\gamma y + 1 + (1-q^2)\gamma^2 + \\ - 2q\gamma\sqrt{y^2-1}(y^2+2\gamma y+1)\Big]/(2y^2(y+\gamma)^2).$$

Multiplying by $2y^3(y+\gamma)^2$ and simplifying, one gets

$$2\gamma q y^4 + (\gamma^2(1+q)^2+3)y^3 + 6\gamma(2-q)y^2 + (\gamma^2(1-q)(q+11)+1)y + \\ 4\gamma^3(1-q) - 2\gamma q y\sqrt{y^2-1}(y^2+2\gamma y+1) > 0.$$

Since $\sqrt{y^2-1} < y$, the LHS is greater than

$$(\gamma^2(1-q)^2+3)y^3 + 4\gamma(3-2q)y^2 + (\gamma^2(1-q)(q+11)+1)y + 4\gamma^3(1-q).$$

Each of the coefficients of this polynomial is positive and $y > 0$ so the whole polynomial is positive and f reaches its supremum at ∞.

In the case $q < 0$, the conclusion follows immediately from *Proposition* 4.1.

4.5 Symmetric SVI

The Symmetric SVI is the sub-SVI obtained by setting $\rho = 0$ and $m = 0$. The corresponding SVI formula becomes

$$\mathrm{SVI}(k; a, b, 0, 0, \sigma) = a + b\sqrt{k^2+\sigma^2}.$$

With our notations, $N(l) = \gamma + \sqrt{l^2+1}$.

We plot in Figure 4.5 a Symmetric SVI with $a = 0.64$, $b = 1.6$ and $\sigma = 0.4$.

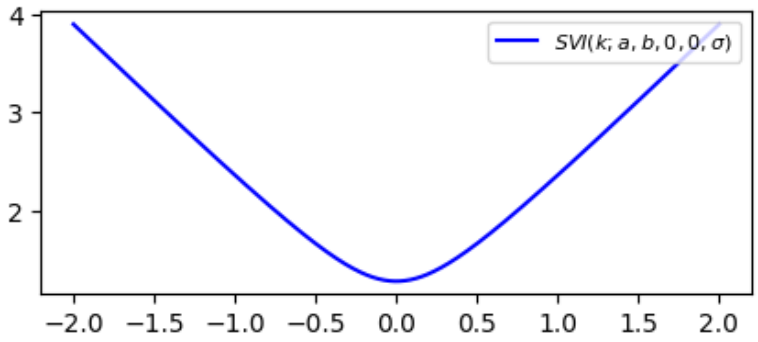

Figure 4.5: Symmetric SVI with $a = 0.64$, $b = 1.6$ and $\sigma = 0.4$.

4.5.1 The Fukasawa conditions

The Roger Lee conditions imply $0 < b \leq 2$. In Appendix B of [53] we prove that the Fukasawa conditions are satisfied iff $\gamma > \tilde{F}(b,0) = g_{-(b,0)}\left(-\frac{6b}{\sqrt{b^4-20b^2+64}}\right) = -\frac{(b^2+32)\sqrt{4-b^2}}{(16-b^2)^{\frac{3}{2}}}$. Since $\rho = 0$, the interval for μ is symmetric and being $\mu = 0$, the Fukasawa condition on μ is automatically satisfied.

Lemma 4.6 (Fukasawa conditions for the Symmetric SVI)**.** *A Symmetric SVI with $0 < b \leq 2$ satisfies the Fukasawa conditions iff $\gamma > \tilde{F}(b,0) = -\frac{(b^2+32)\sqrt{4-b^2}}{(16-b^2)^{\frac{3}{2}}}$.*

For notation simplicity, from now on we drop the dependence of $\tilde{F}$ to $\rho = 0$. For every value of b, we have $\tilde{F}(b) \in]-1, 0]$ and in particular $\tilde{F}(2) = 0$. We can extend the definition of $\tilde{F}$ to the point $b = 0$, setting $\tilde{F}(0) = -1$. Furthermore, $\frac{d\tilde{F}}{db}(b) = \frac{108b^3}{(16-b^2)^{\frac{5}{2}}\sqrt{4-b^2}} > 0$ for positive b. So the inverse function $\tilde{F}^{-1} : [-1, 0] \to [0, 2]$ is well defined. As in Section 4.2.3, we define

$$\tilde{G}(\gamma) = \begin{cases} \tilde{F}^{-1}(\gamma) & \text{if } \gamma \in]-1, 0], \\ 2 & \text{if } \gamma > 0. \end{cases} \tag{4.14}$$

It can be shown, working out the calculation of the roots of a 3rd degree polynomial, that the explicit formula of $\tilde{G}$ in the case $\gamma \leq 0$ is

$$\tilde{G}(\gamma) = 2\sqrt{\frac{6\sqrt{8\gamma^2+1}\cos\left(\frac{1}{3}\arccos\left(-\frac{8\gamma^4+20\gamma^2-1}{(8\gamma^2+1)^{\frac{3}{2}}}\right)\right) - 4\gamma^2 - 5}{1-\gamma^2}}. \tag{4.15}$$

4.5.2 The condition on σ

Before enunciating the main Proposition of this chapter, we need to introduce some notation. In this section, for positive ls we will operate a change of variable setting

$z = \frac{1}{\sqrt{l^2+1}}$, whose domain is $]0,1]$. Then we define the functions $J_1(z) := G_1(l(z))$, $j_2(z) := g_2(l(z))$, $\eta(z) := h(l(z))$ and $j(z) := g(l(z))$. In particular,

$$\begin{aligned} j_2(z;\gamma) &= z\frac{2\gamma z^3 + 3z^2 - 1}{2(\gamma z + 1)}, \\ \eta(z;\gamma) &= 1 - \frac{1 - z^2}{2(1 + \gamma z)}, \\ j(z) &= \frac{1 - z^2}{4}. \end{aligned} \tag{4.16}$$

In the following, the symbol $'$ indicates the derivative with respect to z if not differently specified.

We also define the functions

$$\gamma^*(u) := u\sqrt{6u^3 + 15u^2 + 14u + 6 + (1+u)^2\sqrt{3(12u^2 + 12u + 11)}}, \tag{4.17}$$

$$b^*(z,\gamma) := \sqrt{\frac{\eta(z)(\eta(z)j_2'(z) - 2\eta'(z)j_2(z))}{j(z)(j(z)j_2'(z) - 2j'(z)j_2(z))}}, \tag{4.18}$$

$$\sigma^*(z,\gamma) := -\frac{b^*(z,\gamma)j_2(z)}{2(\eta(z)^2 - b^*(z,\gamma)^2 j(z)^2)}. \tag{4.19}$$

The final result of this subsection will be:

Proposition 4.5 (Fully explicit no arbitrage domain for the Symmetric SVI)**.** *An arbitrage-free Symmetric SVI must have $\gamma > -1$.*

A Symmetric SVI with $b = 2$ is arbitrage-free iff $\gamma > 0$ and $\sigma \geq \frac{1}{\gamma}$.

Call $z^*(\gamma^*(u), 0) = \frac{u}{\gamma^*(u)}$ and $z^*(\gamma, \tilde{G}(\gamma)) = \frac{\sqrt{(4-\tilde{G}(\gamma)^2)(16-\tilde{G}(\gamma)^2)}}{\tilde{G}(\gamma)^2+8}$.

- *If $\gamma = -\sqrt{\frac{9+5\sqrt{3}}{18}}$, the Symmetric SVI is arbitrage-free iff $b < 2\sqrt{3\sqrt{3}-5}$ and $\sigma \geq -\frac{bj_2(\hat{z})}{2(\eta(\hat{z})^2 - b^2 j(\hat{z})^2)}$ where $\hat{z} = \sqrt{\frac{3-\sqrt{3}}{2}}$.*
- *If $\gamma \neq -\sqrt{\frac{9+5\sqrt{3}}{18}}$ and $b < 2$, a Symmetric SVI is arbitrage-free iff it can be parametrized as*

$$SVI(k) = \sigma b^*(z, \gamma^*(u))\left(\gamma^*(u) + \sqrt{\left(\frac{k}{\sigma}\right)^2 + 1}\right) \tag{4.20}$$

where

  - $u > -1$,
  - $z \in]z^*(\gamma^*(u), 0), z^*(\gamma^*(u), \tilde{G}(\gamma^*(u)))[$ *if* $\gamma^*(u) < -\sqrt{\frac{9+5\sqrt{3}}{18}}$ *or* $z \in]z^*(\gamma^*(u), \tilde{G}(\gamma^*(u))), z^*(\gamma^*(u), 0)[$ *if* $\gamma^*(u) > -\sqrt{\frac{9+5\sqrt{3}}{18}}$,
  - $\sigma \geq \sigma^*(z, \gamma^*(u))$.

Proof of Proposition 4.5

When $b = 2$, we are in the case of the Extremal Decorrelated SVI with $q = 0$, so the arbitrage conditions are satisfied iff $\gamma > 0$ and $\sigma \geq \frac{1}{\gamma}$.

From now on we consider $b < 2$. Since $\rho = 0$, the g_2 function is symmetric with respect to l and since $\mu = 0$, the G_1 function is symmetric too. Then $\tilde{f} = -\frac{G_1}{g_2}$ is symmetric as a function of l and we consider $l \geq 0$. Recall that we are interested in the value of $\sigma^* = \sup_{l>l_2} \frac{b}{2\tilde{f}(l)}$.

The function $\tilde{\phi}(z) := \tilde{f}(l(z))$ goes to ∞ at 0 and at the non trivial zero of j_2, which we call $z_2(\gamma)$. Then $\tilde{\phi}$ has at least one point of minimum in $]0, z_2[$.

Study of j_2 To find z_2, we should solve $p_1(z) := 2\gamma z^3 + 3z^2 - 1 = 0$. It can be shown that for γ negative, the polynomial has three roots but only the second one lies in the interval $]0, 1[$. If γ is null, then $z_2(0) = \frac{1}{\sqrt{3}}$. If γ is positive but not greater than 1, the polynomial has three roots and only the third lies in the interval $]0, 1[$. Finally, if $\gamma > 1$, there is only one root. Using the trigonometric notation, $z_2(\gamma)$ is

$$z_2(\gamma) = \begin{cases} -\frac{1}{\gamma}\cos\left(\frac{1}{3}\arccos(1-2\gamma^2) - \frac{2\pi}{3}\right) - \frac{1}{2\gamma} & \text{if } \gamma < 0, \\ \frac{1}{\sqrt{3}} & \text{if } \gamma = 0, \\ \frac{1}{\gamma}\cos\left(\frac{1}{3}\arccos(2\gamma^2-1)\right) - \frac{1}{2\gamma} & \text{if } 0 < \gamma \leq 1, \\ \frac{1}{\gamma}\cosh\left(\frac{1}{3}\operatorname{arccosh}(2\gamma^2-1)\right) - \frac{1}{2\gamma} & \text{if } \gamma > 1. \end{cases}$$

Let us show that j_2 has a single critical point z_{m_2} where it achieves its minimum. It holds

$$j_2'(z) = \frac{6\gamma^2 z^4 + 14\gamma z^3 + 9z^2 - 1}{2(\gamma z + 1)^2},$$
$$j_2''(z) = \frac{6\gamma^3 z^4 + 19\gamma^2 z^3 + 21\gamma z^2 + 9z + \gamma}{(\gamma z + 1)^3}.$$

Since $j_2'(0) = -\frac{1}{2}$ and $j_2'(1) = \frac{3\gamma+4}{\gamma+1}$, which is positive having $\gamma > -1$, then j_2 has at least one critical point in $]0, 1[$.

- The function j_2'' is positive if $\gamma \geq 0$ so j_2 has exactly one critical point in this case, which is a point of minimum.
- Consider the case $\gamma < 0$. The polynomial $p_2(z) := 6\gamma^2 z^4 + 14\gamma z^3 + 9z^2 - 1$ has derivative $p_2'(z) = 2z(12\gamma^2 z^2 + 21\gamma z + 9)$, so its three critical points are 0, $-\frac{3}{4\gamma}$ and $-\frac{1}{\gamma}$. Since $\gamma > -1$, the third critical point, which is a point of minimum, is greater than 1 and, since 0 is the first critical point and since $-1 = p_2(0) < 0 < p_2(1) = 6\gamma^2 + 14\gamma + 8 = 2(\gamma+1)(\gamma+4)$, p_2 has exactly one zero in $]0, 1[$. So again j_2 has only one critical point in this interval.

Because j_2 is negative before z_2 then $j_2'(z_2) > 0$, furthermore $j_2'(0) < 0$ so the critical point z_{m_2} of j_2 (a point of minimum) lies in $]0, z_2[$.

We have already seen that if $\gamma \geq 0$, then j_2 is strictly convex in $]0, 1[$.

We can moreover show that j_2 has a single inflection point if γ is negative. In such case, call the numerator of j_2'' as $p_3(z) := 6\gamma^3 z^4 + 19\gamma^2 z^3 + 21\gamma z^2 + 9z + \gamma$. It holds $p_3(0) = \gamma < 0$ and $p_3(1) = 6\gamma^3 + 19\gamma^2 + 22\gamma + 9$, which is a third degree polynomial in γ with only root equal to -1 and a positive leading order term, so it is positive for $\gamma > -1$. Then p_3 has at least one zero in $]0, 1[$. Moreover, $p_3'(z) = 3(\gamma z + 1)^2(8\gamma z + 3)$, whose zeros are $-\frac{3}{8\gamma}$ and $-\frac{1}{\gamma} > 1$. The former is a point of maximum for p_3 while the latter is an inflection point. Gathering all the informations, it follows that p_3 has exactly one zero in $]0, 1[$. Since $j_2'' > 0$ at the point of minimum of j_2 and $j_2''(0) < 0$, the only inflection point $z_{i_2}(\gamma)$ of j_2 lies in $]0, z_{m_2}[$.

Study of J_1 It holds $J_1(z) = \eta(z)^2 - b^2 j(z)^2 = \left(1 - \frac{1-z^2}{2(1+\gamma z)}\right)^2 - b^2 \frac{1-z^2}{16}$.

We write here a general result which holds for every SVI.

Lemma 4.7. *Under the Fukasawa conditions, G_1 is strictly decreasing at the second zero of g_2.*

Proof. Since $G_1' = G_{1+}' G_{1-} + G_{1+} G_{1-}'$ and $G_{1+}' < G_{1-}'$ for $l > l^*$, then it is enough to prove $G_{1-}'(l_2) < 0$. Remember that $G_{1-}(l_2) = 1 - N'(l_2)\left(\frac{l_2+\mu}{2N} - \frac{b}{4}\right)$. From the formula of g_2, we have that $N''(l_2) = \frac{N'(l_2)^2}{2N(l_2)}$. Then

$$\begin{aligned} G_{1-}'(l_2) &= -\frac{N''(l_2)(l_2+\mu) + N'(l_2)}{2N(l_2)} + \frac{N'(l_2)^2(l_2+\mu)}{2N(l_2)^2} + \frac{bN''(l_2)}{4} \\ &= \frac{N'(l_2)^2(l_2+\mu)}{4N(l_2)^2} + \frac{bN'(l_2)^2}{8N(l_2)} - \frac{N'(l_2)}{2N(l_2)}. \end{aligned}$$

For the absence of arbitrage, $\mu < \inf_{l>l^*} L_+(l) = \inf_{l>l^*}\left(\frac{2N(l)}{N'(l)} - \frac{bN(l)}{2} - l\right)$ so this holds in particular in $l_2 > l^*$. Substituting with the formula of L_+ we obtain

$$G_{1-}'(l_2) < \frac{N'(l_2)^2}{4N(l_2)^2}\left(l_2 + \frac{2N(l_2)}{N'(l_2)} - \frac{bN(l_2)}{2} - l_2\right) + \frac{bN'(l_2)^2}{8N(l_2)} - \frac{N'(l_2)}{2N(l_2)} = 0.$$

□

Observe that

$$\begin{aligned} J_1'(z) &= \frac{\gamma z^4(b^2\gamma^2+4) + z^3(3b^2\gamma^2+8\gamma^2+8) + 3\gamma z^2(b^2+8) + z(b^2+8\gamma^2+8) + 4\gamma}{8(\gamma z+1)^3}, \\ J_1''(z) &= \frac{b^2}{8} + \frac{\gamma^2 z^4 + 4\gamma z^3 + 6z^2 + (4\gamma z+1)(2-\gamma^2)}{2(\gamma z+1)^4}, \\ J_1'''(z) &= \frac{6z(\gamma^2-1)^2}{(\gamma z+1)^5}. \end{aligned}$$

So:

- For $\gamma \geq 0$, J_1' is always greater than 0 so J_1 is increasing and has its minimum in 0.
- If $\gamma < 0$, J_1' is negative for $z = 0$ and we know that $G_1'(l_2) < 0$ so $J_1'(z_2) > 0$. Then J_1 has at least a point of minimum in $]0, z_2[$. Also, the minimum of J_1'' is reached at 0 and it is $\frac{b^2-4\gamma^2+8}{8}$. This quantity is positive iff $\gamma^2 < \frac{b^2+8}{4}$, so for example for every $\gamma < 0$ because γ is greater than -1. In such case J_1 is always convex and it has exactly one critical point z_{m_1}, which is a point of minimum, in $]0, z_2[$.

Note also that when γ goes to its Fukasawa limit, $J_{1+}(z) = 1 - \frac{1-z^2}{2(\gamma z+1)} - \frac{b\sqrt{1-z^2}}{4}$ goes to 0 in $z_+^*(b) := \frac{\sqrt{(4-b^2)(16-b^2)}}{b^2+8}$, which is also a point of minimum for J_1. It is easy to shown that if $\gamma > -\frac{(b^4-38b^2+64)(b^2+8)}{(4-b^2)^{\frac{3}{2}}(16-b^2)^{\frac{3}{2}}} := M(b)$, then $j_2(z_+^*(b))) > 0$, so $z_+^*(b) > z_2(\gamma)$, otherwise $z_+^*(b) \leq z_2(\gamma)$. Note that $M(b) \geq \tilde{F}(b)$ (the equality holding iff $b = 0$), so when γ goes to $\tilde{F}(b)$, J_1 reaches its minimum before $z_2(\gamma)$ (or at $z_2(\gamma)$ if $b = 0$).

Then J_1 has in any case one point of minimum: in the case $\gamma \geq 0$, this is 0, while when $\gamma < 0$, it is $z_{m_1}(\gamma, b)$. Also, J_1 is convex if $\gamma^2 \leq \frac{b^2+8}{4}$ (in particular $\gamma < 0$ given that γ should be larger than -1) while it has a unique inflection point $z_{i_1}(\gamma, b)$ if $\gamma^2 > \frac{b^2+8}{4}$.

Uniqueness of the critical point of $\tilde{\phi}$ for the Symmetric SVI From the formula for $\tilde{\phi}' = \frac{J_1 j_2' - J_1' j_2}{j_2^2}$, given that $J_1 > 0$ and $j_2 < 0$ we have that:

- for $\gamma < 0$: all the critical points of $\tilde{\phi}$ must lie in $[z_{m_1}, z_{m_2}] \cup [z_{m_2}, z_2 \wedge z_{m_1})$;
- for $\gamma \geq 0$: all the critical points of $\tilde{\phi}$ must lie in $]0, z_{m_2}[$ since $J_1' j_2$ is negative.

Considering now the second derivative of $\tilde{\phi}$, so $\tilde{\phi}'' = \frac{J_1 j_2'' - J_1'' j_2}{j_2^2} - 2\tilde{\phi}' \frac{j_2'}{j_2}$ and evaluating it at a point of maximum of $\tilde{\phi}$, the second term becomes null and necessarily $\tilde{\phi}'' \leq 0$. Then it must hold $j_2'' J_1'' \leq 0$ (the equality holding when both factors are 0). In particular:

1. for $\gamma < 0$: if $\tilde{\phi}$ has a point of maximum, it lies in $[z_{m_1}, z_{i_2}(\gamma)[$ because in such case $J_1'' > 0$, $j_2'' < 0$ in $[0, z_{i_2}(\gamma)[$;
2. for $0 \leq \gamma \leq \sqrt{\frac{b^2+8}{4}}$: $\tilde{\phi}$ cannot have a point of maximum, because J_1 and j_2 are strictly convex, so it has one point of minimum which lies in $]0, z_{m_2}[$;
3. for $\gamma > \sqrt{\frac{b^2+8}{4}}$: J_1 has one inflection point z_{i_1} and if $\tilde{\phi}$ has a point of maximum, it lies in $]0, z_{m_2} \wedge z_{i_1}[$.

We consider the first case and show that $z_{m_1} > z_{i_2}(\gamma)$ or equivalently that $J_1'(z_{i_2}) < 0$, so that there is no point of maximum. The quantity $z_{i_2}(\gamma)$ does not depend on b so $\partial_{b^2}(J_1'(z_{i_2})) = \partial_{b^2} J_1'(z_{i_2}) = \frac{z_{i_2}}{8} > 0$. We can then check that for every $-1 < \gamma < 0$ we have indeed $J_1'(z_{i_2}; b = 2) < 0$ in Figure 4.6.

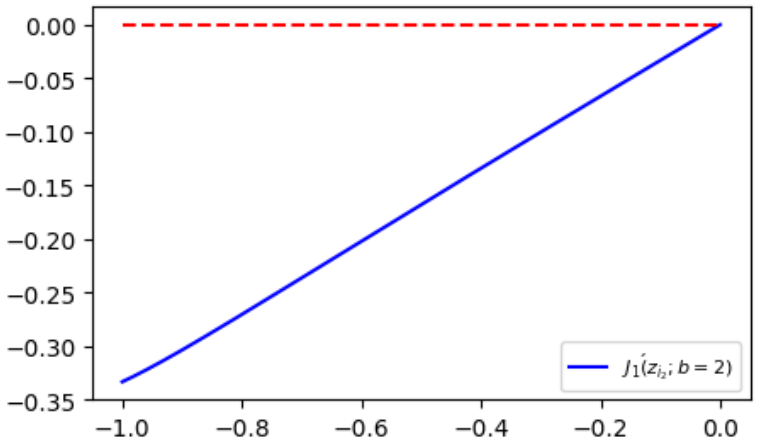

Figure 4.6: $J_1'(z_{i_2}; b = 2)$ as a function of γ.

Let us now consider the last case. Since $\gamma > \sqrt{\frac{b^2+8}{4}}$, then in particular $\gamma^2 > 2$. Suppose we are at a critical point of $\tilde{\phi}$, then $J_1 = \frac{J_1'}{j_2'} j_2$ and its second derivative evaluated at this point is $\tilde{\phi}'' = \frac{J_1 j_2'' - J_1'' j_2}{j_2^2} = \frac{J_1' j_2'' - J_1'' j_2'}{j_2 j_2'}$. In the set $]0, z_{m_2} \wedge z_{i_1}[$, the denominator is always positive. The numerator times $16(\gamma z + 1)^6$ is equal to

$$
\begin{aligned}
& b^2\big(6\gamma^6 z^8 + 36\gamma^5 z^7 + 91\gamma^4 z^6 + 126\gamma^3 z^5 + 3\gamma^2(\gamma^2 + 34)z^4 + 2\gamma(5\gamma^2 + 23)z^3 + \\
& + 3(4\gamma^2 + 3)z^2 + 6\gamma z + 1\big) + 24\gamma^4 z^8 + 96\gamma^3(\gamma^2 + 1)z^7 + 4\gamma^2 z^6(148\gamma^2 + 17) + \\
& + 24\gamma z^5(8\gamma^4 + 48\gamma^2 - 3) + 12z^4(50\gamma^4 + 81\gamma^2 - 6) + 16\gamma z^3(44\gamma^2 + 25) + \\
& + 12z^2(33\gamma^2 + 8) + 120\gamma z + 4\gamma^2 + 8.
\end{aligned}
$$

All coefficients of z are positive quantities, indeed the only minus involved are in the expressions $8\gamma^4+48\gamma^2-3 > 8\times4+48\times2-3 > 0$ and $50\gamma^4+81\gamma^2-6 > 50\times4+81\times2-6 > 0$. Then $\tilde{\phi}''$ is positive and $\tilde{\phi}$ cannot have a point of maximum even in this case.

We denote from now on the well-defined single point of minimum of $\tilde{\phi}$ with $z^*(\gamma, b)$.

Limits of z^* with respect to γ When γ goes to ∞, the ratio between the numerator of $\tilde{\phi}'$ and γ^4 goes to $2z^6(48 + b^2(z^2 - 3))$, which is null for $z = 0$, so $z^*(\infty, b) = 0$.

On the other side, let γ go to $\tilde{F}(b)$. Then, from the above study of J_1, the minimum point of J_1 is $z_+^*(b) = \frac{\sqrt{(4-b^2)(16-b^2)}}{b^2+8} < z_2(\tilde{F}(b))$ and it coincides with the minimum point $z^*(\tilde{F}(b), b)$ of $\tilde{\phi}$, at which $\tilde{\phi}$ is null.

If $b = 0$, it still holds $z^*(\infty, 0) = 0$ and it is easy to show that at the point $z^*(0) = z_2(\tilde{F}(0)) = 1$ the function $\tilde{\phi}$ with $\gamma = \tilde{F}(b)$ vanishes.

The b^* approach for the Symmetric SVI Let $\gamma \in]-1, \infty[$ be fixed. With reference to the definition of $\tilde{G}(\gamma)$ in Equation (4.14), we want to prove that for every $b \in]0, \tilde{G}(\gamma)[$ the corresponding $z^*(\gamma, b)$ lies in $]z^*(\gamma, 0), z^*(\gamma, \tilde{G}(\gamma))[$ (or in $]z^*(\gamma, \tilde{G}(\gamma)), z^*(\gamma, 0)[$).

Remark 4.1. *Note that if $\gamma > 0$, b could actually attain the value $\tilde{G}(\gamma) = 2$ but we have already seen in the Extremal Decorrelated scenario that in such case the function $\tilde{f}$ is increasing and has its minimum (which is not necessarily a critical point) in $z = 0$. So in the following we will always consider $b < \tilde{G}(\gamma)$.*

Remember the definition of η and j in Equation (4.16) and the fact that we use $'$ to indicate the derivative with respect to z.

The proof of the uniqueness of the critical point of $\tilde{\phi}$ still holds for $b = 0$ and $b = \tilde{G}(\gamma)$, and we proved $z^*(\gamma, \tilde{G}(\gamma)) = z^*_+(G(\gamma))$. In particular, the single point of minimum $z^*(\gamma, b)$ defined above of $\tilde{\phi}$ lies in $]0, z_2[$.

In a following paragraph we show that the only solution to $p(z) = q(z) = 0$ is for $\hat{\gamma} = -\sqrt{\frac{9+5\sqrt{3}}{18}}$ at $\hat{z} = \sqrt{\frac{3-\sqrt{3}}{2}}$. Take $\gamma \neq \hat{\gamma}$. With reference to the observations in Section 4.2.3, we can conclude that the set $Z(\gamma)$ of the critical points of $\tilde{\phi}$ is equal to the interval with extrema the only zero of $p = \eta j_2' - 2\eta' j_2$, denoted $z^*(\gamma, 0)$, and the point $z^*(\gamma, \tilde{G}(\gamma))$.

For $\gamma = \hat{\gamma}$, the minimum point of $\tilde{\phi}$ does not depend on b and it is $z^*(\hat{\gamma}, b) = \hat{z}$. In such case, $\tilde{G}(\hat{\gamma}) = 2\sqrt{3\sqrt{3} - 5}$. In the following we see what happens for the other values of γ.

It holds $q(z) = -\frac{2z^4\gamma^2(z^2-3)+4z^3\gamma(z^2-3)+3z^4-8z^2+1}{8\sqrt{1-z^2}(\gamma z+1)^2}$. Suppose $z_q(\gamma)$ is a zero of q, then solving the equation $q(z_q) = 0$ in γ, the only solution greater than -1 is $\gamma = \gamma_q(z_q) := \frac{1-z_q^2}{z_q^2\sqrt{2(3-z_q^2)}} - \frac{1}{z_q}$. It can be shown that $\gamma_q'(z) = \frac{1}{z^2}\left(1 - \frac{z^4-3z^2+6}{z\sqrt{2}(3-z^2)^{\frac{3}{2}}}\right) < 0$ for every $z \in]0, 1[$. This means that γ_q is invertible from $]0, 1[$ to $]-1, \infty[$ and that for fixed γ, q has a unique zero. Observe that $\gamma_q(\hat{z}) = \hat{\gamma}$. Since $q(0) < 0$, on the left of its zero $z_q(\gamma)$, q is negative while on the right it is positive. Also, if $\gamma < \hat{\gamma}$, then $z_q(\gamma) > \hat{z}$. We now look at $p(z) = \frac{2\gamma^2z^6+12\gamma^3z^5+3z^4(10\gamma^2-1)+28\gamma z^3+12z^2-1}{4(\gamma z+1)^3}$. Substituting γ with $\gamma_q(z)$ in the last expression, we find $p\big(z; \gamma = \gamma_q(z)\big) = -\frac{z^2\left(2z^4-3z^2-3+z\sqrt{2}(3-z^2)^{\frac{3}{2}}\right)}{1-z^2}$. We call $r(z) := 2z^4 - 3z^2 - 3 + z\sqrt{2}(3-z^2)^{\frac{3}{2}}$, then $r(\hat{z}) = r(1) = 0$ and $r(0) = -3 < 0$. Furthermore $r'(z) = -(4z^2-3)(\sqrt{2(3-z^2)}-2z)$, where the second factor is positive for $z < 1$, so r is increasing in $\left[0, \frac{\sqrt{3}}{2}\right[\ni \hat{z}$ and decreasing in $\left]\frac{\sqrt{3}}{2}, 1\right[$, with unique zeros at $\hat{z}$ and 1. This means that for the previous choice of $\gamma < \hat{\gamma}$, $p(z_q(\gamma)) < 0$, so $z^*(\gamma, 0) > z_q(\gamma)$ because $z^*(\gamma, 0)$ is the unique zero of p. Remember that q is positive in $]z_q(\gamma), z^*(\gamma, 0)[$. Since b^* defined in Equation (4.5) lives where p and q have the same sign and since the set $Z(\gamma)$ is an interval, the latter will lie on the right of $z^*(\gamma, 0)$, indeed on the immediate left we have seen that p is negative while q is positive. The reasoning is similar for $\gamma > \hat{\gamma}$, with the conclusion that the previous set is on the left of $z^*(\gamma, 0)$.

Limits of z^* with respect to b When b goes to 0 and γ is not $\hat{\gamma}$, the value of $z^*(\gamma, b)$ is the only solution smaller than $z_2(\gamma)$ to $p(z) = 0$, or

$$2\gamma^2z^6 + 12\gamma^3z^5 + 3z^4(10\gamma^2 - 1) + 28\gamma z^3 + 12z^2 - 1 = 0. \tag{4.21}$$

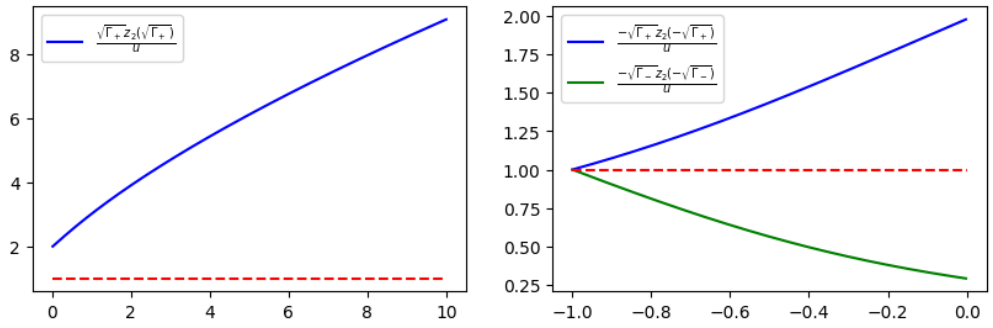

Figure 4.7: Study of $\frac{\pm\sqrt{\Gamma_\pm}z_2(\pm\sqrt{\Gamma_\pm})}{u}$ *as functions of* u.

Since the equation is of sixth degree, the solution cannot be written explicitly, but it can be found by a numeric routine in the interval $]0, z_2(\gamma)[$.

Another cunning way to find $z^*(\gamma, 0)$ is to use again the trick of changing the role between parameters and variables. Firstly, note that if $\gamma = 0$, the only solution to Equation (4.21) is $z = \frac{1}{\sqrt{6+\sqrt{33}}}$. If $\gamma \neq 0$, substitute $z = \frac{u}{\gamma}$ in the above equation, then we look for u solving

$$\gamma^4 - 2u^2(6u^3 + 15u^2 + 14u + 6)\gamma^2 + u^4(3 - 2u^2) = 0. \tag{4.22}$$

The interval where u lives is such that $0 < \frac{u}{\gamma} < z_2(\gamma)$, or equivalently $\frac{\gamma z_2(\gamma)}{u} > 1$. If u is negative, it must hold $0 > u > \gamma z_2(\gamma) > \gamma > -1$, so for sure $u > -1$. Imagine to fix u, then equation Equation (4.22) is quadratic in γ^2 and has four (eventually complex) solutions $\pm\sqrt{\Gamma_\pm}$, with $\Gamma_\pm = u^2\big(6u^3 + 15u^2 + 14u + 6 \pm (1+u)^2\sqrt{3(12u^2 + 12u + 11)}\big)$. The solution γ must have the same sign of u, since $z > 0$. If u is positive, the solution $\sqrt{\Gamma_-}$ is smaller than u, so also $\sqrt{\Gamma_-}z_2(\sqrt{\Gamma_-})$ is smaller than u. Then, the only possible solution could be $\sqrt{\Gamma_+}$. We plot in Figure 4.7 (left) the graph of $\frac{\sqrt{\Gamma_+}z_2(\sqrt{\Gamma_+})}{u}$ to check whether it is greater than 1. If u is negative, we plot in Figure 4.7 (right) the graphs of $\frac{-\sqrt{\Gamma_\pm}z_2(-\sqrt{\Gamma_\pm})}{u}$ and check again whether these quantities are greater than 1.

We can see that for every $u > 0$, the solution $\sqrt{\Gamma_+}$ is admissible and that for $-1 < u < 0$, the only admissible solution is $-\sqrt{\Gamma_+}$. To sum up, for $u > -1$, the solution to Equation (4.22) in terms of γ is Equation (4.17) and it ranges from -1 to ∞. Observe that if $u = 0$, the value of $\frac{u}{\gamma^*(u)}$ is exactly $\frac{1}{\sqrt{6+\sqrt{33}}}$, so the definition is well posed even in this case.

Numerical illustration of the interval for z^*

We report in Figure 4.8 the graphs of $z^*(\gamma, 0)$, which is the only zero of $p(z, \gamma)$ in $\big]0, z_2(\gamma)\big[$, and of $z^*(\gamma, \tilde{G}(\gamma))$.

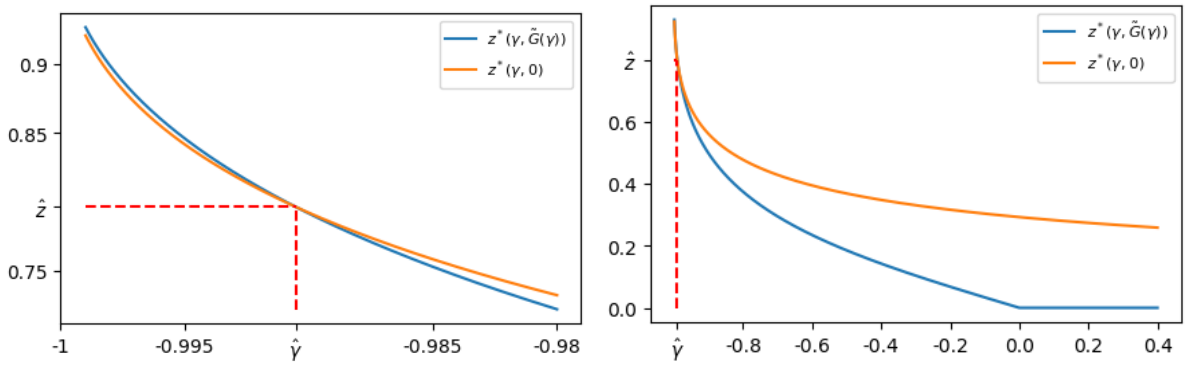

Figure 4.8: $z^*(\gamma, 0)$ and $z^*(\gamma, \tilde{G}(\gamma))$ as functions of γ ranging from -1 to -0.98 (left) and from -1 to 0.4 (right).

4.6 SSVI

An SSVI slice is a sub-SVI classically parametrized as

$$\mathrm{SSVI}(k; \theta, \varphi, \rho) = \frac{\theta}{2}\left(1 + \rho\varphi k + \sqrt{(\varphi k + \rho)^2 + 1 - \rho^2}\right). \tag{4.23}$$

It can be easily checked that the corresponding SVI parameters are $(a, b, m, \rho, \sigma) = \left(\frac{\theta(1-\rho^2)}{2}, \frac{\theta\varphi}{2}, -\frac{\rho}{\varphi}, \rho, \frac{\sqrt{1-\rho^2}}{\varphi}\right)$, so that $\gamma = \sqrt{1-\rho^2}$ and the special property $\mu = -\frac{\rho}{\sqrt{1-\rho^2}} = l^*$ hold, where l^* is the unique critical point of the smile, which is a point of minimum, of N. The corresponding SVI formula becomes

$$\mathrm{SVI}\left(k; b\sigma\sqrt{1-\rho^2}, b, \rho, -\frac{\rho\sigma}{\sqrt{1-\rho^2}}, \sigma\right) = b\sigma\left(\sqrt{1-\rho^2} + \rho\left(\frac{k}{\sigma} + \frac{\rho}{\sqrt{1-\rho^2}}\right) + \sqrt{\left(\frac{k}{\sigma} + \frac{\rho}{\sqrt{1-\rho^2}}\right)^2 + 1}\right).$$

We recall the general notations defined in Section 4.2.1 in the specific case of SSVI:

$$\begin{aligned} N(l; \rho) &= \sqrt{1-\rho^2} + \rho l + \sqrt{l^2+1} \\ N'(l; \rho) &= \rho + \frac{l}{\sqrt{l^2+1}} \\ N''(l) &= \frac{1}{(l^2+1)^{\frac{3}{2}}} \end{aligned}$$

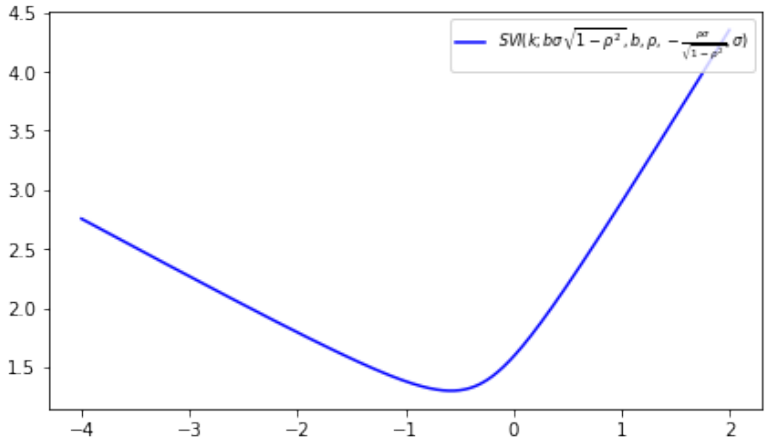

Figure 4.9: SSVI with $b = 1$, $\rho = \frac{1}{2}$ *and* $\sigma = \frac{1}{2}$.

from which we find

$$\begin{aligned} h(l;\rho) &= 1 - \frac{N'(l)}{2N(l)}\Big(l - \frac{\rho}{\sqrt{1-\rho^2}}\Big) = \frac{1}{2}\Big(1 + \frac{1}{\sqrt{1-\rho^2}\sqrt{l^2+1}}\Big) \\ g(l;\rho) &= \frac{N'(l)}{4} = \frac{1}{4}\Big(\rho + \frac{l}{\sqrt{l^2+1}}\Big) \\ g_2(l;\rho) &= N''(l) - \frac{N'(l)^2}{2N(l)} = \frac{1}{(l^2+1)^{\frac{3}{2}}} - \frac{1}{2\big(\sqrt{1-\rho^2} + \rho l + \sqrt{l^2+1}\big)}\Big(\rho + \frac{l}{\sqrt{l^2+1}}\Big)^2 \\ G_1(l;b,\rho) &= (h(l) - bg(l))(h(l) + bg(l)) \end{aligned} \tag{4.24}$$

where the explicit expressions have been computed via simple substitutions and simplifications.

We plot in Figure 4.9 an SSVI with $b = 1$, $\rho = \frac{1}{2}$ and $\sigma = \frac{1}{2}$.

Absence of weak arbitrage in SSVI

In this section we prove that SSVI smiles authomatically satisfy the absence of weak arbitrage condition, i.e. that SSVI smiles always have decreasing functions $d_1(k)$ and $d_2(k)$. To this aim, we use results in Theorem 4.1.

Since $\gamma > 0$, we always have $\gamma > \tilde{F}(b,\rho)$. Thanks to the following lemma, we get that also the condition on μ is verified for $\gamma > 0$.

Lemma 4.8. *Assume* $b(1+|\rho|) \leq 2$ *and* $\gamma > 0$. *Then for every SVI,* $l^* \in I_{\gamma,b,\rho}$.

Proof. Consider the case $\rho \geq 0$, so that $l^* \leq 0$. Note that $N(l) - \gamma = lN'(l) + N''(l)(l^2+1)$. Then the condition $L_-(l) < l^*$ for $l < l^*$ becomes $\frac{\gamma + lN'(l) + N''(l)(l^2+1)}{2N'(l)}(4 + bN'(l)) < l + l^*$ or equivalently $\gamma > \frac{N'(l)}{4+bN'(l)}(2l^* - l(2 + bN'(l))) - N''(l)(l^2+1)$.

The term $-N''(l)(l^2+1)$ is strictly negative and if $2l^* - l(2 + bN'(l)) \geq 0$, also the first term is negative, so that if $\gamma \geq 0$ we automatically have $L_-(l) < l^*$ for every $l < l^*$. When $\rho > 0$ (or $\rho = 0$ and $b < 2$), since $b(1+\rho) \leq 2$, it holds $b(1-\rho) \neq 2$. Because $L_-(-\infty) = L_-(l^{*-}) = -\infty$ and L_- is continuous, it holds $\sup_{l<l^*} L_-(l) = L_-(l_-)$ where $l_- \in]-\infty, l^*[$. Then if we show $L_-(l) < l^*$ for every $l < l^*$, it follows $l^* > \sup_{l<l^*} L_-(l)$. If $\rho = 0$ and $b = 2$, it holds $b(1-\rho) = 2$, so $l_- = -\infty$ and $L_-(l_-) = -\gamma$. In such case, the only annoying situation arises for $\gamma = 0$, because it holds $l^* = 0 = \sup_{l<l^*} L_-(l)$. Taking $\gamma > 0$ this case does not arise.

Consider the quantity $2l^* - l(2 + bN'(l))$. Since $b(1+\rho) \leq 2$, it is greater than $-2\Big(-l^* + l\Big(1 + \frac{N'(l)}{1+\rho}\Big)\Big)$, which is positive iff, substituting and dividing by -2, the quantity $c(l) := \frac{\rho}{\sqrt{1-\rho^2}} + \frac{l}{1+\rho}\Big(1 + 2\rho + \frac{l}{\sqrt{l^2+1}}\Big)$ is negative. Observe that for l going to $-\infty$, c goes to $-\infty$ when $\rho > 0$ and to 0^- when $\rho = 0$. Furthermore $c(l^*) = 0$. The derivative of c is $c'(l) = \frac{1}{1+\rho}\Big(1 + 2\rho + \frac{l(l^2+2)}{(l^2+1)^{\frac{3}{2}}}\Big)$. We would like c' to be positive, so that c would be increasing and, from the previous analysis, negative for $l < l^*$. However, this does not always happen. Indeed, the minimum point of c' is reached at $-\sqrt{2}$ and it is $\frac{1}{1+\rho}\Big(1 + 2\rho - \frac{4}{3}\sqrt{\frac{2}{3}}\Big)$, which is non-negative iff $\rho \geq \Big(\frac{2}{3}\Big)^{\frac{3}{2}} - \frac{1}{2} = \bar{\rho}$. Therefore, we are annoyed by the cases where $\rho < \bar{\rho}$. Under this hypothesis, call $\bar{l}(\rho)$ the first zero of c', that is also the only zero of c' smaller than $-\sqrt{2}$. We need to prove that $d(\rho) := c(\bar{l}(\rho); \rho) < 0$. Observe that $\frac{d}{d\rho} d(\rho) = \partial_\rho c(\bar{l}(\rho); \rho) + c'(\bar{l}(\rho); \rho)\frac{d}{d\rho}\bar{l}(\rho) = \partial_\rho c(\bar{l}(\rho); \rho)$. The latter quantity is $\frac{1}{(1-\rho^2)^{\frac{3}{2}}} + \frac{l}{(1+\rho)^2}\Big(1 - \frac{l}{\sqrt{l^2+1}}\Big)$, which is increasing in l and so it is smaller than its value at $l = -\sqrt{2}$. Then $\partial_\rho c(\bar{l}(\rho); \rho) \leq \frac{1}{(1-\rho^2)^{\frac{3}{2}}} - \frac{\sqrt{2}}{(1+\rho)^2}\Big(1 - \sqrt{\frac{2}{3}}\Big)$ and it is negative for $\rho < \bar{\rho}$. Then, it is enough to show $d(0) < 0$. For $\rho = 0$, we have $\bar{l}(0) = -\infty$ and it is easy to see that $d(0) = 0^-$ as desired.

Remember that if $l^* > 0$ (i.e. $\rho < 0$), then $l^* > L_-(l_-)$ because $L_-(l_-) \leq 0$ for Lemma 4.4 in [53]. Then we have proven that $\sup_{l<l^*} L_-(l) < l^*$ for every $\gamma > 0$ and ρ.

Using the symmetries, we have that for every ρ

$$l^*(\rho) > L_-(l_-(\rho); \rho) = -L_+(-l_-(\rho); -\rho) = -L_+(l_+(-\rho); -\rho)$$

or equivalently $l^*(-\rho) < L_+(l_+(-\rho); -\rho)$.

□

In particular this lemma holds for any SSVI with $b \leq \frac{2}{1+|\rho|}$, or equivalently, with $\theta\varphi \leq \frac{4}{1+|\rho|}$. Then, we have as an immediate consequence the following:

Corollary 4.1 (Absence of weak arbitrage for SSVI)**.** *An SSVI with $\theta\varphi \leq \frac{4}{1+|\rho|}$ is always free of weak arbitrage, i.e. the corresponding functions*

$$d_{1,2}(k) = -\frac{k}{\sqrt{SSVI(k;\theta,\varphi,\rho)}} \pm \frac{\sqrt{SSVI(k;\theta,\varphi,\rho)}}{2}$$

are automatically decreasing.

The above corollary entails that any SSVI smile with $\theta\varphi \leq \frac{4}{1+|\rho|}$ has a parametrization of the form Equation (4.1). In particular, $\phi(k)$ can be defined by $\phi(k) = -d_1(k)|d_1(k)|$, as Lucic does in the proof of Theorem 2.2 of [51]. At this point, the quantities k^* and k_* can be found setting $\phi(k^*) = 0$ and $\phi(k_*) = 2k_*$. In Section 6.5.3 it will be shown

$$k_* = -\frac{8\theta}{(4+\theta\varphi(1+\rho))(4-\theta\varphi(1-\rho))}.$$

Easily, it can be shown

$$k^* = \frac{8\theta}{(4+\theta\varphi(1-\rho))(4-\theta\varphi(1+\rho))}.$$

No Butterfly arbitrage domain for SSVI

Since the Fukasawa conditions for the SSVI are automatically satisfied as soon as $b \leq \frac{2}{1+|\rho|}$, the only condition required for the absence of Butterfly arbitrage is $\sigma \geq \sigma^*(\gamma, b, \rho, \mu)$, as stated in Theorem 4.2. In the case of SSVI, the function

$$\sigma^*(\gamma, b, \rho, \mu) = \sup_{l<l_1(\gamma,\rho)\vee l>l_2(\gamma,\rho)} -\frac{bg_2(l;\gamma,\rho)}{2G_1(l;\gamma,b,\rho,\mu)}$$

can be slightly simplified as it will be done in this section. Firstly observe that the dependence on γ and μ are superfluous since $\gamma = \sqrt{1-\rho^2}$ and $\mu = -\frac{\rho}{\sqrt{1-\rho^2}}$.

Secondly, the following remark holds.

Remark 4.2. *From Proposition 4.1, the smile $SVI(a, b, -\rho, -m, \sigma)$ is arbitrage-free iff $SVI(a, b, \rho, m, \sigma)$ is arbitrage-free. For SSVI, it holds $SVI(k; a, b, -\rho, -m, \sigma) = SVI(-k; a, b, \rho, m, \sigma)$ where $a = b\sigma\sqrt{1-\rho^2}$ and $m = -\sigma\frac{\rho}{\sqrt{1-\rho^2}}$. So an SSVI with $\rho < 0$ is arbitrage-free iff the SSVI with parameter $-\rho > 0$ is arbitrage-free.*

As a consequence, the quantity σ^* becomes

$$\sigma^*(b, \rho) = \sup_{l<l_1(|\rho|)\vee l>l_2(|\rho|)} -\frac{bg_2(l;|\rho|)}{2G_1(l;b,|\rho|)}.$$

Thirdly, the set where to look for the point of maximum of $f = -b\frac{g_2}{2G_1}$ can be restrained thanks to the following remark.

Remark 4.3. *Suppose $\rho \geq 0$, then for Lemma 4.1, $l_2 \leq -l_1$. Fix $l > l_2 > 0$. For the particular case of SSVI, $h(-l) - h(l) = 0$ so the difference between $G_1(l)$ and $G_1(-l)$ is given by $-b^2\frac{\rho l}{4\sqrt{l^2+1}}$, which is negative. Since $g_2(l) \leq g_2(-l)$, then $f(l) \geq f(-l)$, so $\sup_{l>l_2} f(l) \geq \sup_{l<l_1} f(l)$.*

The immediate consequence is that we can write

$$\sigma^*(b, \rho) = \sup_{l>l_2(|\rho|)} -\frac{bg_2(l;|\rho|)}{2G_1(l;b,|\rho|)}.$$

Finally, we can explicitly compute the quantity $l_2(|\rho|)$. Indeed, we will prove in Proposition 4.6 that the only positive zero of g_2 for $\rho \geq 0$ is

$$l_2(\rho) = \frac{1}{\tan\left(\frac{\arccos(-\rho)}{3}\right)}. \tag{4.25}$$

From now on, we will also work with $x = \frac{l}{\sqrt{l^2+1}}$, whence

$$\begin{aligned}\sqrt{1-x^2}N(l(x)) &= 1 + \rho x + \sqrt{1-\rho^2}\sqrt{1-x^2},\\ N'(l(x)) &= x + \rho,\\ N''(l(x)) &= (1-x^2)^{\frac{3}{2}}.\end{aligned}$$

Remark 4.4. *We redefine the useful functions when using the x variable and set:*

$$\Pi(x) := N(l(x)),\ j_2(x) := g_2(l(x)),\ j(x) := g(l(x)),\ J_1(x) := G_1(l(x)),\ \tilde{\phi}(x) := \tilde{f}(l(x)).$$

The derivative with respect to l is indicated with $'$, so for example the function $j_2'(x)$ corresponds to $g_2'(l(x))$ rather than $\frac{dj_2}{dx}(x)$.

We can now state the first important result for SSVI, which is the necessary and sufficient requirements on parameters for no Butterfly arbitrage. This corresponds to Theorem 4.2 applied to the SSVI case.

Proposition 4.6 (No arbitrage domain for SSVI)**.** *Let g_2, G_1 given by Equation (4.24). The following statements hold:*

- *(SVI parametrization) If an SSVI is arbitrage-free then $b \leq \frac{2}{1+|\rho|}$. In such case, the SSVI is arbitrage-free iff $\sigma \geq \sup_{l>l_2(|\rho|)} -\frac{bg_2(l;|\rho|)}{2G_1(l;b,|\rho|)}$ where $l_2(|\rho|) = \frac{1}{\tan\left(\frac{\arccos(-|\rho|)}{3}\right)}$.*
- *(Native parametrization) If an SSVI is arbitrage-free then $\theta\varphi \leq \frac{4}{1+|\rho|}$. In such case, the SSVI is arbitrage-free iff $(\theta\varphi)^2 \leq \inf_{l>l_2(|\rho|)} \frac{4\theta\sqrt{1-\rho^2}h(l;|\rho|)^2}{\theta\sqrt{1-\rho^2}g(l;|\rho|)^2 - g_2(l;|\rho|)}$ where $l_2(|\rho|) = \frac{1}{\tan\left(\frac{\arccos(-|\rho|)}{3}\right)}$.*

Proof. Thanks to Remark 4.2, we can restrain the attention to the case $\rho \geq 0$. In the study of g_2 in [53], we recall that the function has one positive zero $l_2 > l^*$. Thanks to Remark 4.3, the function σ^* is obtained with a superior taken for $l > l_2(|\rho|)$.

Let us study the location of the unique positive zero x_2 of $j_2(x) = (1-x^2)^{\frac{3}{2}} - \frac{(x+\rho)^2\sqrt{1-x^2}}{2(1+\rho x+\sqrt{1-\rho^2}\sqrt{1-x^2})}$. Note that $j_2(1) = 0$; so for $x < 1$, $j_2(x) = 0$ iff

$$2(1-x^2)(1+\rho x+\sqrt{1-\rho^2}\sqrt{1-x^2}) = (x+\rho)^2. \tag{4.26}$$

Isolating the radical and squaring yields that the zeros of j_2 solve the polynomial root equation

$$4(1-\rho^2)(1-x^2)^3 = \big((x+\rho)^2 - 2(1+\rho x)(1-x^2)\big)^2. \tag{4.27}$$

The key observation is that $-\rho$ is a root of Equation (4.27) and not of Equation (4.26). This leads to the following factorization of the polynomial $(x+\rho)(4x^5+8\rho x^4+(4\rho^2-3)x^3-5\rho x^2-\rho^2 x+\rho^3)$. Now $-\rho$ is again a root of the rightmost factor which factors in turn in $(x+\rho)(4x^4+4\rho x^3-3x^2-2\rho x+\rho^2)$... and the miracle continues! Indeed, $-\rho$ is again a root of this later factor, leading to the fact that the zeros of j_2 solve $\rho = x(3-4x^2)$.

In particular:

- $\rho = 0$, assuming we can exclude $x = 0$, gives $x = \frac{\sqrt{3}}{2} < 1$
- $\rho = 1$ gives the polynomial equation $4x^3-3x+1 = 0$ which reads $4(x+1)\big(x-\frac{1}{2}\big)^2 = 0$, whence $x = \frac{1}{2}$.

Taking the derivative with respect to ρ gives $1 = 3x'(\rho)(1-4x(\rho)^2)$ which gives in turn that $\rho \to x(\rho)$ is decreasing when $x > \frac{1}{2}$. We can now either solve the 3rd degree polynomial root problem, or take x as a parameter varying in the range $\big[\frac{1}{2}, \frac{\sqrt{3}}{2}\big]$, and backup ρ by the formula $\rho(x) = 3x - 4x^3$.

Going the latter route, we note that the polynomial $Q_\rho(x) := 4x^3 - 3x + \rho$ satisfies $Q_\rho\big(\frac{1}{2}\big) = -1+\rho < 0$, $Q_\rho\big(\frac{\sqrt{3}}{2}\big) = \rho \geq 0$ and also $Q'_\rho(x) = 3(4x^2-1)$ which is positive in the range $\big]\frac{1}{2}, \frac{\sqrt{3}}{2}\big]$, so it has a unique real root in this range. Its unique local maximum is located at $x = -\frac{1}{2}$ where $Q_\rho\big(-\frac{1}{2}\big) = 1+\rho > 0$, with, together with the observation that $Q_\rho(-1) = \rho - 1 < 0$ and $Q_\rho(\frac{1}{2}) < 0$, gives that Q_ρ has 2 other real roots located in $\big]-1, -\frac{1}{2}\big[$ and $\big]-\frac{1}{2}, \frac{1}{2}\big[$, so that $x_2(\rho)$ is the largest of the roots of Q_ρ. The three solutions have explicit formula

$$\cos\left(\frac{\arccos(-\rho) - 2\pi k}{3}\right), \quad k = 0, 1, 2$$

and the greatest is $x_2(\rho) = \cos\big(\frac{\arccos(-\rho)}{3}\big)$. It follows that $l_2(\rho)$ is given by Equation (4.25).

The proof in the native parametrization follows from the above proof. Indeed, the first statement is immediate from the identity $b = \frac{\theta\varphi}{2}$. The second statement follows from the fact that $\sigma = \frac{\sqrt{1-\rho^2}}{\varphi}$. Also, G_1 is positive thanks to the Fukasawa conditions, then looking at the expression for $\sigma^*(b,\rho)$, there is no arbitrage iff $4G_1(l;b,|\rho|)\frac{\sqrt{1-\rho^2}}{\varphi} \geq -\theta\varphi g_2(l;|\rho|)$ for every $l > l_2(|\rho|)$. The function $G_1(l;b,|\rho|)$ is equal to $h(l;|\rho|)^2 - \frac{(\theta\varphi)^2}{4}g(l;|\rho|)^2$, so that the previous condition holds iff

$$4\theta\sqrt{1-\rho^2}h(l;|\rho|)^2 \geq \Big(\theta\sqrt{1-\rho^2}g(l;|\rho|)^2 - g_2(l;|\rho|)\Big)(\theta\varphi)^2$$

for every $l > l_2(|\rho|)$. Since the function g_2 is negative in this domain, the conclusion immediately follows.

□

4.6.1 Arbitrage-free parametrization for SSVI

Proposition 4.6 characterizes the no Butterfly arbitrage domain for SSVI. The characterization requires the computation of the point of minimum of the function $f_{\text{SSVI}}(l;\theta,|\rho|) := \frac{4\theta\sqrt{1-\rho^2}h(l;|\rho|)^2}{\theta\sqrt{1-\rho^2}g(l;|\rho|)^2 - g_2(l;|\rho|)}$ for l bigger than an explicit quantity $l_2(|\rho|)$. The point of maximum can also be computed looking at the zeros of the derivative of f_{SSVI}. These points depend on the two parameters θ and ρ, so that, when calibrating a free of arbitrage SSVI slice, the no arbitrage domain must be computed at every new choice of either θ or ρ.

Computations can be actually simplified and in this section we provide a new parametrization of the SSVI smile requiring a root-finding depending on the unique parameter ρ. This means that for a fixed ρ, the root-finding can be performed once for all and the calibration algorithm can then proceed with extremely fast computations, simply requiring explicit function evaluations.

The new arbitrage-free SSVI parametrization requires the definition of some useful functions:

$$b^*(l,\rho) := \sqrt{\frac{h(l,\rho)\big(h(l,\rho)g_2'(l,\rho) - 2h'(l,\rho)g_2(l,\rho)\big)}{g(l,\rho)\big(g(l,\rho)g_2'(l,\rho) - 2g'(l,\rho)g_2(l,\rho)\big)}}, \tag{4.28}$$

$$\varphi^*(l,\rho) := -\frac{2(h(l,\rho)^2 - b^*(l,\rho)^2 g(l,\rho)^2)}{b^*(l,\rho)g_2(l,\rho)}\sqrt{1-\rho^2}. \tag{4.29}$$

Proposition 4.7 (No arbitrage parametrization for SSVI)**.** *Consider an SSVI smile as in Equation (4.23). If $b < \frac{2}{1+|\rho|}$ and $\rho \geq 0$, define $\bar{l}(0,\rho)$ the only root in $]l_2(\rho),\infty[$ of $b^*(l,\rho) = 0$. Let G_1, g_2 given by Equation (4.24). Then, under the assumption that the function $\tilde{f} = -\frac{G_1}{g_2}$ has a unique critical point (that we sustain numerically in Section 4.A), an SSVI with $\rho \in [-1,1]$ is arbitrage-free iff it can be parametrized as*

$$SSVI(k) = \frac{b^*(l,|\rho|)}{\varphi}\left(1 + \rho\varphi k + \sqrt{(\varphi k + \rho)^2 + 1 - \rho^2}\right)$$

where $l \geq \bar{l}(0,|\rho|)$ and $\varphi \leq \varphi^*(l,|\rho|)$.

We stress again the fact that the no arbitrage domain is here a product of intervals for fixed ρ. In Section 4.6.1 we will see how to implement this domain into a calibration algorithm.

Proof of Proposition 4.7

As in the proof of Proposition 4.6, we restrain the attention to the case $\rho \geq 0$.

Study of j_2: uniqueness of x_{m_2} In the proof of Proposition 4.6, we have shown that the function g_2 has one positive zero $l_2 > l^*$ which coincides with $\frac{1}{\tan\left(\frac{\arccos(-\rho)}{3}\right)}$. We now prove that g_2 has unique point of minimum m_2 at the right of l_2.

Using the parametrization in x, we have to prove that j_2 has a unique point of minimum x_{m_2} at the right of x_2. The zeros of j_2' solve

$$(x+\rho)^3 = 2(1-x^2)\big(\rho x + \sqrt{1-\rho^2}\sqrt{1-x^2}+1\big)\Big(3x\big(\rho x+\sqrt{1-\rho^2}\sqrt{1-x^2}+1\big)+x+\rho\Big). \tag{4.30}$$

In particular for $\rho = 0$ this reads $x^2 = 2(1-x^2)\big(1+\sqrt{1-x^2}\big)\big(4+3\sqrt{1-x^2}\big)$. Isolating the radical and squaring yields the polynomial equation $\big(x^2+2(1-x^2)(3x^2-7)\big)^2 = 4\times 49(1-x^2)^3$; 0 is a root, whereas it is not a solution of Equation (4.30). Letting $X = 1-x^2$, the polynomial factorizes into $(1-X)^2(36X^2-16X+1)$, yielding two roots in $]0,1[$, $\frac{4+\sqrt{7}}{18}$ and $\frac{4-\sqrt{7}}{18}$, with only the latter one solving Equation (4.30), giving in turn $x_{m_2}(\rho=0) = \sqrt{\frac{\sqrt{7}}{18}+\frac{7}{9}}$.

Let us turn now to $\rho = 1$. In this case Equation (4.30) simplifies to $1 = 2(1-x)(3x+1)$, or yet $6x^2-4x-1=0$, yielding $x_{m_2}(\rho=1) = \frac{2+\sqrt{10}}{6}$.

Before investigating the general case, we can observe that if we set $x=\rho$ in Equation (4.30) we get an equation in ρ which is $8\rho^3 = 4(1-\rho^2)\cdot 8\rho$, or yet $\rho^2 = 4(1-\rho^2)$ which gives $x_{m_2}\big(\rho=\frac{2}{\sqrt{5}}\big) = \frac{2}{\sqrt{5}}$.

Let us prove now that in the general case g_2 has a unique critical point.

Note that $\partial_\rho N(l) = l+l^*$, so $\partial_\rho g_2(l) = -\partial_\rho \frac{N'(l)^2}{2N(l)} = -\frac{N'(l)}{N(l)}\Big(1-\frac{N'(l)}{2N(l)}(l+l^*)\Big) = -\frac{N'(l)}{N(l)}h(l) < 0$ because $h(l) = (G_{1+}(l)+G_{1-}(l))/2 > 0$. From the formula $g_2'(l) = N'''(l) - \frac{N'(l)}{N(l)}g_2(l)$, we find $\partial_\rho g_2'(l) = -\Big(\partial_\rho\frac{N'(l)}{N(l)}\Big)g_2 - \frac{N'(l)}{N(l)}\partial_\rho g_2$ since N''' does not depend on ρ. The second term is positive while the first has the factor $\partial_\rho\frac{N'(l)}{N(l)} = \frac{1}{N(l)}\Big(1-\frac{N'(l)}{N(l)}(l+l^*)\Big) = \frac{1}{N(l)\sqrt{l^2+1}\sqrt{1-\rho^2}} > 0$. Then $\partial_\rho g_2'(l) > 0$.

We have seen that for $\rho = 0$ and $\rho = 1$, the function g_2 has a unique critical point m_2. Since $\partial_\rho g_2'(l) > 0$, then for every $\rho < 1$, the critical points of g_2 must be greater than $m_2(\rho=1) = \frac{2\sqrt{2}+\sqrt{5}}{3}$. A critical point in the l-variable is still a critical point in the x-variable and viceversa, because $\frac{df}{dx}(x) = f'(l(x))\frac{dl}{dx}(x) = \frac{f'(l(x))}{(1-x^2)^{\frac{3}{2}}}$ with the convention $' = \frac{d}{dl}$. Then, showing the convexity of g_2 in the x-variable for $x > x_{m_2}(\rho=1)$ automatically proves the uniqueness of its critical point in any variable (even if g_2 is not convex in the l-variable).

We have $\frac{d^2 j_2}{dx^2}(x) = \big(\frac{dl}{dx}(x)\big)^2 j_2''(x) + j_2'(x)\frac{d^2 l}{dx^2}(x) = (j_2''(x)+3x\sqrt{1-x^2}j_2'(x))(1-x^2)^{-3}$

and

$$j_2''(x) + 3x\sqrt{1-x^2}j_2'(x) = \left(\Pi^{iv}(x) - \frac{\Pi''(x)^2}{\Pi(x)} - \frac{\Pi'(x)\Pi'''(x)}{\Pi(x)} + \frac{5\Pi'(x)^2\Pi''(x)}{2\Pi(x)^2} + \right.$$
$$\left. - \frac{\Pi'(x)^4}{\Pi(x)^3}\right) + 3x\sqrt{1-x^2}\left(\Pi'''(x) - \frac{\Pi'(x)\Pi''(x)}{\Pi(x)} + \frac{\Pi'(x)^3}{2\Pi(x)^2}\right). \tag{4.31}$$

Since $3x\sqrt{1-x^2} = -\frac{\Pi'''(x)}{\Pi''(x)}$, the terms $-\frac{\Pi'(x)\Pi'''(x)}{\Pi(x)}$ and $-3x\sqrt{1-x^2}\frac{\Pi'(x)\Pi''(x)}{\Pi(x)}$ simplify. Also, $\Pi^{iv}(x)+3x\sqrt{1-x^2}\Pi'''(x) = 3(1-x^2)^{\frac{5}{2}}(2x^2-1)$, which is positive since $x > \frac{2+\sqrt{10}}{6} > \frac{1}{\sqrt{2}}$. The term $-\frac{\Pi''(x)^2}{\Pi(x)}$ becomes positive with the sum of $\frac{\Pi'(x)^2\Pi''(x)}{2\Pi(x)^2}$, indeed it becomes $-\frac{\Pi''(x)}{\Pi(x)}\left(\Pi''(x) - \frac{\Pi'(x)^2}{2\Pi(x)}\right) = -\frac{\Pi''(x)}{\Pi(x)}j_2(x) > 0$. Note that the remaining $4\frac{\Pi'(x)^2\Pi''(x)}{2\Pi(x)^2}$ is positive. Finally, $-\frac{\Pi'(x)^4}{\Pi(x)^3}$ compensates with $3x\sqrt{1-x^2}\frac{\Pi'(x)^3}{2\Pi(x)^2}$ summing to

$$\frac{\Pi'(x)^3}{\Pi(x)^2}\left(\frac{3}{2}x\sqrt{1-x^2} - \frac{\Pi'(x)}{\Pi(x)}\right) = \frac{\Pi'(x)^3}{2\Pi(x)^3}\big(x + 3\rho x^2 - 2\rho + 3x\sqrt{1-\rho^2}\sqrt{1-x^2}\big).$$

Now $3\rho x^2 - 2\rho$ is positive iff $x > \sqrt{\frac{2}{3}}$ and this is true, so also the previous quantity is positive. Then, since the negative terms of $\frac{d^2 j_2}{dx^2}$ are smaller in magnitude than its positive terms, j_2 is convex in the x-variable.

As a consequence, j_2 has a unique critical point for $x > x_2$ and it lies between $x_{m_2}(\rho = 1)$ and $x_{m_2}(\rho = 0)$.

In order to study x_{m_2} in the general case, let us isolate the radical of Equation (4.30) and square. We get the daunting polynomial root equation

$$\big((\rho+x)^3 - 2(1-x^2)\big(3(1-\rho^2)x(1-x^2) + (\rho x+1)(3x(\rho x+1) + (\rho+x))\big)\big)^2 =$$
$$4(1-\rho^2)(1-x^2)^3\big(6x(\rho x+1) + \rho + x\big)^2.$$

There again, we observe that $-\rho$ is a root, whereas it does not solve the initial equation. This leads to an iterative factorization where eventually $-\rho$ is a root of order 4. The remaining factor is

$$\rho^2 + 2\rho(12x^5 - 16x^3 + 5x) + (36x^6 - 56x^4 + 21x^2).$$

For $x \in [x_{m_2}(\rho = 1), x_{m_2}(\rho = 0)]$, the only positive ρ solution is

$$\rho(x) = x(-12x^4 + 16x^2 - 5 + 2(1-x^2)\sqrt{36x^4 - 24x^2 + 1}).$$

Since $\partial_\rho j_2' < 0$, also $\partial_\rho \frac{dj_2}{dx} < 0$ and the function $\rho(x)$ is invertible in $[x_{m_2}(\rho = 1), x_{m_2}(\rho = 0)] = \left[\frac{2+\sqrt{10}}{6}, \sqrt{\frac{\sqrt{7}}{18} + \frac{7}{9}}\right]$ and its inverse gives the value of x_{m_2} for fixed ρ.

Study of J_1 From Remark 4.4, $J_1(x) = \eta^2(x) - b^2 j^2(x)$ where $\eta(x) = \frac{1}{2}\big(1 + \sqrt{\frac{1-x^2}{1-\rho^2}}\big)$ and $j(x) = \frac{x+\rho}{4}$. The formulas for η and j can be easily recovered from Equation (4.24). The function j is of course positive and increasing in x, while η is positive decreasing in x, since $\frac{d\eta}{dx}(x) = -\frac{x}{2\sqrt{1-\rho^2}\sqrt{1-x^2}} < 0$. Then J_1 is decreasing and it attains its minimum at 1, equal to $\frac{1}{16}\big(2 - b(1+\rho)\big)\big(2 + b(1+\rho)\big)$.

Looking at the condition $\sigma \geq \sup_{x>x_2} -\frac{bj_2(x)}{2J_1(x)}$, we then have that for every $x > x_2$, it holds $-\frac{bj_2(x)}{2J_1(x)} < -\frac{bj_2(x_{m_2})}{2J_1(1)} = -\frac{8bj_2(x_{m_2})}{4-b^2(1+\rho)^2}$, from which one immediately finds a no arbitrage subdomain for SSVI:

Lemma 4.9 (Simple no arbitrage subdomain for SSVI)**.** *An SSVI with $b(1+|\rho|) \leq 2$ and $\sigma \geq -\frac{8bj_2(x_{m_2}(|\rho|),|\rho|)}{4-b^2(1+|\rho|)^2}$ is arbitrage-free.*

Equivalently, an SSVI with

$$\theta\varphi \leq \min\left(\frac{4}{1+|\rho|}, \sqrt{\frac{16\theta\sqrt{1-\rho^2}}{\theta(1+|\rho|)^2\sqrt{1-\rho^2} - 16j_2(x_{m_2}(|\rho|),|\rho|)}}\right)$$

is arbitrage-free.

The above lemma is of high interest since the subdomain is easy to be computed. Furthermore, we already know a no arbitrage subdomain for SSVI given by the conditions of Gatheral and Jacquier in [36], so we compare the two subdomains in Section 4.6.1. We will see that the subdomain of Lemma 4.9 is generally less stringent than the Gather-Jacquier subdomain.

Uniqueness of the critical point of $\tilde{f}$ for SSVI Observe that Remark 4.3 can be re-written in terms of the function $\tilde{f} = -\frac{G_1}{g_2}$. Indeed, it holds $\sigma^* = \frac{b}{2\inf_{l>l_2}\tilde{f}(l)}$ for $\rho \geq 0$.

Remember that we consider the case $\rho \geq 0$. Then, we look at the infimum of $\tilde{f}(l)$ for $l > l_2$, letting b be eventually 0. For $b < \frac{2}{1+\rho}$, it holds $\tilde{f}(l_2^+) = \infty = \tilde{f}(\infty)$ and $\tilde{f} > 0$, so $\tilde{f}$ must have a critical point which is a point of minimum. We want to prove that in such case $\tilde{f}$ has a unique critical point in $]l_2, \infty[$. When $b = \frac{2}{1+\rho}$, it can be shown that the derivative of $\tilde{f}$ vanishes at ∞. We prove that even in this case this is the unique point of minimum of $\tilde{f}$.

To show that $\tilde{f}$ has a unique critical point, we can prove that $\frac{d\tilde{\phi}}{dx}$ has a unique zero, where we recall from Remark 4.4 $\tilde{\phi}(x) = \tilde{f}(l(x)) = -\frac{J_1(x)}{j_2(x)}$.

We have seen that J_1 is decreasing as a function of x. A critical point must satisfied $\frac{d\tilde{\phi}}{dx} = 0$ or $J_1\frac{dj_2}{dx} - \frac{dJ_1}{dx}j_2 = 0$. Looking at the sign of the previous functions, it must hold $\frac{dj_2}{dx} > 0$ so the critical point is on the right of x_{m_2}. At the critical point, it holds $j_2^2\frac{d^2\tilde{\phi}}{dx^2} = J_1\frac{d^2j_2}{dx^2} - \frac{d^2J_1}{dx^2}j_2 := n$, so in order to prove the uniqueness of the critical point, we could show that n is positive for every $x > x_{m_2}$.

We have $\partial_{b^2} n = -j^2\frac{d^2j_2}{dx^2} + 2(\frac{dj}{dx}^2 + j\frac{d^2j}{dx^2})j_2 = -\frac{\Pi'^2}{16}\frac{d^2j_2}{dx^2} + \frac{j_2}{8} < 0$, so it is enough to show the positivity of $n_{|_{b=\frac{2}{1+\rho}}}$ for $x > x_{m_2}$ or, more generally, for $x > x_{m_2}(\rho = 1) = \frac{2+\sqrt{10}}{6}$.

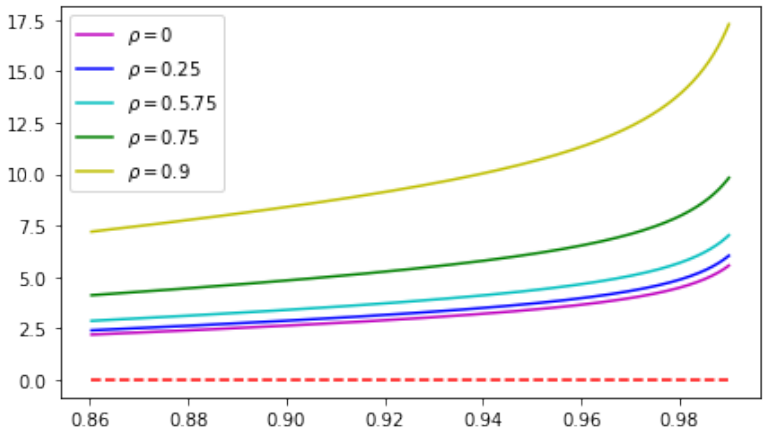

Figure 4.10: Plots of $n|_{b=\frac{2}{1+\rho}}$ for different ρ values.

We show it numerically in Section 4.A. We plot in Figure 4.10 the function $n_{|_{b=\frac{2}{1+\rho}}}$ for $x > x_{m_2}(\rho = 1)$ for different values of ρ.

Proposition 4.8. *Let G_1, g_2 given by Equation (4.24). Then, under the hypothesis that $\tilde{f} = -\frac{G_1}{g_2}$ has a unique critical point in $]l_2(\rho), \infty]$ (that we sustain numerically in Section 4.A), an SSVI with $b = \frac{2}{1+|\rho|}$ is arbitrage-free iff $\sigma \geq \sqrt{1-\rho^2}$.*

Equivalently, under the hypothesis that $\tilde{f}$ has a unique critical point, an SSVI with $\theta\varphi = \frac{4}{1+|\rho|}$ is arbitrage-free iff $\varphi \leq 1$.

Proof. The limit of $\tilde{f}$ at ∞ is finite iff $b = \frac{2}{1+\rho}$ and, for any SVI, it equals $\frac{1}{2}\left(\frac{\gamma}{1+\rho} - \mu\right)$, which corresponds to $\frac{1}{2\sqrt{1-\rho^2}}$ for SSVI. Under the hypothesis that $\tilde{f}$ has a unique critical point in $]l_2(\rho), \infty]$ and since for $b = \frac{2}{1+\rho}$ the derivative of $\tilde{f}$ at ∞ is null, the critical point of $\tilde{f}$ must be ∞. Also, $\tilde{f}(l_2^+) = \infty$, so the critical point is a point of minimum and the function is strictly decreasing. Then, $\sigma^* = \sqrt{1-\rho^2}$.

Using the SSVI notation, $b = \frac{\theta\varphi}{2}$ and $\sigma = \frac{\sqrt{1-\rho^2}}{\varphi}$ and the conclusion follows immediately.

□

Application of the b^* approach to SSVI We have numerically shown that $\tilde{f}$ has at most one critical point in $]l_2, \infty]$ and this is a point of minimum. The case $b = \frac{2}{1+\rho}$ has already been treated in Proposition 4.8. Consider now $b < \frac{2}{1+\rho}$ and the reasoning in Section 4.2.3 regarding the b^* approach which allows to reparametrize the no arbitrage domain from the critical point condition. The idea is to reparametrize the arbitrage bound on σ discussed in Proposition 4.6. The bound requires to compute the minimum of $\tilde{f} = -\frac{G_1}{g_2}$, on an interval which depends on ρ. The strategy is to study the critical

points of $\tilde{f}$, knowing that $\tilde{f}$ goes to infinity at the bounds of this interval. The equation characterizing those critical points $\bar{l}$ reads

$$p(\bar{l};\rho) = b^2 q(\bar{l};\rho).$$

Using $\bar{l}$ as a parameter we can obtain b^2 as $\frac{p(\bar{l};\rho)}{q(\bar{l};\rho)}$. Since $\tilde{f}$ has a unique critical point $\bar{l} \in]l_2, \infty[$, it satisfies $b^2 = b^*(\bar{l})^2$ or $p(\bar{l}) = q(\bar{l}) = 0$. In the latter case, we should have $\frac{2h'(\bar{l})}{h(\bar{l})} = \frac{2g'(\bar{l})}{g(\bar{l})}$ but the right hand side is always positive while we have shown that the left hand side is negative. Then we can apply the observation in Section 4.2.3 and obtain that the set of the critical points of $\tilde{f}$ for ρ fixed is equal to the interval starting from the maximum between all the zeros of p, coinciding with $\bar{l}(0,\rho)$, and ending at $\bar{l}\big(\frac{2}{1+\rho},\rho\big) = \infty$. Fix $\rho \geq 0$, $b < \frac{2}{1+\rho}$, choose $l \in [\bar{l}(0,\rho), \infty[$. Then $\sigma^* = \frac{b^*(l)}{2\tilde{f}(l;\rho)}$. Also, the requirement $\sigma \geq \sigma^*$ coincides with the requirement $\varphi \leq \frac{\sqrt{1-\rho^2}}{\sigma^*} = \varphi^*$.

Limits of $\bar{l}$ Because of the uniqueness of the critical points of $\tilde{\phi}$, when b goes to 0, the critical point satisfies $\eta j_2' - 2\eta' j_2 = 0$. The numerator of $\eta j_2' - 2\eta' j_2$ is

$$\begin{aligned}\sqrt{1-\rho^2}\big(-4\rho x^6 + 2(6\rho^2-5)x^5 + 24\rho x^4 + (31-14\rho^2)x^3 - 13\rho x^2 + 5(\rho^2-4)x + \rho(\rho^2-4)\big) + \\ + \sqrt{1-x^2}\big(2(2\rho^2-1)x^5 + 4\rho(4-3\rho^2)x^4 + (21-22\rho^2)x^3 + \\ + \rho(8\rho^2-15)x^2 + 5(3\rho^2-4)x + \rho(3\rho^2-4)\big).\end{aligned}$$

If $\rho = 0$, possible solutions are $x = 0$ and $x = \sqrt{\sqrt{6} - \frac{3}{2}}$ but since x must be greater than the unique zero of j_2, $x_2 = \cos\big(\frac{\arccos(-\rho)}{3}\big) = \frac{\sqrt{3}}{2}$, the latter is the searched solution. If $\rho = 1$, the only possible solution is $x = 1$. In terms of l, these two points correspond to $\sqrt{9+4\sqrt{6}}$ and ∞.

The above expression can also be written as

$$2u\sqrt{1-x^2}\big[2x(1-x^2)u - x(x+\rho)^2 - \sqrt{1-x^2}(x+\rho+3xu)v\big] + (x+\rho)^3 v$$

or

$$\varphi(x) := -2u^2\sqrt{1-x^2}\big[x\sqrt{1-x^2}\big(3\sqrt{1-\rho^2} + \sqrt{1-x^2}\big) + x + \rho\big] + (x+\rho)^3 v$$

where $u = 1+\rho x+\sqrt{1-x^2}\sqrt{1-\rho^2}$ and $v = \sqrt{1-x^2}+\sqrt{1-\rho^2}$. Observe that $(u,v)(x = 1) = \big(1+\rho, \sqrt{1-\rho^2}\big)$ and $(u,v)(x=\rho) = \big(2, 2\sqrt{1-\rho^2}\big)$ so that

- $\varphi(1) = (1+\rho)^3\sqrt{1-\rho^2} > 0$;
- $\varphi(\rho) = -48\rho(1-\rho^2)^{\frac{3}{2}} \leq 0$;

and φ has a zero in the range $]x_2(\rho) \vee \rho, 1[$.

Lemma 4.10 (Computation of $\bar{l}(0,\rho)$)**.** *The only root of $b^*(l)=0$ is $\sqrt{9+4\sqrt{6}}$ if $\rho=0$, ∞ if $\rho=1$ and, in the general case, it lies in $]\min(l_2(\rho),-l^*(\rho)),\infty[$. It can be computed as $\bar{l}(0,\rho)=\frac{\bar{x}(0,\rho)}{\sqrt{1-\bar{x}(0,\rho)^2}}$, where $\bar{x}(0,\rho)$ is the only zero of*

$$\begin{aligned}\sqrt{1-\rho^2}\big(-4\rho x^6+2(6\rho^2-5)x^5+24\rho x^4+(31-14\rho^2)x^3-13\rho x^2+5(\rho^2-4)x+\rho(\rho^2-4)\big)+\\ +\sqrt{1-x^2}\big(2(2\rho^2-1)x^5+4\rho(4-3\rho^2)x^4+(21-22\rho^2)x^3+\\ +\rho(8\rho^2-15)x^2+5(3\rho^2-4)x+\rho(3\rho^2-4)\big),\end{aligned}$$

which lies in $]\min(x_2(\rho),\rho),1[$ where $x_2(\rho)=\cos\big(\frac{\arccos(-\rho)}{3}\big)$.

The Gatheral-Jacquier sufficient conditions

How does the boundary in Lemma 4.9 compare with the sufficient conditions found by Gatheral and Jacquier in Theorem 4.2 of [36]? The theorem asserts that an SSVI is free of Butterfly arbitrage if $\theta\varphi(1+|\rho|)\leq 4$ and $\theta\varphi^2(1+|\rho|)\leq 4$. Equivalently, an SSVI is free of Butterfly arbitrage if

$$\theta\varphi\leq\min\left(\frac{4}{1+|\rho|},\sqrt{\frac{4\theta}{1+|\rho|}}\right).$$

We can compare with some plots the Gatheral-Jacquier sufficient condition with the sufficient condition of Lemma 4.9 and the necessary and sufficient condition of Proposition 4.6:

$$\theta\varphi\leq\min\left(\frac{4}{1+|\rho|},\inf_{l>l_2(|\rho|)}\sqrt{\frac{4\theta\sqrt{1-\rho^2}h(l;|\rho|)^2}{\theta\sqrt{1-\rho^2}g(l;|\rho|)^2-g_2(l;|\rho|)}}\right).$$

We plot the levels below which the product $\theta\varphi$ must lie. In the first couple of graphs in Figure 4.11, ρ is fixed while θ ranges in $]0,\frac{1}{2}[$; in the second couple of graphs in Figure 4.12, θ is fixed and ρ ranges in $]-1,0[$ as markets generally reflect negative ρs.

We can see that the sufficient condition of Lemma 4.9 is generally weaker than the Gatheral-Jacquier's one, which means that it is less strict and it allows the parameters to live into a larger space, which is still free of arbitrage. In this sense, a calibration performed with Gather-Jacquier sufficient conditions could fit much worse market data, because it obliges parameters to live into a smaller space than the one allowed by Lemma 4.9.

Calibration algorithm

Proposition 4.7 allows us to design an algorithm for the calibration of arbitrage-free SSVI slices. Indeed, for a fixed maturity T, suppose we have market Call prices for different strikes: $\{(K,C^{\mathrm{mkt}}(K,T))\}$. We want to find the best triplet (θ,φ,ρ) which fits with a good calibration error market prices. The calibration error can be computed as squared

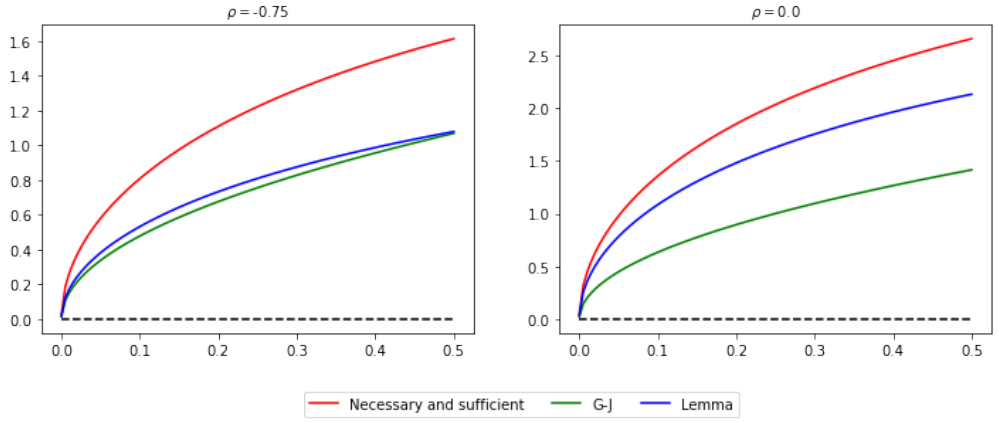

Figure 4.11: Comparison of Gatheral-Jacquier sufficient conditions, Lemma 4.9 sufficient conditions and Proposition 4.6 necessary and sufficient conditions as functions of θ. The area below each line represents the (sub)domain of no arbitrage where $\theta\varphi$ can lie.

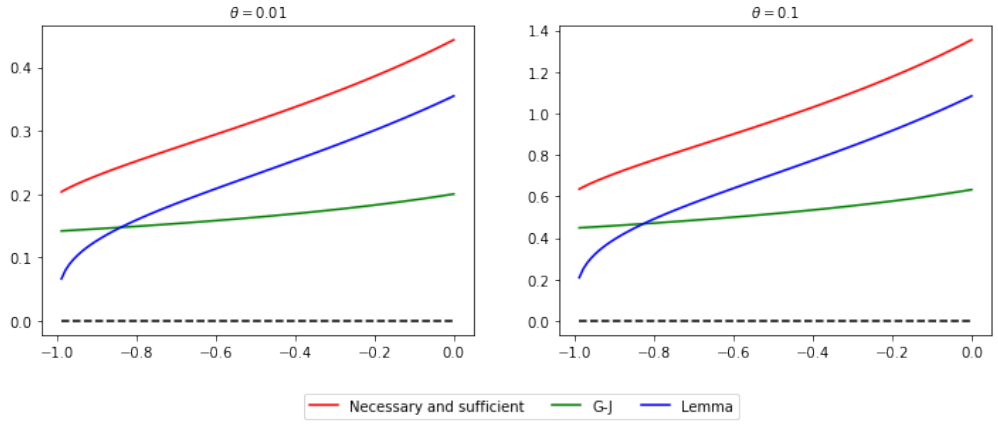

Figure 4.12: Comparison of Gatheral-Jacquier sufficient conditions, Lemma 4.9 sufficient conditions and Proposition 4.6 necessary and sufficient conditions as functions of ρ. The area below each line represents the (sub)domain of no arbitrage where $\theta\varphi$ can lie.

differences and its minimization can be performed via least squares. In other words, we want to minimize the function

$$\varepsilon(\theta,\varphi,\rho)=\sum_K w(K,T)\big(C^{\text{mod}}\big(K,T,\text{SSVI}(k;\theta,\varphi,\rho)\big)\big)-C^{\text{mkt}}(K,T)\big)^2$$

where $k=\log\frac{K}{F}$ is the log-forward moneyness and

$$\begin{aligned} C^{\text{mod}}(K,T,\omega)&=D(T)\big(N(d_1)F(T)-N(d_2)K\big)\\ d_{1,2}&=-\frac{k}{\sqrt{\omega}}\pm\frac{\sqrt{\omega}}{2}. \end{aligned} \tag{4.32}$$

The weights $w(K,T)$ can be arbitrarily chosen and they could be, for example, the inverse vegas. In this way, the calibration would be on implied volatilities rather then on prices.

Using Proposition 4.7, we look for parameters (l,φ,ρ) satisfying $\rho\in[-1,1]$, $l\geq\bar{l}(0,|\rho|)$ and $\varphi\leq\varphi^*(l,|\rho|)$. Then, the corresponding θ parameter is equal to $\frac{2b^*(l,|\rho|)}{\varphi}$. Calibration routines such as the `least_squares` in the `scipy.optimize` Python library only allow for constant bounds on the parameters. Then, a reparametrization is needed. It is easy to see that defining

$$\Lambda=\frac{\bar{l}(0,|\rho|)}{l}\qquad\qquad\Phi=\frac{\varphi}{\varphi^*(l,|\rho|)}$$

allows to set constant bounds on the new parameters (Λ,Φ,ρ), i.e.

$$\begin{aligned}\Lambda&\in\,]0,1]\\ \Phi&\in[0,1]\\ \rho&\in[-1,1].\end{aligned}$$

Given a triplet (Λ,Φ,ρ), the original SSVI parameters can be recovered sequentially computing

$$\begin{aligned} l(\Lambda,\rho)&=\frac{\bar{l}(0,|\rho|)}{\Lambda}\\ \varphi(\Phi,\Lambda,\rho)&=\Phi\varphi^*(l(\Lambda,\rho),|\rho|)\\ \theta(\Phi,\Lambda,\rho)&=\frac{2b^*(l(\Lambda,\rho),|\rho|)}{\varphi(\Phi,\Lambda,\rho)}. \end{aligned} \tag{4.33}$$

In particular, the function to be minimized which has to be given as input to the calibration algorithm is

$$\tilde{\varepsilon}(\Lambda,\Phi,\rho)=\sum_K w(K,T)\big(C^{\text{mod}}\big(K,T,\text{SSVI}(k;\theta(\Phi,\Lambda,\rho),\varphi(\Phi,\Lambda,\rho),\rho)\big)\big)-C^{\text{mkt}}(K,T)\big)^2. \tag{4.34}$$

and parameters (Λ,Φ,ρ) can live in the product of intervals $]0,1]\times[0,1]\times[-1,1]$.

Observe that the computation of $\varphi(\Phi, \Lambda, \rho)$ and $\theta(\Phi, \Lambda, \rho)$ requires to know the quantity $l(\Lambda, \rho)$, which in turn has to be computed from $\bar{l}(0, |\rho|)$. The latter quantity is the only one to require more complex computations. Indeed, it involves a root-finding algorithm which can be settled using Lemma 4.10. In particular, $\bar{l}(0, |\rho|)$ is the only root in the interval $]\min(x_2(\rho), \rho), 1[$ of

$$\begin{aligned}\sqrt{1-\rho^2}\big(-4\rho x^6+2(6\rho^2-5)x^5+24\rho x^4+(31-14\rho^2)x^3-13\rho x^2+5(\rho^2-4)x+\rho(\rho^2-4)\big)+\\ +\sqrt{1-x^2}\big(2(2\rho^2-1)x^5+4\rho(4-3\rho^2)x^4+(21-22\rho^2)x^3+\\ +\rho(8\rho^2-15)x^2+5(3\rho^2-4)x+\rho(3\rho^2-4)\big) \quad (4.35)\end{aligned}$$

A possible root-finding algorithm can be the `brentq` function in the `scipy.optimize` Python library. In Algorithm 1 we describe how to perform the computation.

Algorithm 1 Computation of $\bar{l}(0, \rho)$

Input Parameter ρ.
Output Quantity $\bar{l}(0, \rho)$.

1: Give to the built-in root-finding function the target function in Equation (4.35);
2: Give to the built-in root-finding function the bound $]\min(x_2(\rho), \rho), 1[$ where $x_2(\rho) = \cos\big(\frac{\arccos(-\rho)}{3}\big)$;
3: Let the built-in root-finding function run and find the only zero $\bar{x}(0, \rho)$;
4: Compute $\bar{l}(0, \rho) = \frac{\bar{x}(0,\rho)}{\sqrt{1-\bar{x}(0,\rho)^2}}$.

At this point, we detail the calibration algorithm in Algorithm 2.

Algorithm 2 Calibration algorithm for arbitrage-free SSVI slices

Input Set of market Calls and strikes for a fixed maturity T: $\{(K, C^{\text{mkt}}(K, T))\}$.
Output SSVI parameters (θ, φ, ρ) which best fit market prices.

1: Give to the built-in calibration function the target function $\tilde{\varepsilon}(\Lambda, \Phi, \rho)$ in Equation (4.34);
2: Give to the built-in calibration function the parameters bounds $]0, 1] \times [0, 1] \times [-1, 1]$;
3: Let the built-in calibration function run and find the best fit parameters (Λ, Φ, ρ);
4: Compute (θ, φ, ρ) from (Λ, Φ, ρ) as in Equation (4.33).

4.6.2 The Long Term Heston SVI is an SSVI

In this section, we look at the Long Term Heston SVI introduced by Gatheral and Jacquier in [35]. The authors prove that, under an appropriate hypothesis on the Heston parameters, the Heston implied volatility converges to an SVI implied volatility when the time-to-maturity goes to infinity. Here we show that it actually converges to an SSVI.

In the Heston model, the underlying process S_t follows the stochastic differential equation of the form

$$\begin{aligned} dS_t &= \sqrt{v_t} S_t dW_t, & S_0 > 0 \\ dv_t &= \tilde{\kappa}(\tilde{\theta} - v_t)dt + \tilde{\sigma}\sqrt{v_t} dZ_t, & v_0 > 0 \\ < W, Z >_t &= \tilde{\rho} dt. \end{aligned}$$

We use the $\sim$ hat to distinguish Heston parameters from SVI and SSVI parameters.

In [35] it is shown that setting

$$\begin{aligned} \theta &:= \frac{4\tilde{\kappa}\tilde{\theta}T}{\tilde{\sigma}^2(1-\tilde{\rho}^2)}\Big(\sqrt{(2\tilde{\kappa} - \tilde{\rho}\tilde{\sigma})^2 + \tilde{\sigma}^2(1-\tilde{\rho}^2)} - (2\tilde{\kappa} - \tilde{\rho}\tilde{\sigma})\Big), \\ \varphi &:= \frac{\tilde{\sigma}}{\tilde{\kappa}\tilde{\theta}T}, \end{aligned}$$

the Heston model converges to an SVI with parameters

$$a = \frac{\theta}{2}(1-\rho^2), \qquad b = \frac{\theta\varphi}{2}, \qquad \rho = \tilde{\rho}, \qquad m = -\frac{\rho}{\varphi}, \qquad \sigma = \frac{\sqrt{1-\rho^2}}{\varphi}. \tag{4.36}$$

In the following proposition we state our main result of this section: the Long Term Heston SVI is an SSVI and we find a lower bound for T which grants no arbitrage. Intuitively, choosing T large enough grants no arbitrage in the SVI Long Term Heston parametrization since it approximates the Heston smile, which is itself arbitrage-free. Now the convergence in the proof of Gatheral and Jacquier is pointwise, so the no arbitrage property at a fixed T is still to be proven.

Proposition 4.9 (No arbitrage subdomain for the Long Term Heston SVI)**.** *The Long Term Heston SVI approximation defined by Equation (4.36) is an SSVI. There is no Butterfly arbitrage in the Long Term Heston SVI approximation as soon as*

$$T \geq -\frac{8b\tilde{\sigma}g_2(m_2(|\rho|), |\rho|)}{\tilde{\kappa}\tilde{\theta}\sqrt{1-\rho^2}(4 - b^2(1+|\rho|)^2)}$$

where $m_2(\rho)$ is the only positive point of minimum of the function $g_2(\cdot, \rho)$ defined in Equation (4.24), and $b = \frac{2}{\tilde{\sigma}(1-\tilde{\rho}^2)}\Big(\sqrt{(2\tilde{\kappa} - \tilde{\rho}\tilde{\sigma})^2 + \tilde{\sigma}^2(1-\tilde{\rho}^2)} - (2\tilde{\kappa} - \tilde{\rho}\tilde{\sigma})\Big)$.

Proof. Recall the canonical SSVI Equation (4.23). From Equation (4.36), observe that it holds $\gamma = \sqrt{1-\rho^2}$ and $\mu = -\frac{\rho}{\sqrt{1-\rho^2}} = l^*$. In this way it is evident that the Long Term Heston corresponds to a sub-SVI with three parameters and, in particular, it corresponds to an SSVI with parameters (θ, ρ, φ).

From Corollary 4.1, an SSVI always satisfies the Fukasawa conditions. The only requirement needed for a no arbitrage parametrization is $\sigma > \sigma^*(b, \rho)$. Then, a Long Term Heston is free of arbitrage iff it satisfies Proposition 4.6. Since σ depends linearly to T, indeed $\sigma = \frac{\tilde{k}\tilde{\theta}\sqrt{1-\rho^2}}{\tilde{\sigma}}T$, it increases with T. On the other hand, b and ρ are constant

with respect to T so $\sigma^*(b,\rho)$ is also constant and σ will eventually become greater than the arbitrage bound for T going to ∞.

Suppose $\rho \geq 0$, then from Remark 4.3, f reaches its superior on the right of l_2, the second zero of g_2. The requirement $\sigma \geq -\frac{bg_2(l)}{2(h^2(l)-b^2g^2(l))}$ for every $l > l_2$ can be rewritten substituting σ with its expressions in terms of the Heston parameters, as

$$T \geq A \sup_{l>l_2} -\frac{g_2(l)}{h^2(l) - b^2g^2(l)} \tag{4.37}$$

with $A := \frac{b\tilde{\sigma}}{2\tilde{\kappa}\tilde{\theta}\sqrt{1-\rho^2}} = \left(\sqrt{(2\tilde{\kappa} - \tilde{\rho}\tilde{\sigma})^2 + \tilde{\sigma}^2(1-\tilde{\rho}^2)} - (2\tilde{\kappa} - \tilde{\rho}\tilde{\sigma})\right) / \left(\tilde{\kappa}\tilde{\theta}(1-\tilde{\rho}^2)^{\frac{3}{2}}\right)$. Note that the argument of the supremum does only depend on ρ and on $b = \frac{2}{\tilde{\sigma}(1-\tilde{\rho}^2)}\left(\sqrt{(2\tilde{\kappa} - \tilde{\rho}\tilde{\sigma})^2 + \tilde{\sigma}^2(1-\tilde{\rho}^2)} - (2\tilde{\kappa} - \tilde{\rho}\tilde{\sigma})\right)$.

The numerator attains its minimum at the locus of the unique minimum of g_2 on the right of l_2, that we denoted $m_2(\rho)$. We have proven that for SSVI, h is a decreasing function so it is always greater than $h(\infty) = \frac{1}{2}$, while g is increasing with limit $g(\infty) = \frac{(1+\rho)}{4}$. In this way, for $T \geq -\frac{8b\tilde{\sigma}g_2(m_2(\rho))}{\tilde{\kappa}\tilde{\theta}\sqrt{1-\rho^2}(4-b^2(1+\rho)^2)}$, the inequality in Equation (4.37) is satisfied.

In the case $\rho < 0$, from Proposition 4.1 it follows that the previous discussion still holds substituting ρ with $-\rho$.

□

It is worth observing that Proposition 4.9 actually corresponds to Lemma 4.9 using the Long Term Heston notations.

4.A Numerical evidence of the uniqueness of the critical point of $\tilde{f}$ in SSVI

As discussed in Section 4.6.1, to prove the uniqueness of the critical point of $\tilde{f}$ for SSVI, it is enough to show that the function $n = J_1 \frac{d^2 j_2}{dx^2} - \frac{d^2 J_1}{dx^2} j_2$ is positive for every $x > x_{m_2}(\rho = 1) = \frac{2+\sqrt{10}}{6}$ setting $b = \frac{2}{1+\rho}$.

Observe that from Remark 4.4, $\Pi(x) = N(l(x))$. In the following code, we call `Pik` the k-th derivative of $N(l)$ with respect to l, evaluated in $l(x)$. In Section 4.6.1 we denoted such quantity as $\Pi^{(k)}$. Then,

$$j_2(x) = g_2(l(x)) = \Pi''(x) - \frac{\Pi'(x)^2}{2\Pi(x)},$$

and the derivative of j_2 with respect to x can be computed starting from these quantities. In particular, using Equation (4.31), it holds $\frac{d^2 j_2}{dx^2}(x) = (j_2''(x) + 3x\sqrt{1-x^2} j_2'(x))(1-x^2)^{-3}$ and

$$\begin{aligned} j_2''(x) + 3x\sqrt{1-x^2} j_2'(x) = &\left(\Pi^{iv}(x) - \frac{\Pi''(x)^2}{\Pi(x)} - \frac{\Pi'(x)\Pi'''(x)}{\Pi(x)} + \frac{5\Pi'(x)^2\Pi''(x)}{2\Pi(x)^2} - \right. \\ &\left. + \frac{\Pi'(x)^4}{\Pi(x)^3} \right) + 3x\sqrt{1-x^2} \left(\Pi'''(x) - \frac{\Pi'(x)\Pi''(x)}{\Pi(x)} + \frac{\Pi'(x)^3}{2\Pi(x)^2} \right). \end{aligned}$$

The check consists into evaluating the target function at ρ spanning from 0 to 0.999 and x spanning from $\frac{2+\sqrt{10}}{6}$ to 0.999, choosing for each variable 1000 points between the extrema. The algorithm is the following:

```
import numpy as np

def Pi_fun(x, rho): return (1.+rho*x)/np.sqrt(1.-x**2)+np.sqrt(1.-rho
    **2)

def Pi1_fun(x, rho): return x+rho

def Pi2_fun(x): return (1.-x**2)**(3./2.)

def Pi3_fun(x): return -3.*x*(1-x**2)**2

def Pi4_fun(x): return 3.*(5.*x**2-1.)*(1.-x**2)**(5./2.)


def eta_fun(x, rho): return (1. + np.sqrt((1.-x**2)/(1.-rho**2)))/2.

def eta_der_fun(x, rho): return -x/(2.*np.sqrt(1.-rho**2)*np.sqrt(1.-
    x**2))

def eta_der2_fun(x, rho): return -1./(2.*np.sqrt(1.-rho**2)*(1.-x**2)
    **(3./2.))
```

```
def j_fun(x, rho): return (x+rho)/4.

def j_der_fun(): return 1./4.

def j2_fun(x, rho): return Pi2_fun(x)-Pi1_fun(x,rho)**2/(2.*Pi_fun(x,
    rho))

def j2_der2_fun(x, rho):

Pi = Pi_fun(x, rho)

Pi1 = Pi1_fun(x, rho)

Pi2 = Pi2_fun(x)

Pi3 = Pi3_fun(x)

j22 = Pi4_fun(x) - Pi2**2/Pi - Pi1*Pi3/Pi + 5.*Pi1**2*Pi2/(2.*Pi**2)
    - Pi1**4/Pi**3

j21 = Pi3 - Pi1*Pi2/Pi + Pi1**3/(2.*Pi**2)

return j22 + 3.*x*np.sqrt(1.-x**2)*j21


def n_fun(x, b, rho):

eta = eta_fun(x,rho)

return -2*(eta_der_fun(x,rho)**2+eta*eta_der2_fun(x,rho)-b**2*
    j_der_fun()**2)*j2_fun(x,rho) +\\
(eta**2-b**2*j_fun(x,rho)**2)*j2_der2_fun(x,rho)

def n_fun_b_fixed(x, rho):

return n_fun(x,2./(1.+rho),rho)


def check_unicity():

rho_check = np.linspace(0.,0.999,1000)
x_check = np.linspace((2.+np.sqrt(10.))/6.,0.999,1000)

rho_v, x_v = np.meshgrid(rho_check,x_check)

x =  np.sum([n_fun_b_fixed(x_v,rho_v)<0.])

if x == 0.: return 'There is unicity'
```

```
else: return 'No-unicity'
```

The result is that at the chosen points, the function is positive, indeed the command `check_unicity()`

returns:

```
'There-is-unicity'
```

Chapter 5

No arbitrage global parametrization for the eSSVI volatility surface

Abstract

This chapter describes a global and arbitrage-free parametrization of the eSSVI implied volatility surfaces introduced by [Hendriks, Martini, J. Comput. Finance, 2019]. A calibration of such surfaces has already been proposed by the quantitative research team at Zeliade Systems [Cohort, Corbetta, Laachir, Martini, Decis. Econ. Finance, 2019], but the calibration algorithm is sequential in expiries (one maturity is calibrated after the other) and lacks of a global view on the surface. The alternative calibration suggested in this chapter targets all maturities at once and always guarantees an arbitrage-free fit of market data.

From:
A. Mingone, *No arbitrage global parametrization for the eSSVI volatility surface*, Quantitative Finance, 22(12), 2205-2217, 2022.

5.1 Structure of the chapter

This chapter focuses on the SSVI model studied in Chapter 4 and [54], enriched to a surface as in [44]. In particular, we firstly analyze the conditions of absence of Calendar spread and Butterfly arbitrage for the SSVI model in Section 5.2. Then, in Section 5.3, we define the new global parametrization and prove it automatically satisfies the above conditions. In Section 5.4 we detail a calibration algorithm and show calibration results. Numerical results are on data from the Israel index TA35 and the exchange rate NIS/USD between the Israeli shekel and the American dollar and they are the outcome of a collaboration between the research team at Zeliade Systems and Tel Aviv Stock Exchange (TASE), to calibrate end-of-day implied volatility surfaces with no arbitrage. We compare these results with calibration results output by another model (that we

dub Carr-Pelts-Tehranchi and present in Section 5.4.3). Finally, we describe a way to interpolate and extrapolate SSVI parameters between different maturities, in Section 5.5.

5.2 No arbitrage conditions for the eSSVI model

The SSVI model introduced by Gatheral and Jacquier [36] is a model for the implied total variance $\sigma^2_{\text{imp}}(K,T)T$ where $\sigma_{\text{imp}}(T,K)$ is the implied volatility for a vanilla option with strike K and maturity T:

$$\text{SSVI}(K,T) = \frac{\theta(T)}{2}\big(1+\rho\varphi(T)k + \sqrt{(\varphi(T)k+\rho)^2+(1-\rho^2)}\big),$$

where k is the log-forward moneyness $k = \log\frac{K}{F_0(T)}$ and $F_0(T)$ is the forward.

It has been extended to the eSSVI model by Hendriks and Martini [44]:

$$\text{eSSVI}(K,T) = \frac{\theta(T)}{2}\big(1+\rho(T)\varphi(T)k + \sqrt{(\varphi(T)k+\rho(T))^2+(1-\rho(T)^2)}\big).$$

From this formula, the option prices can be recovered with the classic Black-Scholes formula

$$\begin{aligned} C(K,T) &= D_0(T)(F_0(T)\Phi(d_1) - K\Phi(d_2)),\\ d_1 &= \frac{\sigma_{\text{imp}}(K,T)\sqrt{T}}{2} - \frac{k}{\sigma_{\text{imp}}(K,T)\sqrt{T}},\\ d_2 &= d_1 - \sigma_{\text{imp}}(K,T)\sqrt{T}, \end{aligned}$$

setting $\sigma_{\text{imp}}(K,T) = \sqrt{\frac{\text{eSSVI}(K,T)}{T}}$, where $D_0(T)$ is the discount factor and Φ the cumulative standard normal distribution function.

Setting $\psi := \theta\varphi$, the eSSVI formula in terms of the parameters (θ,ρ,ψ) becomes

$$\text{eSSVI}(K,T) = \frac{1}{2}\big(\theta(T)+\rho(T)\psi(T)k + \sqrt{(\psi(T)k+\theta(T)\rho(T))^2+\theta(T)^2(1-\rho(T)^2)}\big).$$

5.2.1 Calendar spread arbitrage

A Calendar spread arbitrage occurs when two Calls with different maturities $T_1 < T_2$ and same moneyness are such that $C(T_1) > C(T_2)$. In general, the absence of such arbitrage is guaranteed by the requirement for the Call price function to be non-decreasing in time-to-maturity for fixed moneyness.

Given two maturities with SSVI parameters (θ_1,ρ_1,ψ_1) for the first and (θ_2,ρ_2,ψ_2) for the second, the no Calendar spread arbitrage conditions have been characterized by Hendriks and Martini in Proposition 3.5 of [44]:

- necessary conditions: $\theta_2 > \theta_1$; $\psi_2 > \psi_1 \max\big(\frac{1+\rho_1}{1+\rho_2}, \frac{1-\rho_1}{1-\rho_2}\big) \geq \psi_1$;

- sufficient conditions: the necessary conditions above and one between condition $\psi_2 \leq \frac{\psi_1}{\theta_1}\theta_2$ or condition $\left(\rho_1 - \frac{\psi_2}{\psi_1}\rho_2\right)^2 \leq \left(\frac{\theta_2}{\theta_1} - 1\right)\left(\frac{\psi_2^2\theta_1}{\psi_1^2\theta_2} - 1\right)$.

We will consider the condition $\psi_2 \leq \frac{\psi_1}{\theta_1}\theta_2$ rather than $\left(\rho_1 - \frac{\psi_2}{\psi_1}\rho_2\right)^2 \leq \left(\frac{\theta_2}{\theta_1} - 1\right)\left(\frac{\psi_2^2\theta_1}{\psi_1^2\theta_2} - 1\right)$, since it is more tractable and a natural candidate for a global parametrization. We leave anyways open the possibility of studying a new parametrization using the second condition.

5.2.2 Butterfly arbitrage

A Butterfly arbitrage arises when it is possible to build a portfolio of three Calls with same maturity and different strikes such that

$$C((1-\alpha)K_1 + \alpha K_2) > (1-\alpha)C(K_1) + \alpha C(K_2).$$

Indeed, the absence of Butterfly arbitrage is guaranteed if and only if the Call price function coming from the model is convex and bounded between the discounted Call payoff evaluated at the forward value $D_0(T)(F_0(T) - K)^+$ and the discounted forward $D_0(T)F_0(T)$.

In the case of Call prices obtained injecting an implied volatility in the Black-Scholes formula, in order to avoid Butterfly arbitrage it is of course sufficient to satisfy the requirement of convexity, see section 2.1 of [53].

Given a maturity with SSVI parameters (θ, ρ, ψ), the necessary and sufficient no Butterfly arbitrage conditions have been described in [54] and will be explained in Section 5.2.2. For efficiency reasons, it is also possible to consider a set of sufficient but not necessary no-arbitrage conditions, which are easier to compute and implement and will be presented in Section 5.2.2.

The Gatheral-Jacquier (GJ) no Butterfly arbitrage sufficient conditions

These conditions are so named since they consist of the sufficient (but not necessary) no Butterfly arbitrage conditions by Gatheral and Jacquier [36] in Theorem 4.2, together with the necessary (but not sufficient) asymptotic conditions related to the Roger Lee Moment formula [48]:

- necessary conditions: $\psi \leq \frac{4}{1+|\rho|}$;
- sufficient conditions: the necessary conditions above and $\psi^2 \leq \frac{4\theta}{1+|\rho|} := \mathfrak{f}_{GJ}(\theta, |\rho|)$.

The Martini-Mingone (MM) no Butterfly arbitrage necessary and sufficient conditions

The name of these conditions comes from the explicit no Butterfly arbitrage conditions in Proposition 2.11 of Martini and Mingone [54]. In the article, the parametrization used

is the SVI one, with implied total variance

$$\mathrm{SVI}(k) = a + b(\rho(k-m) + \sqrt{(k-m)^2 + \sigma^2}).$$

The SVI parameters are mapped to the SSVI parameters through

$$a = \frac{\theta(1-\rho^2)}{2}, \qquad b = \frac{\theta\varphi}{2} = \frac{\psi}{2},$$
$$m = -\frac{\rho}{\varphi} = -\frac{\theta\rho}{\psi}, \qquad \sigma = \frac{\sqrt{1-\rho^2}}{\varphi} = \frac{\theta\sqrt{1-\rho^2}}{\psi},$$

and viceversa

$$\varphi = \frac{\sqrt{1-\rho^2}}{\sigma}, \quad \theta = \frac{2b\sigma}{\sqrt{1-\rho^2}}$$

so that $\psi = 2b$.

The authors show that an SSVI with $b(1+|\rho|) \leq 2$, corresponding to $\psi \leq \frac{4}{1+|\rho|}$, automatically satisfies the necessary Fukasawa conditions of monotonicity of the functions

$$k \to f_{1,2}(k) := \frac{k}{\sqrt{\mathrm{SSVI}(k)}} \mp \frac{\sqrt{\mathrm{SSVI}(k)}}{2},$$

see [32]. Then, the only additional required condition is

$$\sigma \geq -\frac{b g_2(l,|\rho|)}{2(h^2(l,|\rho|) - b^2 g^2(l,|\rho|))} \tag{5.1}$$

for every $l > l_2(|\rho|) = \tan\left(\frac{\arccos(-|\rho|)}{3}\right)^{-1}$, where

$$g(l,\rho) := \frac{N'(l,\rho)}{4},$$
$$h(l,\rho) := 1 - \Big(l - \frac{\rho}{\sqrt{1-\rho^2}}\Big)\frac{N'(l,\rho)}{2N(l,\rho)},$$
$$g_2(l,\rho) := N''(l,\rho) - \frac{N'(l,\rho)^2}{2N(l,\rho)},$$
$$N(l,\rho) := \sqrt{1-\rho^2} + \rho l + \sqrt{l^2+1},$$

and derivatives are taken with respect to l. The denominator in Equation (5.1) for σ is positive thanks to the Fukasawa conditions, so that using the SSVI parameters, the inequality becomes $(\theta\sqrt{1-\rho^2}g^2(l,|\rho|) - g_2(l,|\rho|))\psi^2 \leq 4\theta\sqrt{1-\rho^2}h^2(l,|\rho|)$, and since g_2 is negative in the considered domain for l, the inequality can be written as

$$\psi^2 \leq \frac{4\theta\sqrt{1-\rho^2}h^2(l;|\rho|)}{\theta\sqrt{1-\rho^2}g^2(l;|\rho|) - g_2(l;|\rho|)}$$

for all $l > l_2(|\rho|)$. Eventually, the necessary and sufficient no Butterfly arbitrage conditions for SSVI read

$$\psi \leq \frac{4}{1+|\rho|},$$
$$\psi^2 \leq \inf_{l>l_2(|\rho|)} \frac{4\theta\sqrt{1-\rho^2}h^2(l;|\rho|)}{\theta\sqrt{1-\rho^2}g^2(l;|\rho|) - g_2(l;|\rho|)} := \mathfrak{f}_{MM}(\theta,|\rho|).$$

5.2.3 Final no Calendar spread and no Butterfly arbitrage conditions for eSSVI

All in all, the Calendar spread and Butterfly constraints for successive SSVI slices (θ_1,ρ_1,ψ_1) and (θ_2,ρ_2,ψ_2) can be summed up to:

$$\begin{aligned}
&\theta_2 > \theta_1 > 0,\\
&\psi_1 \leq \min\left(\frac{4}{1+|\rho_1|}, \sqrt{\mathfrak{f}(\theta_1,|\rho_1|)}\right),\\
&0 < \psi_1 \max\left(\frac{1+\rho_1}{1+\rho_2}, \frac{1-\rho_1}{1-\rho_2}\right) < \psi_2 \leq \min\left(\frac{\psi_1}{\theta_1}\theta_2, \frac{4}{1+|\rho_2|}, \sqrt{\mathfrak{f}(\theta_2,|\rho_2|)}\right)
\end{aligned} \tag{5.2}$$

where the function $\mathfrak{f}$ can be either from the MM model ($\mathfrak{f} = \mathfrak{f}_{MM}$) or from the GJ model ($\mathfrak{f} = \mathfrak{f}_{GJ}$).

Since the MM conditions are less strict then the GJ conditions, it could seem natural to implement the former in a calibration routine. However, in contrast with the latter ones, they are not explicit and require to use a minimization algorithm to evaluate $\mathfrak{f}_{MM}(\theta,|\rho|)$, causing an increase in calibration time.

5.3 A global parametrization

5.3.1 The case with two maturities

Let us consider the model with only two SSVI slices and in particular the conditions on the second maturity parameters. In order to achieve the condition on θ_2, we could write $\theta_2 = \theta_1 + \tilde{a}_2$ and choose $\tilde{a}_2 > 0$. The condition on ψ_2 requires that ψ_2 lies in an interval $]A_{\psi_2}, C_{\psi_2}[$ where

$$\begin{aligned}
A_{\psi_2} &:= \psi_1 p_2,\\
C_{\psi_2} &:= \min\left(\frac{\psi_1}{\theta_1}\theta_2, f_2\right),\\
p_2 &:= \max\left(\frac{1+\rho_1}{1+\rho_2}, \frac{1-\rho_1}{1-\rho_2}\right),\\
f_2 &:= \min\left(\frac{4}{1+|\rho_2|}, \sqrt{\mathfrak{f}(\theta_2,|\rho_2|)}\right),
\end{aligned}$$

so that setting $\psi_2 = c_2(C_{\psi_2} - A_{\psi_2}) + A_{\psi_2}$ with $c_2 \in]0,1[$ would guarantee the absence of arbitrage, if indeed $C_{\psi_2} > A_{\psi_2}$. However, the requirement $C_{\psi_2} > A_{\psi_2}$ is not automatically guaranteed and it depends on the conditions of the first maturity parameters. Indeed, one needs $p_2 < \frac{\theta_2}{\theta_1}$ and $\psi_1 p_2 < f_2$. The first requirement is easily attained setting $\theta_2 = p_2\theta_1 + a_2$ with $a_2 > 0$, while the second requirement is guaranteed if and only if in the calibration of ψ_1 we also impose $\psi_1 < \frac{f_2}{p_2}$. More specifically, ψ_1 can be calibrated as $\psi_1 = c_1 C_{\psi_1}$ where $c_1 \in]0,1[$ and

$$C_{\psi_1} := \min\left(\frac{4}{1+|\rho_1|}, \sqrt{\mathfrak{f}(\theta_1, |\rho_1|)}, \frac{f_2}{p_2}\right).$$

5.3.2 The general case

The Global eSSVI is a model which uses a new global parametrization for a set of consecutive SSVI slices, satisfying the above no-arbitrage conditions. Given N maturities, the parametrization involves $3 \times N$ parameters as the eSSVI classical one (since it is a more practical re-parametrization of it).

The new parameters are

$$\rho_1, \dots, \rho_N, \theta_1, a_2, \dots, a_N, c_1, \dots, c_N \in]-1,1[^N \times]0,\infty[^N \times]0,1[^N \tag{5.3}$$

where the ρ_i are the original eSSVI parameters while the a_i and c_i are defined as

$$\begin{aligned} a_i &= \theta_i - \theta_{i-1} p_i, \\ c_i &= \frac{\psi_i - A_{\psi_i}}{C_{\psi_i} - A_{\psi_i}}, \end{aligned}$$

and

$$\begin{aligned} p_i &:= \max\left(\frac{1+\rho_{i-1}}{1+\rho_i}, \frac{1-\rho_{i-1}}{1-\rho_i}\right) && \text{if } i > 1, \\ f_i &:= \min\left(\frac{4}{1+|\rho_i|}, \sqrt{\mathfrak{f}(\theta_i, |\rho_i|)}\right) && \text{if } i \geq 1, \\ A_{\psi_1} &:= 0 && \text{if } i = 1, \\ A_{\psi_i} &:= \psi_{i-1} p_i && \text{if } i > 1, \\ C_{\psi_1} &:= \min\left(f_1, \frac{f_2}{p_2}, \dots, \frac{f_N}{\prod_{j=2}^N p_j}\right) && \text{if } i = 1, \\ C_{\psi_i} &:= \min\left(\frac{\psi_{i-1}}{\theta_{i-1}}\theta_i, f_i, \frac{f_{i+1}}{p_{i+1}}, \dots, \frac{f_N}{\prod_{j=i+1}^N p_j}\right) && \text{if } i > 1. \end{aligned} \tag{5.4}$$

The original SSVI parameters are sequentially obtained through the ordered relations

$$\begin{aligned} &\theta_2 = \theta_1 p_2 + a_2, \quad \dots, \quad \theta_N = \theta_{N-1} p_N + a_N, \\ &\psi_1 = c_1(C_{\psi_1} - A_{\psi_1}) + A_{\psi_1}, \quad \dots, \quad \psi_N = c_N(C_{\psi_N} - A_{\psi_N}) + A_{\psi_N}. \end{aligned} \tag{5.5}$$

It can be useful to note that:

- $p_i \geq 1$ with $p_i = 1$ iff $\rho_i = \rho_{i-1}$;
- $0 < f_i \leq 4$.

The following result guarantees that the Global eSSVI parametrization has no arbitrage.

Proposition 5.1. *For any integer $N > 0$ and any parameters*

$$\rho_1, \dots, \rho_N, \theta_1, a_2, \dots, a_N, c_1, \dots, c_N \in]-1, 1[^N \times]0, \infty[^N \times]0, 1[^N,$$

the set of N SSVI slices

$$eSSVI_i(K, T_i) = \frac{1}{2}\big(\theta_i + \rho_i\psi_i k + \sqrt{(\psi_i k + \theta_i\rho_i)^2 + \theta_i^2(1-\rho_i^2)}\big),$$

with today's forward $F_0(T_i)$ for increasing maturities T_i, log-forward moneyness $k = \log\frac{K}{F_0(T_i)}$ and parameters $(\theta_i, \rho_i, \psi_i)$ defined through Equation (5.4) and Equation (5.5), is free of Butterfly and Calendar spread arbitrage.

Proof. We now prove that with such parametrization the arbitrage constraints in Equation (5.2) are satisfied. Indeed,

- $\theta_1 > 0$ is chosen from the start, and $\theta_i = \theta_{i-1}p_i + a_i > \theta_{i-1}p_i \geq \theta_{i-1}$;
- $\psi_1 = c_1 C_{\psi_1} > 0$ and $\psi_1 \leq c_1 f_1 < f_1$ (by assumption on the domain of c_1), which is the no Butterfly arbitrage for the first maturity.

We need to show that if $\psi_i = c_i(C_{\psi_i} - A_{\psi_i}) + A_{\psi_i}$, then

$$A_{\psi_i} < \psi_i \leq \min\left(\frac{\psi_{i-1}}{\theta_{i-1}}\theta_i, f_i\right).$$

We have already seen that this holds true for $i = 1$, where it holds even that $A_{\psi_1} < \psi_1 < C_{\psi_1}$. We now show by induction that if $A_{\psi_{i-1}} < \psi_{i-1} < C_{\psi_{i-1}}$ then $A_{\psi_i} < C_{\psi_i}$. This will ensure the no-arbitrage condition, since then, by definition, $\psi_i = c_i(C_{\psi_i} - A_{\psi_i}) + A_{\psi_i}$ so that $A_{\psi_i} < \psi_i < C_{\psi_i}$ and also $C_{\psi_i} = \min\left(\frac{\psi_{i-1}}{\theta_{i-1}}\theta_i, f_i, \frac{f_{i+1}}{p_{i+1}}, \dots, \frac{f_N}{\prod_{j=i+1}^N p_j}\right) \leq \min\big(\frac{\psi_{i-1}}{\theta_{i-1}}\theta_i, f_i\big)$.

For induction, we suppose we have proven $A_{\psi_{i-1}} < C_{\psi_{i-1}}$. If $C_{\psi_i} = \frac{\psi_{i-1}}{\theta_{i-1}}\theta_i$, the requirement $\theta_i > \theta_{i-1}p_i$ implies $A_{\psi_i} = \psi_{i-1}p_i < C_{\psi_i}$. Otherwise, the inequality holds true iff $\psi_{i-1} < \frac{1}{p_i}\min\big(f_i, \frac{f_{i+1}}{p_{i+1}}, \dots, \frac{f_N}{\prod_{j=i+1}^N p_j}\big)$. By the above consequence of the induction hypothesis,

$$\begin{aligned}
\psi_{i-1} < C_{\psi_{i-1}} &= \min\left(\frac{\psi_{i-2}}{\theta_{i-2}}\theta_{i-1}, f_{i-1}, \frac{f_i}{p_i}, \dots, \frac{f_N}{\prod_{j=i}^N p_j}\right) \\
&\leq \min\left(\frac{f_i}{p_i}, \frac{f_{i+1}}{p_i p_{i+1}}, \dots, \frac{f_N}{\prod_{j=i}^N p_j}\right) \\
&= \frac{1}{p_i}\min\left(f_i, \frac{f_{i+1}}{p_{i+1}}, \dots, \frac{f_N}{\prod_{j=i+1}^N p_j}\right).
\end{aligned}$$

This shows the induction step. Observe that the case $i = N$ is easily proven by observing that $C_{\psi_N} = \min\Big(\frac{\psi_{N-1}}{\theta_{N-1}}\theta_N, \frac{4}{1+|\rho_N|}, \sqrt{\mathfrak{f}(\theta_N, |\rho_N|)}\Big)$. □

The above proof shows that consecutive slices are free of arbitrage and the result can be immediately extended from two consecutive slices to any pair of slices.

5.4 Calibration strategy

The Global eSSVI parametrization can be easily implemented for calibration purposes. The calibration function can either target the market total implied variance or the market option prices. Indeed, model prices can be easily recovered through the Black-Scholes formula, given implied volatility $\sqrt{\frac{\text{eSSVI}(K,T)}{T}}$.

We performed tests on real market data on two different assets: the Israel index TA35 and the exchange rate NIS/USD between the Israeli shekel and the American dollar. We describe the procedure which led to the calibration results shown in Section 5.4.3.

5.4.1 Routine

The eSSVI calibration consists of finding the parameters $\{\rho_i\}_{i=1}^N, \theta_1, \{a_i\}_{i=2}^N, \{c_i\}_{i=1}^N$ such that the model eSSVI prices best match the basket of available option prices. In our experiments, parameters are chosen to minimize the possibly weighted squared difference between market option prices $\tilde{C}(K,T)$ and model option prices $C(K,T)$:

$$\sum_{K,T}\big(\tilde{C}(K,T) - C(K,T)\big)^2\omega(K,T).$$

Weights ω can be arbitrarily chosen by the user. In particular, they can be chosen to match the inverse of squared market Black-Scholes vegas, in order to have a calibration in implied volatilities at the first order, instead of a calibration in prices.

Recall that the minimum search can be performed over the product of intervals Equation (5.3), which is an appealing feature when using standard minimization routines such as the `least_squares` function. In our tests, the calibration is performed using the `least_squares` function in the `scipy.optimize` library. The maximum number of function evaluations is set at 1000 and the argument for convergence `ftol` is set at its default value 10^{-8}.

Observe that in general, trades in the market have different timestamps, so that available option prices are not simultaneous. Then when computing model prices starting from eSSVI new global parameters as described in Algorithm 3, these prices are evaluated using the forward computed at the timestamp of the corresponding market option, so it could be possibly different for different options. This is linked to the fact that model parameters are supposed to be constant in logforward-moneyness. In particular, given a market option with strike K and maturity T traded (or quoted) at timestamp t, the

corresponding model price is:

$$\begin{aligned} C(K,T) &= D_C(T)(F_t(T)\Phi(d_1) - K\Phi(d_2)), \\ d_1 &= -\frac{k}{\sigma_{\text{imp}}(K,T)\sqrt{T-t}} + \frac{\sigma_{\text{imp}}(K,T)\sqrt{T-t}}{2}, \\ d_2 &= d_1 - \sigma_{\text{imp}}(K,T)\sqrt{T-t}, \\ \sigma_{\text{imp}}(K,T)\sqrt{T-t} &= \sqrt{\frac{\theta(T) + \rho(T)\psi(T)k + \sqrt{(\psi(T)k + \theta(T)\rho(T))^2 + \theta(T)^2(1-\rho(T)^2)}}{2}}, \\ k &= \log\frac{K}{F_t(T)}, \end{aligned} \tag{5.6}$$

where $D_C(T)$ is the discount factor at closing time (here we suppose it does not change a lot during the day) and $F_t(T) = F_C(T)\frac{S_t}{S_C}$ is the current forward.

Algorithm 3 Computation of model prices from eSSVI new global parameters

Input Parameters $\{\rho_i\}_{i=1}^N, \theta_1, \{a_i\}_{i=2}^N, \{c_i\}_{i=1}^N$; maturities $\{T_i\}_{i=1}^N$; set of strikes $\{\mathcal{K}_i\}_{i=1}^N$.

Output Model prices $\{C(K,T_i)\}_{K\in\mathcal{K}_i, i=1\dots N}$.

1: With input parameters $\rho_1,\dots,\rho_N$ compute intermediate quantities $p_2,\dots,p_N$ in Equation (5.4);
2: With input parameters $\theta_1, a_2,\dots,a_N$ compute from Equation (5.5) and the p_i the model parameters $\theta_2,\dots,\theta_N$;
3: Compute from the θ_i and the ρ_i the intermediate quantities $f_1,\dots,f_N$;
4: Compute the intermediate quantity C_{ψ_1};
5: With input parameter c_1 compute the model parameter ψ_1 from c_1, C_{ψ_1} and $A_{\psi_1} = 0$;
6: **for** $i = 2,\dots,N$ **do**:
7: With input parameter c_i compute the intermediate quantities A_{ψ_i}, C_{ψ_i},
8: Compute the model parameter ψ_i;
9: **for** $K \in \mathcal{K}_i$ **do**
10: Compute the model price $C(K,T_i)$ from Equation (5.6).
11: **end for**
12: **end for**

5.4.2 Parameters domain and initial conditions

We set the following initial conditions:

- the a parameters are obtained guessing initial values for the θs from the ATM total implied variances;
- initial ρs are set to the intermediate value 0;

- initial cs are set to 0.5.

The a parameters could lie in an infinite range but, for optimization reasons, it is a good practice to bound them. After many tests on TA35 and NIS/USD data, we chose to impose a to be smaller than 0.05. If the maximum of the initial values for the as is larger than the fixed upper bound, we double the latter bound. The other parameters are already bounded from the definition of the Global eSSVI model, indeed $\rho \in]-1, 1[$ and $c \in]0, 1[$.

The calibration weights $\omega(K, T)$ are taken to be constant.

5.4.3 Numerical experiments

We show here the numerical results obtained with the GJ conditions for the Global eSSVI parametrization and compare them with the well-know rich and flexible parametric price surface of Carr and Pelts, which we describe in the following section.

The Carr-Pelts-Tehranchi model

Carr and Pelts presented in 2015, at a conference in honour of Steven Shreve at Purdue university (see e.g. [15]), an explicit arbitrage-free parametrization for FX option prices. In 2019, in a deep and brilliant paper on a subtle property of the Black-Scholes formula [66], Mike Tehranchi re-discovered independently this family of models, with the more mathematical perspective of semi-groups acting on sets of convex functions (Calls and Puts normalized prices). Therefore, from now on we name this model using the acronym CPT.

Unlike eSSVI, CPT gives a direct formula for the vanilla price, not its implied volatility. The implied volatility is not a natural object in the CPT family (except of course in the case of the Black-Scholes model itself, which indeed belongs to this family); if one needs to get the implied volatility, it is required to resort to numerical algorithm like the excellent rationale approach by Jaeckel.

In order to perform calibrations, we use the approach accurately described in section 3 of [5]. Denote S_t the current value of the underlyer (in case of dividends and rates not null, S_t should be replaced by the Forward F_t) and pick up a log-concave density $f = e^{-h}$ on $\mathbb{R}$. Then, under the risk-neutral measure, the law of the underlyer S at fixed maturity will be given by

$$S^{(\tau)} := S_t \frac{f(\tau + Z)}{f(Z)}$$

where Z is a random variable with law f, and τ some real positive parameter. It holds

$$E[(S^{(\tau)} - K)^+] = S_t \int_R f(\tau + z)dz - K \int_R f(z)dz$$

where $R := \left\{ z \,\middle|\, \frac{f(\tau+z)}{f(z)} \geq \frac{K}{S_t} \right\}$. Observe now that $z \to \frac{f(\tau+z)}{f(z)}$ is non-increasing, indeed $\frac{d}{dz}\left(\log \frac{f(\tau+z)}{f(z)}\right) = -h'(\tau + z) + h'(z)$ where $h = -\log f$ is convex (so its second derivative

is positive). It follows that

$$d_f(\tau, k) := \sup\Big\{z \;\Big|\; \log \frac{f(\tau + z)}{f(z)} = k\Big\} = \sup\big\{z \;\big|h(\tau + z) - h(z) = -k\big\}$$

is well defined and that $R = \{z \leq d_f(\tau, k)\}$, where $k = \log\big(\frac{K}{S_t}\big)$. Eventually

$$E[(S^{(\tau)} - K)^+] = S_t\Omega(d_f(\tau, k) + \tau) - K\Omega(d_f(\tau, k))$$

where Ω is the cumulative density function of f.

All the above generalizes the Black-Scholes formula, which corresponds to the particular case $f(x) = \frac{1}{\sqrt{2\pi}} \exp\big(\frac{-x^2}{2}\big)$.

The second ingredient to CPT is based on the remarks that:

1. if $\tau < z$, then for every $K > 0$, $E[(S^{(\tau)} - K)^+] < E[(S^{(z)} - K)^+]$.
2. $S_t\Omega(d_f(\tau, k) + \tau) - K\Omega(d_f(\tau, k)) = (S_t - K)^+$ iff $\tau = 0$.

It follows that if one chooses any non-decreasing continuous function $T \to \tau(T)$ such that $\tau(0) = 0$, then the price surface:

$$(K, T) \to S_t\Omega(d_f(\tau(T), k) + \tau(T)) - K\Omega(d_f(\tau(T), k))$$

is free of arbitrage.

In Antonov et al. specification, the function τ is a piecewise-linear function and h a piecewise-quadratic differentiable convex function. It is calibrated using a grid of $2N_{CPT}$ node points. Then, the algorithm requires to make a choice on the number of nodes used to calibrate the model density. There is always a trade-off between taking N_{CPT} large, which could allow for an increased fitting ability at the price of more instability in the results if there are too few options in the calibration basket, and choosing a smaller N_{CPT} with the opposite benefits or issues. A rational start is to compare eSSVI and CPT number of parameters; eSSVI has $3N$ parameters, while CPT has $N + 2N_{CPT}$ parameters, so equating them gives $N_{CPT} = N$. In practice, on both the TA35 market and the NISUSD one, the choice of $N_{CPT} = 6$ gives very good fit results and it is coherent with the fact that data always consists of no more than 7 maturities.

Price and volatility plots

The calibration algorithm has as a target the market prices. First, we show the Call and Put prices on the date 2021/10/26 for both the TA35 index (spot of 1871.67) and the NIS/USD Forex (spot of 319.98). The calibration basket is composed of both trade and quote prices in the last 10 minutes of trading, filtered to remove noisy data and aggregated to a synthetic market price per option. The GJ Global eSSVI and the CPT models are calibrated and the corresponding model prices are shown in Figures 5.1 and 5.2 with the following notation:

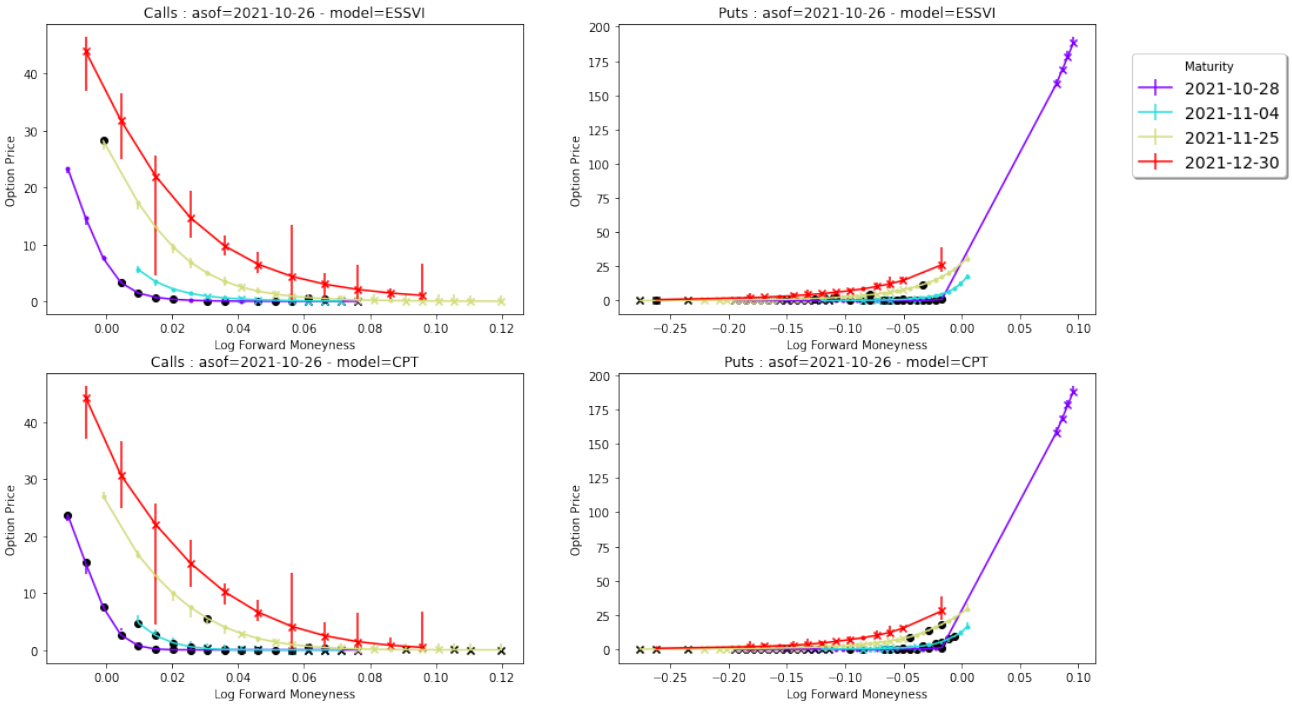

Figure 5.1: Comparison of calibrated Call and Put prices on TA35 with the eSSVI and the CPT models. Markers × indicate quotes, markers · indicate trades, black markers indicate model prices outside the bid-ask.

- the marker × indicates a quote;
- the marker · indicates a trade;
- the vertical line indicates the bid-ask prices;
- a black marker indicates a model price outside the bid-ask.

Figures 5.1 and 5.2 can be summarized into Tables 5.1 and 5.2 comparing the percentage of calibrated prices falling outside the bid-ask spread and outside twice the bid-ask spread. The number of Calls, respectively Puts, to be calibrated is 64, respectively 107, for TA35 and 16, respectively 12, for NISUSD. Observe that, for the TA35 index, the number of calibrated prices falling outside the bid-ask spread is primarily due to the first slice, whose time to maturity is 2 days and whose bid-ask spreads are very small. Indeed, among the prices outside the bid-ask spread for the eSSVI model, 81% of Calls and 63% of Puts come from the first maturity, and among the prices outside twice the bid-ask spread these percentages arise to 83% and 89%. We retrieve the well-known fact that very short maturities are difficult to be calibrated with parametric models.

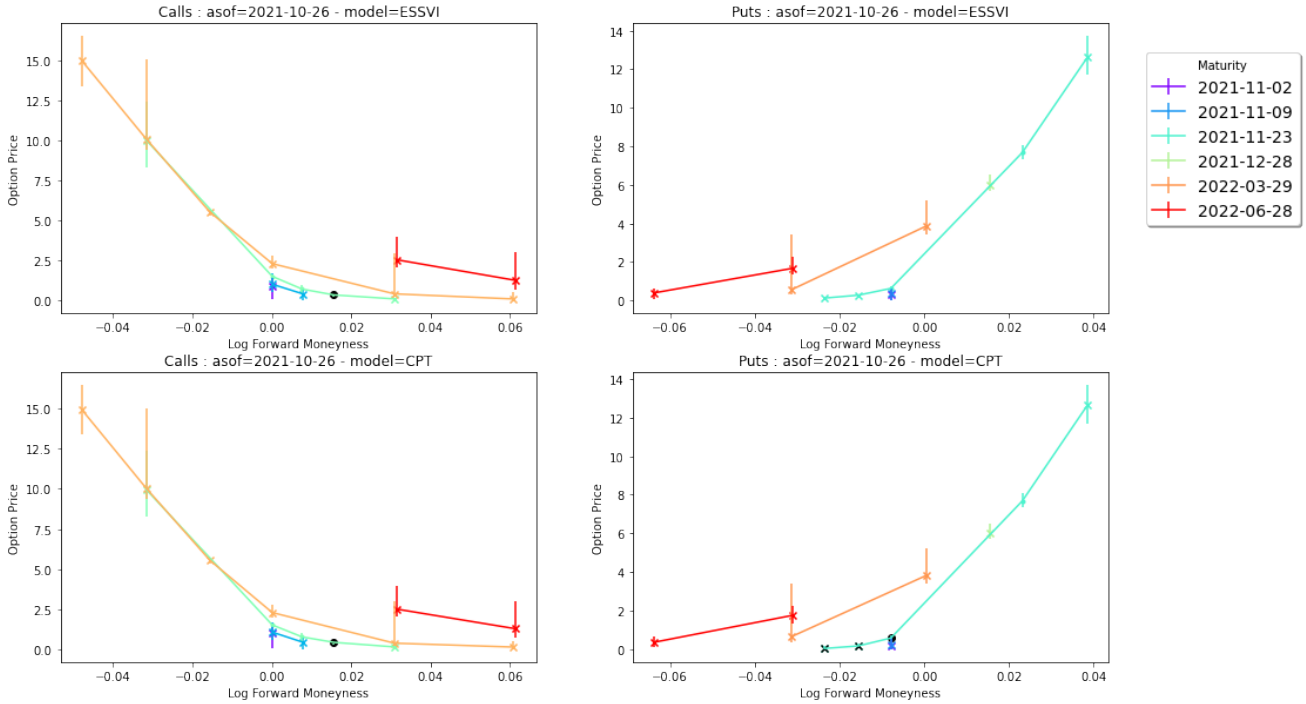

Figure 5.2: Comparison of calibrated Call and Put prices on NIS/USD with the eSSVI and the CPT models. Markers × indicate quotes, markers · indicate trades, black markers indicate model prices outside the bid-ask.

TA35	Calls		Puts	
eSSVI	25.00	9.38	37.38	17.76
CPT	57.81	26.56	34.58	14.02

Table 5.1: Percentage of calibrated Calls and Puts outside the bid-ask spread and outside twice the bid-ask spread for TA35.

NIS/USD	Calls		Puts	
eSSVI	6.25	0.00	0.00	0.00
CPT	6.25	6.25	25.00	8.33

Table 5.2: Percentage of calibrated Calls and Puts outside the bid-ask spread and outside twice the bid-ask spread for NIS/USD.

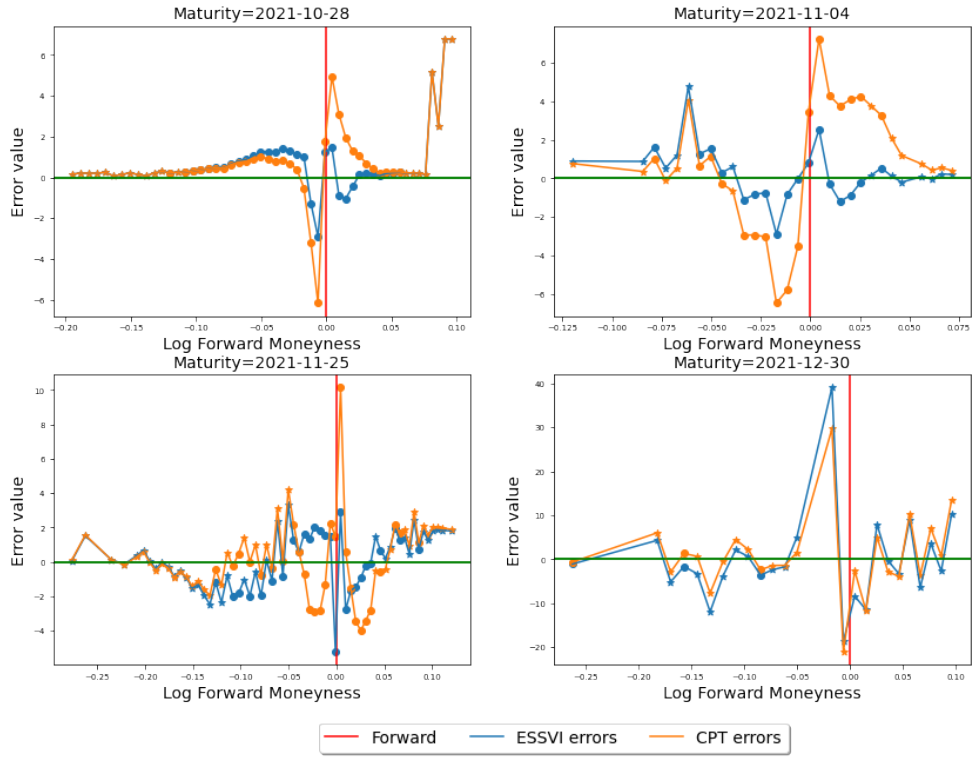

Figure 5.3: Comparison of calibration errors in basis points to the Forward for OTM options, for TA35 in the eSSVI and the CPT models.

The corresponding absolute errors in basis point to the Forward are shown in Figures 5.3 and 5.4.

For those plots, we also report the corresponding implied volatilities in Figures 5.5 and 5.6.

Calibration results are very satisfactory for the GJ Global eSSVI model since they are comparable if not even better than the CPT model's ones. In particular, typically the tested markets have an average calibration error (in absolute value) of 10 basis points to the Forward (for more liquid markets, the average calibration error can decrease to 4 basis points to the Forward). In the above tests, the eSSVI average calibration errors are 1.92 basis points to the Forward for TA35 and 1.29 for NIS/USD, while the CPT ones are 2.19 for TA35 and 1.86 for NIS/USD. The calibration of the shortest maturity is typically more difficult in terms of error reduction, which explains the high number of black dots in the Call and Put prices graphs in Figures 5.1 and 5.2, as also quantified in Tables 5.1 and 5.2. However, the ATM prices are always well calibrated.

In general, the GJ Global eSSVI calibration is preferable to the CPT for four reasons:

- from a theoretical point of view, it is easier to understand the role of its parameters

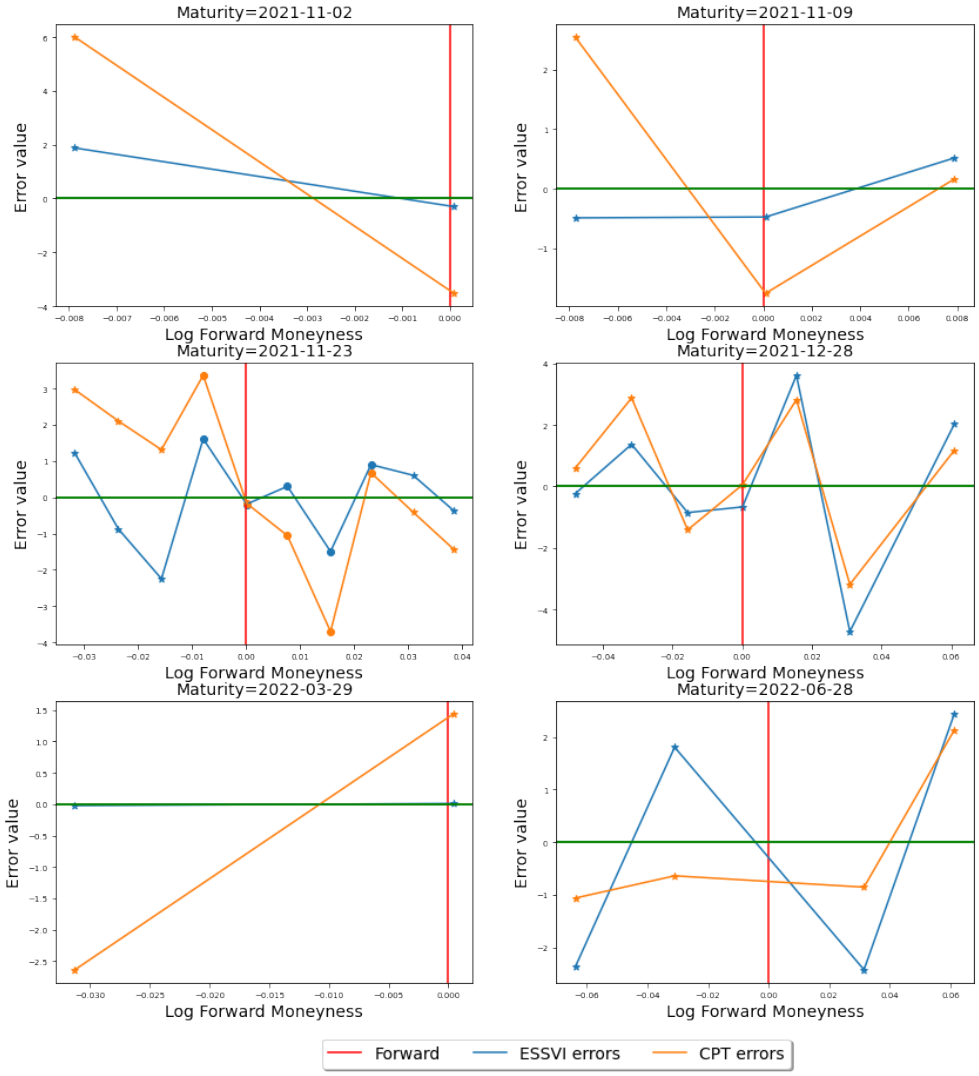

Figure 5.4: Comparison of calibration errors in basis points to the Forward for OTM options, for NIS/USD in the eSSVI and the CPT models.

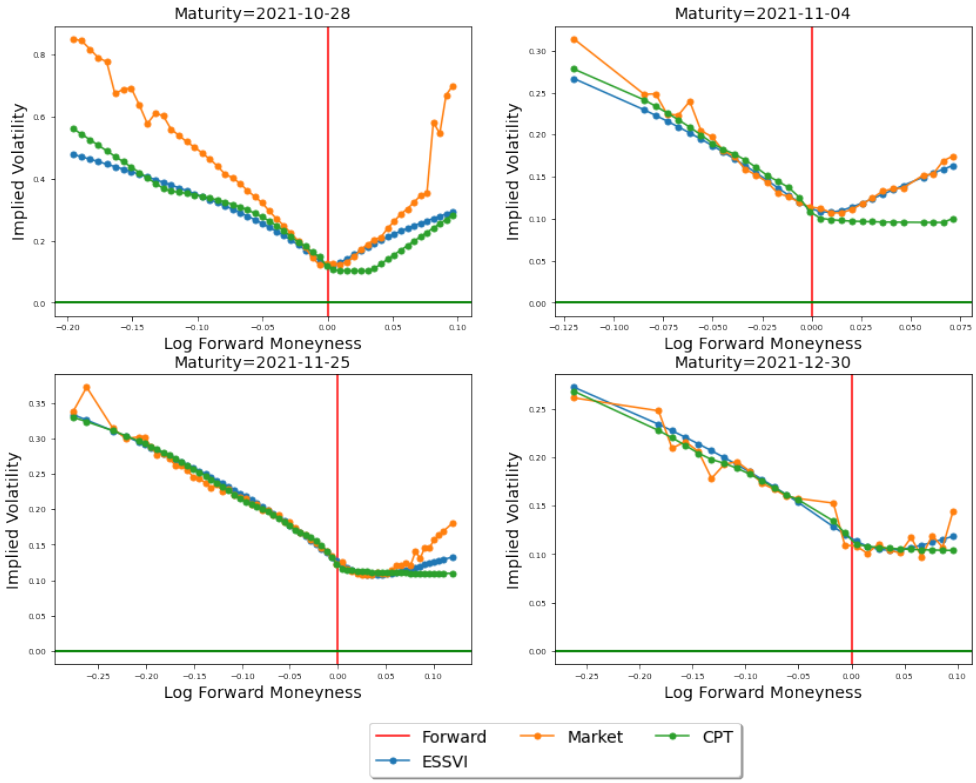

Figure 5.5: Comparison of calibrated smiles for TA35 in the eSSVI and the CPT models.

and how to tune them in order to change the smile shape;

- it directly models a volatility surface, so that the calibrated implied smiles have more natural and desirable shapes. Also, working with implied volatilities allows to compare between underlyers and dates, while directly comparing prices is less obvious;
- it is much easier and straightforward to be coded;
- calibration times are 10 times smaller than CPT calibration times.

5.4.4 No-arbitrage check

As a sanity check of the code implementation, we implemented an arbitrage identifier routine as described in Section 3 of [17]. It is worth noticing that such routine is not necessary since models have been shown to be arbitrage-free. However, numerical approximations could arise in computed prices, even though calibration parameters live in the no-arbitrage domain.

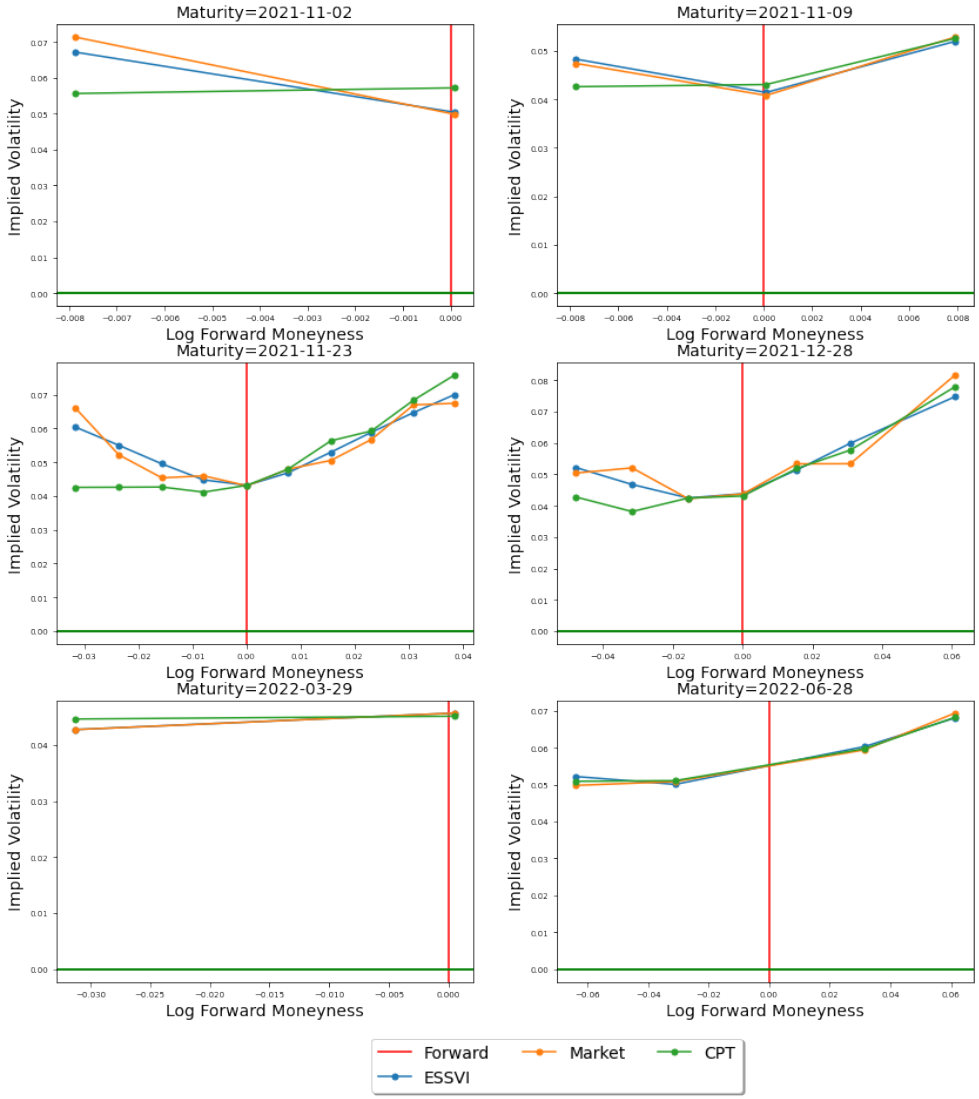

Figure 5.6: Comparison of calibrated smiles for NIS/USD in the eSSVI and the CPT models.

We take all model prices resulting from the calibration routine and verify (with linear constraints) whether there is any kind of arbitrage. In particular, we check the positivity of prices and look for what Reisinger et al. [17] call Vertical Spread, Vertical Butterfly, Calendar Spread, Calendar Vertical Spread, and Calendar Butterfly arbitrage.

We run the Reisenger algorithm on all calibration results for trading dates from 2019/01/01 to 2021/10/27 on both TA35 and NIS-USD. Arbitrage check outputs confirm the lack of arbitrage in prices resulting from both the Global eSSVI and the CPT models. Some little arbitrage caused by numerical approximations could arise with the former model for values of ρ near ± 1. A way to avoid this is to bound ρ in a smaller interval, such as $]-0.95, 0.95[$. Calibration results in terms of overall calibration error do not suffer from this choice.

5.5 Interpolation and extrapolation

This section describes how to interpolate and extrapolate eSSVI parameters on maturities which differ from the ones used in the calibration, in order to guarantee the absence of arbitrage. These methodologies are taken from [20].

5.5.1 Interpolation

Suppose we have calibrated the model on two maturities $T_1 < T_2$ with eSSVI parameters $(\theta_1, \rho_1, \psi_1)$ for the first one and $(\theta_2, \rho_2, \psi_2)$ for the second. How could we interpolate arbitrage-free parameters for a maturity $t \in [T_1, T_2]$? Similarly to what is done in section 5.1.2 of [20], we define the new parameters $(\theta_t, \rho_t, \psi_t)$ through the scheme:

- $\theta_t = (1-\lambda)\theta_1 + \lambda\theta_2$
- $\psi_t = (1-\lambda)\psi_1 + \lambda\psi_2$
- $\psi_t\rho_t = (1-\lambda)\psi_1\rho_1 + \lambda\psi_2\rho_2$

where $\lambda = \frac{t-T_1}{T_2-T_1}$. We now show that in such way arbitrage conditions are satisfied.

We look at the Calendar Spread conditions and take $T_1 \leq t < u \leq T_2$ with $\lambda = \frac{t-T_1}{T_2-T_1}$ and $\mu = \frac{u-T_1}{T_2-T_1}$. First, it is immediate that $\theta_u > \theta_t$. Second, the sufficient condition $\psi_u\theta_t - \psi_t\theta_u \leq 0$ can be rewritten as $(\mu-\lambda)(\psi_2\theta_1 - \psi_1\theta_2) \leq 0$, which is verified since the two calibrated slices are arbitrage-free. Last, the second necessary condition is equivalent to $\psi_u(1 \pm \rho_u) - \psi_t(1 \pm \rho_t) > 0$. Substituting and simplifying as above, we find $(\mu-\lambda)(\psi_2(1 \pm \rho_2) - \psi_1(1 \pm \rho_1)) > 0$, which again holds true.

The Butterfly arbitrage conditions at time t are proven in [20].

5.5.2 Extrapolation

The extrapolation procedure is also taken from [20], sections 5.2 and 5.3. In the following, for completeness we give full proofs of the consistency of this choice.

Before the first maturity

For t smaller than the first calibrated maturity T_1, we set

- $\theta_t = \lambda\theta_1$
- $\psi_t = \lambda\psi_1$
- $\rho_t = \rho_1$

where $\lambda = \frac{t}{T_1} < 1$.

It is easy to verify that the fact that the parameters θ and ψ are increasing while ρ is constant, combined with the arbitrage-free calibration of these parameters on the first maturity, guarantees the absence of arbitrage for the extrapolated triple $(\theta_t, \rho_t, \psi_t)$. Indeed, the absence of arbitrage is easily checked observing that the implied volatility on maturities t and T_1 is the same for fixed k, so that total variances are increasing with respect to maturity and cannot have Butterfly arbitrage, given that the smile on T_1 is Butterfly arbitrage-free.

To show that the implied volatilities coincide, we look at the eSSVI formula on maturity T_1

$$\text{eSSVI}(K, T_1) = \frac{1}{2}\big(\theta_1 + \rho_1\psi_1 k + \sqrt{(\psi_1 k + \theta_1\rho_1)^2 + \theta_1^2(1-\rho_1^2)}\big)$$

and on maturity t

$$\text{eSSVI}(\tilde{K}, t) = \frac{1}{2}\big(\lambda\theta_1 + \rho_1\lambda\psi_1\tilde{k} + \sqrt{(\lambda\psi_1\tilde{k} + \lambda\theta_1\rho_1)^2 + (\lambda\theta_1)^2(1-\rho_1^2)}\big).$$

It is easy to see that if $k = \tilde{k}$, corresponding to $\tilde{K} = K\frac{F_0(t)}{F_0(T_1)}$, then $\text{eSSVI}(\tilde{K}, t) = \lambda\text{eSSVI}(K, T_1) = \frac{t}{T_1}\text{eSSVI}(K, T_1)$ and the conclusion immediately follows.

After the last maturity

Extrapolation on the right of the last calibrated maturity T_N is performed setting

- $\theta_t = \theta_N + \frac{\theta_N - \theta_{N-1}}{T_N - T_{N-1}}(t - T_N)$
- $\psi_t = \psi_N$
- $\rho_t = \rho_N$

where $t > T_N$. The first bullet point is such that θ_t preserves the last slope available, but it can be replaced with any positive angular coefficient.

Since ψ and ρ are constant, θ is increasing and the parameters satisfy no-arbitrage conditions at maturity T_N, it is straightforward to show that the extrapolated parameters have no arbitrage.

TA35	2021/10/28	2021/11/04	2021/11/25	2021/12/30
Excluded	77.36	40.00	52.38	28.00
Not excluded	71.70	16.67	19.05	4.00

Table 5.3: Percentage of calibrated prices (Calls and Puts) outside the bid-ask spread for TA35. Comparison between the calibration where the maturity is excluded (hence interpolated or extrapolated) and the calibration where it is not excluded.

TA35	2021/10/28	2021/11/04	2021/11/25	2021/12/30
Excluded	1.29	2.57	4.76	10.64
Not excluded	0.93	0.92	1.33	6.70

Table 5.4: Average calibration absolute error in basis points to the Forward for OTM options for TA35. Comparison between the calibration where the maturity is excluded (hence interpolated or extrapolated) and the calibration where it is not excluded.

5.5.3 Numerical experiments on interpolation and extrapolation

This section aims to verify the efficiency of the interpolation and extrapolation methodologies. We remark that it is out of the main scope of this paper and that results can be improved by choosing other interpolation and extrapolation routines still verifying the no-arbitrage conditions.

For each of the 4 maturities of the TA35 index on day 2021/10/26, we calibrate eSSVI parameters excluding such maturity and then interpolate/extrapolate them for the considered maturity. In order to numerically appreciate the interpolation/extrapolation capacity of the model, we use two metrics:

- the percentage of prices (both Calls and Puts) outside the bid-ask spread on the excluded maturity (Table 5.3);
- the average calibration absolute error in basis points to the Forward for OTM options on the excluded maturity (Table 5.4).

We compare these metrics with the corresponding metrics computed on the same maturity but using the original calibration of Section 5.4.3, with no maturity excluded.

We can see from Tables 5.3 and 5.4 that interpolation and extrapolation methodologies can sometimes significantly decrease the precision of the calibration, which is somehow expected. However, the result highly depends on data and in particular on the number of maturities, on the number of traded options of non-excluded maturities, on the number of traded options of the excluded maturity. Overall, the methodologies reported in Sections 5.5.1 and 5.5.2 provide reasonable results, while being very simple. This is confirmed by looking at interpolated/extrapolated smiles plots in Figure 5.7, whose shapes are well-captured and still consistent with the market ones. As expected,

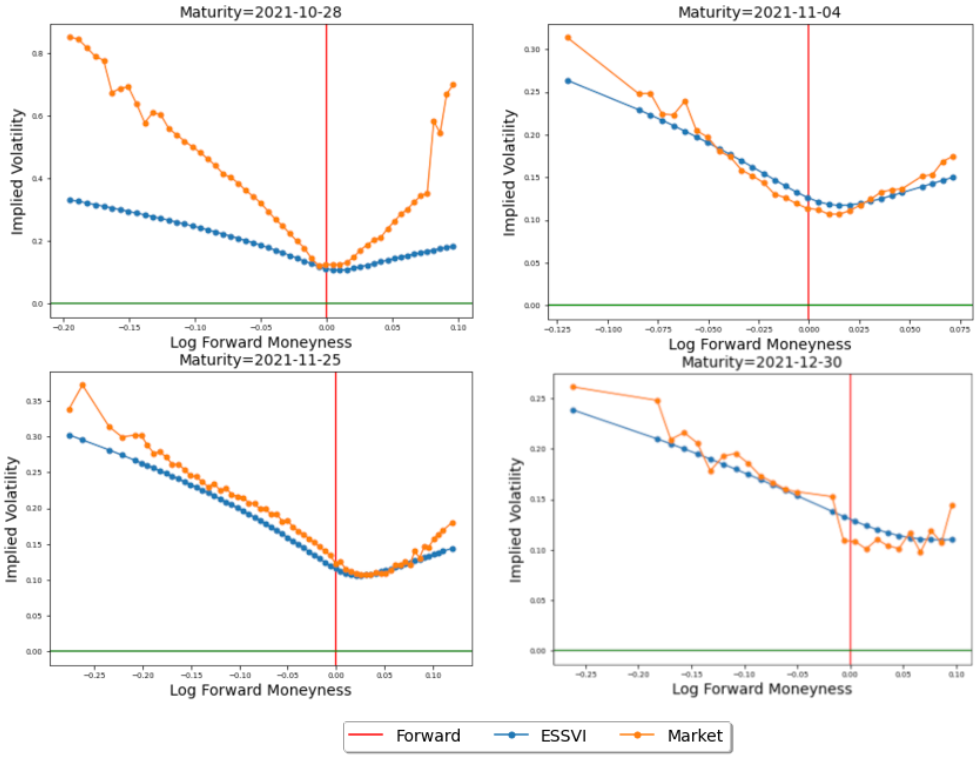

Figure 5.7: Plot of the results of the interpolation/extrapolation procedures described in Sections 5.5.1 and 5.5.2 for TA35. More precisely, we exclude one maturity from the calibration set, calibrate no arbitrage eSSVI on the remaining maturities, and plot the smile generated by the calibrated model for the excluded maturity.

the performance of the eSSVI model in the maturity that has been excluded is worse than the calibration in Figure 5.5 but we can observe that eSSVI model manages to provide reasonable interpolation/extrapolation procedures for smiles slices. It is likely that these methodologies could be further improved by choosing other interpolation/extrapolation schemes.

Chapter 6

Smiles in delta

Abstract

Fukasawa introduced in [Fukasawa, Math Finance, 22(4):753-762, 2012] two necessary conditions for no Butterfly arbitrage on a given implied volatility smile which require that the functions d_1 and d_2 of the Black-Scholes formula have to be decreasing. In this chapter, we characterize the set of smiles satisfying these conditions, using the parametrization of the smile in delta. We obtain a parametrization of the set of such smiles via one real number and three positive functions, which can be used by practitioners to calibrate a weak arbitrage-free smile. We also show that such smiles and their symmetric smiles can be transformed into smiles in the strike space by a bijection. Our result motivates the study of the challenging question of characterizing the subset of Butterfly arbitrage-free smiles using the parametrization in delta.

From:
A. Mingone, *Smiles in delta*, Quantitative Finance https://doi.org/10.1080/14697688.2023.2258932, 2023.

6.1 Structure of the chapter

In this chapter we look at the notion of absence of Butterfly arbitrage in the delta notation. We start in Section 6.2 with a description of the delta notation and a detailed discussion of the no Butterfly arbitrage conditions in the delta space.

In Section 6.3.1 we characterize the set of smiles in delta that can be converted to smiles in log-forward moneyness, i.e. the set of smiles that allow to unambiguously define a delta function $k \to \delta(k)$ from the relation $\delta(k) = N(d_1(k, \sigma(\delta(k)))$. Looking at a similar question but from the strike perspective, Section 6.3.2 achieves the characterization of the set of smiles in delta which correspond to an existing smile in strike, i.e. smiles that are defined from a strike function $\delta \to k(\delta)$ satisfying $\delta = N(d_1(k(\delta), \hat{\sigma}(k(\delta)))$. The two sets are shown to coincide, so that a smile in delta belonging to them can be transformed into a smile in strike and re-transformed in the original smile in delta. We synthetize in Section 6.3.4 a practical methodology to calibrate smiles satisfying this property.

Section 6.4 summarizes some properties of the smiles in delta which have a corresponding smile in strike. In particular, we write in the delta notation the notions of maximum and minimum points in the smiles, the second order approximation of the smiles around 0 and the Lee asymptotic conditions in [48].

Section 6.5 deals with the set of smiles in delta which satisfy the weak no arbitrage conditions of monotonicity of the functions d_1 and d_2. These smiles contain the subset of the Butterfly arbitrage-free smiles and are for this reason of interest for practitioners calibrating smiles in the delta notation. In particular, the result in Theorem 6.1 shows that such set can be parametrized by a real number and three positive functions:

$$\sigma(\delta)\sqrt{T} = \begin{cases} N^{-1}(\delta) + \sqrt{N^{-1}(\delta)^2 + 2\left(\int_\delta^{\frac{1}{2}} \lambda(x)\,dx + \int_{\frac{1}{2}}^{\tilde{\delta}} \mu(x)\,dx\right)} & \text{if } \delta \leq \frac{1}{2}, \\ N^{-1}(\delta) + \sqrt{2\int_\delta^{\tilde{\delta}} \mu(x)\,dx} & \text{if } \frac{1}{2} < \delta \leq \tilde{\delta}, \\ N^{-1}(\delta) - \sqrt{2\int_{N^{-1}(\tilde{\delta})}^{N^{-1}(\delta)} x\beta(x)\,dx} & \text{if } \delta > \tilde{\delta}. \end{cases}$$

The functions λ, μ and β are required to satisfy some weak conditions (positiveness and diverging integrals) that can be easily achieved. We also compare our result with the parametrization in the strike space obtained by Lucic in Theorem 2.2 of [51]. In Section 6.5.2 we describe a calibration routine which can be adopted by practitioners to fit market smiles in delta with weak arbitrage-free functions.

The section ends with practical examples of smiles in the weak no arbitrage set. In particular, both skew-shaped smiles and W-shaped smiles can be easily obtained with appropriate choices of the functions λ, μ and β (see Section 6.5.3).

6.2 Notations and preliminaries

In Table 6.1 we summarize the notations that will be used in the chapter.

For a fixed maturity T, we denote by $C_{\mathrm{BS}}(k, \hat{\sigma}(k))$ the Black-Scholes pricing formula for a Call option with maturity T, strike $F_0(T)e^k$, forward value $F_0(T)$, discounting factor $D_0(T)$, and implied volatility $\hat{\sigma}(k)$:

$$\begin{aligned} C_{\mathrm{BS}}(k, \hat{\sigma}(k)) &= D_0(T)F_0(T)\big(N(d_1(k, \hat{\sigma}(k)) - e^k N(d_2(k, \hat{\sigma}(k))\big), \\ d_{1,2}(k, \hat{\sigma}(k)) &= -\frac{k}{\hat{\sigma}(k)\sqrt{T}} \pm \frac{\hat{\sigma}(k)\sqrt{T}}{2}. \end{aligned}$$

For easier notation, we will sometimes denote with $d_1(k)$ and $d_2(k)$ the functions $d_1(k, \hat{\sigma}(k))$ and $d_2(k, \hat{\sigma}(k))$ respectively.

Let a number $C(k)$ lie strictly between $D_0(T)(F_0(T) - K)^+$ and $D_0(T)F_0(T)$. Then the implied volatility $\hat{\sigma}(k)$ is well defined in $]0, \infty[$ by the equation $C_{\mathrm{BS}}(k, \hat{\sigma}(k)) = C(k)$. In turn, this defines the quantity *delta* by

$$\delta(k) := N(d_1(k, \hat{\sigma}(k))). \tag{6.1}$$

Remark 6.1. *Observe that the pair* $(\delta(k), \hat{\sigma}(k))$ *allows to recover the log-forward moneyness* k*, indeed*

$$k = \Big(-N^{-1}(\delta(k)) + \frac{\hat{\sigma}(k)\sqrt{T}}{2}\Big)\hat{\sigma}(k)\sqrt{T}.$$

6.2.1 No Butterfly arbitrage and convexity of Call prices

Under the hypothesis of a perfect market for the underlying asset and for the Call options, a Call price function with respect to the strike is free of Butterfly arbitrage if and only if (cf. e.g. [66]) it is

1. convex,
2. non-increasing,
3. contained in the interval $[D_0(T)(F_0(T) - K)^+, D_0(T)F_0(T)]$.

In the case of Call prices specified with the Black-Scholes formula through an implied volatility, the third property is automatically satisfied. Indeed, the Black-Scholes formula is increasing with respect to the implied volatility and it tends to the two bounds when the implied volatility goes to 0 and ∞ respectively. Given that the third property is granted, then the first property implies the second one since an increasing and convex function cannot be bounded.

Therefore, in our context, there is no Butterfly arbitrage if and only if Call prices are convex.

As shown in Theorem 2.9 condition (IV3) of [62], in the case of twice differentiable implied volatility functions, the requirement of convexity corresponds to the requirement that the function

$$\hat{\sigma}''(k) + d_1'(k, \hat{\sigma}(k))d_2'(k, \hat{\sigma}(k))\hat{\sigma}(k) \tag{6.2}$$

Symbol	Meaning
σ	Smile in delta
$\hat{\sigma}$	Smile in log-forward moneyness, here abbreviated as smile in strike
Σ	Set of smiles in delta
$\mathbb{R}^+$	Real numbers strictly larger than 0
$\mathbb{R}_*^+$	Non-negative real numbers
$\mathcal{C}(\mathcal{A}, \mathcal{B})$	Set of continuous functions from the open interval $\mathcal{A}$ to the open interval $\mathcal{B}$
D	Subscript for sets related to the transformation of a smile in delta into a smile in strike
K	Subscript for sets and functions related to the transformation of a smile in strike into a smile in delta
a.e.	almost everywhere

Table 6.1: Symbols used in the chapter and their meaning.

is non-negative. This requirement is sometimes called the Durrleman condition [28].

6.2.2 Behavior of d_1 and d_2 under convexity assumptions

As shown in section 2.3 of [53], the condition of vanishing Call prices for increasing strikes is not necessary for the absence of arbitrage. Such condition is one-to-one with the behavior of the function d_1 at ∞. In particular, it holds true if and only if the property

P1. $\lim_{k\to\infty} d_1(k, \hat{\sigma}(k)) = -\infty$

is satisfied. This follows from the fact that the arithmetic mean exceeds the geometric mean, so that the function $d_2(k, \hat{\sigma}(k))$ is smaller than $-\sqrt{2k}$ and goes to $-\infty$ at ∞. For negative z, the Mills ratio $\frac{N(z)}{n(z)}$, where n is the Gaussian pdf, satisfies $\frac{N(z)}{n(z)} < \frac{1}{|z|}$. Observe that $n(d_2(k, \hat{\sigma}(k)) < n(-\sqrt{2k}) = \frac{e^{-k}}{\sqrt{2\pi}}$, so that the quantity $e^k N(d_2(k, \hat{\sigma}(k))$ is smaller than the Mills ratio applied to $z = d_2(k, \hat{\sigma}(k))$ (divided by the constant $\sqrt{2\pi}$), which in turn is smaller than $(|d_2(k, \hat{\sigma}(k))|\sqrt{2\pi})^{-1}$. Then, the value of a Call price at ∞ is

$$C(\infty) = D_0(T)F_0(T)N(d_1(\infty)). \tag{6.3}$$

From now on and only in this section, we work under the assumption that Call prices are convex, i.e. that the function $K \to C(K) = C_{\mathrm{BS}}(k, \hat{\sigma}(k))\big|_{k=\log \frac{K}{F_0(T)}}$ is convex. Under the hypothesis of convex prices and **P1**, for every $K \geq 0$, Call prices have the representation

$$C(K) = \int_{\mathbb{R}_*^+} (x - K)^+ \mu(dx)$$

where μ is a probability measure supported by $\mathbb{R}_*^+$. Then, using the identity $(x-K)^+ = \int_K^\infty \mathbb{1}_{y<x}\, dy$ and Fubini's theorem, the following formula holds:

$$C(K) = \int_K^\infty (1 - F_\mu(x))\, dx,$$

where $F_\mu(K) = \mu([0, K])$ is the CDF of the measure μ. Taking the right derivative of the above formula, it holds $C'_+(K) = -1 + F_\mu(K)$, so that $\mu(\{0\}) = F_\mu(0) = 1 + C'_+(0)$. Observe that the bounds of the Call prices function imply that its slope for K going to 0 lies between -1 and 0. Then, the limit corresponds to -1 if and only if there is no mass of the underlying in 0. This in turn is equivalent to the fact that the left limit of the function d_2 is ∞:

P2. $\lim_{k\to-\infty} d_2(k, \hat{\sigma}(k)) = \infty$,

which follows from the fact that $d_2(-\infty) = -N^{-1}(\mu(\{0\}))$ as shown in proposition 2.4 of [31].

Remark 6.2. ***P1*** *and* ***P2*** *are not necessary conditions for Butterfly arbitrage-free Call prices.* ***P1*** *implies that Call prices vanish for increasing strikes. Under the assumption of convex prices,* ***P2*** *implies that Call prices have a slope of* -1 *for null strikes and that there is no mass of the underlying in* 0*.*

P1 and **P2** imply some easy consequences that we state in the following proposition. The Lee conditions can be found in [48], Lemmas 3.1 and 3.3.

Proposition 6.1. ***P1*** *holds if and only if* $d_1(k)$ *is surjective, and* ***P2*** *holds if and only if* $d_2(k)$ *is surjective. Furthermore,*

- ***P1*** *implies the Lee right wing condition, i.e.* $\hat{\sigma}(k)\sqrt{T} < \sqrt{2k}$ *for* k *large enough;*
- ***P2*** *implies the Lee left wing condition, i.e.* $\hat{\sigma}(k)\sqrt{T} < \sqrt{-2k}$ *for* k *small enough.*

Proof. Since the arithmetic mean exceeds the geometric mean, the function $k \to d_1(k, \hat{\sigma}(k))$ is greater than $\sqrt{-2k}$ for every $k \leq 0$ (see Lemma 3.5 of [32]). As a consequence, $d_1(k, \hat{\sigma}(k))$ always goes to ∞ as k goes to $-\infty$. Similarly with the same proof, the function $d_2(k, \hat{\sigma}(k))$ always goes to $-\infty$ at ∞ since it is smaller than $-\sqrt{2k}$. Then, under **P1** and **P2**, the functions $d_1(k)$ and $d_2(k)$ are surjective. The if implication is trivial.

Since the function $k \to d_1(k, \hat{\sigma}(k))$ goes to $-\infty$ on the right under **P1**, it must be negative for k large enough, which implies the right wing Lee condition. Similarly, since the function $k \to d_2(k, \hat{\sigma}(k))$ explodes on the left, it must be positive for k small enough and the left wing Lee condition holds.

□

Observe that Lee shows that the left wing condition holds for every arbitrage-free smile if and only if $P(S_T = 0) < \frac{1}{2}$, and in particular that no mass in 0 implies the left wing condition, as in the above proposition. Furthermore, the proof for the only if implication can also be found in Lemmas 3.2 and 3.5 of [32].

We will often use necessary conditions for the absence of Butterfly arbitrage found by Fukasawa in Theorem 2.8 of [32]. We recall them in the following generalized remark from Lemma 3.1 of [53].

Remark 6.3 (Weak no Butterfly arbitrage conditions/Fukasawa necessary no Butterfly arbitrage conditions)**.** *If* $k \to C(k)$ *is convex and* $k \to \hat{\sigma}(k)$ *is differentiable, then*

F1 $k \to d_1(k, \hat{\sigma}(k))$ *is strictly decreasing,*

F2 $k \to d_2(k, \hat{\sigma}(k))$ *is strictly decreasing.*

6.2.3 No Butterfly arbitrage in the delta notation

As discussed above, a difficult challenge is to parametrize the set of implied volatility functions corresponding to functions $k \to C(k)$ with no Butterfly arbitrage. We will denote with Σ_{A} the set of such implied volatility functions (in the delta notation $\delta \to \sigma(\delta)$) satisfying also **P1** and **P2**. In the following, we will show that these delta smiles

can be transformed into strike smiles and, given the above study, they will guarantee the monotonicity and surjectivity of functions $d_1(k, \hat{\sigma}(k))$ and $d_2(k, \hat{\sigma}(k))$. Even though we will not reach the aim of parametrizing the set Σ_{A}, we will be able to achieve the parametrization of the larger set of smiles satisfying the two monotonicity conditions on the functions $d_1(k, \hat{\sigma}(k))$ and $d_2(k, \hat{\sigma}(k))$.

The range of the functions δ and $\bar{\delta}$

In this section we consider properties **P1** and **P2** in the delta notation, since they influence the range of the function $\delta(k)$ in Equation (6.1) and of its symmetric delta $\bar{\delta}(k)$ defined as

$$\bar{\delta}(k) := N(d_1(k, \hat{\bar{\sigma}}(k))) \tag{6.4}$$

where

$$\hat{\bar{\sigma}}(k) = \hat{\sigma}(-k).$$

Proposition 6.2. *The function $\delta(k)$ defined in Equation (6.1) has range $\left]\frac{C(\infty)}{D_0(T)F_0(T)}, 1\right[$, and this coincides with $]0,1[$ if and only if **P1** holds.*

*The function $\bar{\delta}(k)$ defined in Equation (6.4) has range $\left]1+C'(0), 1\right[$, and this coincides with $]0,1[$ if and only if **P2** holds.*

Proof. It is easy to see that the function $\delta(k)$ goes to 0 as k goes to ∞ if and only if **P1** holds. The condition $\delta(-\infty) = 1$ corresponds to the fact that $d_1(k, \hat{\sigma}(k))$ goes to ∞ as k decreases, but this is always true. More precisely, from Equation (6.3) it follows

$$C(\infty) = D_0(T)F_0(T)\delta(\infty)$$

so that the range of $\delta(k)$ is $\left]\frac{C(\infty)}{D_0(T)F_0(T)}, 1\right[$.

Secondly, the assumption **P2** is the equivalent of the assumption **P1** for the symmetric smile $\hat{\bar{\sigma}}(k) = \hat{\sigma}(-k)$. Indeed, it holds $d_1(k, \hat{\bar{\sigma}}(k)) = -d_2(-k, \hat{\sigma}(-k))$, so that condition **P2** is equivalent to requiring that the d_1 function for the symmetric smile, i.e. $d_1(k, \hat{\bar{\sigma}}(k))$, satisfies **P1**. It is immediate that

$$\bar{\delta}(\infty) = 1 + C'(0)$$

since $d_2(-\infty) = -N^{-1}(1 + C'(0))$. On the other hand, the limit of $\bar{\delta}(k)$ for k decreasing is still 1. Then, the range of the symmetric delta $\bar{\delta}(k)$ is $]1 + C'(0), 1[$.

□

6.3 From a smile in delta to a smile in strike and vice versa

6.3.1 From a smile in delta to a smile in strike

In the industry, it is not an unusual practice to calibrate volatility smiles in the delta parametrization, instead of the usual strike one, especially when dealing with FX products. When options are quoted on a grid of maturities and deltas, such choice is natural

and easy to be exploited. When, on the other hand, options are quoted on a grid of maturities and strikes, the permutation between delta and strike smiles is not straightforward.

Indeed, it is firstly necessary to transform data in strikes into data in deltas. For a fixed maturity, the procedure reads:

1. compute volatilities $\hat{\sigma}(k)$ for the quoted strikes $F_0(T)e^k$ using the inversion of the Black-Scholes pricing formula;
2. compute corresponding deltas $\delta(k)$ with Equation (6.1) and uniquely associate them to the smile values;
3. interpolate pairs $(\delta(k), \hat{\sigma}(k)) = (\delta, \sigma(\delta))$ with a chosen method in order to recover the continuous smile $\delta \to \sigma(\delta)$.

At this point, different operations can be done using the smiles in delta, such as collecting historical values, stressing data, doing statistics, and so on. When it is necessary to come back to the strike notation, for example to compute a stressed option price with known strike, the smile in delta must have the ability to be converted into a smile in strike.

In this section we aim to find conditions under which any positive and continuous function $\delta \to \sigma(\delta)$ defined on $]0,1[$ allows to recover a function $k \to \hat{\sigma}(k)$.

Definition 6.1. *We call Σ_D the set of positive and continuous delta smiles $\delta \to \sigma(\delta)$ defined on $]0,1[$ for which there exists a surjective mapping $k \to \delta(k)$ defined on $\mathbb{R}$ satisfying $\delta(k) = N(d_1(k, \sigma(\delta(k))))$.*

In particular,

$$\Sigma_D := \Big\{\delta \to \sigma(\delta) \in \mathcal{C}(]0,1[,\mathbb{R}^+) \,\Big|\, \forall k \in \mathbb{R}\, \exists! \delta(k) \,|\, \delta(k) = N(d_1(k, \sigma(\delta(k)))), \\ \{\delta(k) \,|\, k \in \mathbb{R}\} =]0,1[\Big\}.$$

For a smile in Σ_D, the corresponding smile in strike is defined as $\hat{\sigma}(k) := \sigma(\delta(k))$. Note that there is indeed a question, because in the above procedure it could happen that two different pairs $(\delta, \sigma(\delta))$ produce the same strike, meaning that there is no way to define the value $\hat{\sigma}(k)$. The second condition defining Σ_D is added in order to avoid degenerate cases.

Let us define the function

$$l(\delta) := \Big(N^{-1}(\delta) - \frac{\sigma(\delta)\sqrt{T}}{2}\Big)\sigma(\delta)\sqrt{T}. \tag{6.5}$$

In the following Lemma, we characterize the set Σ_D through the function $l(\delta)$.

Lemma 6.1. *A continuous and positive delta smile $\delta \to \sigma(\delta)$ belongs to Σ_D if and only if the function $\delta \to l(\delta)$ is strictly increasing and surjective onto $\mathbb{R}$.*

In other words,

$$\Sigma_D = \Big\{\delta \to \sigma(\delta) \in \mathcal{C}(]0,1[,\mathbb{R}^+) \,|\, l \text{ strictly increasing and surjective onto } \mathbb{R}\Big\}.$$

Proof. Reformulating the conditions of Σ_D, we ask that given a continuous positive function $\delta \to \sigma(\delta)$ defined on $]0,1[$,

- for every $k \in \mathbb{R}$ there exists a unique $\delta(k)$ such that $-k = l(\delta(k))$,
- the mapping $k \to \delta(k)$ is surjective in $]0,1[$.

This implies that the function $k \to l(\delta(k))$ must be well-defined, monotonic and surjective. For a delta smile in Σ_D, the function $k \to \delta(k)$ is monotonic and surjective, so that we are requiring $\delta \to l(\delta)$ to be monotonic and surjective from the interval $]0,1[$ to $]-\infty, +\infty[$. Since $l(0)$ is negative, l must be strictly increasing.

These conditions are necessary but also sufficient. Indeed, the uniqueness of $\delta(k)$ follows from the fact that if for a fixed k there are two different $\delta(k)$, say δ_1 and δ_2, then $l(\delta_1) \neq l(\delta_2)$ since l is monotonic. However, $l(\delta(k)) = -k$, so the two values of l should be the same. The existence of $\delta(k)$ is guaranteed by the surjectivity of l.

For the second condition, firstly observe that from the relation $-k = l(\delta(k))$, it follows $l(\delta(\infty)) = -\infty$ and since l is injective, it must hold $\delta(\infty) = 0$. Similarly we can show $\delta(-\infty) = 1$. These observations guarantee the full range of $\delta(k)$.

□

6.3.2 From a smile in strike to a smile in delta

Starting from a smile in strike $k \to \hat{\sigma}(k)$ and making k move in $\mathbb{R}$, one obtains a collection of pairs $(\delta(k), \hat{\sigma}(k))$. There is no guarantee that such a collection allows to define a function $\delta \to \sigma(\delta)$. Indeed, in order to define a function in δ, for every $\delta \in]0,1[$, there must exist a unique k such that $\delta = N(d_1(k, \hat{\sigma}(k))$. In such a way one can define unambiguously a function σ by the equality $\sigma(\delta) = \hat{\sigma}(k)$ for all ks. We define Σ_K the set of delta smiles obtained from a smile in strike when this condition holds.

Definition 6.2. *We call Σ_K the set of positive and continuous delta smiles $\delta \to \sigma(\delta)$ defined on $]0,1[$ for which there exist a positive and continuous strike smile $k \to \hat{\sigma}(k)$ and a surjective mapping $\delta \to k(\delta)$ such that $\delta = N(d_1(k(\delta), \hat{\sigma}(k(\delta))))$ and $\hat{\sigma}(k(\delta)) = \sigma(\delta)$.*

In particular,

$$\Sigma_K := \Big\{\delta \to \sigma(\delta) \in \mathcal{C}(]0,1[, \mathbb{R}^+) \,\Big|\, \exists k \to \hat{\sigma}(k) \in \mathcal{C}(\mathbb{R}, \mathbb{R}^+) \text{ s.t. } \forall \delta \in]0,1[\, \exists! k(\delta) \text{ s.t.} \\ \delta = N(d_1(k(\delta), \hat{\sigma}(k(\delta)))), \\ \hat{\sigma}(k(\delta)) = \sigma(\delta), \\ \big\{k(\delta) \,|\, \delta \in]0,1[\big\} = \mathbb{R}\Big\}.$$

This definition can be simplified thanks to the following Lemma:

Lemma 6.2. *A continuous and positive delta smile $\delta \to \sigma(\delta)$ belongs to Σ_K if and only if there exists a continuous and positive strike smile $k \to \hat{\sigma}(k)$ with $\hat{\sigma}(k) = \sigma(N(d_1(k, \hat{\sigma}(k))))$ and such that the function $k \to d_1(k, \hat{\sigma}(k))$ is strictly decreasing and surjective.*

In other words,

$$\Sigma_K = \Big\{\delta \to \sigma(\delta) \in \mathcal{C}(]0,1[,\mathbb{R}^+) \,\Big|\, \exists k \to \hat{\sigma}(k) \in \mathcal{C}(\mathbb{R},\mathbb{R}^+) \text{ s.t.} \\ k \to d_1(k,\hat{\sigma}(k)) \text{ is strictly decreasing and surjective,} \\ \hat{\sigma}(k) = \sigma(N(d_1(k,\hat{\sigma}(k))))\Big\}.$$

Proof. The existence and uniqueness of a $k(\delta)$ in the first condition defining Σ_K can be translated in requiring that the map $k \to N(d_1(k,\hat{\sigma}(k)))$ is strictly monotonic. The second condition defining Σ_K requires the surjectivity of $\delta \to k(\delta)$. Equivalently, the two conditions hold if and only if the map $k \to d_1(k,\hat{\sigma}(k))$ is strictly monotonic and surjective in $]-\infty,\infty[$. The decreasing behavior is due to the formula defining the map $k \to d_1(k,\hat{\sigma}(k))$.

□

Remark 6.4. *In Lemma 6.2 we have shown that all smiles in delta living in Σ_K are such that their smile in strike satisfies **P1** and **F1**, i.e. $d_1(k,\hat{\sigma}(k))$ is decreasing and surjective.*

6.3.3 Reversibility of smiles in delta to smiles in strike

We now look at the relation between smiles in delta that can be transformed into smiles in strike and smiles in strike that can be transformed in smiles in delta. It turns out that the image of the latter set in the space of smiles in delta actually coincides with the former set. In other words, a smile in delta obtained through a smile in strike can be re-transformed into the original smile in strike. Also, any smile in delta that can be transformed in a smile in strike can be recovered from its transformation into a smile in strike.

Proposition 6.3. *It holds $\Sigma_D = \Sigma_K$ and $\delta \to k(\delta)$ is the inverse of $k \to \delta(k)$.*

Proof. We firstly prove $\Sigma_K \subset \Sigma_D$. For a function $\delta \to \sigma(\delta)$ in Σ_K and a fixed $k \in \mathbb{R}$, there exists a unique $\delta_K(k)$ such that $\delta_K(k) = N(d_1(k,\hat{\sigma}(k))$. Given the definition of $\sigma(\delta)$, it holds $\hat{\sigma}(k) = \sigma(\delta_K(k))$, so that $\delta_K(k) = N(d_1(k,\sigma(\delta_K(k)))$. Suppose there is a second $\tilde{\delta}$ such that $\tilde{\delta} = N(d_1(k,\sigma(\tilde{\delta})))$. For such $\tilde{\delta}$ there is a unique $k(\tilde{\delta})$ such that $\tilde{\delta} = N(d_1(k(\tilde{\delta}),\hat{\sigma}(k(\tilde{\delta}))))$, furthermore $\hat{\sigma}(k(\tilde{\delta})) = \sigma(\tilde{\delta})$. Then, $N(d_1(k,\sigma(\tilde{\delta}))) = \tilde{\delta} = N(d_1(k(\tilde{\delta}),\sigma(\tilde{\delta})))$, from which it immediately follows $k = k(\tilde{\delta})$. Then for k, it both holds $\delta_K(k) = N(d_1(k,\hat{\sigma}(k))$ and $\tilde{\delta} = N(d_1(k,\hat{\sigma}(k)))$, so that $\delta_K(k) = \tilde{\delta}$. The function $\delta_K(k)$, in particular, corresponds to the function $\delta(k)$ defining the set Σ_D.

On the other hand, we now look at the relation $\Sigma_D \subset \Sigma_K$. Let $\delta \to \sigma(\delta)$ in Σ_D, then we can define a function $k \to \hat{\sigma}(k)$ such that $\hat{\sigma}(k) := \sigma(\delta(k))$, where $\delta(k)$ is the only δ satisfying $\delta = N(d_1(k,\sigma(\delta))$. If two values k_1 and k_2 have the same delta $\delta(k_1) = \delta(k_2)$, then they also have the same volatility since $\hat{\sigma}(k_1) = \sigma(\delta(k_1)) = \sigma(\delta(k_2)) = \hat{\sigma}(k_2)$. From Remark 6.1, the pair $(\delta(k_1),\hat{\sigma}(k_1)) = (\delta(k_2),\hat{\sigma}(k_2))$ is associated to a unique log-forward moneyness, so that $k_1 = k_2$. Then $k \to \delta(k)$ is injective. The function $d_1(k) = d_1(k,\hat{\sigma}(k))$ is surjective since for any $\delta \in]0,1[$ there exists a k such that $\delta = \delta(k)$

so that $\delta = N(d_1(k, \sigma(\delta)) = N(d_1(k, \hat{\sigma}(k))$. To prove the monotonicity, firstly observe that for any k there is a unique $\delta(k)$ such that $\delta(k) = N(d_1(k, \sigma(\delta(k))))$, which is equivalent to write

$$l(\delta(k)) = -k. \tag{6.6}$$

Secondly, from the definition of $\hat{\sigma}(k)$, it holds

$$\delta(k) = N(d_1(k, \hat{\sigma}(k))). \tag{6.7}$$

The functions $l(\delta)$, $\delta(k)$ and $d_1(k)$ are monotonic into an open interval, so they are a.e. differentiable. Taking derivatives with respect to k (in the complementary of the zero measure set where these derivatives are not defined) in both Equation (6.6) and Equation (6.7), we find $l'(\delta(k))\frac{d\delta}{dk}(k) = -1$ and $\frac{d\delta}{dk}(k) = n(d_1(k))d_1'(k)$, so in particular

$$d_1'(k)l'(\delta(k)) = -\frac{1}{n(d_1(k))}.$$

Since $\delta \to \sigma(\delta)$ lives in Σ_D, l is increasing so d_1 is decreasing and $\sigma(\delta)$ lives also in Σ_K. The inverse of the function $k \to \delta(k)$, in particular, corresponds to the function $\delta \to k(\delta)$ defined in Σ_K.

□

Remark 6.5. *From now on, we will generally denote the identical sets Σ_D and Σ_K as Σ.*

Remark 6.6. *As a side-product of the proof of Proposition 6.3, the functions $k \to \delta(k)$ and $\delta \to k(\delta)$ are both a.e. differentiable. Furthermore, in the proof we have shown that a smile in strike transformed into a smile in delta, can be re-transformed into a smile in strike and such smile necessarily coincides with the initial one.*

6.3.4 Calibration of a delta smile in Σ

Suppose we want to calibrate a smile in delta to market quotes given in either the strike or the delta variable, in such a way that the calibrated smile can be converted to a smile in strike and vice versa, i.e. that the smile lives in Σ. We could go through the following steps:

1. a. consider the market discrete pillars in delta notation $\{(\delta_i, \sigma_i)\}_i$;
   b. if the market discrete pillars are in strike notation $\{(k_i, \sigma_i)\}_i$, convert them to the pillars in delta notation $\{(\delta_i, \sigma_i)\}_i$ by defining $\delta_i = N(d_1(k_i, \sigma_i))$;
2. compute the pillars $\{(\delta_i, l_i)\}_i$ with $l_i = \left(N^{-1}(\delta_i) - \frac{\sigma_i\sqrt{T}}{2}\right)\sigma_i\sqrt{T}$;
3. interpolate/extrapolate an increasing and surjective function $\delta \to l(\delta)$, given pillars in point 2.

The last natural point would be to recover a function $\delta \to \sigma(\delta)$ by the calibrated function $l(\delta)$. However, this reduces to solve the equation

$$\frac{\sigma^2 T}{2} - N^{-1}(\delta)\sqrt{T}\sigma + l(\delta) = 0,$$

which could have none, one or two solutions. In the following section we study the problem of existence and uniqueness of the solution σ.

Conditions on l for the existence of σ

Proposition 6.4. *Let $l :]0,1[\to \mathbb{R}$ an increasing and surjective function. The equation*

$$\frac{\sigma^2 T}{2} - N^{-1}(\delta)\sqrt{T}\sigma + l(\delta) = 0 \tag{6.8}$$

has at least one solution σ for every $\delta \in]0,1[$ if and only if $l\left(\frac{1}{2}\right) < 0$ and $l(\delta) \leq \frac{N^{-1}(\delta)^2}{2}$ for all $\delta > \frac{1}{2}$. The continuous solution $\delta \to \sigma(\delta)$ is

$$\left(N^{-1}(\delta) + \sqrt{N^{-1}(\delta)^2 - 2l(\delta)}\right)/\sqrt{T} \tag{6.9}$$

for $\delta \leq \frac{1}{2}$ and it could switch between Equation (6.9) and

$$\left(N^{-1}(\delta) - \sqrt{N^{-1}(\delta)^2 - 2l(\delta)}\right)/\sqrt{T} \tag{6.10}$$

at any $\tilde{\delta} > \frac{1}{2}$ such that $l(\tilde{\delta}) = \frac{N^{-1}(\tilde{\delta})^2}{2}$.

Proof. For a fixed $\delta \in]0,1[$, the two admissible solutions to Equation (6.8) are

$$\sigma_\pm(\delta) = \frac{N^{-1}(\delta)}{\sqrt{T}} \pm \frac{\sqrt{N^{-1}(\delta)^2 - 2l(\delta)}}{\sqrt{T}}.$$

For the existence, the delta of the equation must be non-negative and at least one of the two solutions must be positive, so

$$\begin{aligned} &N^{-1}(\delta)^2 - 2l(\delta) \geq 0, \\ &N^{-1}(\delta) \pm \sqrt{N^{-1}(\delta)^2 - 2l(\delta)} > 0, \end{aligned}$$

where the sign depends on the chosen solution.

Firstly, when $\delta \leq \frac{1}{2}$, the quantity $N^{-1}(\delta)$ is non-positive and the σ_- solution is negative, so it can be discarded. Instead, the σ_+ solution is well-defined and positive if and only if $l(\delta) < 0$. Since l is increasing, the latter condition is equivalent to $l\left(\frac{1}{2}\right) < 0$.

When $\delta > \frac{1}{2}$, then $N^{-1}(\delta)$ is positive and both the solutions could be valid. Under the requirement $l(\delta) \leq \frac{N^{-1}(\delta)^2}{2}$, the σ_+ solution is always positive while the σ_- solution becomes positive when l becomes positive, and it will stay positive since l is increasing.

The possibility to pass from the σ_+ to the σ_- solution at a point $\tilde{\delta}$ must be guaranteed by continuity of the volatility on $\tilde{\delta}$. Then, it should hold $\sigma_+(\tilde{\delta}) = \sigma_-(\tilde{\delta})$, or $2l(\tilde{\delta}) = N^{-1}(\tilde{\delta})^2$ and $\sigma(\tilde{\delta})\sqrt{T} = N^{-1}(\tilde{\delta})$. Rewriting equation Equation (6.8) as

$$l(\delta) = \Big(N^{-1}(\delta) - \frac{\sigma(\delta)\sqrt{T}}{2}\Big)\sigma(\delta)\sqrt{T}$$

and evaluating in $\tilde{\delta}$, one finds $l(\tilde{\delta}) = \frac{\sigma(\tilde{\delta})^2 T}{2}$. If there exists such a point $\tilde{\delta}$, then either the solution keeps being σ_+ after $\tilde{\delta}$, or it switches to σ_-. If there is no such point $\tilde{\delta}$, the solution remains σ_+.

□

As an immediate consequence of Proposition 6.4, under the requirements $l\big(\frac{1}{2}\big) < 0$ and $l(\delta) \leq \frac{N^{-1}(\delta)^2}{2}$, if there are no $\tilde{\delta}$ such that $l(\tilde{\delta}) = \frac{N^{-1}(\tilde{\delta})^2}{2}$, the σ solution is unique and it coincides with Equation (6.9). If there are one or more points $\tilde{\delta}$, the uniqueness is not guaranteed since the σ solution could either switch between Equation (6.9) and Equation (6.10) or not do the switch. For the uniqueness of the solution, more requirements on the solution itself are needed.

In the following lemma we show that smiles in delta living in Σ actually satisfy the requirements $l\big(\frac{1}{2}\big) < 0$ and $l(\delta) \leq \frac{N^{-1}(\delta)^2}{2}$, so that calibrating a function $l(\delta)$ which satisfies such conditions and is increasing and surjective guarantees the existence of at least one solution to Equation (6.8), and this solution lives in Σ. We will show in Lemma 6.4 that if we require the solution σ to satisfy **P1**, **P2**, **F1** and **F2**, then the uniqueness is satisfied.

Lemma 6.3. *For every $\delta \to \sigma(\delta) \in \Sigma$, the function l satisfies $l\big(\frac{1}{2}\big) < 0$ and $l(\delta) \leq \frac{N^{-1}(\delta)^2}{2}$ for all $\delta > \frac{1}{2}$.*

Proof. From the definition of Σ_K, it follows that its smiles are such that for any $\delta \in]0,1[$ there is a unique $k(\delta)$ satisfying $\delta = N(d_1(k(\delta), \hat{\sigma}(k(\delta))))$. In particular, by the definition of the smile $\delta \to \sigma(\delta)$, it holds $\delta = N(d_1(k(\delta), \sigma(\delta)))$, or $\frac{\sigma(\delta)^2 T}{2} - N^{-1}(\delta)\sqrt{T}\sigma(\delta) - k(\delta) = 0$. Then, $\sigma = \sigma(\delta)$ is a solution of Equation (6.8) and the conclusion follows from Proposition 6.4.

The statement can also be proven by hand. Indeed, $l\big(\frac{1}{2}\big)$ is equal to $-\frac{\sigma(\frac{1}{2})^2 T}{2}$, which is always negative, and the second condition reads

$$\Big(N^{-1}(\delta) - \frac{1}{2}\sigma(\delta)\sqrt{T}\Big)\sigma(\delta)\sqrt{T} \leq \frac{N^{-1}(\delta)^2}{2},$$

which simplifying becomes $(N^{-1}(\delta) - \sigma(\delta)\sqrt{T})^2 \geq 0$, which is always verified.

□

Application: calibration of delta smiles in Σ

We now reconsider the calibration of a smile in delta started in the introduction to this section. The aim here is to calibrate a delta smile which can be transformed into a smile in strike and vice versa, and in particular the smile in strike satisfies **F1** and **P1** (i.e. the function $d_1(k)$ is decreasing and surjective). We have seen that belonging to Σ guarantees the existence of a solution $\sigma(\delta)$ to Equation (6.8) but not its uniqueness. In the following calibration procedure, we target one of the possible solutions. The reason behind the choice of this particular solution is that it approaches the weak arbitrage-free solution as we will see in Lemma 6.4.

The calibration methodology follows the steps:

1. a. consider the market discrete pillars in delta notation $\{(\delta_i, \sigma_i)\}_i$;
   b. if the market discrete pillars are in strike notation $\{(k_i, \sigma_i)\}_i$, convert them to the pillars in delta notation $\{(\delta_i, \sigma_i)\}_i$ by defining $\delta_i = N(d_1(k_i, \sigma_i))$;
2. compute the pillars $\{(\delta_i, l_i)\}_i$ with $l_i = \big(N^{-1}(\delta_i) - \frac{\sigma_i\sqrt{T}}{2}\big)\sigma_i\sqrt{T}$;
3. given the pillars in point 2., interpolate/extrapolate a function $\delta \to l(\delta)$ such that
   a. $l(0) = -\infty, l(1) = +\infty$,
   b. l strictly increasing,
   c. $l\big(\frac{1}{2}\big) < 0$,
   d. $l(\delta) \leq \frac{N^{-1}(\delta)^2}{2}\ \forall \delta > \frac{1}{2}$,
   e. $\exists! \tilde{\delta} |\, l(\tilde{\delta}) = \frac{N^{-1}(\tilde{\delta})^2}{2}$.

This would guarantee that the smile $\delta \to \sigma(\delta)$ defined as

$$\sigma(\delta)\sqrt{T} = \begin{cases} N^{-1}(\delta) + \sqrt{N^{-1}(\delta)^2 - 2l(\delta)} & \text{if } \delta \leq \tilde{\delta}, \\ N^{-1}(\delta) - \sqrt{N^{-1}(\delta)^2 - 2l(\delta)} & \text{if } \delta > \tilde{\delta}. \end{cases}$$

lives in Σ, i.e. that the smile in delta can be transformed into a smile in strike and reverted to the original smile in delta.

Remark 6.7. *The calibration here described requires the knowledge of market points $\{(\delta_i, \sigma_i)\}_i$ (or equivalently $\{(k_i, \sigma_i)\}_i$). However, in the FX market smiles are quoted in terms of ATM volatility, risk reversals and strangles associated to a specific delta.*

The market point $(\delta_{ATM}, \sigma_{ATM})$ is easily retrieved by $\delta_{ATM} = N\big(\frac{\sigma_{ATM}\sqrt{T}}{2}\big)$. Under the approximation that the δ risk reversal volatility is the difference between the volatility at δ and the volatility at $1-\delta$, while the strangle volatility is the average between these two volatilities, the market points $(\delta, \sigma(\delta))$ and $(1-\delta, \sigma(1-\delta))$ are easily calculated.

If the approximation is not used, there is ambiguity in the definition of market points associated to δ and $1-\delta$. Indeed, it is known (cf. e.g. [16]) that there are 4 different

strikes associated to risk reversals and strangles (and so 4 different deltas), so that the situation is much more intricate.

To our knowledge, market providers such as Bloomeberg make available $\{(\delta_i, \sigma_i)\}_i$ *market quotes (which the provider deducts under its internal models from standard FX quotes), so that no ambiguity arises.*

Application: inversion of delta smiles in Σ into strike smiles

We have seen in Proposition 6.3 that smiles in delta living in Σ can be converted into smiles in strike satisfying **P1** and **F1**, using the relation $\hat{\sigma}(k) = \sigma(\delta(k))$. The function $\delta(k)$ is the inverse of $k(\delta) = -l(\delta)$. Then, if we are able to calibrate the inverse of $-l(\delta)$ for a given smile σ in delta living in Σ, we can directly calibrate a smile $\hat{\sigma}$ in strike satisfying **P1** and **F1**. To do so, we can use a monotonic preserving interpolation and a well-chosen fine enough grid for the δ_i points.

The inversion of the smile σ in delta living in Σ can then be performed following the steps:

1. consider the fine enough discrete pillars in delta notation $\{(\delta_i, \sigma_i)\}_i$;
2. compute the pillars $\{(-l_i, \delta_i)\}_i$ with $l_i = \big(N^{-1}(\delta_i) - \frac{\sigma_i\sqrt{T}}{2}\big)\sigma_i\sqrt{T}$;
3. given the pillars in point 2., interpolate/extrapolate a function $-l \to \delta(-l)$ such that
   a. $\delta(-\infty) = 1$, $\delta(\infty) = 0$,
   b. δ strictly decreasing.

If the delta grid is fine enough and in particular it includes the points $\big(-l\big(\frac{1}{2}\big), \frac{1}{2}\big)$ and $\big(-\frac{N^{-1}(\tilde{\delta})^2}{2}, \tilde{\delta}\big)$, the interpolated function $\delta(-l)$ is the inverse of $l(\delta)$. Then, the smile defined as

$$\hat{\sigma}(k) = \sigma(\delta(k))$$

is the corresponding smile in strike of the original smile in delta.

Following the steps above we can transform any smile in delta living in Σ into its smile in strike satisfying **P1** and **F1** in an efficient way and without involving any optimization routine.

6.4 Qualitative properties of the smile: from k to δ

In this section we look at the qualitative properties of the smile in delta resulting from the transformation of a smile in strike with decreasing $k \to d_1(k, \hat{\sigma}(k))$ function.

We start with an easy consequence to Proposition 6.3.

Proposition 6.5. *For every smile* $\sigma(\delta) \in \Sigma$ *it holds*

- *$\sigma'(\delta) \geq 0$ if and only if $\hat{\sigma}'(k(\delta)) \leq 0$ and points of minimum (respectively maximum) $\bar{\delta}$ for $\sigma(\delta)$ are points of minimum (resp.maximum) $k(\bar{\delta})$ for $\hat{\sigma}(k)$;*
- *$\hat{\sigma}'(k) \geq 0$ if and only if $\sigma'(\delta(k)) \leq 0$ and points of minimum (respectively maximum) $\bar{k}$ for $\hat{\sigma}(k)$ are points of minimum (resp.maximum) $\delta(\bar{k})$ for $\sigma(\delta)$.*

Proof. Firstly observe that $\hat{\sigma}(k) = \sigma(\delta(k))$ is differentiable being composition of differentiable functions.

Since $\hat{\sigma}(k(\delta)) = \sigma(\delta)$, taking derivatives with respect to δ implies

$$\begin{aligned}\sigma'(\delta) &= \hat{\sigma}'(k(\delta))k'(\delta) \\ &= -\hat{\sigma}'(k(\delta))l'(\delta)\end{aligned}$$

where we have used $l(\delta) = -k(\delta)$ as in Equation (6.6). Smiles in Σ have increasing $l(\delta)$, so the sign of $\sigma'(\delta)$ is opposite to the sign of $\hat{\sigma}'(k(\delta))$. Furthermore, if $\bar{\delta}$ is a point of minimum for $\sigma(\delta)$, then for every δ in a neighborhood of $\bar{\delta}$, it holds $\sigma(\bar{\delta}) < \sigma(\delta)$. Using the relation $\hat{\sigma}(k(\delta)) = \sigma(\delta)$, it follows $\hat{\sigma}(k(\bar{\delta})) < \hat{\sigma}(k(\delta))$. Since the function $d_1(k, \hat{\sigma}(k))$ is monotonic and surjective, it is continuous, so also $k(\delta) = d_1^{-1}(N^{-1}(\delta))$ is continuous. Then for every k in a neighborhood of $k(\bar{\delta})$ it holds $\hat{\sigma}(k(\bar{\delta})) < \hat{\sigma}(k)$, so $k(\bar{\delta})$ is a point of minimum for $\hat{\sigma}(k)$. Similarly for points of maximum.

The proof is similar for the second point, using the relation $\hat{\sigma}(k) = \sigma(\delta(k))$ and the fact that $\delta'(k) = \frac{1}{k'(\delta(k))} = -\frac{1}{l'(\delta(k))}$.

□

We have already seen in Proposition 6.1 that under **P1** and **P2**, the left and right wing Lee conditions hold, i.e. that the limits of $\frac{\hat{\sigma}(k)^2 T}{k}$ at $\pm\infty$ are bounded by 2. We now look at what these limits correspond to in the δ notation.

Proposition 6.6 (Lee moment formula in delta)**.** *Let $\sigma(\delta) \in \Sigma$ and $\hat{\sigma}(k)$ the corresponding smile in strike. The left wing Lee condition $\frac{\hat{\sigma}(k)^2 T}{k} > -2$ for sufficiently large $-k$ holds if and only if $\frac{\sigma(\delta)\sqrt{T}}{N^{-1}(\delta)} < 1$ for δ near 1. The right wing Lee condition $\frac{\hat{\sigma}(k)^2 T}{k} < 2$ for sufficiently large k holds.*

Furthermore,

$$\limsup_{k\to-\infty} \frac{\hat{\sigma}(k)^2 T}{k} = a \qquad \Longleftrightarrow \qquad \liminf_{\delta\to 1} \frac{\sigma(\delta)\sqrt{T}}{N^{-1}(\delta)} = -\frac{2a}{2-a},$$

and

$$\limsup_{k\to\infty} \frac{\hat{\sigma}(k)^2 T}{k} = b \qquad \Longleftrightarrow \qquad \liminf_{\delta\to 0} \frac{\sigma(\delta)\sqrt{T}}{N^{-1}(\delta)} = -\frac{2b}{2-b}.$$

Proof. Since $\sigma(\delta) \in \Sigma$, there exists a smile in strike $\hat{\sigma}(k)$ with strictly decreasing surjective function $k \to d_1(k, \hat{\sigma}(k))$. Thanks to the surjectivity of d_1, the right wing Lee condition is satisfied as shown in Proposition 6.1.

For every k it holds $\frac{\hat{\sigma}(k)^2 T}{k} = \frac{\sigma(\delta(k))^2 T}{k(\delta(k))}$, where $\delta(k) = N(d_1(k, \hat{\sigma}(k))$ and $k(\delta)$ is its inverse. Since $\delta(k)$ is surjective decreasing, it goes to 1 when k goes to $-\infty$ and to 0 when k goes to ∞. Also, since it is monotonic continuous, the left wing Lee condition holds if and only if $\frac{\sigma(\delta)^2 T}{k(\delta)} > -2$ for every δ near 1, and it holds

$$\limsup_{k \to -\infty} \frac{\hat{\sigma}(k)^2 T}{k} = \limsup_{\delta \to 1} \frac{\sigma(\delta)^2 T}{k(\delta)}$$

and similarly for $k \to \infty$. From Equation (6.6), we can substitute $k(\delta)$ with $-l(\delta)$, which is defined in Equation (6.5), so that we are now studying the quantity $-\frac{\sigma(\delta)\sqrt{T}}{N^{-1}(\delta) - \frac{\sigma(\delta)\sqrt{T}}{2}}$. Since $l(\delta)$ is increasing surjective, the denominator is negative for small δ and positive for large δ. Then, it is easy to see that the left wing Lee condition holds if and only if $\frac{\sigma(\delta)\sqrt{T}}{N^{-1}(\delta)} < 1$.

The argument of the limit becomes $\left(\frac{1}{2} - \frac{N^{-1}(\delta)}{\sigma(\delta)\sqrt{T}}\right)^{-1}$, so that the two limits superior are equal to

$$\left(\frac{1}{2} - \limsup_{\delta} \frac{N^{-1}(\delta)}{\sigma(\delta)\sqrt{T}}\right)^{-1}.$$

This quantity is equal to c (either a for $\delta \to 1$ or b for $\delta \to 0$) if and only if $\limsup_\delta \frac{N^{-1}(\delta)}{\sigma(\delta)\sqrt{T}} = \frac{c-2}{2c}$ and the conclusion follows. The reasoning still holds for $c = 0$.

□

From the above proposition it follows that Lee conditions applied to the total variance limits (i.e. $a \in [-2, 0]$ and $b \in [0, 2]$), translate in the delta notation into the requirement that the limit at 0 of $\frac{\sigma(\delta)\sqrt{T}}{N^{-1}(\delta)}$ is negative while the limit at 1 is positive and smaller than 1. The first condition, i.e. the right wing Lee condition, is automatically granted by the sign of the function $N^{-1}(\delta)$. This is what we expected using Proposition 6.1 since a smile in delta which can be transformed in a smile in strike has surjective function $k \to d_1(k, \hat{\sigma}(k))$. The second condition, i.e. the left wing Lee condition, is not automatically granted.

Remark 6.8. *In section 4.1 of [64], it is shown that the high Gaussian quantile can be asymptotically written as* $N^{-1}(\delta) = \sqrt{-2\log(1-\delta)} + o(N^{-1}(\delta))$. *With a similar reasoning, it is easy to prove that for low Gaussian quantiles it holds* $N^{-1}(\delta) = -\sqrt{-2\log(\delta)} + o(N^{-1}(\delta))$. *Then, the limit inferior in Proposition 6.6 can be substituted with*

$$\liminf_{\delta \to 1} \frac{\sigma(\delta)\sqrt{T}}{\sqrt{-2\log(1-\delta)}} \qquad \textit{and} \qquad \liminf_{\delta \to 0} \frac{\sigma(\delta)\sqrt{T}}{-\sqrt{-2\log(\delta)}}.$$

We now look at the expansion of a smile in delta around the ATM point for a given expansion of the corresponding smile in strike.

Proposition 6.7. *Let $\sigma(\delta) \in \Sigma$ and $\hat{\sigma}(k)$ the corresponding smile in strike, and suppose $\hat{\sigma}(k)$ is twice differentiable in a neighborhood of $k = 0$. If*

$$\hat{\sigma}(k)\sqrt{T} = a_0 + a_1 k + a_2 k^2 + o(k^2)$$

then

$$\begin{aligned}\sigma(\delta)\sqrt{T} = a_0 &- \frac{2a_0 a_1}{n\left(\frac{a_0}{2}\right)(2 - a_0 a_1)}(\delta - \delta_{ATM}) + \\ &+ \frac{a_0(-a_0^3 a_1^3 + 4a_0^2 a_1^2 - 4a_0 a_1 + 16a_0 a_2 + 16a_1^2)}{2n\left(\frac{a_0}{2}\right)^2(2 - a_0 a_1)^3}(\delta - \delta_{ATM})^2 + \\ &+ o((\delta - \delta_{ATM})^2)\end{aligned}$$

where $\delta_{ATM} = N\left(\frac{a_0}{2}\right)$ is the delta ATM point.

Proof. The delta ATM point is $\delta_{\text{ATM}} = \delta(0) = N(d_1(0, \hat{\sigma}(0)) = N\left(\frac{a_0}{2}\right)$. Observe that from Equation (6.6), it holds $l(\delta) = -k(\delta)$. We will use the following relations:

$$\sigma(\delta) = \hat{\sigma}(k(\delta)) \tag{6.11}$$

$$\sigma'(\delta) = -\hat{\sigma}'(k(\delta))l'(\delta) \tag{6.12}$$

$$\sigma''(\delta) = \hat{\sigma}''(k(\delta))l'(\delta)^2 - \hat{\sigma}'(k(\delta))l''(\delta). \tag{6.13}$$

The first and second derivatives of $l(\delta)$ can be computed from the definition in Equation (6.5) as

$$l'(\delta) = \left(\frac{1}{n(N^{-1}(\delta))} - \frac{\sigma'(\delta)\sqrt{T}}{2}\right)\sigma(\delta)\sqrt{T} + l(\delta)\frac{\sigma'(\delta)}{\sigma(\delta)}$$

$$l''(\delta) = \left(\frac{N^{-1}(\delta)}{n(N^{-1}(\delta))^2} - \frac{\sigma''(\delta)\sqrt{T}}{2}\right)\sigma(\delta)\sqrt{T} + \left(2l'(\delta) - l(\delta)\frac{\sigma'(\delta)}{\sigma(\delta)}\right)\frac{\sigma'(\delta)}{\sigma(\delta)} + l(\delta)\frac{d}{d\delta}\frac{\sigma'(\delta)}{\sigma(\delta)}.$$

From Equation (6.11), it follows that the constant coefficient of the expansion of $\sigma(\delta)\sqrt{T}$ is a_0.

Observe that $l(\delta(0)) = -k(\delta(0)) = 0$, so that $l'(\delta(0))$ has only one term. For the first order coefficient, from Equation (6.12) we obtain $\sigma'(\delta(0))\sqrt{T} = -a_1 l'(\delta(0))$. Substituting in the expression for the derivative of $l(\delta)$ and solving for $l'(\delta(0))$, it follows

$$l'(\delta(0)) = \frac{2a_0}{n\left(\frac{a_0}{2}\right)(2 - a_0 a_1)},$$

so that we find the first order coefficient of the expansion of $\sigma(\delta)\sqrt{T}$.

Finally, the second order coefficient can be found from the expression for the second derivative of $l(\delta)$ evaluated in $\delta(0)$. As in the previous steps, substituting $\sigma''(\delta(0))\sqrt{T}$ with $2a_2 l'(\delta(0))^2 - a_1 l''(\delta(0))$ from Equation (6.13) and solving for $l''(\delta(0))$, we find

$$l''(\delta(0)) = \frac{a_0\left(a_0(2 - a_0 a_1)^2 - 8(a_0^2 a_2 + 2a_1)\right)}{n\left(\frac{a_0}{2}\right)^2(2 - a_0 a_1)^3}$$

and from this the expression for the second order coefficient of the expansion of $\sigma(\delta)\sqrt{T}$. □

6.5 Weak Arbitrage-free smiles in delta

The parametrization of the set Σ_{A} of Butterfly arbitrage-free smiles is hard. Indeed, twice differentiable functions $\sigma(\delta)$ belong to Σ_{A} if and only if they satisfy the delta version of the requirement of positivity of Equation (6.2). Observe that such condition can be written only for the subset of Σ_{A} of twice differentiable functions since it involves second derivatives of the smile.

Consider the smile inversion $k \to \hat{\sigma}(-k) =: \bar{\hat{\sigma}}(k)$. Arbitrage-free smiles are such that their inverse smile is still arbitrage-free. Even though we do not parametrize the set Σ_{A}, we show that the subset of Σ which is closed under symmetry can be parametrized, and this could make a step forward to the search of arbitrage-free smiles in delta. We define such set Σ_{WA} since it is the set of smiles satisfying the Weak Arbitrage-free conditions **F1** and **F2** (plus the surjectivity of such functions, i.e. properties **P1** and **P2**). As we showed before, the requirement defining Σ is that there exists a smile $k \to \hat{\sigma}(k)$ such that the function $d_1(k) = d_1(k, \hat{\sigma}(k))$ is decreasing and surjective. For the inverse smile, we are asking that the function $\bar{d}_1(k) = d_1(k, \bar{\hat{\sigma}}(k))$ is decreasing and surjective. It is easy to see that $\bar{d}_1(k) = -d_2(-k)$.

Taking the function $\sigma(\delta) = \hat{\sigma}(k(\delta))$ where $k(\delta) = d_1^{-1}(N^{-1}(\delta))$, the requirement that $d_2(k)$ is decreasing and surjective corresponds to the requirement that $\delta \to d_2(k(\delta))$ is increasing and surjective. It holds $d_2(k) = d_1(k) - \hat{\sigma}(k)\sqrt{T}$, so in the delta notation the requirement is that the function

$$m(\delta) := N^{-1}(\delta) - \sigma(\delta)\sqrt{T}$$

is increasing and surjective.

We can then define the subset of Σ closed for smile inversion as in the following:

Definition 6.3. *We call Σ_{WA} the set of Weak Arbitrage-free delta smiles, i.e. the set of continuous and positive delta smiles $\delta \to \sigma(\delta)$ for which there exist a continuous and positive strike smile $k \to \hat{\sigma}(k)$ with $\hat{\sigma}(k(\delta)) = \sigma(\delta)$ where $k(\delta) = d_1^{-1}(N^{-1}(\delta))$, which satisfies **P1**, **P2**, **F1**, **F2**, i.e. the functions $k \to d_{1,2}(k, \hat{\sigma}(k))$ are strictly decreasing surjective.*

In particular,

$$\begin{aligned}\Sigma_{WA} := \big\{\delta \to \sigma(\delta) \in \mathcal{C}(]0,1[, \mathbb{R}^+) \,\big|\, \exists k \to \hat{\sigma}(k) \in \mathcal{C}(\mathbb{R}, \mathbb{R}^+) \text{ s.t.} \\ k \to d_1(k, \hat{\sigma}(k)) \text{ is strictly decreasing surjective}, \\ k \to d_2(k, \hat{\sigma}(k)) \text{ is strictly decreasing surjective}, \\ \hat{\sigma}(k(\delta)) = \sigma(\delta), k(\delta) = d_1^{-1}(N^{-1}(\delta))\big\}.\end{aligned}$$

It is easy to see that if $\sigma(\delta)$ belongs to Σ_{WA}, then it is possible to define the smile $\bar{\sigma}(\delta) := \bar{\hat{\sigma}}(\bar{k}(\delta))$ where $\bar{k}(\delta) = \bar{d}_1^{-1}(N^{-1}(\delta)) = -d_2^{-1}(-N^{-1}(\delta))$, so that $\bar{\sigma}(\delta) = \hat{\sigma}(d_2^{-1}(-N^{-1}(\delta)))$ and $\bar{\sigma}(\delta)$ belongs to Σ_{WA}.

Remark 6.9. *For smiles living in Σ_{WA}, the functions $d_1(k, \hat{\sigma}(k))$ and $d_2(k, \hat{\sigma}(k))$ are strictly decreasing and surjective, so they are a.e. differentiable. As a consequence,*

the smile $\hat{\sigma}(k) = d_1(k, \hat{\sigma}(k)) - d_2(k, \hat{\sigma}(k))$ *is a.e. differentiable and so is* $\sigma(\delta) = \hat{\sigma}(d_1^{-1}(N^{-1}(\delta)))$.

6.5.1 Parametrization of $\Sigma_{\mathbf{WA}}$

We showed in Lemma 6.3 that the properties **P1** and **F1** for a smile guarantee the existence of a σ solution to Equation (6.8). We now see that smiles satisfying also **P2** and **F2**, i.e. living in Σ_{WA}, guarantee the existence and uniqueness of a σ solution.

Lemma 6.4. *For every* $\delta \to \sigma(\delta) \in \Sigma_{WA}$*, the function* l *satisfies* $l\left(\frac{1}{2}\right) < 0$, $l(\delta) \leq \frac{N^{-1}(\delta)^2}{2}$ *for all* $\delta > \frac{1}{2}$ *and there exists a unique* $\tilde{\delta}$ *such that* $l(\tilde{\delta}) = \frac{N^{-1}(\tilde{\delta})^2}{2}$*. Furthermore*

$$\sigma(\delta)\sqrt{T} = \begin{cases} N^{-1}(\delta) + \sqrt{N^{-1}(\delta)^2 - 2l(\delta)} & \text{if } \delta \leq \tilde{\delta}, \\ N^{-1}(\delta) - \sqrt{N^{-1}(\delta)^2 - 2l(\delta)} & \text{if } \delta > \tilde{\delta}. \end{cases} \tag{6.14}$$

Proof. Since Σ_{WA} is a subset of Σ, the conditions $l\left(\frac{1}{2}\right) < 0$ and $l(\delta) \leq \frac{N^{-1}(\delta)^2}{2}$ for all $\delta > \frac{1}{2}$ are satisfied by Lemma 6.3.

Since $d_2(k)$ is a decreasing and surjective function, the left wing Lee condition $\hat{\sigma}(k)\sqrt{T} < \sqrt{-2k}$ for sufficiently large $-k$ holds. Since $l(\delta(k)) = -k$, the condition becomes $\sigma(\delta(k))\sqrt{T} < \sqrt{2l(\delta(k))}$ for k negative enough. Given the monotonicity of the function $\delta(k)$, this is equivalent to requiring $\sigma(\delta)\sqrt{T} < \sqrt{2l(\delta)}$ for sufficiently large δ, in particular $\delta > \frac{1}{2}$. The σ solution of Equation (6.8) can be either of the form σ_+ Equation (6.9) or σ_- Equation (6.10). Computing the square of $\sigma_\pm$, the Lee condition is

$$N^{-1}(\delta)\left(N^{-1}(\delta) \pm \sqrt{N^{-1}(\delta)^2 - 2l(\delta)}\right) < 2l(\delta).$$

Since $\delta > \frac{1}{2}$, the quantity $N^{-1}(\delta)$ is positive and dividing we get

$$\pm\sqrt{N^{-1}(\delta)^2 - 2l(\delta)} < \frac{2l(\delta)}{N^{-1}(\delta)} - N^{-1}(\delta).$$

The right hand side is positive if and only if $2l(\delta) > N^{-1}(\delta)^2$, which cannot hold. Then the σ_+ solution does not satisfy the left wing Lee condition for small k. On the other hand, the σ_- solution always satisfies it since the above inequality holds true if and only if $2l(\delta) < N^{-1}(\delta)^2$.

Then, the σ solution is equal to σ_+ for δ smaller than a certain $\tilde{\delta}$ such that $l(\tilde{\delta}) = \frac{N^{-1}(\tilde{\delta})^2}{2}$ and then it switches to σ_-. The uniqueness of such point $\tilde{\delta}$ follows from the monotonicity of $d_2(k)$. Indeed, if two points $\tilde{\delta}$ and $\hat{\delta}$ satisfy $l(\delta) = \frac{N^{-1}(\delta)^2}{2}$, then they also satisfy $\sigma(\delta)\sqrt{T} = N^{-1}(\delta)$. In the log-forward moneyness notation, there exist $\tilde{k} = k(\tilde{\delta})$ and $\hat{k} = k(\hat{\delta})$ which satisfy $\hat{\sigma}(k)\sqrt{T} = d_1(k)$, or $d_2(k) = 0$. Since d_2 is one-to-one, $\tilde{k} = \hat{k}$ and $\tilde{\delta} = \hat{\delta}$.

□

Thanks to the above result, it is possible to parametrize the smile σ living in Σ_{WA} using a parametrization of the function l, which has to be increasing and surjective and has to satisfy the conditions of existence and uniqueness of Lemma 6.4.

We finally state how to parametrize the set Σ_{WA}, which is our main result.

Theorem 6.1 (Parametrization of weak arbitrage-free smiles)**.** *A smile $\sigma(\delta)$ belongs to Σ_{WA} if and only if it can be parametrized as*

$$\sigma(\delta)\sqrt{T} = \begin{cases} N^{-1}(\delta) + \sqrt{N^{-1}(\delta)^2 + 2\left(\int_\delta^{\frac{1}{2}} \lambda(x)\,dx + \int_{\frac{1}{2}}^{\tilde{\delta}} \mu(x)\,dx\right)} & \text{if } \delta \le \frac{1}{2}, \\ N^{-1}(\delta) + \sqrt{2\int_\delta^{\tilde{\delta}} \mu(x)\,dx} & \text{if } \frac{1}{2} < \delta \le \tilde{\delta}, \\ N^{-1}(\delta) - \sqrt{2\int_{N^{-1}(\tilde{\delta})}^{N^{-1}(\delta)} x\beta(x)\,dx} & \text{if } \delta > \tilde{\delta}, \end{cases} \tag{6.15}$$

where

- $\tilde{\delta} \in \left]\frac{1}{2}, 1\right[$;
- λ *is an a.e. continuous positive function defined on* $\left]0, \frac{1}{2}\right]$ *such that* $\int_0^{\frac{1}{2}} \lambda(x)\,dx = \infty$;
- μ *is an a.e. continuous positive function defined on* $\left[\frac{1}{2}, \tilde{\delta}\right[$;
- β *is an a.e. continuous function defined on* $]N^{-1}(\tilde{\delta}), \infty[$ *such that* $\beta(x) \in]0, 1[$ *a.e. and* $\int_{N^{-1}(\tilde{\delta})}^{\infty} x\beta(x)\,dx = \int_{N^{-1}(\tilde{\delta})}^{\infty} x(1-\beta(x))\,dx = \infty$.

Proof. Let $\sigma(\delta) \in \Sigma_{\text{WA}}$, then $\sigma(\delta)$ is a.e. differentiable for Remark 6.9, and has the form in Equation (6.14) for Lemma 6.4. The function $l(\delta)$ is increasing and surjective and it satisfies $l\left(\frac{1}{2}\right) < 0$, $l(\delta) \le \frac{N^{-1}(\delta)^2}{2}$ for all $\delta > \frac{1}{2}$, and there exists a unique $\tilde{\delta}$ such that $l(\tilde{\delta}) = \frac{N^{-1}(\tilde{\delta})^2}{2}$. Since $d_2(k)$ is decreasing and surjective, the function $m(\delta) = N^{-1}(\delta) - \sigma(\delta)\sqrt{T}$ is increasing and surjective. Substituting with the expression for $\sigma(\delta)$, $m(\delta)$ can be re-written as $\mp\sqrt{N^{-1}(\delta)^2 - 2l(\delta)}$, where the sign is negative for $\delta \le \tilde{\delta}$ and positive otherwise. Its derivative, when it is defined, is $\mp\frac{\frac{N^{-1}(\delta)}{n(N^{-1}(\delta))} - l'(\delta)}{\sqrt{N^{-1}(\delta)^2 - 2l(\delta)}}$, and it is positive. Equivalently, the derivative of l satisfies

$$\begin{aligned} l'(\delta) &> \frac{N^{-1}(\delta)}{n(N^{-1}(\delta))} \text{ a.e.} \quad \text{if } \delta < \tilde{\delta}, \\ l'(\delta) &< \frac{N^{-1}(\delta)}{n(N^{-1}(\delta))} \text{ a.e.} \quad \text{if } \delta > \tilde{\delta} \end{aligned} \tag{6.16}$$

The first inequality is weaker than $l'(\delta) > 0$ a.e. if $\delta < \frac{1}{2}$. Consider then $\delta \in \left[\frac{1}{2}, \tilde{\delta}\right[$. The first inequality implies that there is a positive and a.e. continuous function $\mu(\delta)$ on $\left[\frac{1}{2}, \tilde{\delta}\right[$, such that $l'(\delta) = \frac{N^{-1}(\delta)}{n(N^{-1}(\delta))} + \mu(\delta)$. Taking the integral from δ to $\tilde{\delta}$ results

into $l(\delta) = \frac{N^{-1}(\delta)^2}{2} - \int_\delta^{\tilde{\delta}} \mu(x)\,dx$. If $\delta \leq \frac{1}{2}$, the fact that $l'(\delta)$ is positive can be written as $l'(\delta) = \lambda(\delta)$ where $\lambda(\delta)$ is an a.e. continuous positive function defined on $]0, \frac{1}{2}]$. Taking the integral between δ and $\frac{1}{2}$ implies $l(\delta) = l\left(\frac{1}{2}\right) - \int_\delta^{\frac{1}{2}} \lambda(x)\,dx$. Substituting with the value of $l\left(\frac{1}{2}\right)$ in the expression with μ, it holds $l(\delta) = -\int_\delta^{\frac{1}{2}} \lambda(x)\,dx - \int_{\frac{1}{2}}^{\tilde{\delta}} \mu(x)\,dx$. Since $l(\delta)$ is surjective and $l\left(\frac{1}{2}\right)$ is finite, then $\int_0^{\frac{1}{2}} \lambda(x)\,dx = \infty$. Similarly, for $\delta > \tilde{\delta}$, the property of the derivative of $l(\delta)$ implies $l(\delta) = \frac{N^{-1}(\delta)^2}{2} - \int_{\tilde{\delta}}^{\delta} \eta(x)\,dx$ for an a.e. continuous positive function $\eta(\delta)$ defined on $]\tilde{\delta}, 1[$. Also, since $l'(\delta) > 0$, $\eta(\delta)$ is smaller than $\frac{N^{-1}(\delta)}{n(N^{-1}(\delta))}$, so $\eta(\delta) = \alpha(\delta)\frac{N^{-1}(\delta)}{n(N^{-1}(\delta))}$ where $\alpha(\delta)$ is a function strictly bounded between 0 and 1 and a.e. continuous on $]\tilde{\delta}, 1[$. Since $\int_{\tilde{\delta}}^{\delta} \alpha(x)\frac{N^{-1}(x)}{n(N^{-1}(x))}\,dx = \int_{N^{-1}(\tilde{\delta})}^{N^{-1}(\delta)} x\alpha(N(x))\,dx$ and $\frac{N^{-1}(\delta)^2}{2} = \int_{N^{-1}(\tilde{\delta})}^{N^{-1}(\delta)} x\,dx + \frac{N^{-1}(\tilde{\delta})^2}{2}$, it holds $l(\delta) = \frac{N^{-1}(\tilde{\delta})^2}{2} + \int_{N^{-1}(\tilde{\delta})}^{N^{-1}(\delta)} x(1-\alpha(N(x))\,dx$ for $\delta > \tilde{\delta}$. The function $\beta(x) = \alpha(N(x))$ is strictly bounded between 0 and 1 and a.e. continuous on $]N^{-1}(\tilde{\delta}), \infty[$. Furthermore, $l(1) = \infty$, then $\int_{N^{-1}(\tilde{\delta})}^{\infty} x(1-\beta(x))\,dx = \infty$. In order to have $m(1) = \infty$, it must hold $\int_{N^{-1}(\tilde{\delta})}^{\infty} x\beta(x)\,dx = \infty$.

On the other hand, if a function $\sigma(\delta)$ has the form in Equation (6.15), the function $l(\delta)$ is

$$l(\delta) = \begin{cases} -\int_\delta^{\frac{1}{2}} \lambda(x)\,dx - \int_{\frac{1}{2}}^{\tilde{\delta}} \mu(x)\,dx & \text{if } \delta \leq \frac{1}{2}, \\ \frac{N^{-1}(\delta)^2}{2} - \int_\delta^{\tilde{\delta}} \mu(x)\,dx & \text{if } \frac{1}{2} < \delta \leq \tilde{\delta}, \\ \frac{N^{-1}(\tilde{\delta})^2}{2} + \int_{N^{-1}(\tilde{\delta})}^{N^{-1}(\delta)} x(1-\beta(x))\,dx & \text{if } \delta > \tilde{\delta}. \end{cases}$$

Given the hypothesis on the parameters, it is easy to show that $l(\delta)$ is a.e. differentiable and $l'(\delta) > 0$ for every δ where the derivative is defined. Also, $l(0) = -\int_0^{\frac{1}{2}} \lambda(x)\,dx = -\infty$ and $l(1) = \infty$. So far, we have proven $\sigma(\delta) \in \Sigma$. The only requirement left is $m(\delta)$ increasing and surjective. The monotonicity holds since inequalities in Equation (6.16) are verified. For the surjectivity, $m(0) = -\sqrt{N^{-1}(0)^2 + 2\left(\int_0^{\frac{1}{2}} \lambda(x)\,dx + \int_{\frac{1}{2}}^{\tilde{\delta}} \mu(x)\,dx\right)}$ which is $-\infty$, while $m(1)$ is equal to $\sqrt{2\int_{N^{-1}(\tilde{\delta})}^{\infty} x\beta(x)\,dx}$ which diverges.

□

A first consequence of Theorem 6.1 is that the given reparametrization of the smiles in delta guarantees a simpler calibration than the one in Section 6.3.4 as we will see in Section 6.5.2. Indeed, conditions on λ, μ and β are more easily verified in practice.

Interestingly, in the attempt of characterizing the Fukasawa conditions **F1** and **F2** considering the parametrization in log-forward moneyness, Lucic has found similar results in Theorem 2.2 of [51]. In particular, Lucic's theorem states that an a.e. differentiable smile $\hat{\sigma}(k)$ has both $d_1(k, \hat{\sigma}(k))$ and $d_2(k, \hat{\sigma}(k))$ strictly decreasing if and only if it

can be parametrized as

$$\hat{\sigma}(k)\sqrt{T} = \begin{cases} -\sqrt{2k-\phi(k)} + \sqrt{-\phi(k)} & \text{if } k \leq k_*, \\ \sqrt{2k-\phi(k)} + \sqrt{-\phi(k)} & \text{if } k_* < k \leq k^*, \\ \sqrt{2k+\phi(k)} - \sqrt{\phi(k)} & \text{if } k > k^* \end{cases} \tag{6.17}$$

where $k_* < 0 < k^*$ and $\phi(k)$ is a continuous increasing function such that $\phi(k_*) = 2k_*$, $\phi(k^*) = 0$, and $\phi'(k) > 2$ for $k < k_*$, $\phi'(k) < 2$ for $k_* < k \leq k^*$.

Note that results in Theorem 6.1 have been achieved independently and looking solely at the delta parametrization. It can be shown that Theorem 6.1 and Lucic's theorem are equivalent when some requirements on the limits at infinity of the functions ϕ and $2k - \phi(k)$ are added. Indeed, the additional conditions are needed under **P1** and **P2** and guarantee the surjectivity of functions $d_1(k, \hat{\sigma}(k))$ and $d_2(k, \hat{\sigma}(k))$. In the following proposition we explain how to pass from the parametrization in Theorem 6.1 to Lucic's parametrization as in Equation (6.17). The proof can be found in Section 6.A.

Proposition 6.8. *A smile $\sigma(\delta) \in \Sigma_{WA}$ with parametrization as in Theorem 6.1 has corresponding smile in strike $\hat{\sigma}(k)$ with parametrization as in Equation (6.17) where $\phi(k) := -N^{-1}(l^{-1}(-k))|N^{-1}(l^{-1}(-k))|$ and*

$$l(\delta) := \begin{cases} -\int_\delta^{\frac{1}{2}} \lambda(x)\,dx - \int_{\frac{1}{2}}^{\tilde{\delta}} \mu(x)\,dx & \text{if } \delta \leq \frac{1}{2}, \\ \frac{N^{-1}(\delta)^2}{2} - \int_\delta^{\tilde{\delta}} \mu(x)\,dx & \text{if } \frac{1}{2} < \delta \leq \tilde{\delta}, \\ \frac{N^{-1}(\tilde{\delta})^2}{2} + \int_{N^{-1}(\tilde{\delta})}^{N^{-1}(\delta)} x(1-\beta(x))\,dx & \text{if } \delta > \tilde{\delta}. \end{cases} \tag{6.18}$$

Vice versa, consider a smile $\hat{\sigma}(k)$ with parametrization as in Equation (6.17) and surjective function $\phi(k)$ such that $2k - \phi(k)$ goes to ∞ for k going to $-\infty$. Then, it has corresponding smile in delta $\sigma(\delta)$ with parametrization as in Theorem 6.1 where

$$\begin{aligned} \lambda(\delta) &:= -\frac{2\chi(\delta)}{\phi'\big(\phi^{-1}(\chi(\delta)^2)\big)n(\chi(\delta))} \\ \mu(\delta) &:= \frac{\chi(\delta)}{n(\chi(\delta))}\left(\frac{2}{\phi'\big(\phi^{-1}(-\chi(\delta)^2)\big)} - 1\right) \\ \beta(x) &:= 1 - \frac{2}{\phi'\big(\phi^{-1}(-x^2)\big)} \end{aligned} \tag{6.19}$$

and $\chi(\delta) = N^{-1}(\delta)$.

Requirements on the parameters

The relation of the functions λ, μ and β with l and m is detailed in Section 6.B.

In Theorem 6.1, the positivity of parameter λ and the requirement $\beta < 1$ are directly linked to the fact that the function $l(\delta)$ must be increasing. The positivity of μ and β is instead connected with the monotonicity of the function $m(\delta)$.

The requirement $\int_0^{\frac{1}{2}} \lambda(x)\,dx = \infty$ comes from the fact that $l(0) = -\infty$ and it also implies $m(0) = -\infty$.

The requirement $\int_{N^{-1}(\tilde{\delta})}^{\infty} x(1-\beta(x))\,dx$ arises to satisfy $l(1) = \infty$. Observe that if $\lim_{x\to\infty} \beta(x) < 1$, the requirement is automatically satisfied, but this is a sufficient and not necessary property of β. Indeed, the choice of $\beta(x) = 1 - \frac{c}{x}$ for a positive constant $c \leq N^{-1}(\tilde{\delta})$ still satisfies conditions of Theorem 6.1 even though β has right limit equal to 1.

The last requirement $\int_{N^{-1}(\tilde{\delta})}^{\infty} x\beta(x)\,dx = \infty$ originates from $m(1) = \infty$. Similarly as before, taking $\lim_{x\to\infty} \beta(x) > 0$ automatically satisfies this requirement, even though it is not a necessary condition on β. As a counterexample, we could indeed take $\beta(x) = \frac{c}{x}$ with c positive and smaller than $N^{-1}(\tilde{\delta})$.

Remark 6.10. *If we are requiring the delta smile to not have discontinuity points, we should add additional conditions on parameters. In particular, the function $l(\delta)$ should be differentiable also in $\frac{1}{2}$ and $\tilde{\delta}$. Which means that it should hold $\lambda\left(\frac{1}{2}\right) = \mu\left(\frac{1}{2}\right)$ and $\mu(\tilde{\delta}) = -\frac{N^{-1}(\tilde{\delta})}{n(N^{-1}(\tilde{\delta}))}\beta(N^{-1}(\tilde{\delta})) = 0$ since all the functions are positive.*

6.5.2 Application: calibration of weak arbitrage-free smiles in delta

In this section we develop a similar calibration procedure as the one in Section 6.3.4 but for weak arbitrage-free smiles, i.e. smiles living in Σ_{WA}.

There are two methodologies that can be designed. The first one is tricky to be implemented. Indeed, it reconsiders the calibration in Section 6.3.4 and, in order to have that the smile in delta lives in Σ_{WA}, it requires to add in step 3. the conditions:

f. $l'(\delta) > \frac{N^{-1}(\delta)}{n(N^{-1}(\delta))}$ a.e. for $\delta \in \left]\frac{1}{2}, \tilde{\delta}\right[$,

g. $l'(\delta) < \frac{N^{-1}(\delta)}{n(N^{-1}(\delta))}$ a.e. for $\delta > \tilde{\delta}$,

so that the function $m(\delta)$ is increasing and surjective.

Interpolating a function l which satisfies these requirements is not immediate. For this reason, a second more cunning calibration methodology can be implemented, using Theorem 6.1. The target of such calibration routine are the functions λ, μ and β. These functions must satisfy the requirements in Theorem 6.1 in order to guarantee that the smile in Equation (6.15) can be transformed into a smile in strike satisfying the conditions of bijectivity of the functions $d_1(k, \hat{\sigma}(k))$ and $d_2(k, \hat{\sigma}(k))$.

The steps to be performed become:

1. a. consider the market discrete pillars in delta notation $\{(\delta_i, \sigma_i)\}_i$;
   b. if the market discrete pillars are in strike notation $\{(k_i, \sigma_i)\}_i$, convert them to the pillars in delta notation $\{(\delta_i, \sigma_i)\}_i$ by defining $\delta_i = N(d_1(k_i, \sigma_i))$;
2. given the pillars in point 1., interpolate/extrapolate a function $\delta \to \sigma(\delta)$ defined as in Theorem 6.1 such that

a. $\tilde{\delta} \in \left]\frac{1}{2}, 1\right[$;

b. λ is a positive function defined on $\left]0, \frac{1}{2}\right]$ such that $\int_0^{\frac{1}{2}} \lambda(x)\,dx = \infty$;

c. μ is a positive function defined on $\left[\frac{1}{2}, \tilde{\delta}\right[$;

d. β is a function defined on $]N^{-1}(\tilde{\delta}), \infty[$ such that $\beta(x) \in]0, 1[$, and $\int_{N^{-1}(\tilde{\delta})}^{\infty} x\beta(x)\,dx = \int_{N^{-1}(\tilde{\delta})}^{\infty} x(1-\beta(x))\,dx = \infty$.

The requirements on the functions λ, μ and β can be easily achieved parametrizing these functions. As an example, for chosen positive degrees n_λ, $m_\lambda > 1$, n_μ, m_μ and n_β, we could define

$$\begin{aligned}
\lambda(\delta) &= a_{n_\lambda}\delta^{n_\lambda} + \cdots + a_1 x + a_0 + \frac{a_{-1}}{\delta} + \cdots + \frac{a_{-m_\lambda}}{\delta^{m_\lambda}} \\
\mu(\delta) &= c_{n_\mu}\delta^{n_\mu} + \cdots + c_1\delta + c_0 + \frac{c_{-1}}{\delta} + \cdots + \frac{c_{-m_\mu}}{\delta^{m_\mu}} \\
\beta(x) &= \frac{b_{n_\beta}N(x)^{n_\beta} + \cdots + b_1 N(x)}{b_{n_\beta} + \cdots + b_1 + b_0}
\end{aligned}$$

and require the coefficients a_i, c_i and b_i to be non-negative, b_0 to be non zero, and at least one of the a_i for $i < -1$ and one of the b_i with $i > 0$ to be non zero. In alternative, all parameters can be required to be positive.

These definitions satisfy the requirements on the functions λ, μ and β because they are positive, continuous, $\beta(x) < 1$, and $\int_0^{\frac{1}{2}} \lambda(x)\,dx$ diverges because of the terms $\frac{a_{-i}}{\delta^i}$ for $i > 1$. The divergence of the integrals $\int_{N^{-1}(\tilde{\delta})}^{\infty} x\beta(x)\,dx$ and $\int_{N^{-1}(\tilde{\delta})}^{\infty} x(1-\beta(x))\,dx$ is guaranteed by the fact that $\lim_{x\to\infty}\beta(x) \in]0, 1[$.

In this way, the calibration algorithm in point 2. can be implemented as a least squares, i.e. a minimization algorithm on the target function defined as the sum of squared differences between market and model delta volatilities evaluated at market delta points $\{(\delta_i, \sigma_i)\}_i$. In particular, the minimization depends on the positive parameters a_i, c_i and b_i. The `scipy` library of Python can easily perform this kind of calibrations via the function `optimize.least_squares`.

6.5.3 Examples of smiles in $\Sigma_{\mathbf{WA}}$

Bounded smiles

Let us look at the requirement on β detailed in Section 6.5.1. It has been shown that $\beta(x) = 1 - \frac{c_\beta}{x}$ with $c_\beta \leq N^{-1}(\tilde{\delta})$ satisfies conditions of Theorem 6.1. An interesting consequence of this example is that the limit of $\sigma(\delta)$ in 1 is finite in this case. Indeed, for $\delta > \tilde{\delta}$, it holds

$$\sigma(\delta)\sqrt{T} = N^{-1}(\delta) - \sqrt{N^{-1}(\delta)\big(N^{-1}(\delta) - 2c_\beta\big) + d}$$

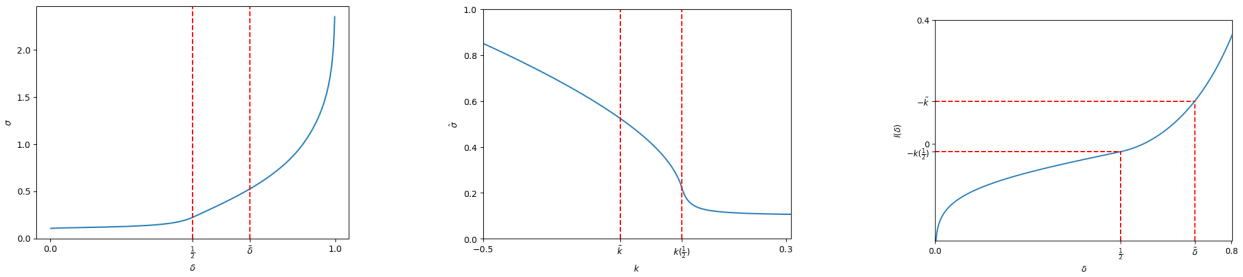

Figure 6.1: Skew shaped smile with bounded left wing in delta (left) and bounded right wing in log-forward moneyness (center) obtained with parameters as in Equation (6.20). On the right, the corresponding function $l(\delta) = -k(\delta)$ for $\delta < 0.8$.

where $d = N^{-1}(\tilde{\delta})^2(2c_\beta - 1)$ is a constant term. In turn, the latter expression coincides with

$$\frac{2c_\beta N^{-1}(\delta) - d}{N^{-1}(\delta) + \sqrt{N^{-1}(\delta)\big(N^{-1}(\delta) - 2c_\beta\big) + d}}$$

which converges to c_β as δ goes to 1. This means that it is possible to obtain bounded smiles on the right appropriately choosing the function $\beta(x)$ in the parametrization of Theorem 6.1.

Similarly, it is possible to have bounded wings on the left choosing a suitable $\lambda(\delta)$ function. For example, we can define $\lambda(\delta) = \frac{c_\lambda}{n(N^{-1}(\delta))}$ to have the convergence of the smile to c_λ on the left.

Figure 6.1 shows a skew-shaped smile, which notably has a bounded left wing of the smile in delta. The function λ is defined as above with $c_\lambda = 0.1$. In order to guarantee continuity of the derivative and a nice shape of the smile, the other parameters have been chosen as

$$\mu(\delta) = \frac{c_\lambda}{n(0)} \frac{\tilde{\delta} - \delta}{\tilde{\delta} - \frac{1}{2}} \qquad \beta(x) = \begin{cases} \frac{n(x)}{x}\mu(2\tilde{\delta} - N(x)) & \text{if } x < N^{-1}(\hat{\delta}) \\ \frac{n(N^{-1}(\hat{\delta}))}{N^{-1}(\hat{\delta})}\mu\big(\frac{1}{2}\big) & \text{if } x \geq N^{-1}(\hat{\delta}) \end{cases} \tag{6.20}$$

where $\hat{\delta} = 2\tilde{\delta} - \frac{1}{2}$ and $\tilde{\delta} = 0.7$. As a consequence, $\tilde{k} = k(\tilde{\delta}) \approx -0.137$ and $k\big(\frac{1}{2}\big) \approx 0.025$.

This example can be further pushed to obtain a flat smile. Indeed, if we want a flat total implied volatility $\sigma(\delta)\sqrt{T}$ at a level c, we can define

$$\lambda(\delta) = \frac{c}{n(N^{-1}(\delta))} \qquad \mu(\delta) = \frac{c - N^{-1}(\delta)}{n(N^{-1}(\delta))} \qquad \beta(x) = 1 - \frac{c}{x}$$

and $\tilde{\delta} = N(c)$.

Remark 6.11. *Smiles of the form in Theorem 6.1 allow for bounded wings and for flat shapes.*

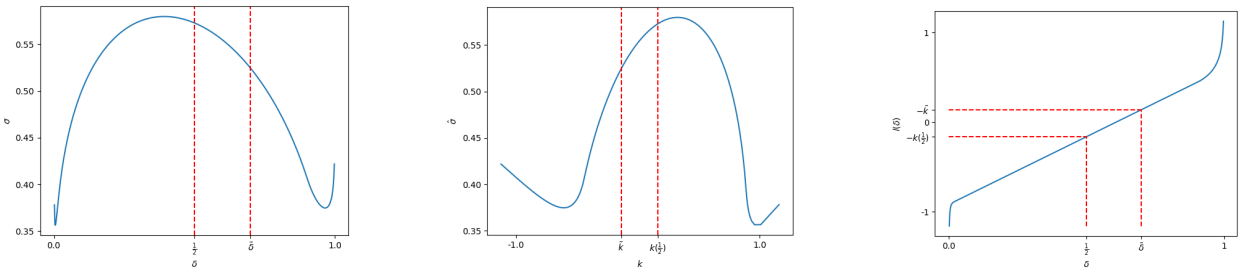

Figure 6.2: W-shaped smile in delta (left) and in log-forward moneyness (center) obtained with parameters as in Equation (6.21). On the right, the corresponding function $l(\delta) = -k(\delta)$.

W-shaped smile

The parametrization in Theorem 6.1 can be used to model very different kind of smiles, and also unusual ones. For example, we can model 'sad smiles' setting $\tilde{\delta} = 0.7$ and

$$\lambda(\delta) = \begin{cases} \frac{\delta_1^2 c}{\delta^2} & \text{if } \delta < \delta_1 \\ c & \text{if } \delta \geq \delta_1 \end{cases} \qquad \mu(\delta) = c - \frac{N^{-1}(\delta)}{n(N^{-1}(\delta))}$$
$$\beta(x) = \begin{cases} 1 - c\frac{n(x)}{x} & \text{if } x < N^{-1}(\delta_2) \\ 1 - c\frac{n(N^{-1}(\delta_2))}{N^{-1}(\delta_2)} & \text{if } x \geq N^{-1}(\delta_2) \end{cases} \tag{6.21}$$

where $c = \frac{N^{-1}(\tilde{\delta})}{n(N^{-1}(\tilde{\delta}))}$, $\delta_1 = 0.02$ and $\delta_2 = 0.9$.

With these parameters, all conditions of Theorem 6.1 are satisfied and the resulting smile in delta $\sigma(\delta)$ has a W-shape as in Figure 6.2.

It is easy to show that the left and right limits of the smile (both in delta and in strike) are infinite. Choosing different values for δ_1 and δ_2 allows to move the location of the two minima and the maxima of the smile. In this way, it is possible to obtain smiles with W-shapes that have been described in the log-normal mixture framework by Glasserman and Pirjol [37] and have been seen, for example, for the Amazon stock on the 26 of April 2018 for options with expiry 27 April 2018, before to the first quarter earnings announcement.

SVI

The SVI model has been introduced by Gatheral at the Global Derivatives conference in Madrid in 2004 [33]. It is a model for the implied total variance $\hat{\omega}(k) = \hat{\sigma}(k)^2 T$ as a function of the log-forward moneyness k and it is defined as

$$\hat{\omega}(k) = a + b\big(\rho(k-m) + \sqrt{(k-m)^2 + \bar{\sigma}^2}\big).$$

We suppose that the SVI parameters under study satisfy **P1**, **P2**, **F1** and **F2**. These conditions have been explicited in [53]. In such way, the corresponding smile in delta

obtained through the definition $\sigma(\delta) = \hat{\sigma}(k(\delta))$ where $k(\delta) = N^{-1}(d_1^{-1}(\delta))$, belongs to Σ_{WA}. In the following, we do not compute explicitly the whole delta parametrization of SVI smiles and we just compute the quantity $\tilde{\delta}$.

By Lemma 6.4, there exists a unique $\tilde{\delta}$ such that $l(\tilde{\delta}) = \frac{N^{-1}(\tilde{\delta})^2}{2}$. In the strike notation, this is equivalent to say that there exists a unique $\tilde{k}$ such that $-\tilde{k} = \frac{d_1(\tilde{k})^2}{2}$, or simplifying $\sigma(\tilde{k})^2 T = -2\tilde{k}$. We now calculate such $\tilde{k}$.

We need to look at the solutions of

$$a + b(\rho(k-m) + \sqrt{(k-m)^2 + \bar{\sigma}^2}) = -2k,$$

or equivalently

$$-(2 + b\rho)k + b\rho m - a = b\sqrt{(k-m)^2 + \bar{\sigma}^2}.$$

Under the Lee moment formula for sufficiently large δ (larger than $\frac{1}{2}$), or sufficiently large $-k$, it holds $b(1-\rho) < 2$, so $2 + b\rho > 2 + b\rho - b > 0$. Then, the above condition is never satisfied if $k \geq \frac{b\rho m - a}{2 + b\rho} := E$. Otherwise, we can take the square and simplifying, one recovers a second-degree equation of the form $Ak^2 + Bk + C = 0$ where

$$\begin{aligned} A &:= (2 + b\rho - b)(2 + b\rho + b) \\ B &:= 2\big(a(2 + b\rho) - m\big(b^2(1-\rho^2) - 2b\rho\big)\big) \\ C &:= (b\rho m - a)^2 - b^2(m^2 + \bar{\sigma}^2) = 0. \end{aligned}$$

The leading coefficient A is positive, and the Delta of such equation is $\Delta = 4b^2((a + 2m)^2 + \bar{\sigma}^2(2 + b\rho - b)(2 + b\rho + b))$, which is also positive since both terms are positive. Let us call k_+ and k_- the two possible solutions, with $k_- < k_+$. They are acceptable if and only if they are smaller than E, or if and only if $\pm\sqrt{\Delta} < 2AE + B$ respectively. The RHS is $\frac{2b^2(a+2m)}{2+b\rho}$, which is positive if and only if $a > -2m$. In such case, the $+$ solution is acceptable if and only if

$$\begin{aligned} 0 < (2AE + B)^2 - \Delta &= 2A(2AE^2 - EB + 2C) \\ &= -\frac{4b^2 A}{(2 + b\rho)^2}\big((a + 2m)^2 + \bar{\sigma}^2(2 + b\rho)^2\big) \end{aligned}$$

which is not possible. On the other hand, the $-$ solution is acceptable if and only if $a > -2m$ or $\Delta - (2AE + B)^2 > 0$, which is always verified as proved above.

In particular,

$$\tilde{k} = \frac{bm(2\rho - b(1-\rho^2)) - a(2 + b\rho) - b\sqrt{(a + 2m)^2 + \bar{\sigma}^2(2 + b(1+\rho))(2 - b(1-\rho))}}{(2 + b(1+\rho))(2 - b(1-\rho))}.$$

The SVI model has given birth to other sub-models, obtained reducing the original 5 parameters model to a model with fewer parameters. Among them, the SSVI model by Gatheral and Jacquier [36] has been largely used in industry. It has the form

$$\hat{\omega}(k) = \frac{\theta}{2}\big(1 + \rho\varphi k + \sqrt{(\varphi k + \rho)^2 + (1-\rho^2)}\big).$$

where the parameters are defined from the SVI ones as $\varphi = \frac{\sqrt{1-\rho^2}}{\sigma}$ and $\theta = \frac{2b\sigma}{\sqrt{1-\rho^2}}$.

In the case of SSVI, the expression for $\tilde{k}$ is easier. Indeed

$$\tilde{k} = -\frac{2\theta}{R}$$

where

$$R = \Big(2 + \frac{\theta\varphi}{2}(1+\rho)\Big)\Big(2 - \frac{\theta\varphi}{2}(1-\rho)\Big)$$

is positive because of the Lee bounds, which require $\frac{\theta\varphi}{2}(1+|\rho|) < 2$. The corresponding delta is $\tilde{\delta} = N\Big(4\sqrt{\frac{\theta}{R}}\Big)$.

6.6 Conclusion

The possibility to pass from a smile in delta to a smile in strike and vice versa has been characterized requiring that the d_1 function of the Black-Scholes formula has to be decreasing and surjective. This condition is one of the two necessary requirements for the absence of Butterfly arbitrage obtained by Fukasawa. The requirement that the d_2 function is decreasing too ensures that also the symmetric smile can be transformed into the delta space.

The requirements that the d_1 and d_2 functions have to be decreasing can be translated into the delta space with specific conditions. These conditions identify a characterization of the set of smiles in delta satisfying the weak no Butterfly arbitrage requirements and allow to parametrize such set. As a consequence, we have obtained a parametrization depending on one real number and three positive functions which guarantees that the resulting smiles in delta satisfy the weak no arbitrage conditions identified by Fukasawa.

Practitioners who use smiles in delta could use those parametrizations to ensure at least weak no Butterfly arbitrage. An open challenging task is to characterize the subfamily of no Butterfly arbitrage smiles in delta. We recall that the task of characterizing the set of Butterfly arbitrage-free smiles is open in both delta and strike spaces. The results in the present chapter give hope of achieving the characterization of such set using the delta parametrization.

6.A Proof of Proposition 6.8

Let $\sigma(\delta)$ be a smile in Σ_{WA}, then it is a.s. differentiable. Then, for Theorem 6.1 it has a parametrization as in Equation (6.15), and for Theorem 2.2 of [51] the corresponding smile $\hat{\sigma}(k)$ has the parametrization of Equation (6.17) since it has decreasing functions $d_1(k, \hat{\sigma}(k))$ and $d_2(k, \hat{\sigma}(k))$. From the proof of Theorem 6.1, the function $l(\delta)$ defined as in Equation (6.18) is strictly increasing and surjective, so that it has an inverse. Given the parametrization of $\sigma(\delta)$, it holds $l(\delta) = \big(N^{-1}(\delta) - \frac{\sigma(\delta)\sqrt{T}}{2}\big)\sigma(\delta)\sqrt{T}$. Since $N^{-1}(\delta(k)) = d_1(k, \hat{\sigma}(k))$ and $\sigma(\delta(k)) = \hat{\sigma}(k)$, it follows $l(\delta(k)) = -k$, or $\delta(k) = l^{-1}(-k)$. In particular, $N^{-1}(l^{-1}(-k)) = d_1(k, \hat{\sigma}(k))$ and $\phi(k) = -d_1(k, \hat{\sigma}(k))|d_1(k, \hat{\sigma}(k))|$. This function coincides with the one given in proof of Theorem 2.2 of [51], so that the definition of $\phi(k)$ is correct. In particular, $\delta(k^*)$ coincides with $N(d_1(k^*, \hat{\sigma}(k^*))) = N(0)$, i.e. $\delta(k^*) = \frac{1}{2}$, and $l(\delta(k_*)) = -k_*$ which is equivalent to $-\frac{\phi(k_*)}{2}$ and, from the definition of $\phi(k)$, to $\frac{N^{-1}(\delta(k_*))^2}{2}$, i.e. $\delta(k_*) = \tilde{\delta}$. The bounds on the slope of $\phi'(k)$ coincide with the requirement that the function $m(\delta)$ is increasing. Indeed, $\phi(k) = -N^{-1}(l^{-1}(-k))^2$ for $k \leq k^*$ so its derivative is

$$\phi'(k) = 2\frac{N^{-1}(l^{-1}(-k))}{n(N^{-1}(l^{-1}(-k)))}\frac{1}{l'(l^{-1}(-k))}$$

or equivalently, substituting k with $k(\delta)$ for $\delta \geq \frac{1}{2}$,

$$\phi'(k(\delta)) = 2\frac{N^{-1}(\delta)}{n(N^{-1}(\delta))}\frac{1}{l'(\delta)}.$$

This quantity is larger than 2 for $k < k_*$ and smaller than 2 for $k_* < k \leq k^*$ iff the quantity $l'(\delta)$ is smaller than $\frac{N^{-1}(\delta)}{n(N^{-1}(\delta))}$ for $\delta > \tilde{\delta}$ and larger than $\frac{N^{-1}(\delta)}{n(N^{-1}(\delta))}$ for $\frac{1}{2} \leq \delta < \tilde{\delta}$. This corresponds to Equation (6.16), i.e. to the fact that $m(\delta)$ is increasing.

Vice versa, let $\hat{\sigma}(k)$ be a smile with parametrization Equation (6.17). Since $\phi(k)$ is surjective and increasing, it is a.s. differentiable, and so is $\hat{\sigma}(k)$. For Theorem 2.2 of [51], $\hat{\sigma}(k)$ has strictly decreasing functions $d_1(k, \hat{\sigma}(k))$ and $d_2(k, \hat{\sigma}(k))$. Furthermore, in the proof of Theorem 2.2, Lucic shows that $d_1(k, \hat{\sigma}(k)) = \pm\sqrt{\mp\phi(k)}$ for $k \lessgtr k^*$ respectively, and

$$d_2(k, \hat{\sigma}(k)) = \begin{cases} \sqrt{2k - \phi(k)} & \text{if } k \leq k_* \\ -\sqrt{2k - \phi(k)} & \text{if } k_* < k \leq k^* \\ -\sqrt{2k + \phi(k)} & \text{if } k > k^*. \end{cases}$$

As a consequence, since $\phi(k)$ is required to be surjective, $d_1(k, \hat{\sigma}(k))$ is also surjective and $d_2(k, \hat{\sigma}(k))$ has infinite right limit. Furthermore, since for hypothesis $2k - \phi(k)$ explodes at $-\infty$, also $d_2(k, \hat{\sigma}(k))$ does. Then, the smile $\hat{\sigma}(k)$ admits a delta smile $\sigma(\delta)$ which lives in Σ_{WA} and it has a parametrization as in Equation (6.15) for Theorem 6.1. It holds $d_1(k(\delta), \sigma(\delta)) = N^{-1}(\delta)$, so that $N^{-1}(\delta) = \mp\sqrt{\pm\phi(k(\delta))}$ for $\delta \lessgtr \frac{1}{2}$ or equivalently $k(\delta) = \phi^{-1}(\pm N^{-1}(\delta)^2)$ for $\delta \lessgtr \frac{1}{2}$. Then $k\big(\frac{1}{2}\big) = \phi^{-1}(0) = k^*$ and $\phi(k(\tilde{\delta})) = -N^{-1}(\tilde{\delta})^2$

which coincides with $-2l(\tilde{\delta})$ and in turn with $2k(\tilde{\delta})$, so $k(\tilde{\delta}) = k_*$. From the definition of $\mu(\delta)$ in Equation (6.19) and from the fact that

$$\frac{N^{-1}(\tilde{\delta})^2}{2} + \phi^{-1}(-N^{-1}(\tilde{\delta})^2) = -\frac{\phi(k(\tilde{\delta}))}{2} + k(\tilde{\delta}) = 0,$$

it is easy to see that

$$\int_{\delta}^{\tilde{\delta}} \mu(x)\,dx = \frac{N^{-1}(\delta)^2}{2} + \phi^{-1}(-N^{-1}(\delta)^2).$$

It holds $\sigma(\delta)\sqrt{T} = \hat{\sigma}(k(\delta))\sqrt{T}$, which has the form $\sqrt{2k(\delta) - \phi(k(\delta))} + \sqrt{-\phi(k(\delta))}$ for $\frac{1}{2} < \delta \leq \tilde{\delta}$. Substituting with the expression of $k(\delta)$, it follows $\sigma(\delta)\sqrt{T} = N^{-1}(\delta) + \sqrt{2\phi^{-1}(-N^{-1}(\delta)^2) + N^{-1}(\delta)^2}$, or $\sigma(\delta)\sqrt{T} = N^{-1}(\delta) + \sqrt{2\int_{\delta}^{\tilde{\delta}} \mu(x)\,dx}$. This is what we looked for, so the definition of $\mu(\delta)$ is correct. The proof is similar for $\lambda(\delta)$ and $\beta(x)$. The conditions on parameters $\lambda(\delta)$, $\mu(\delta)$ and $\beta(x)$ in Theorem 6.1 hold true because $\sigma(\delta)$ lives in Σ_{WA}.

6.B Relations of the parameters with l and m

In this appendix we study the relation between the parameters $\tilde{\delta}$, λ, μ and β and the two functions $l(\delta)$ and $m(\delta)$.

Given a function $\delta \to \sigma(\delta)$ in Σ_{WA}, the point $\tilde{\delta}$ is the only solution (which will be automatically greater than $\frac{1}{2}$) to $l(\delta) = \frac{N^{-1}(\delta)^2}{2}$. Equivalently, the point $\tilde{\delta}$ is the only solution to $m(\delta) = 0$.

The function μ can be recovered from

$$\sigma(\delta)\sqrt{T} = N^{-1}(\delta) + \sqrt{2\int_{\delta}^{\tilde{\delta}} \mu(x)\,dx}$$

for $\delta \in \left[\frac{1}{2}, \tilde{\delta}\right[$. In particular, $\int_{\delta}^{\tilde{\delta}} \mu(x)\,dx = \frac{(\sigma(\delta)\sqrt{T} - N^{-1}(\delta))^2}{2} = \frac{m(\delta)^2}{2}$, and deriving one finds

$$\mu(\delta) = -m(\delta)m'(\delta).$$

In the proof of Theorem 6.1, we showed $\lambda(\delta) = l'(\delta)$ for $\delta \leq \frac{1}{2}$.

Finally, consider $\delta > \tilde{\delta}$. Then

$$\sigma(\delta)\sqrt{T} = N^{-1}(\delta) - \sqrt{2\int_{N^{-1}(\tilde{\delta})}^{N^{-1}(\delta)} x\beta(x)\,dx}$$

and similarly as before

$$\beta(x) = \frac{n(x)}{x} m(N(x))m'(N(x)).$$

Chapter 7

Options are also options on options: how to smile with Black-Scholes

Abstract

We observe that a European Call option with strike $L > K$ can be seen as a Call option with strike $L-K$ on a Call option with strike K. Under no arbitrage assumptions, this yields immediately that the prices of the two contracts are the same, in full generality. We study in detail the relative pricing function which gives the price of the Call on Call option as a function of its underlying Call option, and provide quasi-closed formula for those new pricing functions in the Carr-Pelts-Tehranchi family [Carr and Pelts, Duality, Deltas, and Derivatives Pricing, 2015] and [Tehranchi, A Black-Scholes inequality: applications and generalisations, Finance Stoch, 2020] that includes the Black-Scholes model as a particular case. We also study the properties of the function that maps the price normalized by the underlier, viewed as a function of the moneyness, to the normalized relative price, which allows us to produce several new closed formulas. In connection to the symmetry transformation of a smile, we build a lift of the relative pricing function in the case of an underlier that does not vanish. We finally provide some properties of the implied volatility smiles of Calls on Calls and lifted Calls on Calls in the Black-Scholes model.

From:
C. Martini and **A. Mingone**, *Options are also options on options: how to smile with Black-Scholes*, arXiv preprint https://arxiv.org/abs/2308.04130, 2023.

7.1 Structure of the chapter

In this chapter we look at the relation for which a Call option with strike $L > K$ can be seen as a Call option with strike $L - K$ on a Call option with strike K and same maturity T.

In Section 7.2, we foster this option on option point of view and obtain, in full generality, relationships between the price of options on options and the initial Call or Put prices at other strikes: Calls are Calls on Calls, and Puts are Calls on Puts.

In Section 7.2.2 we define rigorously the *relative* Call on Call pricing function $\hat{C}_K$ and obtain useful properties in Section 7.2.3 in the case of homogeneous pricing function where the option price normalized by the underlier value does depend only of the moneyness.

In Section 7.2.4 we provide interesting properties of the transformation

$$\mathbb{T}_k c(x) := \frac{c\big(k + c(k)x\big)}{c(k)}$$

defined on the space of normalized (by the value of the underlier) Call prices c as functions of the moneyness $k = \frac{K}{S}$, that we call the Tehranchi space. We then extend the transformations $\mathbb{T}_k$, in some sense, to 2-parameter transformations.

Section 7.3 is devoted to the computation of new closed formulas either for pricing functions or for normalized ones. We provide a quasi-closed formula when the initial pricing function belongs to the Carr-Pelts-Tehranchi family, which generalizes the Black-Scholes formula, obtaining along the way an expression for the underlier value viewed as a function of the option price for this family.

In relation to the inversion of the volatility smile in the moneyness space, there is a generic pricing function transformation which consists in working in the numeraire of the underlier. We investigate in detail in Section 7.4 this transformation in the case where the underlier may vanish at maturity, and show that iterating it twice provides a pricing function on an underlier which does not vanish at maturity. We provide a quasi-closed formula for the so lifted pricing function in the case of the Black-Scholes model.

Eventually we provide basic properties of the volatility smiles associated to the Black-Scholes relative function and to the lifted relative one in Section 7.5.

7.2 Pricing functions

We consider general *pricing functions* which give the price $C(S, K)$ of a Call option as a function of the underlier price S and of its strike K. Of course, the option price may depend on other variables as well (like the instantaneous variance in stochastic volatility models as the Heston model), but we will be only interested in this partial dependency in this case.

The partial function $K \to C(S, K)$ gives the Call prices when the strike varies for the current value of the underlier S, and typically will aim at *calibrating* the market quotes, whereas the function $S \to C(S, K)$ is more interesting in a risk and/or sensitivity context, e.g. to get an insight of the order of magnitude of the tail risk of an option portfolio at horizon one day for margining purposes.

7.2.1 Options are options on options

In this section we use the notation $X(T)$ for the value of the contract X at time T. We will drop this notation in the following sections where T will not play any role.

Calls are Calls on Calls

Consider $0 < K < L$ and a *European Call on Call contract*, with strike $L - K$, which delivers at maturity T a Call contract with strike K. The payoff at T of this contract will be $(C(S,K)(T) - (L-K))_+$.

Observe now that $C(S,K)(T) = (S(T) - K)_+$ so that the payoff of the contract is equal to $((S(T) - K)_+ - (L-K))_+$. Now this latter quantity is 0 if $S(T) \leq L$ and $S(T) - L$ otherwise, so it is equal to $(S(T) - L)_+ = C(S,L)(T)$, i.e.

$$\forall 0 < K < L, \;\; C(S,L)(T) = \big(C(S,K)(T) - (L-K)\big)_+.$$

So in full generality under perfect market assumptions for the underlier and the options $C(S,K)$ and $C(S,L)$, *the price of a Call option with strike $L-K$ on a Call option with strike K is the price of a Call option with strike L.*

What happens if L is smaller than K? There is no hope to get any relation in this case, since the option price $C(S,K)(T)$ will vanish in the range $[L,K]$ where the Call $C(S,L)(T)$ will not. So, in terms of smiles, for a fixed value of K, only the part of the smile *on the right of* K will give rise to a new smile.

What happens for Put prices?

Put-Call-Parity and the Put price

Let us denote $\hat{C}_K(X,M)$ and $\hat{P}_K(X,M)$ the Call and Put pricing functions for Calls and Puts with a strike M on a Call option $X = C(S,K)$ with strike K. The Put-Call-Parity reads

$$\hat{C}_K(X, L-K) - \hat{P}_K(X, L-K) = C(S,K) - (L-K).$$

Now $\hat{C}_K(X, L-K) = C(S,L)$ and using the classic Put-Call-Parity at strikes K and L yields, taking the difference: $C(S,L) - C(S,K) = P(S,L) - P(S,K) - (L-K)$. This implies that

$$\hat{P}_K(X, L-K) = P(S,L) - P(S,K).$$

This relation clarifies what the price of the Put is in the new world where the underlier is the option with strike K, but also provides insights on the properties of the difference $P(S,L) - P(S,K)$. Can we prove it directly? Yes, indeed if we look at the difference $(L - S(T))_+ - (K - S(T))_+$, it is constantly equal to $L - K$ below K, and then goes to 0 linearly at point L, where it remains. This can be viewed also as a function of $(S(T) - K)_+$, which is exactly a Put payoff with strike $L - K$. In other words, it holds that

$$(L - S(T))_+ - (K - S(T))_+ = \big((L-K) - (S(T) - K)_+\big)_+,$$

which gives another proof of the relation $\hat{P}_K(X, L-K) = P(S,L) - P(S,K)$.

Eventually, summarizing the Call an Put computations we have the property that

$$\hat{C}_K(X, L-K) = C(S, L)$$
$$\hat{P}_K(X, L-K) = P(S, L) - P(S, K)$$

or yet for any $M \geq 0$

$$\hat{C}_K(X, M) = C(S, K+M)$$
$$\hat{P}_K(X, M) = P(S, K+M) - P(S, K).$$

Calls on Calls: further iterations Considering now a Call option with strike N written on a Call $\hat{C}_K(X, M)$, from the above equation this is equivalent to a Call of the form $\hat{C}_{K+M}(X, N)$. The latter quantity again is equivalent to $C(S, K+M+N)$. Similarly, a Put option with strike N written on a Call $\hat{C}_K(X, M)$ is a Put option written on $C(S, K+M)$, so it equals $\hat{P}_{K+M}(X, N)$, or $P(S, K+M+N) - P(S, K+M)$.

We have therefore a semigroup property, and iterating further does not yield new pricing functions.

Puts are Calls on Puts

Consider now $Y = P(S, K)$ as an underlier. Can we mimic the above approach using Puts as underliers? Observe first that Put prices are bounded by the strike, so that we have an underlier with values in $[0, K]$.

Take now any strike $0 \leq L < K$. Then

$$\big((K - S(T))_+ - L\big)_+ = \big((K-L) - S(T)\big)_+$$

which gives that a Call on $P(S, K)$ with strike L is a Put on S with strike $K - L$.

This entails, if we denote by $\tilde{C}_K(Y, L)$ the price of this Call, that $\tilde{C}_K(Y, L) = P(S, K-L)$. To get the price $\tilde{P}_K(Y, L)$ of the corresponding Put, let us use the Put-Call-Parity as above: $\tilde{C}_K(Y, L) - \tilde{P}_K(Y, L) = P(S, K) - L$. Using the classic Put-Call-Parity at the strikes $K - L$ and K, we find

$$\tilde{P}_K(Y, L) = P(S, K-L) - P(S, K) + L = C(S, K-L) - C(S, K).$$

We get eventually another pair transform: for all $0 \leq L < K$,

$$\tilde{C}_K(Y, L) = P(S, K-L)$$
$$\tilde{P}_K(Y, L) = C(S, K-L) - C(S, K).$$

Puts on Puts: further iterations Consider now a Call option with strike $N < L < K$ written on $\tilde{P}_K(Y, L)$. Looking at the payoff function, it can be easily shown that

$$\big(\big(L - (K - S(T))_+\big)_+ - N\big)_+ = \big(S(T) - (K - (L-N))\big)_+ - \big(S(T) - K\big)_+.$$

In other words, such a Call has the same value as the portfolio $C(S, K-(L-N)) - C(S,K)$, which in turn we have shown to be equal to $\tilde{P}_K(Y, L-N)$. Again, using the Put-Call-Parity, we find that a Put with strike N on $\tilde{P}_K(Y,L)$ is equivalent to the portfolio $P(S, K-(L-N)) - P(S, K-L)$, i.e. to $\tilde{C}_K(Y, L-N) - \tilde{C}_K(Y,L)$.

Summarizing, we get the following relationships:

$$\begin{aligned}\text{Call on } \tilde{P}_K(Y,L)(N) &= C(S, K-(L-N)) - C(S,K)\\ \text{Put on } \tilde{P}_K(Y,L)(N) &= P(S, K-(L-N)) - P(S,K-L).\end{aligned}$$

7.2.2 The Call on Call pricing function

In order to define rigorously the *relative* function $\hat{C}_K$ of Calls on Calls, we need to assume for a while that the function $S \to C(S,K)$ is invertible (it is the case in the Black-Scholes model and many other ones).

Definition 7.1. *Let $X = C(S,K)$ and $M > 0$. We denote with $\hat{C}_K(X,M)$ the Call option with strike M on the Call option with strike K. Then $\hat{C}_K(X,M)$ is the price of a Call option with strike $K+M$ and underlier S. In particular, if the function $S \to C(S,K)$ is invertible it holds*

$$\hat{C}_K(X,M) := C\big(C^{-1}(X,K), K+M\big). \tag{7.1}$$

K is called the relative underlying strike of $\hat{C}_K(X,M)$ and X the underlier of $\hat{C}_K(X,M)$.

Equation (7.1) gives a first representation of $\hat{C}_K$. It is of little practical interest though, since we are not aware of any model where both C^{-1} and C can be computed explicitly. Nevertheless, we will see in Section 7.3.1 that in the vast class of pricing functions of the Carr-Pelts-Tehranchi family a convenient representation formula for the inverse function is available.

We investigate below general properties of the Call on Call pricing function.

Properties of the Call on Call pricing function

From the arguments of Section 7.2.1, we can deduct some first properties of the function $\hat{C}_K$. In particular, Calls on Calls satisfy the usual arbitrage bounds for Call prices, i.e. they are always larger than their intrinsic value and smaller than the underlier. Furthermore, they are convex and non-increasing as functions of the strike. We already expect these properties to hold true for arbitrage arguments, and we show them rigorously in the following proposition.

Proposition 7.1 (Relative pricing function: strike dependence)**.** *The function $M \to \hat{C}_K(X,M)$ satisfies*

$$(X-M)_+ \le \hat{C}_K(X,M) \le X$$

and it is convex, non-increasing, with a slope strictly larger than -1.

Proof. The function $M \to C(S, K+M)$ is convex and non-increasing, and so is $M \to \hat{C}_K(X, M)$. This can be proved also observing that the basic relations $(S-K)_+ \leq C(S,K) \leq S$ translate into $(C(S,K)-M)_+ \leq C(S, K+M) \leq C(S,K)$, which gives in particular that the function $M \to \hat{C}_K(X,M)$ is non-increasing. Furthermore, from Equation (7.1), $\frac{d}{dM}\hat{C}_K(X,M) = \partial_K C(C^{-1}(X,M), K+M) \geq \partial_K C(C^{-1}(X,M),K)$ which is strictly larger than -1.

□

Observe that the above inequality implies $(S-(K+M))_+ \leq \hat{C}_K(X,M) \leq S$ since $(S-(K+M))_+ \leq (C(S,K)-M)_+$ and $S \geq C(S,K) = X$.

Remark 7.1. *Proposition 7.1 implies in particular that the slope of Calls on Calls in* 0 *is stricly larger than* -1*. This is not a problem in terms of arbitrageable prices, but it is an uncommon feature since it is linked to the presence of a positive mass of the underlier in* 0 *(see Theorem 2.1.2. of [66]). This is expected indeed, since the new underlier is a Call option, which has a whole region of null payoff. We will target this point in Section 7.4 where we will define lifted Calls on Calls' prices with derivative equal to* -1 *at* 0*.*

In the following we identify the necessary and sufficient conditions that the Call on Call pricing function must satisfy in order to be monotone as a function of the relative underlying strike and convex as a function of the underlier, i.e. the original Call price.

Lemma 7.1 (Monotonicity with respect to the relative underlying strike)**.** *Assuming the* $\mathcal{C}^1$ *smoothness of* $C(\cdot, K)$ *and* $C(S,\cdot)$*, the function* $K \to \hat{C}_K(X,M)$ *is non-decreasing if and only if the function*

$$L \to \frac{\partial_K C(S,L)}{\partial_S C(S,L)}$$

is non-decreasing for every S*.*

Proof. Firstly observe that $C(C^{-1}(X,K),K) = X$, so that taking the derivative with respect to K we find

$$0 = \partial_S C(C^{-1}(X,K),K)\partial_K C^{-1}(X,K) + \partial_K C(C^{-1}(X,K),K)$$

or

$$\partial_K C^{-1}(X,K) = -\frac{\partial_K C(C^{-1}(X,K),K)}{\partial_S C(C^{-1}(X,K),K)}.$$

We can now consider the relation $\hat{C}_K(X,M) = C(C^{-1}(X,K), K+M)$ and develop the derivative with respect to K:

$$\frac{d}{dK}\hat{C}_K(X,M) = -\frac{\partial_S C(C^{-1}(X,K),K+M)\partial_K C(C^{-1}(X,K),K)}{\partial_S C(C^{-1}(X,K),K)} + \\ + \partial_K C(C^{-1}(X,K),K+M).$$

Then, $\hat{C}_K(X, M)$ is non-decreasing as a function of K iff

$$\frac{\partial_K C(C^{-1}(X,K),K)}{\partial_S C(C^{-1}(X,K),K)} \leq \frac{\partial_K C(C^{-1}(X,K),K+M)}{\partial_S C(C^{-1}(X,K),K+M)},$$

or equivalently iff the function

$$L \to \frac{\partial_K C(S,L)}{\partial_S C(S,L)}$$

is non-decreasing for every S.

□

Lemma 7.2 (Convexity with respect to the underlier)**.** *Assuming the $\mathcal{C}^2$ smoothness of $C(\cdot, K)$, the function $X \to \hat{C}_K(X, M)$ is convex if and only if the function*

$$K \to \frac{\partial_S^2 C(S,K)}{\partial_S C(S,K)}$$

is non-decreasing for every S.

Proof. Let us restart from $\hat{C}_K(X, M) = C(C^{-1}(X, K), K + M)$. Assuming the $\mathcal{C}^2$ smoothness of $C(\cdot, K)$ we get

$$\frac{d}{dX}\hat{C}_K(X,M) = \frac{\partial_S C(C^{-1}(X,K),K+M)}{\partial_S C(C^{-1}(X,K),K)},$$

so that $\frac{d^2}{dX}\hat{C}(X, M)$ has the sign of the quantity

$$\partial_S^2 C(C^{-1}(X,K),K+M)\partial_S C(C^{-1}(X,K),K)+ \\ -\partial_S C(C^{-1}(X,K),K+M)\partial_S^2 C(C^{-1}(X,K),K).$$

As a consequence, the function $X \to \hat{C}_K(X, M)$ is convex for any K, M iff the function

$$L \to \frac{\partial_S^2 C(C^{-1}(X,K),L)}{\partial_S C(C^{-1}(X,K),L)}$$

is non-decreasing for any X, which is equivalent to state the same property for the function

$$K \to \frac{\partial_S^2 C(S,K)}{\partial_S C(S,K)}$$

at any point S.

□

In the next section we will apply Lemmas 7.1 and 7.2 to the case of homogeneous pricing functions, and in particular to the Black-Scholes case for which the properties of monotonicity with respect to the relative underlying strike and of convexity with respect to the underlier are always satisfied.

7.2.3 Normalized Call prices

We now switch from the strike space to the moneyness $k = \frac{K}{S}$ space and consider *normalized* Call pricing functions, i.e. Call prices divided by their underlier.

Definition 7.2. *Let $k = \frac{K}{S}$ the moneyness of the Call option $C(S, K)$, and $m = \frac{M}{C(S,K)}$ the moneyness of the Call option $\hat{C}_K(C(S, K), M)$. We denote with*

$$C_S(k) := \frac{C(S, Sk)}{S}$$

the normalization of C with respect to k, and with

$$\hat{C}_{K,S}(m) := \frac{\hat{C}_K(C(S, K), C(S, K)m)}{C(S, K)}$$

the normalization of $\hat{C}_K$ with respect to m.

Furthermore, we say that C is homogeneous if C_S does not depend on S and define the normalized pricing function c by the relation $c(k) := C_1(k)$ for every k.

Normalized Call prices are particularly interesting when Call prices are homogeneous, since they satisfy key properties as we will show in Section 7.2.3. Furthermore, the most notorious models such as the Black-Scholes, the Heston and the implied volatility models are homogeneous. In this case the function C can be recovered from c through the formula $C(S, K) = Sc\big(\frac{K}{S}\big)$. Not all models are homogeneous: examples of inhomogeneous models include local volatility or local stochastic volatility (except in very few cases).

In order to work with Black-Scholes prices, throughout the rest of the paper we denote with ϕ the standard normal probability density function and with Φ its cumulative density function. Furthermore, we denote with $\mathrm{BS}(S, K, v)$ the traditional Black-Scholes function for Call prices with implied total volatility $v = \sigma\sqrt{T}$:

$$\begin{aligned} \mathrm{BS}(S, K, v) &= S\Phi\big(d_1(S, K, v)\big) - K\Phi\big(d_2(S, K, v)\big) \\ d_{1,2}(S, K, v) &= -\frac{\log \frac{K}{S}}{v} \pm \frac{v}{2}. \end{aligned} \tag{7.2}$$

We will sometimes drop the dependency in v for notation simplicity. When considering normalized Black-Scholes prices, we use the notation

$$\mathrm{BS}(S, K) = S\mathrm{bs}\left(\frac{K}{S}\right).$$

Reconstructed prices obtained from $\mathrm{bs}(k)$ in the Black-Scholes case correspond to the perspective function of section 3.2.6. of [12].

In the following Lemma we show that the normalization $C_S(k)$ as a function of the moneyness k has the same properties as the original price $C(S, K)$ as a function of the strike K.

Lemma 7.3. *Normalized prices $C_S(k)$ are non-increasing and convex functions of k, and satisfy*

$$(1-k)_+ \leq C_S(k) \leq 1.$$

Proof. It holds $C'_S(k) = \frac{d}{dK}C(S,Sk)$ and $C''_S(k) = S\frac{d^2}{dK^2}C(S,Sk)$, so that $C_S(k)$ is non-increasing and convex in k. Also, since Call prices satisfy $(S-K)_+ \leq C(S,K) \leq S$, then dividing by S it holds $(1-k)_+ \leq C_S(k) \leq 1$.

□

It turns out that there is a convenient relationship between the initial normalized Call pricing function and the (normalized) relative Call on Call one. Indeed, observe that $\hat{C}_K(C(S,K),M) = C(S,K+M)$, so that $\hat{C}_K(C(S,K),C(S,K)m) = C\big(S,S\frac{K+C(S,K)m}{S}\big)$. Now from Definition 7.2 it holds

$$C_S(k+C_S(k)m) = \frac{C\big(S,S(k+C_S(k)m)\big)}{S}$$

and consequently

$$\hat{C}_{Sk,S}(m) = \frac{C_S(k+C_S(k)m)}{C_S(k)}.$$

In particular, in case C is homogeneous,

$$\hat{C}_{Sk,S}(m) = \frac{c(k+c(k)m)}{c(k)}. \tag{7.3}$$

We will further exploit the relationship in Equation (7.3) in Section 7.2.4 where we work in the space of normalized homogeneous Call prices, and define transformations in such space.

As in Remark 7.1, observe that by the chain rule it holds that $\frac{d}{dm}\hat{C}_{Sk,S}(0_+) = C'_S(k) = \partial_K C(S,Sk)$, which will be in general (for strictly convex functions) strictly larger than -1.

Homogeneous Call prices

We now look at the properties of the Call on Call pricing functions in Section 7.2.2 in the case of homogeneous Call prices. It turns out that Calls on Calls with homogeneous pricing function are non-decreasing functions of the relative underlying strike and that Black-Scholes Calls on Calls are convex with respect to the underlier. As a consequence, when calibrating Calls on Calls, one should design an algorithm such to satisfy these necessary properties.

In the following proposition we consider conditions of Lemma 7.1 in the case of homogeneous Call prices and show that they are always satisfied, i.e. that Calls on Calls with homogeneous pricing function are non-decresing with respect to the relative underlying strike.

Proposition 7.2 (Monotonicity of Calls on Calls with respect to the relative underlying strike: the homogeneous case)**.** *Let $C(S,K)$ homogeneous and $\mathcal{C}^1$ in both variables. Then the function $K \to \hat{C}_K(X,M)$ is non-decreasing.*

Proof. From Lemma 7.1, we shall prove that the function

$$L \to \frac{\partial_K C(S,L)}{\partial_S C(S,L)}$$

is non-decreasing for every S. In the homogeneous case this can be simplified writing $C(S,L) = c\big(\frac{L}{S}\big)S$ and considering that a function is monotone in L iff it is monotone in $l = \frac{L}{S}$. We then find that the function $K \to \hat{C}_K(X,M)$ is non-decreasing iff the function

$$l \to \frac{c'(l)}{c(l) - lc'(l)}$$

is non-decreasing. This is actually the case since the derivative of the latter function is $\frac{c(l)c''(l)}{(c(l)-lc'(l))^2}$, which is always positive for convex prices.

□

The previous proposition implies in particular that $C(X,M) = \hat{C}_0(X,M) \leq \hat{C}_K(X,M)$ in case of homogeneous Call prices. In other words, Call on Call prices are always not smaller than the original Call prices for fixed moneyness.

Consider now a fixed value of X. The proposition gives that at a fixed moneyness $\frac{M}{X}$, the map $K \to \hat{C}_K(X,M)$ is non-decreasing and so, for any continuous increasing function Y with $Y(0) = 0$, the map $t \to \hat{C}_{Y(t)}(X,M)$ is non-decreasing as well, meaning there is no calendar-spread arbitrage for the price surface $(t,M) \to C_{Y(t)}(X,M)$. Since there is no Butterfly arbitrage in the strike dimension for any t, we have built an arbitrage-free *forward extrapolation* of the pricing function $C_0(X,M) = C(X,M)$. One can see that we treat the strike K here as a shadow parameter, completely forgetting its role in the design of the relative pricing function.

We now pass to the study of Lemma 7.2. Conditions for the convexity of Calls on Calls with respect to the underlier can be re-written in the homogeneous case. Differently from the property of monotonicity with respect to the relative underlying strike, here we do not achieve to show the convexity property for all homogeneous pricing function. However, we prove it for the Black-Scholes case.

Proposition 7.3 (Convexity of Calls on Calls with respect to the underlier: the homogeneous case)**.** *Let $C(S,K)$ homogeneous and $\mathcal{C}^2$ in the first variable. Then the function $X \to \hat{C}_K(X,M)$ is convex if and only if the function*

$$k \to \frac{k^2 c''(k)}{c(k) - kc'(k)}$$

is non-decreasing. In particular this holds true in the Black-Scholes case.

Proof. From Lemma 7.2 we shall prove that the function

$$K \to \frac{\partial_S^2 C(S,K)}{\partial_S C(S,K)}$$

is non-decreasing for every S. We can write $C(S,K) = c\left(\frac{K}{S}\right)S$, develop the derivatives and consider that a function is monotone in K iff it is monotone in $\frac{K}{S}$. We find that in the homogeneous case, $X \to \hat{C}_K(X,M)$ is convex for any K, M iff the function

$$k \to \frac{k^2 c''(k)}{S(c(k) - kc'(k))}$$

is non-decreasing for any S. We can drop S at the denominator and conclude.

In the Black-Scholes case, $\mathrm{bs}''(k) = \frac{\phi(d_2)}{kv}$ where $d_{1,2} = -\frac{\log k}{v} \pm \frac{v}{2}$. Then the above requirement is that

$$k \to \frac{k\phi(d_2)}{v\Phi(d_1)}$$

is non-decreasing. This holds true iff, taking the derivative, the quantity

$$\frac{\phi(d_2)}{v\Phi(d_1)^2}\left(\Phi(d_1) + \frac{\Phi(d_1)d_2 + \phi(d_1)}{v}\right)$$

is positive. Observe that $d_2 = d_1 - v$, so that we are asking the quantity $\Phi(d_1)d_1 + \phi(d_1)$ to be positive. When d_1 is positive this is gained. Otherwise, we can use the upper bound of the Mill's ratio $\frac{1-\Phi(x)}{\phi(x)} < \frac{1}{x}$ for every $x > 0$ with $x = -d_1$ and obtain the desired property.

□

The convexity in the underlier of the option price is a key property from a risk analysis perspective, and allows to study the behavior of the option price dynamic as being locally Black-Scholes-like, with a positive Gamma for Calls and Puts. Combined with the previous proposition and the discussion that follows it, we get a forward extrapolation scheme with nice properties when the convexity property is fulfilled.

Remark 7.2. *It is interesting to observe that the function $\hat{C}_K(X,M)$ cannot be homogeneous when $C(S,K)$ is. Indeed, in order to satisfy such a property, its normalized function $\hat{C}_{K,S}(m) = \frac{\hat{C}_K(X,Xm)}{X}$ where S is recovered from $X = C(S,K)$ should not depend on X, i.e. it should be a function of the form $g(m)$. From Equation (7.3), it should hold $g(m) = \frac{c\left(\frac{K}{S} + c\left(\frac{K}{S}\right)m\right)}{c\left(\frac{K}{S}\right)}$. However the right term depends on X in the S term, so that the equality cannot hold for all X. Indeed, for X moving from its lowest values to ∞, S moves from 0 to ∞, so that $g(m) = \frac{c(\infty)}{c(\infty)} = 1$ and $g(m) = c(m)$ respectively. In non-degenerate cases, normalized prices are not constantly equal to 1 so that Calls on Calls with homogeneous pricing function cannot be homogeneous.*

7.2.4 A transformation in the Tehranchi space

In Lemma 7.3 we have pointed out some necessary properties that normalized Call prices satisfy: monotonicity and convexity with respect to the moneyness, and upper and lower bounds corresponding to the constant function $\mathbb{1}$ and the normalized intrinsic value function $(1-k)_+$. Note that the property of monotonicity is actually implied by the two other properties.

A crucial point here is that the underlier is considered to be frozen (and, given the normalization, with unit value): in other words we only consider the partial dependency in the normalized strike (the moneyness) of the pricing function.

As Tehranchi has deeply studied normalized Call prices in [66], we will name *Tehranchi space* the space $\mathbb{C}$ of such normalized Call prices:

$$\mathbb{C} = \Big\{ c : \mathbb{R}_+ \longrightarrow [0,1] \;\Big|\; c \text{ convex},\ \forall m,\ (1-m)_+ \leq c(m) \leq 1 \Big\}.$$

As an immediate consequence, functions in $\mathbb{C}$ are non-increasing and satisfy $c(0) = 1$. Also, from Lemma 7.3, functions obtained by the normalization $c(k)$ of homogeneous prices defined in Definition 7.2 belong to the Tehranchi space.

Equation (7.3) suggests to define the following transformation on $\mathbb{C}$.

Definition 7.3. *For any $c \in \mathbb{C}$ and $k \geq 0$ with $c(k) > 0$ we define the transformation*

$$\mathbb{T}_k c(\cdot) := \frac{c(k + c(k)\cdot)}{c(k)}.$$

k is called the relative underlying moneyness of $\mathbb{T}_k$.

Observe that functions in $\mathbb{C}$ are either positive, or positive before a threshold a and null beyond a. It is natural if needed to extend the definition of $\mathbb{T}_k$ for $k \geq a$ by $\mathbb{T}_k \equiv \mathbb{1}$, the constant function equal to the normalized underlier.

In relation to Equation (7.3), the transformation $\mathbb{T}_k$ corresponds to the normalization of Calls on Calls with homogeneous pricing function, i.e. $\mathbb{T}_k c(m) = \hat{C}_{Sk,S}(m)$. Also, for a given S and a function $\mathbb{T}_k c(\cdot)$, it is always possible to reconstruct the corresponding non-normalized Call on Call. In particular, the original underlier Call written on S has strike $K = Sk$, and the Call on Call with strike M is

$$\hat{C}_K(C(S,K), M) = \hat{C}_K(Sc(k), M) = \mathbb{T}_k c\Big(\frac{M}{Sc(k)}\Big) Sc(k).$$

Properties of the transformation $\mathbb{T}_k$

The following lemma lists important properties of the transformations $\mathbb{T}_k$. In particular, it states that the new function $\mathbb{T}_k c$ still lives in $\mathbb{C}$ and that it has a derivative in 0 which is larger than -1. Furthermore the lemma gives the limits of the transformation $\mathbb{T}_k$ with respect to k.

Lemma 7.4 (Properties of $\mathbb{T}_k$)**.** *For any $c \in \mathbb{C}$ and $k \geq 0$ with $c(k) > 0$ it holds:*

1. $\mathbb{T}_k c \in \mathbb{C}$;
2. $\frac{d}{dm}\mathbb{T}_k c(0_+) = c'(k)$;
3. $\mathbb{T}_k c(\infty) = \frac{c(\infty)}{c(k)}$;
4. $\mathbb{T}_0 c = c$;
5. $\mathbb{T}_\infty c \equiv \mathbb{1}$.

Proof. The only difficult point is the last one. Observe that at fixed k, m, it holds $c(k + c(k)m) = c(k) + c'(k + uc(k))c(k)m$ for some u in $]0, m[$. Whence $\mathbb{T}_k c(m) = 1 + c'(k + uc(k))m$ and since c' goes uniformly to 0 at infinity this yields $\mathbb{T}_k c(\cdot) \to 1$ as $k \to \infty$.

□

At this point, one can consider the family $\{\mathbb{T}_k c : k \geq 0\}$, as a (one-dimensional) *enrichment* of the price curve c, given that $\mathbb{T}_0 c = c$. The initial forward moneyness k should be considered here as a plain parameter; all the price curves $\mathbb{T}_k c$ are arbitrage-free in the sense that they belong to $\mathbb{C}$.

In relation to this latter point, one can wonder about the *composition* of the above enrichment/extensions, like $\mathbb{T}_{k_n} \cdots \mathbb{T}_{k_2}\mathbb{T}_{k_1}$. The following property corresponds to the image of the semigroup property in the normalized space:

Lemma 7.5 (Iterates of $\mathbb{T}_k$)**.** *It holds* $\mathbb{T}_b\mathbb{T}_a c = \mathbb{T}_{a+c(a)b}c$.

Proof. The following relations hold

$$\begin{aligned}
\mathbb{T}_b\mathbb{T}_a c(m) &= \frac{\mathbb{T}_a c(b + \mathbb{T}_a c(b)m)}{\mathbb{T}_a c(b)} \\
&= \mathbb{T}_a c\Big(b + \frac{c(a + c(a)b)}{c(a)}m\Big)\frac{c(a)}{c(a + c(a)b)} \\
&= \frac{c\Big(a + c(a)\big(b + \frac{c(a+c(a)b)}{c(a)}m\big)\Big)}{c(a)}\frac{c(a)}{c(a + c(a)b)} \\
&= \frac{c(a + c(a)b + c(a + c(a)b)m}{c(a + c(a)b)} \\
&= \mathbb{T}_{a+c(a)b}c(m).
\end{aligned}$$

□

This means that the range of $\mathbb{T}_\cdot$ is the same as the range of its iterates, and there is no *additional* enrichment to hope for from performing those iterations.

We shall now consider Proposition 7.2 where we proved that Calls on Calls with homogeneous pricing function are non-decreasing with respect to the relative underlying strike. We expect to find a similar property for $k \to \mathbb{T}_k c(m) = \hat{C}_{Sk,S}(m)$.

Proposition 7.4 (Monotonicity of $\mathbb{T}_k$ with respect to the relative underlying moneyness)**.** *For any $c \in \mathbb{C}$ and $m \geq 0$, the map $k \to \mathbb{T}_k c(m)$ is non-decreasing.*

Proof. It holds

$$\frac{d}{dk}\mathbb{T}_k c(m) = \frac{c'(k + c(k)m)(1 + c'(k)m)c(k) - c(k + c(k)m)c'(k)}{c(k)^2}.$$

Doing the derivative with respect to m, one finds $\frac{d}{dm}\frac{d}{dk}\mathbb{T}_k c(m) = c''(k + c(k)m)(1 + c'(k)m)$ which is positive iff $m < -\frac{1}{c'(k)}$. Then, the function $\frac{d}{dk}\mathbb{T}_k c(m)$ with variable m is increasing up to $-\frac{1}{c'(k)}$ and then starts decreasing. To show that it is non-negative for every k and m, it is enough to show that it is non-negative for every k and $m \in \{0, \infty\}$.

At $m = 0$, it is easy to see $\frac{d}{dk}\mathbb{T}_k c(0) = 0$. From Theorem 2.1.2 of [66], there exists a random variable S such that $c(y) = 1 - E[S \wedge y]$ and $-c'(y) = P(S > y)$. Then

$$\begin{aligned} c(y) - yc'(y) &= 1 - E[S \wedge y] + yP(S > y) \\ &= 1 - \int (s \wedge y) f_S(s)\, ds + y \int_y^\infty f_S(s)\, ds \\ &= 1 - \int_0^y s f_S(s)\, ds - y \int_y^\infty f_S(s)\, ds + y \int_y^\infty f_S(s)\, ds \\ &= 1 - \int_0^y s f_S(s)\, ds \end{aligned}$$

and this goes to $1 - E[S] = c(\infty) \geq 0$ as y goes to ∞. As a consequence, $yc'(y)$ goes to 0 and the result holds.

□

Note that the above proposition implies in particular

$$c(m) = \mathbb{T}_0 c(m) \leq \mathbb{T}_k c(m) = \frac{c(k + c(k)m)}{c(k)}.$$

In Figure 7.1 we plot the function $\mathbb{T}_k c(m)$ with respect to k for different fixed ms. The function c is a normalized Black-Scholes Call function with implied total volatility equal to 0.2. It can be seen that $\mathbb{T}_k c(m)$ is non-decreasing in k (as shown in Proposition 7.4) and non-increasing in m, as expected since c is a non-increasing function.

A slight generalization

From Lemma 7.4, the function $m \to \mathbb{T}_k c(m)$ has a particular feature at 0. Indeed, its right derivative is $\frac{d}{dm}\mathbb{T}_k c(0_+) = c'(k)$ which for $k > 0$ is in general larger than -1. As already seen in Remark 7.1, this feature might be annoying since it implies the presence of a mass in 0 of the probability density function associated to the underlier of prices.

We are then interested in generalizing suitably the transformation $\mathbb{T}_k$ in order to get rid of this mass at 0 phenomenon. This generalization is formulated on the Tehranchi

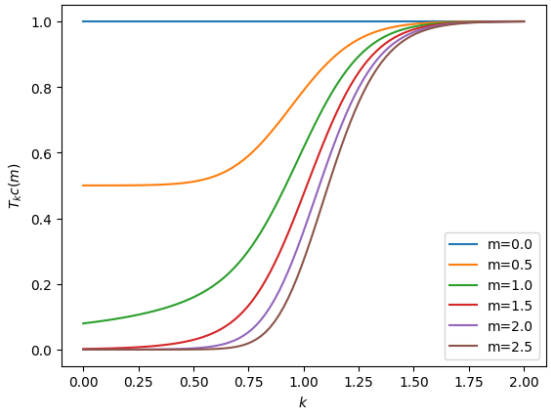

Figure 7.1: Function $k \to \mathbb{T}_k c(m)$ for different levels of m. The original Black-Scholes implied total volatility is set at 0.2.

space here. In Section 7.4.1 we will see how to change the probability measure in order to lift Calls on Calls to Calls on Calls with no mass in 0, and will provide the connection with the generalized transformation of this section.

To introduce the generalized transformation, firstly consider $\alpha \geq 0$ and the quantity $\mathbb{V}_{k,\alpha}$ defined as follows.

Definition 7.4. *For any $c \in \mathbb{C}$ and $\alpha, k \geq 0$ with $c(k) > 0$, we define the transformation*

$$\mathbb{V}_{k,\alpha} c(\cdot) := \frac{c(k + \alpha \cdot)}{c(k)}.$$

The transformation $\mathbb{T}_k$ can be written as a function of $\mathbb{V}_{k,\alpha}$ in the sense that $\mathbb{T}_k c = \mathbb{V}_{k,c(k)} c$.

It is easy to see that functions $m \to \mathbb{V}_{k,\alpha} c(\cdot)$ are convex and bounded by 1. Adding the requirement that $\alpha(k) \leq -\frac{c(k)}{c'(k)}$ also guarantees the lower bound $(1-m)_+$, so that the transformations $\mathbb{V}_{k,\alpha}$ can be viewed as operating on the Tehranchi space.

Lemma 7.6 (Properties of $\mathbb{V}_{k,\alpha}$)**.** *For any $c \in \mathbb{C}$, $\alpha, k \geq 0$ with $c(k) > 0$ and $\alpha \leq -\frac{c(k)}{c'(k)}$, it holds*

1. $\mathbb{V}_{k,\alpha} c \in \mathbb{C}$;
2. $\frac{d}{dm} \mathbb{V}_{k,\alpha} c(0_+) = \frac{c'(k)}{c(k)} \alpha \geq 1$;
3. $\mathbb{V}_{k,\alpha} c(\infty) = \frac{c(\infty)}{c(k)}$;
4. $\mathbb{V}_{0,1} c = c$.

Proof. The derivative and second derivative of $\mathbb{V}_{k,\alpha} c(m)$ with respect to m are respectively $\frac{c'(k+\alpha m)}{c(k)} \alpha$ and $\frac{c'(k+\alpha m)}{c(k)} \alpha^2$. Then $\mathbb{V}_{k,\alpha} c(m)$ is convex in m. Since $c \in \mathbb{C}$, it is

non-increasing and $\mathbb{V}_{k,\alpha}c(m) \leq 1$. The inequality $\mathbb{V}_{k,\alpha}c(m) \geq (1-m)_+$ amounts to $c(k+\alpha m) - c(k) \geq -c(k)m$ for $m < 1$; by the mean value theorem the LHS writes $\alpha c'(k+\alpha u)m$ for some u in $]0, m[$ where the derivative is negative. Since c is convex, the latter quantity is larger than $\alpha c'(k)m$ and $\mathbb{V}_{k,\alpha}c(m) \geq (1-m)_+$ holds true as soon as $\alpha \leq -\frac{c(k)}{c'(k)}$.

The other points follow immediately.

□

As the transformation $\mathbb{T}_k$, also its generalization $\mathbb{V}_{k,\alpha}$ satisfies a semigroup property and, as a consequence, iterations of this transformation do not further enrich the family $\{\mathbb{V}_{k,\alpha} : k \geq 0, \alpha \geq 0\}$:

Lemma 7.7 (Iterates of $\mathbb{V}_{k,\alpha}$)**.** *It holds* $\mathbb{V}_{b,\beta}\mathbb{V}_{a,\alpha}c = \mathbb{V}_{a+\alpha b,\alpha\beta}c$.

Furthermore, if $\alpha \leq -\frac{c(a)}{c'(a)}$ *and* $\beta \leq -\frac{\mathbb{V}_{a,\alpha}c(b)}{\frac{d}{dm}\mathbb{V}_{a,\alpha}c(b)}$ *then* $\alpha\beta \leq -\frac{c(a+\alpha b)}{c'(a+\alpha b)}$.

Proof. The proof of the first statement is similar to the proof of Lemma 7.5.

Since $\frac{d}{dm}\mathbb{V}_{a,\alpha}c(b) = \frac{c'(a+\alpha b)}{c(a)}\alpha$, if $\beta \leq -\frac{\mathbb{V}_{a,\alpha}c(b)}{\frac{d}{dm}\mathbb{V}_{a,\alpha}c(b)}$ then $\beta \leq -\frac{c(a+\alpha b)}{\alpha c'(a+\alpha b)}$ and the second statement follows.

□

The second statement of Lemma 7.7 implies that the family $\left\{\mathbb{V}_{k,\alpha}c : c \in \mathbb{C}, k \geq 0, 0 \leq \alpha \leq -\frac{c(k)}{c'(k)}\right\}$ (where $c \in \mathbb{C}$ is also a parameter) is closed under iterations.

From Lemma 7.6, we see that the critical case $\alpha = -\frac{c(k)}{c'(k)}$ is of particular interest since it will entail the property $\frac{d}{dm}\mathbb{V}_{k,\alpha}c(0_+) = -1$. This gives rise to a new transform on the Tehranchi space:

Definition 7.5. *For any* $c \in \mathbb{C}$ *and* $k \geq 0$ *with* $c(k) > 0$ *and* $c'(k) \neq 0$ *we define the transformation*

$$\mathbb{U}_k c(\cdot) := \frac{c\big(k - \frac{c(k)}{c'(k)}\cdot\big)}{c(k)}.$$

The transformation $\mathbb{U}_k$ can be written as a function of $\mathbb{V}_{k,\alpha}$ in the sense that $\mathbb{U}_k c = \mathbb{V}_{k,-\frac{c(k)}{c'(k)}}c$, so properties of the latter transformation (for fixed c) still hold for the former one.

Lemma 7.8 (Properties of $\mathbb{U}_k$)**.** *For any* $c \in \mathbb{C}$ *and* $k \geq 0$ *with* $c(k) > 0$ *and* $c'(k) \neq 0$, *it holds:*

1. $\mathbb{U}_k c \in \mathbb{C}$;
2. $\frac{d}{dm}\mathbb{U}_k c(0_+) = -1$;
3. $\mathbb{U}_k c(\infty) = \frac{c(\infty)}{c(k)}$;
4. *If* $c'(0_+) = -1$, *then* $\mathbb{U}_0 c = c$.

Lemma 7.7 can be applied to $\mathbb{U}_k$ but it does not automatically guarantee that iterates of $\mathbb{U}_k$ are still functions in the family $\{\mathbb{U}_k c : k \geq 0\}$, even though they certainly live in $\{\mathbb{V}_{k,\alpha} : k \geq 0, \alpha \geq 0\}$. In the following lemma we prove this point.

Lemma 7.9 (Iterates of $\mathbb{U}_k$)**.** *It holds* $\mathbb{U}_b \mathbb{U}_a c = \mathbb{U}_{a - \frac{c(a)}{c'(a)} b} c$.

Proof. The proof can be shown directly as in Lemma 7.5. Alternatively, applying Lemma 7.7, we have $\mathbb{U}_b \mathbb{U}_a c = \mathbb{V}_{b,\beta} \mathbb{V}_{a,\alpha} c$ where $\alpha = -\frac{c(a)}{c'(a)}$ and $\beta = -\frac{\mathbb{V}_{a,\alpha} c(b)}{\frac{d}{dm} \mathbb{V}_{a,\alpha} c(b)} = -\frac{c(a+\alpha b)}{\alpha c'(a+\alpha b)}$. Then $a + \alpha b = a - \frac{c(a)}{c'(a)} b$ and $\alpha\beta = -\frac{c(a+\alpha b)}{c'(a+\alpha b)}$, so $\mathbb{V}_{a+\alpha b, \alpha\beta} c = \mathbb{U}_{a - \frac{c(a)}{c'(a)} b} c$. □

The transformation $\mathbb{U}_k$ has a derivative in 0_+ equal to -1 and is then linked to probabilitiees with no mass in 0. This will allow us to define new closed pricing formulas in Section 7.4, that we will call lifted Calls on Calls.

7.3 New closed formulas

In this section, we provide a quasi-closed formula for the pricing function within the Carr-Pelts-Tehranchi family (see [14, 15, 66]), which generalizes the Black-Scholes pricing function associated to the standard normal density to any log-concave (and even, unimodal, as shown by Vladimir Lucic in [50]) density function. This includes the Black-Scholes case as a particular case. Those pricing functions are the pricing functions of options on option, where the price of the latter option is viewed as the underlier.

The reason to work with this family of pricing function is that a variational formula for the option price, reminiscent of a dual transform, is available, and it turns out that this variational formula can be inverted to get an expression for the underlier value in terms of the option price and the other parameters.

In the second section below, we derive new closed formulas from the normalized price transformations - these formulas will yield (new) homogeneous pricing functions when de-normalized.

7.3.1 The Carr-Pelts-Tehranchi family

Knowing the expression of the underlier S as a function of the option price $X := C(S, K)$ yields a closed formula for the option price which is given by $C(S, K + M)$, as a pricing function of X and M. In general, such an expression is unavailable, even if one can resort to straightforward numerical procedures like a basic dichotomy to compute it numerically, given the monotonicity of the map $S \to C(S, K)$.

It turns out that one can say more in the case of the Carr-Pelts-Tehranchi family, due to the availability of a particular variational formulation for the option price.

We dub Carr-Pelts-Tehranchi (CPT) model the explicit arbitrage-free parametrization for FX option prices introduced by Carr and Pelts in 2015 at a conference in honor of Steven Shreve at Purdue university (see [14]). The model has then been independently

rediscovered by Tehranchi in [66] while studying advanced properties of the Black-Scholes formula.

In the CPT model, the family of Call prices is indexed by log-concave densities $f : \mathbb{R} \to [0, \infty[$ and increasing functions $y : [0, \infty[\to \mathbb{R}$ (which correspond to the total implied volatility in the Black-Scholes framework). The Black-Scholes model is a special case of CPT choosing f to be the standard normal probability density function ϕ and $y(t) = v = \sigma\sqrt{t}$ with reference to Equation (7.2).

Similarly to Black-Scholes, the CPT model has the nice feature that option prices have a closed quasi-explicit formula. Indeed, the CPT Call price is

$$C^{\mathrm{CPT}}(S, K; f, y(t)) := \int_{-\infty}^{\infty} \Big(Sf(z + y(t)) - Kf(z) \Big)_+ dz. \tag{7.4}$$

Tehranchi shows in section 3.2 of [66] that if f is log-concave and y is increasing, then prices in Equation (7.4) represent a Call price surface of the form $E[(S_T - K)_+]$ for a certain non-negative supermartingale S_t such that $E[S_T] = S$. Equivalently, Call prices are non-decreasing in t, convex in K and equal to $(S - K)_+$ for $t = 0$.

Remarkably, prices in Equation (7.4) can actually be represented with a formulation very close to the Black-Scholes one as

$$C^{\mathrm{CPT}}(S, K; f, y(t)) = SF(d(K, y(t); f) + y(t)) - KF(d(K, y(t); f)$$

where

$$d(K, y; f) := \sup \left\{ z : \frac{f(z + y)}{f(z)} \geq K \right\}$$

and z lives in the support of f, and F is the cumulative density function associated with f. In the Black-Scholes case, the function $d(K, y(t); f)$ can be obtained explicitly and is given by the classical expression $d_2(S, K, v) = -\frac{\log \frac{K}{S}}{v} - \frac{v}{2}$.

Note that the CPT pricing functions are homogeneous ones.

Furthermore, Lucic has shown in [50] that under the more general hypothesis that f is unimodal, i.e. it has a single peak (point of maximum), and y is increasing, prices in Equation (7.4) are still a Call price surface.

Since in the present article we are considering smiles of the Call surface, i.e. for fixed time-to-maturity, we will drop the dependence to t of y.

One of the important properties of the CPT family is the availability of a variational formula for the option price (Theorem 4.1.2 of [66]):

$$C^{\mathrm{CPT}}(S, K; f, y) = \sup_{p \in]0,1[} SF(F^{-1}(p) + y) - pK.$$

This formula is the key of the following result.

Lemma 7.10 (Inversion of the CPT formula)**.** *Let f a unimodal probability density function and F its cumulative density function. For $K, y \geq 0$ let $X = C^{CPT}(S, K; f, y)$, then it holds*

$$S = K\psi\left(\frac{X}{K}, y; F\right)$$

where

$$\psi(a, y; F) := \inf_{r\in\mathbb{R}} \frac{a + F(r - y)}{F(r)}.$$

Proof. It holds

$$X = \sup_{p\in]0,1[} SF(F^{-1}(p) + y) - pK$$

so that for every p, $X \geq SF(F^{-1}(p) + y) - pK$ or yet $S \leq \frac{X+pK}{F(F^{-1}(p)+y)}$, and also

$$S = \inf_{p\in]0,1[} \frac{X + pK}{F(F^{-1}(p) + y)}.$$

Set $r := F^{-1}(p) + y$, then $p = F(r - y)$ and, given that the range of r when p runs into $]0, 1[$ is $\mathbb{R}$ irrespective of y, the conclusion follows.

□

We can easily apply the result from this lemma to the relation $\hat{C}_K(C(S, K), M) = C(S, K + M)$ and obtain the following.

Proposition 7.5 (Quasi-closed formula for the Call on Call pricing function in the CPT family)**.** *Let f a unimodal probability density function. For $K, M, y \geq 0$ it holds*

$$\begin{aligned}\hat{C}_K(X, M) &= \int_{-\infty}^{\infty} \left(K\psi\left(\frac{X}{K}, y; F\right) f(z + y) - (K + M)f(z)\right)_+ dz \\ &= K\psi\left(\frac{X}{K}, y; F\right) F(d(K + M, y; f) + y) - (K + M)F(d(K + M, y; f)).\end{aligned}$$

In particular, in the Black-Scholes case with $v = \sigma\sqrt{T}$

$$\begin{aligned}\hat{C}_K(X, M) = K\psi\Big(\frac{X}{K}, v; \Phi\Big)\Phi\Big(d_1\Big(K\psi\Big(\frac{X}{K}, v; \Phi\Big), K + M, v\Big)\Big) + \\ - (K + M)\Phi\Big(d_2\Big(K\psi\Big(\frac{X}{K}, v; \Phi\Big), K + M, v\Big)\Big)\end{aligned}$$

where $d_{1,2}(a, b, v) = -\frac{\log \frac{b}{a}}{v} \pm \frac{v}{2}$.

Observe that in the Black-Scholes case of the above proposition, choosing $M = 0$, we find the expression

$$\hat{C}_K(X, 0) = K\psi\Big(\frac{X}{K}, v; \Phi\Big)\Phi\Big(d_1\Big(K\psi\Big(\frac{X}{K}, v; \Phi\Big), K, v\Big)\Big) - K\Phi\Big(d_2\Big(K\psi\Big(\frac{X}{K}, v; \Phi\Big), K, v\Big)\Big)$$

which is the classic Black-Scholes formula for the Call $C(\tilde{S}, K)$ where $\tilde{S} = K\psi\big(\frac{X}{K}, v; \Phi\big)$. This was indeed expected since Call prices with null strike coincide with the value of their underlier. Furthermore, by the definition of ψ in Lemma 7.10, the underlier of the Call option X with strike K is $S = K\psi\big(\frac{X}{K}, v; \Phi\big)$, so that $\tilde{S} = S$ and the above expression is the Call price of an option with strike K and underlier S, i.e. it coincides with X:

$$X = K\psi\Big(\frac{X}{K}, v; \Phi\Big)\Phi\Big(d_1\Big(K\psi\Big(\frac{X}{K}, v; \Phi\Big), K, v\Big)\Big) - K\Phi\Big(d_2\Big(K\psi\Big(\frac{X}{K}, v; \Phi\Big), K, v\Big)\Big).$$

Numerical computation of ψ

The function ψ in Lemma 7.10 is still not explicit. However, one can study more precisely ψ in the case of a strictly log-concave f, which covers the Black-Scholes case in particular.

Indeed, let us look at the function $\psi(a, y)$. We consider $a > 0$ and call

$$\gamma(r; a, y) := \frac{a + F(r - y)}{F(r)}$$

the argument of the infimum. Then $\gamma(\infty; a, y) = 1 + a$ and γ is non-increasing iff $\gamma'(r; a, y) = \frac{F(r)f(r-y)-f(r)(a+F(r-y))}{F(r)^2} \leq 0$, or iff $a \geq \delta(r; y)$ where

$$\delta(r; y) := \frac{f(r - y)}{f(r)} F(r) - F(r - y).$$

We have the following:

Lemma 7.11. *Let f a strictly log-concave probability density function and F its cumulative density function, and let $a, y > 0$. Then $\delta(r; y)$ is strictly increasing in r and:*

- *If $\delta(\infty; y) \leq a$ then $\psi(a, y) = 1 + a$;*
- *If $\delta(\infty; y) > a$ then $\psi(a, y) = \frac{a+F(r^*-y)}{F(r^*)} = \frac{f(r^*-y)}{f(r^*)}$ where $r^* := \delta^{-1}(a; y)$.*
- *In the Black-Scholes case $\delta(\infty; v) = \infty$ for every v.*

Proof. Observe that if γ has a finite limit at $-\infty$, then $\delta(r; y) = F(r)\big(\frac{f(r-y)}{f(r)} - \frac{F(r-y)}{F(r)}\big)$ has a limit equal to 0 due to l'Hï¿½pital's rule. Also, if γ explodes at $-\infty$, then its derivative must be non-positive at $-\infty$, i.e. a is always larger or equal than $\delta(-\infty; y)$, which does not depend on a. As a consequence, $\delta(-\infty; y) = 0$ in any case.

If $\delta(r; y)$ is increasing, two scenarios are possible:

- $\delta(\infty; y) > a$, then a cannot be always larger than $\delta(r, y)$ and the function γ has at least one point of minimum. Also, since δ is monotonous in r, there is a unique r^* such that $a = \delta(r^*; y)$ and this point is also the point of minimum of γ, i.e. $\psi(a, y) = \frac{a+F(r^*-y)}{F(r^*)} = \frac{f(r^*-y)}{f(r^*)}$;
- $\delta(\infty; y) \leq a$ (in particular δ is not surjective), then γ is decreasing and therefore $\psi(a, y) = 1 + a$.

Now since f is strictly log-concave, then the function $\frac{f'}{f}$ is decreasing. The derivative of δ with respect to r is $\frac{F(r)}{f(r)^2}(f(r)f'(r - y) - f(r - y)f'(r))$ and this is positive iff $\frac{f'(r-y)}{f(r-y)} > \frac{f'(r)}{f(r)}$, which holds true.

Then, if $\delta(\infty; y) > a$ there exists a unique $r^* := \delta^{-1}(a; y)$ and $\psi(a, y) = \frac{a+F(r^*-y)}{F(r^*)}$. Otherwise, if $\delta(\infty; y) \leq a$, it holds $\psi(a, y) = 1 + a$.

In the Black-Scholes case, f is strictly log-concave and $\frac{f(r-v)}{f(r)} = \exp\big(-\frac{(r-v)^2-r^2}{2}\big) = \exp\big(\frac{v(2r-v)}{2}\big)$ which explodes for r going to ∞. Then $\delta(\infty; v) = \infty > a$.

□

We can actually give more details on the bounds of the point of minimum r^* of the function γ, when it exists (i.e. when $\delta(\infty; y) > a$). In the following lemma we find a lower bound r^l and an upper bound r^u for r^* under the hypothesis of surjectivity of the function $\frac{f'}{f}$, and in the successive corollary we study the Black-Scholes case, where the bounds are actually explicit. In this way, the point r^* can be numerically computed in a very quick way inverting the function δ in the interval $[r^l, r^u]$.

Lemma 7.12. *Let f a strictly log-concave probability density function and F its cumulative density function, and let $a, y > 0$. Then $\frac{f'(r)}{f(r)}$ is decreasing and f is unimodal. Let s the unique point of maximum of f.*

In the case $\delta(\infty; y) > a$, if $\frac{f'(r)}{f(r)}$ is surjective, it holds

- *If $\delta(s; y) \geq a$ then $\tilde{r} < r^* \leq s$, where $\tilde{r}$ is the unique $r \leq s$ solving $a = F(r) - F(r-y)$.*
- *If $\delta(s; y) < a$ and $\delta(s+y; y) \geq a$ then $s < r^* \leq s + y$.*
- *If $\delta(s+y; y) < a$ then $s + y < r^* < \hat{r}$, where $\hat{r}$ is the unique r solving $\frac{f(r-y)}{f(r)} = \frac{a}{F(s)} + 1$.*

Proof. Firstly observe that since f is strictly log-concave, then the function $r \to \frac{f'(r)}{f(r)}$ is decreasing while the function $r \to \frac{f(r-y)}{f(r)}$ is increasing, given that its derivative is $\frac{f'(r-y)f(r)-f(r-y)f'(r)}{f(r)^2}$. Secondly, from the proof of Theorem 4.1.6. of [66], it holds

$$\frac{f'(r)}{f(r)} \leq \frac{1}{y} \log \frac{f(r)}{f(r-y)} \leq \frac{f'(r-y)}{f(r-y)}.$$

Then if $\frac{f'(r)}{f(r)}$ goes to $-\infty$ at ∞, the function $\frac{f(r-y)}{f(r)}$ explodes at ∞, while if $\frac{f'(r)}{f(r)}$ goes to ∞ at $-\infty$, then $\frac{f(r-y)}{f(r)}$ goes to 0 at $-\infty$.

In the case $\delta(\infty; y) > a$, from Lemma 7.11 there exists a unique r^* such that $a = \delta(r^*; y)$. Since $F(r) > F(r-y)$, it holds

$$a = \delta(r^*; y) > \left(\frac{f(r^*-y)}{f(r^*)} - 1\right) F(r^*-y).$$

If $\delta(s; y) \geq a$ then $r^* \leq s$. Otherwise $r^* > s$. If $\delta(s+y; y) \geq a$ then $r^* \leq s+y$. Otherwise $r^* > s + y$, so $F(r^*-y) > F(s)$, $\frac{f(r^*-y)}{f(r^*)} > 1$ and $\frac{a}{F(s)} + 1 > \frac{f(r^*-y)}{f(r^*)}$. Since the function $r \to \frac{f(r-y)}{f(r)}$ is increasing and explodes at ∞ under the hypothesis that $\frac{f'(r)}{f(r)}$ goes to $-\infty$, then there exists a unique $\hat{r}$ such that $\frac{f(\hat{r}-y)}{f(\hat{r})} = \frac{a}{F(0)} + 1$. Furthermore, $r^* < \hat{r}$.

In the previous step we have already found a lower bound (either s or $s+y$) in the case $\delta(s;y) < a$. If $\delta(s;y) \geq a$, then $r^* \leq s$ and $\frac{f(r^*-y)}{f(r^*)} < 1$, so

$$a = \delta(r^*;y) < F(r^*) - F(r^*-y).$$

The function $r \to F(r) - F(r-y)$ has derivative $f(r) - f(r-y)$, which is positive for $r \leq s$. Then there is a unique $\tilde{r} \leq s$ solving $a = F(\tilde{r}) - F(\tilde{r}-y)$ and $\tilde{r} < r^*$.

□

Lemma 7.12 can be further exploit in the Black-Scholes case. It turns out that the bounds for r^* are explicit and do not need any inversion algorithm.

Corollary 7.1 (Explicit bounds for r^* in the Black-Scholes case)**.** *In the Black-Scholes case:*

- *If* $a \leq \sqrt{\frac{\pi}{2}}\phi(v) + \Phi(v) - 1$ *then* $-\sqrt{-2\log\big(\frac{a}{v}\sqrt{2\pi}\big)} < r^* \leq 0$.
- *If* $\sqrt{\frac{\pi}{2}}\phi(v) + \Phi(v) - 1 < a \leq \frac{1}{\sqrt{2\pi}}\frac{\Phi(v)}{\phi(v)} - \frac{1}{2}$ *then* $0 < r^* \leq v$.
- *If* $a > \frac{1}{\sqrt{2\pi}}\frac{\Phi(v)}{\phi(v)} - \frac{1}{2}$ *then* $v < r^* < \frac{1}{2}\big(v + \frac{2}{v}\log(2a+1)\big)$.

Proof. In the Black-Scholes case, $f = \phi$, $s = 0$ and $\phi'(r) = -r\phi(r)$, so that $\frac{\phi'(r)}{\phi(r)} = -r$ which is surjective. In Lemma 7.11 we showed that $\delta(\infty;v) = \infty > a$, so three scenarios are possible applying Lemma 7.12.

In the first scenario, the condition $\delta(s;v) \geq a$ reads $\frac{\phi(-v)}{\phi(0)}\Phi(0) - \Phi(-v) \geq a$. Since $\Phi(-v) = 1 - \Phi(v)$, $\phi(-v) = \phi(v)$, $\Phi(0) = \frac{1}{2}$ and $\phi(0) = \frac{1}{\sqrt{2\pi}}$, the condition is equivalent to $a \leq \sqrt{\frac{\pi}{2}}\phi(v) + \Phi(v) - 1$. In this case the point $\tilde{r} \leq 0$ is the only r satisfying $a = \int_{r-v}^{r} \phi(z)\,dz$ and it holds $r^* > \tilde{r}$. Since $\tilde{r} \leq 0$, it holds $\int_{r-v}^{r} \phi(z)\,dz < \int_{r-v}^{r} \phi(\tilde{r})\,dz = \phi(\tilde{r})v$, so that $\frac{a}{v}\sqrt{2\pi} < \exp\big(-\frac{\tilde{r}^2}{2}\big)$. As a consequence, the LHS is smaller than 1 and we can solve $\tilde{r}^2 < -2\log\big(\frac{a}{v}\sqrt{2\pi}\big)$, which implies in particular $\tilde{r} > -\sqrt{-2\log\big(\frac{a}{v}\sqrt{2\pi}\big)}$.

In the second scenario, the condition $\delta(s+v;v) \geq a$ is $\frac{\phi(0)}{\phi(v)}\Phi(v) - \Phi(0) \geq a$, or equivalently $a \leq \frac{1}{\sqrt{2\pi}}\frac{\Phi(v)}{\phi(v)} - \frac{1}{2}$.

In the last scenario, the point $\hat{r}$ solves $\frac{\phi(\hat{r}-v)}{\phi(\hat{r})} = 2a+1$. The LHS is $\exp\big(\frac{v(2r-v)}{2}\big)$ and solving we find $\hat{r} = \frac{1}{2}\big(v + \frac{2}{v}\log(2a+1)\big)$.

□

Thanks to Lemma 7.12 and Corollary 7.1 we have found specific bounds for r^*. Then extremely fast numerical implementations based on standard methods such as the `brentq` function of the `scipy.optimize` library in Python can be obtained using these bounds.

7.3.2 Formulas from normalized transformations

The transformation $\mathbb{V}_{k,\alpha}$ of Section 7.2.4, and so $\mathbb{T}_k$ and $\mathbb{U}_k$, allow to generate new pricing formulas using the following recipe: start from a Call pricing function with closed formula, normalize it, apply the transformation and de-normalize to get another closed formula. This allows to extend any closed formula to a 2-parameter family of closed formulas.

In other words, we look at the pricing formula in the new world, but consider eventually applying it to the usual world: we take a financial engineer point of view here, where any pricing function depending on some parameters is a useful candidate to calibrate the current state of the market (in the usual world).

So we go through the following pipeline of transformations:

1. Start from any homogeneous arbitrage-free Call pricing function $K \to C(S,K)$;
2. Normalize it and get a function $c \in \mathbb{C}$ defined as $c(k) := \frac{C(S,kS)}{S}$;
3. Apply, for any $k, \alpha \geq 0$ such that $c(k) > 0$ and $\alpha \leq -\frac{c(k)}{c'(k)}$, the transformation $\mathbb{V}_{k,\alpha}$ to c;
4. Get a new arbitrage-free Call pricing function given by the formula $M \to X\mathbb{V}_{k,\alpha}c\big(\frac{M}{X}\big)$, where $X > 0$ represents the value of the new underlier.

The above steps can be applied in particular for the two special choices of α defining $\mathbb{T}_k$ and $\mathbb{U}_k$: $\alpha = c(k)$ and $\alpha = -\frac{c(k)}{c'(k)}$.

Observe that if we choose $k = \frac{K}{S}$ and $X = C(S,K)$ the new Call pricing function obtained using the transformation $\mathbb{V}_{k,c(k)} = \mathbb{T}_k$ coincides with $M \to \hat{C}_K(C(S,K), M) = C(S, K+M)$.

Let us compute the new closed formulas we obtain implementing the above pipeline for the known families of Black-Scholes, SVI (composed with the Black-Scholes pricing function), and CPT, that all provide closed-form formulas.

A 2-parameter enrichment of the Black-Scholes formula

Let $\mathrm{BS}(S,K)$ the Black-Scholes function defined in Equation (7.2). Then $\mathrm{BS}(S,K) = S\mathrm{bs}\big(\frac{K}{S}\big)$ where the normalized Black-Scholes pricing function bs belongs to the Tehranchi space. We can therefore consider the two families of functions indexed by k:

$$\mathrm{bs}_k^{\mathbb{T}}\left(\frac{M}{X}\right) := \frac{\mathrm{bs}\big(k + \mathrm{bs}(k)\frac{M}{X}\big)}{\mathrm{bs}(k)}$$
$$\mathrm{bs}_k^{\mathbb{U}}\left(\frac{M}{X}\right) := \frac{\mathrm{bs}\big(k - \frac{\mathrm{bs}(k)}{\mathrm{bs}'(k)}\frac{M}{X}\big)}{\mathrm{bs}(k)}$$

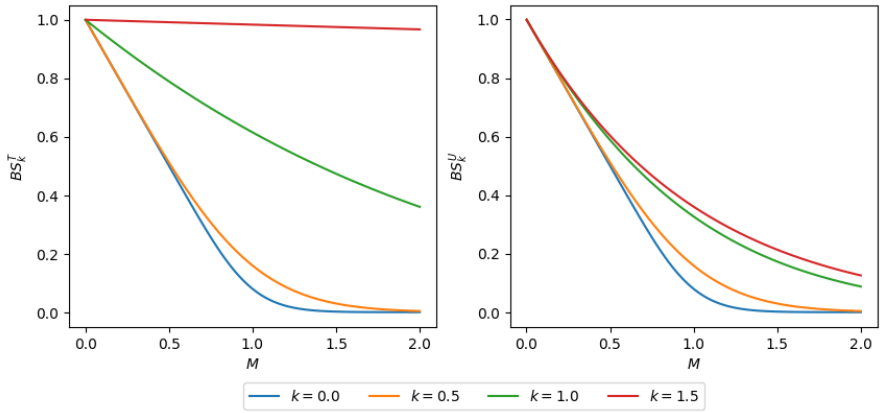

Figure 7.2: Prices $BS_k^{\mathbb{T}}(X, M)$ and $BS_k^{\mathbb{U}}(X, M)$ obtained from Black-Scholes formula applying transformations $\mathbb{T}_k$ and $\mathbb{U}_k$ respectively. The original Black-Scholes implied total volatility is set at 0.2.

leading to the *enriched Black-Scholes models*

$$\mathrm{BS}_k^{\mathbb{T}}(X, M) := X\mathrm{bs}_k^{\mathbb{T}}\left(\frac{M}{X}\right) = X\frac{\mathrm{bs}\left(k + \mathrm{bs}(k)\frac{M}{X}\right)}{\mathrm{bs}(k)}$$
$$\mathrm{BS}_k^{\mathbb{U}}(X, M) := X\mathrm{bs}_k^{\mathbb{U}}\left(\frac{M}{X}\right) = X\frac{\mathrm{bs}\left(k - \frac{\mathrm{bs}(k)}{\mathrm{bs}'(k)}\frac{M}{X}\right)}{\mathrm{bs}(k)}.$$

In Figure 7.2 we plot the $\mathrm{BS}_k^{\mathbb{T}}(X, M)$ and $\mathrm{BS}_k^{\mathbb{U}}(X, M)$ prices (for fixed maturity) for different values of k and a fixed value of $X = 1$. The original implied total volatility is fixed at 0.2.

In order to define the 2-parameter enrichment of the Black-Scholes formula $\mathrm{BS}_{k,z}^{\mathbb{V}}$ as suggested in the introduction of this section, let us make explicit the second parameter of the function bs, the total implied volatility $v = \sigma\sqrt{T}$. As seen in Section 7.2.4, the families $\mathrm{BS}_k^{\mathbb{T}}$ and $\mathrm{BS}_k^{\mathbb{U}}$ are a particular choice in the more generic set of families $X\frac{\mathrm{bs}\left(k+\alpha\frac{M}{X},v\right)}{\mathrm{bs}(k,v)}$. In particular, $\mathrm{BS}_k^{\mathbb{T}}$ corresponds to the choice $\alpha = \mathrm{bs}(k)$ and $\mathrm{BS}_k^{\mathbb{U}}$ to the choice $\alpha = -\frac{\mathrm{bs}(k)}{\mathrm{bs}'(k)}$.

More generally, we can represent α as the value of a normalized Black-Scholes pricing function at k: $\alpha = \mathrm{bs}(k, z)$ where the variability is given by the choice of the implied total volatility z. We have then identified a *2-parameters enrichment* of the Black-Scholes pricing function that satisfies properties of Proposition 7.1:

Proposition 7.6 (2-parameter enrichment of the Black-Scholes formula from the normalized pricing function)**.** *For any $k \geq 0$ and $z > 0$ we define the 2-parameter enrich-*

ment of the Black-Scholes pricing function with $v = \sigma_1\sqrt{T}$ *and* $z = \sigma_2\sqrt{T}$

$$BS^{\mathbb{V}}_{k,z}(X, M, v) := X\frac{bs\big(k + bs(k, z)\frac{M}{X}, v\big)}{bs(k, v)}.$$

Then $BS^{\mathbb{V}}_{k,z}(X, M, v)$ *is a non-increasing and convex function of* M *with* $BS^{\mathbb{V}}_{k,z}(X, M, v) \le X$. *If* $z \le v$, *then* $(X - M)_+ \le BS^{\mathbb{V}}_{k,z}(X, M, v)$.

In the case $z = v$, *the function* $k \to BS^{\mathbb{V}}_{k,v}(X, M, v)$ *is non-decreasing.*

In particular $BS^{\mathbb{V}}_{0,z}(X, M, v) = X\,bs\big(\frac{M}{X}, v\big) = BS(X, M, v)$.

Proof. The monotonicity and convexity with respect to M are easy to be proved. Since $k + \mathrm{bs}(k, z)\frac{M}{X} \ge k$ and the function $\mathrm{bs}(k, v)$ is decreasing in k, we obtain the first inequality. If $z \le v$, then $\mathrm{bs}(k, z) \le \mathrm{bs}(k, v)$ and $\mathrm{BS}^{\mathbb{V}}_{k,z}(X, M, v) \ge X\frac{\mathrm{bs}\left(k+\mathrm{bs}(k,v)\frac{M}{X}, v\right)}{\mathrm{bs}(k,v)}$ which is larger than $(X - M)_+$ from Lemma 7.4.

Finally if $z = v$ we apply Proposition 7.4.

□

The enriched SVI models

The Stochastic Volatility Inspired model (SVI) is a model for the implied total variance with formulation

$$\mathrm{SVI}(h) = a + b(\rho(h - m) + \sqrt{(h - m)^2 + \sigma^2})$$

where $h = \log\frac{K}{S}$ is the log-forward moneyness. It was firstly proposed by Jim Gatheral in 2004 at the Global Derivatives and Risk Management Madrid conference [33].

In this model, Call prices at moneyness k are Black-Scholes prices with implied total variance $\sqrt{\mathrm{SVI}(\log k)}$. We denote these prices with

$$\mathrm{BS}^{\mathrm{SVI}}(S, K) := \mathrm{BS}\left(S, K, \sqrt{\mathrm{SVI}\left(\log\frac{K}{S}\right)}\right).$$

Applying $\mathbb{T}_k$ and $\mathbb{U}_k$, we obtain the *enriched SVI models*:

$$\mathrm{BS}^{\mathrm{SVI},\mathbb{T}}_k(X, M) := X\frac{\mathrm{bs}^{\mathrm{SVI}}\big(k + \mathrm{bs}^{\mathrm{SVI}}(k)\frac{M}{X}\big)}{\mathrm{bs}^{\mathrm{SVI}}(k)}$$

$$\mathrm{BS}^{\mathrm{SVI},\mathbb{U}}_k(X, M) := X\frac{\mathrm{bs}^{\mathrm{SVI}}\big(k - \frac{\mathrm{bs}^{\mathrm{SVI}}(k)}{(\mathrm{bs}^{\mathrm{SVI}})'(k)}\frac{M}{X}\big)}{\mathrm{bs}^{\mathrm{SVI}}(k)}.$$

We plot prices $\mathrm{BS}^{\mathrm{SVI},\mathbb{T}}_k(X, M)$ and $\mathrm{BS}^{\mathrm{SVI},\mathbb{U}}_k(X, M)$ of the enriched SVI models in Figure 7.3, where we choose as initial arbitrage-free parameters for the SVI smile $a =$

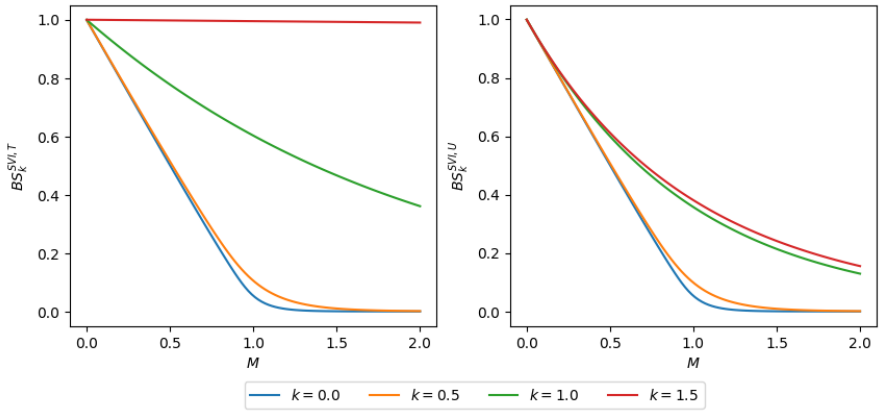

Figure 7.3: Prices $BS_k^{SVI,\mathbb{T}}(X,M)$ and $BS_k^{SVI,\mathbb{U}}(X,M)$ obtained from SVI prices applying transformations $\mathbb{T}_k$ and $\mathbb{U}_k$ respectively. The original arbitrage-free SVI parameters are $a = 0.01$, $b = 0.1$, $rho = -0.6$, $m = -0.05$, and $sigma = 0.1$.

0.01, $b = 0.1$, $\rho = -0.6$, $m = -0.05$, and $\sigma = 0.1$ (see Table 2 of [53]). We take a fixed value $X = 1$.

Again, we can enrich also the SVI model to a 2-parameter family indexed by k and α:

$$\mathrm{BS}^{\mathrm{SVI},\mathbb{V}}_{k,\alpha}(X,M) := X\frac{\mathrm{bs}^{\mathrm{SVI}}\left(k+\alpha\frac{M}{X}\right)}{\mathrm{bs}^{\mathrm{SVI}}(k)}.$$

The enriched CPT models

The CPT prices are defined in Equation (7.4) and have corresponding normalized prices

$$c^{\mathrm{CPT}}(k;f,y) := \frac{C^{\mathrm{CPT}}(S,Sk;f,y)}{S} = \int_{-\infty}^{\infty}\left(f(z+y)-kf(z)\right)_+ dz.$$

Even the CPT model can be extended via $\mathbb{T}_k$ and $\mathbb{U}_k$ to get the *enriched CPT models*:

$$C^{\mathrm{CPT},\mathbb{T}}_k(X,M) := X\frac{c^{\mathrm{CPT}}\left(k+c^{\mathrm{CPT}}(k;f,y)\frac{M}{X};f,y\right)}{c^{\mathrm{CPT}}(k;f,y)}$$
$$C^{\mathrm{CPT},\mathbb{U}}_k(X,M) := X\frac{c^{\mathrm{CPT}}\left(k-\frac{c^{\mathrm{CPT}}(k;f,y)}{(c^{\mathrm{CPT}})'(k;f,y)}\frac{M}{X};f,y\right)}{c^{\mathrm{CPT}}(k;f,y)}$$

and via $\mathbb{V}_{k,\alpha}$ to get the 2-parameter extension of the CPT model:

$$C^{\mathrm{CPT},\mathbb{V}}_{k,\alpha}(X,M) := X\frac{c^{\mathrm{CPT}}\left(k+\alpha\frac{M}{X};f,y\right)}{c^{\mathrm{CPT}}(k;f,y)}.$$

7.4 Smile symmetry and a lift of the relative pricing function

As already pointed out in Remark 7.1, Calls on Calls are contracts written on underliers with a positive mass in 0, i.e. that can become null with positive probability. This implies some unusual features such as a derivative of the pricing function with respect to the strike which is larger than -1 at the origin. However, it is possible to change the probability measure in order to obtain new contracts that do not present this feature anymore. Moreover, here is a tight relationship with the symmetry operation applied to the smile, as is well-known and detailed in section 2.2 of [52]. An analogous transformation is performed in section 2.2 of [66] with the involution of Call prices.

7.4.1 The change of probability with a mass at 0

We firstly start with the definition of the probability P^* associated to a non-negative non-constantly zero random variable X_T. Note that we re-introduce the maturity T here, in order to convey some financial context, on one hand, and also to distinguish those random variables from constant quantities known at the current time (assumed to be time 0).

Definition 7.6. *Let X_T a non-negative non-constantly zero random variable on the probability space (P, Ω), with finite expectation $E[X_T]$. We define*

$$P^*(A) := \frac{E\big[\mathbb{1}_A X_T \mathbb{1}_{\{X_T>0\}}\big]}{E[X_T]}$$

for every $A \in \Omega$. We also denote $X := E[X_T]$ and $P_0 := P(X_T = 0)$.

An immediate consequence of the above definition is that P^* is actually a probability measure on a subset of the original Ω. The proof of the following lemma is elementary and so omitted.

Lemma 7.13. *P^* is a probability measure on $\Omega^* = \{X_T > 0\}$. Any random variable Z on (P, Ω) can be restricted to a random variable on (P^*, Ω^*) (that we still denote with Z). Then $E^*[Z] = \frac{1}{X} E[Z X_T \mathbb{1}_{\{X_T>0\}}]$.*

Let $X_T^* := \frac{1}{X_T}$. Then

$$E^*[X_T^*] = \frac{1 - P_0}{X}.$$

This lemma suggests to consider contracts under the probability P^* on the underlier $X_T^* = \frac{1}{X_T}$. Indeed, it holds

$$E^*\left[\left(\frac{1}{X_T} - \frac{1}{K}\right)_+\right] = \frac{E\big[(K - X_T)_+ \mathbb{1}_{\{X_T>0\}}\big]}{XK}$$
$$E^*\Big[\Big(\frac{1}{K} - \frac{1}{X_T}\Big)_+\Big] = \frac{E\big[(X_T - K)_+\big]}{XK}$$

which suggests that a Call price under P is also a Put price on the underlier X_T^* under the new probability P^*. Furthermore $E[(K-X_T)_+] = E\big[(K-X_T)_+ \mathbb{1}_{\{X_T>0\}} + K\mathbb{1}_{\{X_T=0\}}\big] = E\big[(K-X_T)_+\mathbb{1}_{\{X_T>0\}}\big] + KP_0$ so that the Put-Call-Parity

$$\frac{1-P_0}{X} - \frac{1}{K} = E^*\left[\left(\frac{1}{X_T} - \frac{1}{K}\right)_+\right] - E^*\left[\left(\frac{1}{K} - \frac{1}{X_T}\right)_+\right]$$

allows us to express the price of Calls on $\frac{1}{X_T}$ with strike $\frac{1}{K}$ under P^* with respect to the original Call prices as

$$\begin{aligned} E^*\left[\left(\frac{1}{X_T} - \frac{1}{K}\right)_+\right] &= E^*\left[\left(\frac{1}{K} - \frac{1}{X_T}\right)_+\right] + \frac{1-P_0}{X} - \frac{1}{K} \\ &= \frac{E\big[(X_T-K)_+\big]}{XK} + \frac{1-P_0}{X} - \frac{1}{K}. \end{aligned}$$

We can re-apply the same procedure to the underlier X_T^* under P^*, defining a probability measure P^{**} and an underlier X_T^{**}.

Definition 7.7. *Let X_T a non-negative non-constantly zero random variable on the probability space (P,Ω), with finite expectation $E[X_T]$. We define*

$$P^{**}(A) := \frac{E^*\big[\mathbb{1}_A X_T^*\big]}{E^*[X_T^*]}$$

for every $A \in \Omega^*$.

The random variable $\frac{1}{X_T^*}$ is well defined and does not vanish under P^* since it is defined on the space Ω^*, so that the corresponding of Lemma 7.13 for the function P^{**} becomes the following.

Lemma 7.14. *P^{**} is a probability measure on $\Omega^* = \{X_T > 0\}$. Any random variable Z on (P,Ω) can be restricted to a random variable on (P^{**},Ω^*) (that we still denote with Z). Then $E^{**}[Z] = \frac{X}{1-P_0}E^*[Z] = \frac{1}{1-P_0}E[Z\mathbb{1}_{\{X_T>0\}}]$.*

Let $X_T^{**} := \frac{1}{X_T^*}$. Then

$$E^{**}[X_T^{**}] = \frac{X}{1-P_0}.$$

We can now consider Call prices written on X_T^{**} using again the Put-Call-Parity:

$$\begin{aligned}
E^{**}\left[\left(\frac{1}{X_T^*}-\frac{1}{K}\right)_+\right] &= E^{**}\left[\left(\frac{1}{K}-\frac{1}{X_T^*}\right)_+\right]+\frac{X}{1-P_0}-\frac{1}{K}\\
&= \frac{1}{1-P_0}E\left[\left(\frac{1}{K}-X_T\right)_+\mathbb{1}_{\{X_T>0\}}\right]+\frac{X}{1-P_0}-\frac{1}{K}\\
&= \frac{1}{1-P_0}\left(E\left[\left(X_T-\frac{1}{K}\right)_+\right]-X+\frac{1-P_0}{K}\right)+\frac{X}{1-P_0}-\frac{1}{K}\\
&= \frac{1}{1-P_0}E\left[\left(X_T-\frac{1}{K}\right)_+\right].
\end{aligned}$$

In other words, Calls on X_T^{**} are still Calls on X_T, with an apposite rescaling. The main difference between the two type of Calls is that while Calls on X_T might have a derivative larger than -1 in 0 because of the non-null probability of X_T to nullify, the derived Calls on X_T^{**} will anyways have a derivative in 0 strictly equal to -1.

7.4.2 In terms of pricing functions

We have introduced the change of probability measure to avoid mass in zero, now we can consider contracts written in the new probability spaces. From the previous discussion in Section 7.4.1 we get the following.

Lemma 7.15. *Assume there is a function C of two variables such that $C(X,K) = E[(X_T-K)_+]$. Then*

$$\begin{aligned}
E^*\left[\left(X_T^*-K\right)_+\right] &= \frac{K}{X}C\left(X,\frac{1}{K}\right)+\frac{1-P_0}{X}-K\\
E^{**}\left[\left(X_T^{**}-K\right)_+\right] &= \frac{1}{1-P_0}C(X,K).
\end{aligned}$$

We are interested by necessary and sufficient conditions on C such that there exist functions C^* and C^{**} satisfying

$$\begin{aligned}
C^*(X^*,K) &= E^*\left[\left(X_T^*-K\right)_+\right]\\
C^{**}(X^{**},K) &= E^{**}\left[\left(X_T^{**}-K\right)_+\right]
\end{aligned} \tag{7.5}$$

where $X^* := \frac{1-P_0}{X}$ and $X^{**} := \frac{X}{1-P_0}$.

The usual situation where $P_0 = 0$ constantly leads, given that $X^* = \frac{1}{X}$ and $X^{**} = X$, to the formulas

$$\begin{aligned}
C^*(X^*,K) &= KX^*C\left(\frac{1}{X^*},\frac{1}{K}\right)+X^*-K\\
C^{**}(X^{**},K) &= C(X^{**},K).
\end{aligned}$$

which are described in Theorem 2.2.2. of [66].

Note that in the degenerate case where $C(X,K) = (X-(1-P_0)K)_+$, it holds $\frac{K}{X}C\big(X,\frac{1}{K}\big) + \frac{1-P_0}{X} - K = (K-X^*)_+ + X^* - K = (X^*-K)_+$ and $\frac{1}{1-P_0}C(X,K) = (X^{**}-K)_+$, so that the required property in Equation (7.5) holds.

Going back to the general case, the following lemma finds a sufficient condition for the existence of the functions C^* and C^{**} in case P_0 is given by some function $\hat{P}_0(X) = P_0$.

Lemma 7.16. *A sufficient condition for the existence of functions C^* and C^{**} satisfying Equation (7.5) is the existence of a function f such that $X = f\big(\frac{X}{1-P_0}\big)$ where $P_0 = \hat{P}_0(X)$. Then*

$$C^*(X^*,K) := \frac{K}{f\big(\frac{1}{X^*}\big)}C\left(f\left(\frac{1}{X^*}\right),\frac{1}{K}\right) + X^* - K$$
$$C^{**}(X^{**},K) := \frac{X^{**}}{f(X^{**})}C(f(X^{**}),K).$$

This lemma generalizes the no-mass at 0 situation where f is the identity function.

In terms of normalized pricing functions

It is also interesting to look at the normalized partial pricing functions (for a fixed pair X, P_0) where the strike is replaced by the moneyness m, and the price is scaled by the underlier as in Section 7.2.3.

Lemma 7.17. *It holds*

$$c^*(m) = mc\left(\frac{1}{-c'(0_+)m}\right) + 1 - m$$
$$c^{**}(m) = c\left(\frac{m}{-c'(0_+)}\right).$$

In particular

$$c^{*'}(0_+) = c(\infty) - 1$$
$$c^{**'}(0_+) = -1.$$

Proof. From definitions, it holds

$$c^*(m) = \frac{m}{X}C\left(X,\frac{X}{(1-P_0)m}\right) + 1 - m$$
$$c^{**}(m) = \frac{1}{X}C\left(X,\frac{X}{1-P_0}m\right).$$

The first conclusion follows from the definition of $c(m) = \frac{C(X,Xm)}{X}$ and from the fact that $P_0 = P(X_T = 0) = 1 + c'(0_+)$.

It is easy to show that the derivative of c^{**} in 0_+ is -1, while the derivative of c^* in 0_+ is $c(\infty) - 1 + \lim_{m\to 0_+} \frac{c'\left(\frac{1}{-c'(0_+)m}\right)}{-c'(0_+)m}$. Now by Fubini, $\frac{d}{dK}C(X,K) = -P(X_T > K) = -E[\mathbb{1}_{\{X_T>K\}}]$ and it holds

$$\begin{aligned} X &= E[X_T\mathbb{1}_{\{X_T>K\}}] + E[X_T\mathbb{1}_{\{X_T\leq K\}}] \\ &\geq KE[\mathbb{1}_{\{X_T>K\}}] + E[X_T\mathbb{1}_{\{X_T\leq K\}}]. \end{aligned}$$

Letting K to ∞, the term $E[X_T\mathbb{1}_{\{X_T\leq K\}}]$ goes to $E[X_T] = X$, so that the term $KE[\mathbb{1}_{\{X_T>K\}}] = -K\frac{d}{dK}C(X,K)$ must go to 0. As a consequence, $kc'(k) = \frac{K}{X}\frac{d}{dK}C(X,K)$ goes to 0 as $k = \frac{1}{-c'(0_+)m}$ goes to ∞.

□

7.4.3 The lifted Calls on Calls

Let us go back now to the case of the relative Call on Call pricing function which satisfies

$$C(S, K+M) = \hat{C}_K\big(C(S,K), M\big).$$

This means that we set the random variable X_T in Definition 7.6 to be $(S_T - K)_+$, so that $X_T = 0$ iff $S_T \leq K$.

In this contest, the random variables X_T^* and X_T^{**} correspond to $\frac{1}{(S_T-K)_+}$ and $(S_T - K)_+$ seen as random variables in the set $\{S_T > K\}$, so that

$$\begin{aligned} X &= C(S,K) \\ X^* &= \frac{1 - P(S_T \leq K)}{C(S,K)} \\ X^{**} &= \frac{C(S,K)}{1 - P(S_T \leq K)}. \end{aligned}$$

Then we want to find functions $\hat{C}_K^*$ and $\hat{C}_K^{**}$ such that

$$\begin{aligned} \hat{C}_K^*(X^*, M) &= \frac{M}{C(S,L)}\hat{C}_K\left(C(S,K), \frac{1}{M}\right) + \frac{1 - P(S_T \leq K)}{C(S,L)} - M \\ \hat{C}_K^{**}(X^{**}, M) &= \frac{1}{1 - P(S_T \leq K)}\hat{C}_K\big(C(S,K), M\big) \end{aligned} \tag{7.6}$$

and we call *lifted* Calls on Calls the prices $\hat{C}_K^{**}$.

In the following proposition we reconsider Lemma 7.16 and state a sufficient condition for the existence of such functions $\hat{C}_K^*$ and $\hat{C}_K^{**}$.

Proposition 7.7 (Sufficient condition for the existence of $\hat{C}_K^*$ and $\hat{C}_K^{**}$)**.** *A sufficient condition for the existence of functions $\hat{C}_K^*$ and $\hat{C}_K^{**}$ satisfying Equation (7.6) is that the function*

$$S \to -\frac{d}{dS}C(S,K)\frac{d}{dK}C(S,K) + C(S,K)\frac{d^2}{dSdK}C(S,K)$$

has constant sign for fixed K.

In the homogeneous case this is equivalent to the condition that the function

$$kc'(k)^2 - c(k)c'(k) - kc(k)c''(k) \tag{7.7}$$

has constant sign.

In the Black-Scholes case, the condition holds true.

Proof. From Lemma 7.16, a sufficient condition for the existence of $\hat{C}_K^*$ and $\hat{C}_K^{**}$ is the existence of a function f such that $f\left(\frac{C(S,K)}{1-P(S_T\leq K)}\right) = C(S,K)$. This condition is one to one with the fact that the function $g(S) = \frac{C(S,K)}{1-P(S_T\leq K)}$ is monotone. Indeed, if $g(S^1) = g(S^2)$, then $f(g(S^1)) = f(g(S^2))$, i.e. $C(S^1,K) = C(S^2,K)$ and $S^1 = S^2$ since $C(\cdot,K)$ is a monotone function. On the other hand, if g is monotone, then the function $f(x) = C(g^{-1}(x),K)$ is the required function.

Observe that $1 - P(S_T \leq K) = -\frac{d}{dK}C(S,K)$. Then we should prove that $g(S) = \frac{C(S,K)}{-\frac{d}{dK}C(S,K)}$ is monotone. The derivative of g has the sign of

$$-\frac{d}{dS}C(S,K)\frac{d}{dK}C(S,K) + C(S,K)\frac{d^2}{dSdK}C(S,K).$$

In the homogeneous case

$$g(S) = \frac{c(k)S}{-c'(k)}$$

and it is monotone iff Equation (7.7) has constant sign.

In the Black-Scholes case, we have

$$\begin{aligned}
\mathrm{bs}(k) &= \Phi(d_1) - k\Phi(d_2)\\
\mathrm{bs}'(k) &= -\Phi(d_2)\\
\mathrm{bs}''(k) &= \frac{\phi(d_2)}{kv}
\end{aligned}$$

where $d_{1,2} = -\frac{\log k}{v} \pm \frac{v}{2}$ and $v = \sigma\sqrt{T}$. The expression in Equation (7.7) becomes $\Phi(d_1)\Phi(d_2) - (\Phi(d_1) - k\Phi(d_2))\frac{\phi(d_2)}{v}$ and, using the equality $k\phi(d_2) = \phi(d_1)$, the latter expression becomes $\frac{1}{v}(\phi(d_1)\Phi(d_2) + v\Phi(d_1)\Phi(d_2) - \phi(d_2)\Phi(d_1))$. Since $d_1 = d_2 + v$, we shall study the sign of $\phi(x+v)\Phi(x) + v\Phi(x+v)\Phi(x) - \phi(x)\Phi(x+v)$ for $v > 0$. Dividing by $\phi(x+v)\phi(x)$, this reduces to study

$$R(x) + vR(x+v)R(x) - R(x+v) \tag{7.8}$$

where $R(x) = \frac{\Phi(x)}{\phi(x)}$. Let us consider the function $\mathcal{R}_x(v) = \frac{R(x+v)}{1+vR(x+v)}$ for a fixed x. Its derivative can be simplified to $\frac{R'(x+v)-R^2(x+v)}{(1+vR(x+v))^2}$, which has the sign of $\phi(z)^2 + z\Phi(z)n(z) - \Phi(z)^2 = \phi(z)^2\left(1 + zR(z) - R(z)^2\right)$ where $z = x+v$.

Observe now that $R(z) = r(-z)$ where r denotes the Mill's ratio, whence $1+zR(z)-R(z)^2 = 1-tr(t)-r(t)^2$ with $t=-z$. We know that $1-tr(t) = -r'(t)$, and also from Sampford [63] that $1 > \frac{1}{r}' = \frac{-r'}{r^2}$, which proves that $1+zR(z)-R(z)^2 < 0$.

Therefore $\mathcal{R}_x(v)$ attains its maximum for v going to 0, where $\mathcal{R}_x(0) = R(x)$. As a consequence the expression Equation (7.8) is always positive and so is Equation (7.7). □

Proposition 7.7 is of particular interest since it shows that in the Black-Scholes case there exist functions $\hat{C}_K^*$ and $\hat{C}_K^{**}$ satisfying Equation (7.6), meaning that it is possible to define lifted Call on Call pricing functions, i.e. Calls on Calls with the property of a derivative equal to -1 at 0. In particular, we have proved that in the Black-Scholes case the function $g(X) = \frac{X}{\Phi(d_2(C^{-1}(X,K),K,v))}$ where $d_2(a,b,v) = -\frac{\log\frac{b}{a}}{v} - \frac{v}{2}$ is invertible in the variable X and its inverse is $f(X^{**}) := g^{-1}(X^{**})$ which satisfies

$$f\left(\frac{X}{\Phi(d_2(C^{-1}(X,K),K,v))}\right) = X.$$

The function f, however, does not present an explicit form.

Using Proposition 7.5 we can write the formula for the lifted Call on Call pricing function in the Black-Scholes case:

Proposition 7.8 (Formula for the lifted Call on Call pricing function in the Black-Scholes model)**.** *For $K, M, v = \sigma\sqrt{T} \geq 0$, let $\psi(a,v;\Phi) = \inf_{r\in\mathbb{R}} \frac{a+\Phi(r-v)}{\Phi(r)}$. Call $\tilde{f}(\cdot,v)$ the inverse of the function*

$$z \to \frac{z}{\Phi\big(d_2(\psi(z,v;\Phi),1,v)\big)}.$$

Then for Black-Scholes prices it holds

$$\hat{C}_K^{**}(X^{**},M) = \frac{1}{\Phi(d_2(\hat{s},1,v))}\Big(K\hat{s}\Phi\big(d_1(K\hat{s},K+M,v)\big) - (K+M)\Phi\big(d_2(K\hat{s},K+M,v)\big)\Big)$$

where $d_{1,2}(a,b,v) = -\frac{\log\frac{b}{a}}{v} \pm \frac{v}{2}$ and $\hat{s} = \psi\big(\tilde{f}\big(\frac{X^{**}}{K},v\big),v;\Phi\big)$.

Relation with the transformation $\mathbb{U}_k$

Let us consider the lifted Call on Call $\hat{C}_K^{**}(X^{**},M)$ and its normalized function

$$\hat{c}_K^{**}(m) = \frac{\hat{C}_K^{**}(X^{**},X^{**}m)}{X^{**}}$$

where $X^{**} = \frac{X}{1-P(X_T=0)}$. Then in turn it holds

$$\begin{aligned}\hat{c}_K^{**}(m) &= \frac{1}{X}\hat{C}_K\left(X, \frac{X}{1-P(X_T=0)}m\right)\\ &= \frac{1}{X}C\left(C^{-1}(X,K), K + \frac{X}{1-P(X_T=0)}m\right).\end{aligned}$$

We write $k = \frac{K}{C^{-1}(X,K)}$ and define the function $c(l) := \frac{C(C^{-1}(X,K),lC^{-1}(X,K))}{C^{-1}(X,K)}$, then $X = c(k)C^{-1}(X,K)$ and

$$\hat{c}_K^{**}(m) = \frac{c\big(k + \frac{c(k)}{1-P(X_T=0)}m\big)}{c(k)}$$

where $1 - P(X_T = 0) = -\partial_K C(C^{-1}(X,K),K) = -c'(k)$. Then

$$\hat{c}_K^{**}(m) = \mathbb{U}_{k(X,K)}c(m)$$

where $k(X,K) = \frac{K}{C^{-1}(X,K)}$.

While in Section 7.2.4 we have proved that the transformation $\mathbb{T}_k$ corresponds to the normalization of the relative Call on Call $\hat{C}_K$, here we showed that the transformation $\mathbb{U}_k$ actually corresponds to the normalization of the lifted Call on Call $\hat{C}_K^{**}$. In other words we showed the following.

Lemma 7.18. *The transformations $\mathbb{T}_k$ and $\mathbb{U}_k$ are linked to relative Calls on Calls $\hat{C}_K$ and lifted Calls on Calls $\hat{C}_K^{**}$ via*

$$\hat{C}_K(X,M) = \mathbb{T}_{k(X,K)}c\Big(\frac{M}{X}\Big)X$$
$$\hat{C}_K^{**}(X^{**},M) = \mathbb{U}_{k(X,K)}c\Big(\frac{M}{X^{**}}\Big)X^{**}$$

where $k(X,K) = \frac{K}{C^{-1}(X,K)}$.

7.5 Implied volatility of the relative pricing functions

The Calls on Calls prices and their lifted prices can be studied from an implied volatility point of view. It is of extreme interest that, even in the case of underlying Black-Scholes Calls, the relative prices actually hide smile shapes which are not constant. Also, we can state something more on these smiles, as we will see that they always explode for small strikes while behave as the original smiles at ∞.

We define the implied total volatility $\hat{v}_K(X,M)$ of the relative Call on Call $\hat{C}_K(X,M)$ and the implied total volatility $\hat{v}_K^{**}(X^{**},M)$ of the lifted Call on Call $\hat{C}_K^{**}(X^{**},M)$ to be the value of $v = \sigma\sqrt{T}$ satisfying

$$\hat{C}_K(X,M) = \mathrm{BS}(X,M,v)$$

and

$$\hat{C}_K^{**}(X^{**},M) = \mathrm{BS}(X^{**},M,v)$$

respectively, where BS is defined in Equation (7.2).

We are interested in the functions $M \to \hat{v}_K(X, M)$ and $M \to \hat{v}_K^{**}(X, M)$ and in particular to their asymptotics for M going to ∞ and 0. We can look at the limits of this function studying the relations

$$\hat{C}_K(X, M) = C(C^{-1}(X, K), K + M)$$
$$\hat{C}_K^{**}(X^{**}, M) = \frac{X^{**}}{f(X^{**})}\hat{C}_K(f(X^{**}), M)$$

where we suppose that there exists f such that $f(X^{**}) = X$.

We denote with $v(S, L)$ the implied total volatility of the Call option with strike L and underlier S, so that we shall study the relation between $\hat{v}_K(X, M)$ (or $\hat{v}_K^{**}(X^{**}, M)$) and $v(C^{-1}(X, K), K + M)$). We will also denote $\hat{d}_{1,2}(X, M) := d_{1,2}\big(X, M, \hat{v}_K(X, M)\big)$ and $\hat{d}_{1,2}^{**}(X^{**}, M) := d_{1,2}\big(X^{**}, M, \hat{v}_K^{**}(X^{**}, M)\big)$. When it is clear, we will drop the dependence to X and X^{**}.

Remark 7.3. *In the framework of arbitrage-free prices the functions d_1 and d_2 are monotone for the results found by Fukasawa in [32]. The condition of surjectivity is not granted a priori. In general, it always holds that the function d_1 goes at $+\infty$ for small strikes and the function d_2 goes at $-\infty$ for large strikes. However, the function d_1 goes to $-\infty$ for large strikes iff Call prices vanish at ∞, and d_2 goes to ∞ for small strikes iff there is no mass at 0, i.e. the derivative of Call prices is -1 in 0.*

Furthermore, Proposition 2.1 in [56] shows that if d_1 is surjective, then the Lee right wing condition holds: $v(S, K) < \sqrt{2 \log \frac{K}{S}}$ for K large enough; while if d_2 is surjective, then the Lee left wing condition holds: $v(S, K) < \sqrt{-2 \log \frac{K}{S}}$ for K small enough.

Remark 7.4. *For two functions f and g we write $f \sim g$ iff $\lim_x \frac{f(x)}{g(x)} = 1$, where the limit of x will depend on the situation.*

Integrating by parts the expression for $1 - \Phi(x)$ we find

$$1 - \Phi(x) = \frac{\phi(x)}{x} - \int_x^\infty \frac{\phi(t)}{t^2}\, dt = \frac{\phi(x)}{x} - \frac{\phi(x)}{x^3} + \int_x^\infty \frac{\phi(t)}{t^4}\, dt$$

so that for $x > 0$

$$\frac{\phi(x)}{x} - \frac{\phi(x)}{x^3} < 1 - \Phi(x) < \frac{\phi(x)}{x}.$$

Then, for x going to ∞, $\Phi(-x) \sim \frac{\phi(x)}{x}$.

7.5.1 Calls on Calls' implied volatility

In this section we prove that the Calls on Calls' smiles behave as the original smiles at ∞, while at small strikes they always explode with a rate of $\sqrt{-2 \log \frac{M}{X}}$.

Lemma 7.19 (Asymptotic behavior of the Calls on Calls' total implied volatility)**.** *The Calls on Calls' total implied volatility $\hat{v}_K(X, M)$ behaves asymptotically as*

- $\sqrt{-2\log\frac{M}{X}} - \hat{d}_2(X,0)$ *for small strikes;*
- *the underlying total implied volatility* $v(C^{-1}(X,K),K+M)$ *for large strikes. If it exists, the limit of* $\hat{v}_K(X,M)$ *is equal to the limit of* $v(C^{-1}(X,K),K+M)$.

Proof. Firstly, we observe that $-\frac{d}{dM}\hat{C}_K(X,M) = -\partial_K C(C^{-1}(X,K),K+M)$. In $M=0$, this means that Calls on Calls' prices have a derivative in 0 equal to $-\partial_K C(C^{-1}(X,K),K)$ which, for $K>0$, is strictly larger than -1. Then, the function $\hat{d}_2(M)$ is not surjective and $\hat{v}_K \sim \sqrt{-2\log\frac{M}{X}}$. More precisely, we can write $\hat{v}_K = \gamma\sqrt{-2\log\frac{M}{X}}$ where $\gamma\sim 1$ at ∞. Substituting in the expression for $\hat{d}_2(M)$, we find

$$\hat{d}_2(M) = \frac{\sqrt{2}}{2}\sqrt{-\log\frac{M}{X}}\left(\frac{1}{\gamma}-\gamma\right) \sim \hat{d}_2(0) \tag{7.9}$$

where $\hat{d}_2(0)$ is the finite limit of $\hat{d}_2$ for $M=0$. Observe that $\frac{1}{1-\varepsilon} = 1-\varepsilon+o(\varepsilon)$ so that $\frac{1}{1-\varepsilon}-(1-\varepsilon) = 2\varepsilon+o(\varepsilon)$. Setting $\gamma = 1-\epsilon$, we find from Equation (7.9) that it must hold $\varepsilon \sim \frac{\hat{d}_2(0)}{\sqrt{-2\log\frac{M}{X}}}$, or $\gamma\sim 1-\frac{\hat{d}_2(0)}{\sqrt{-2\log\frac{M}{X}}}$. All in all, we find $\hat{v}_K \sim \sqrt{-2\log\frac{M}{X}} - \hat{d}_2(0)$.

Secondly, from the definition of Calls on Calls' prices, it follows that Calls on Calls vanish at increasing strikes iff the underlying Calls vanish. In particular, if d_1 has finite limit, then its implied total volatility must explode at ∞, and similarly for $\hat{d}_1$ and $\hat{v}_K$. Otherwise, if d_1 is surjective, then $\hat{d}_1$ is surjective and in particular for Remark 7.3 the Lee right wing condition holds, i.e. $v(C^{-1}(X,K),M), \hat{v}_K(X,M) < \sqrt{2\log\frac{M}{X}}$ for large M. We can then write the asymptotics of the Call on Call price using Remark 7.4 as

$$C_{\mathrm{BS}}(X,M,\hat{v}_K) \sim X\frac{\phi(\hat{d}_1(M))}{-\hat{d}_1(M)} - M\frac{\phi(\hat{d}_2(M))}{-\hat{d}_2(M)}$$

and since $M\phi(\hat{d}_2(M)) = X\phi(\hat{d}_1(M))$, developing the expressions for $\hat{d}_1(M)$ and $\hat{d}_2(M)$, the right hand side becomes

$$C_{\mathrm{BS}}(X,M,\hat{v}_K) \sim X\phi(\hat{d}_1(M))\frac{\hat{v}_K^3}{\left(\log\frac{M}{X}\right)^2 - \frac{\hat{v}_K^4}{4}}.$$

Taking the logarithm in the above expression and looking at the dominating terms, we find

$$\log C_{\mathrm{BS}}(X,M,\hat{v}_K) \sim -\frac{\hat{d}_1(M)^2}{2}.$$

If $\hat{v}_K \in o\left(\sqrt{\log\frac{M}{X}}\right)$ (this includes the case where it goes to a finite limit) then

$$\log C_{\mathrm{BS}}(X,M,\hat{v}_K) \sim -\frac{1}{2\hat{v}_K^2}\left(\log\frac{M}{X}\right)^2,$$

otherwise, if $\hat{v}_K \sim \hat{c}\sqrt{\log \frac{M}{X}}$ with $0 < \hat{c} < \sqrt{2}$, then

$$\log C_{\mathrm{BS}}(X, M, \hat{v}_K) \sim -\frac{\log \frac{M}{X}}{8\hat{c}^2}(2 - \hat{c}^2)^2.$$

Similarly, we develop the asymptotics of the logarithm of $C_{\mathrm{BS}}(C^{-1}(X, K), K + M, v)$ and get that if $v \in o\big(\sqrt{\log \frac{K+M}{C^{-1}(X,K)}}\big)$ these coincide with

$$\log C_{\mathrm{BS}}(C^{-1}(X, K), K + M, v) \sim -\frac{1}{2v^2}\left(\log \frac{K + M}{C^{-1}(X, K)}\right)^2.$$

Then, also the logarithm of Call on Call prices must have a similar behavior, so that since $\log \frac{M}{X} \sim \log \frac{K+M}{C^{-1}(X,K)}$ the only possible case is $\hat{v}_K \in o\big(\sqrt{\log \frac{M}{X}}\big)$ and equating the asymptotic behaviors it follows $\hat{v}_K \sim v$. In particular, if it exists, the limit of $\hat{v}_K(X, M)$ at ∞ is equal to the limit of $v(C^{-1}(X, K), K + M)$.

Otherwise, if $v \sim c\sqrt{\log \frac{K+M}{C^{-1}(X,K)}}$ with $0 < c < \sqrt{2}$, then

$$\log C_{\mathrm{BS}}(X, M, \hat{v}_K) \sim -\frac{\log \frac{K+M}{C^{-1}(X,K)}}{8c^2}(2 - c^2)^2.$$

Now in order to have equivalent asymptotic behaviors, then there must be a positive $\hat{c} < \sqrt{2}$ such that $\hat{v}_K \sim \hat{c}\sqrt{\log \frac{M}{X}}$ and $\frac{(2-\hat{c}^2)^2}{\hat{c}^2} = \frac{(2-c^2)^2}{c^2}$. Solving, the only positive solution is $\hat{c} = c$, so that again $\hat{v}_K \sim v$.

□

Relation with the underlying implied volatility

In Lemma 7.1 we show that, in the homogeneous case, Calls on Calls' prices are increasing as functions of the relative underlying strike. Then, $\hat{C}_{K_1}(X, M) < \hat{C}_{K_2}(X, M)$ for every $K_1 < K_2$, which implies $\hat{v}_{K_1}(X, M) < \hat{v}_{K_2}(X, M)$. In particular, for $K_1 = 0$ and $K_2 = K$, it holds $\hat{C}_0(X, M) = C(X, M) = C_{\mathrm{BS}}(X, M, v(X, M))$ and $\hat{C}_K(X, M) = C_{\mathrm{BS}}(X, M, \hat{v}_K(X, M))$, so that looking at the implied total volatilities it follows $v(X, M) < \hat{v}(X, M)$. This means that the Calls on Calls's implied total volatility is always larger than the original implied total volatility for fixed moneyness.

7.5.2 Lifted Calls on Calls' implied volatility

We now look at the behavior of the smiles of lifted Calls on Calls. In this case, we expect $\hat{d}_2^{**}$ to be surjective since prices have derivative equal to -1 in 0. However, we will show that the smiles still explode for small strikes, while the behavior at ∞ is as for the original smile.

Lemma 7.20 (Asymptotic behavior of the lifted Calls on Calls' total implied volatility)**.** *The lifted Calls on Calls' total implied volatility $\hat{v}_K^{**}(X^{**}, M)$ behaves asymptotically as*

- $(2-\sqrt{2})\sqrt{-\log\frac{M}{X^{**}}}$ *for small strikes;*
- *the underlying total implied volatility* $v(C^{-1}(f(X^{**}),K),K+M)$ *for large strikes. If it exists, the limit of* $\hat{v}_K^{**}(X^{**},M)$ *is equal to the limit of* $v(C^{-1}(f(X^{**}),K),K+M)$.

Proof. It holds $-\partial_M\hat{C}_K^{**}(X^{**},M) = -\frac{X^{**}}{f(X^{**})}\partial_M\hat{C}_K(f(X^{**}),M)$ and $-\partial_M\hat{C}_K(f(X^{**}),M) = 1-P(f(X^{**})=0) = \frac{f(X^{**})}{X^{**}}$, then lifted Calls on Calls' prices have derivative equal to -1 at null strikes. In particular, the Lee left wing condition holds: $\hat{v}_K^{**}(X^{**},M) < \sqrt{-2\log\frac{M}{X^{**}}}$ for small strikes.

From the Put-Call-Parity, it holds

$$P_{\mathrm{BS}}^{**}(X^{**},M,\hat{v}_K^{**}) = \frac{X^{**}}{f(X^{**})}C(C^{-1}(f(X^{**}),K),K+M) - X^{**} + M \tag{7.10}$$

where we dropped the dependence of the implied total variance from the underlier and the strike for notation simplicity. For M going to 0, since both $\hat{d}_1^{**}$ and $\hat{d}_2^{**}$ go to ∞, the left hand side behaves as

$$\begin{aligned} P_{\mathrm{BS}}^{**}(X^{**},M,\hat{v}_K^{**}) &\sim -X^{**}\frac{\phi(\hat{d}_1^{**})}{\hat{d}_1^{**}} + M\frac{\phi(\hat{d}_2^{**})}{\hat{d}_2^{**}} \\ &= X^{**}\phi(\hat{d}_1^{**})\frac{\hat{v}_K^{**3}}{\left(\log\frac{M}{X^{**}}\right)^2 - \frac{\hat{v}_K^{**4}}{4}}. \end{aligned}$$

As in the proof of Lemma 7.19, we take the logarithm and consider the dominant terms, so that

$$\log P_{\mathrm{BS}}^{**}(X^{**},M,\hat{v}_K^{**}) \sim -\frac{\hat{d}_1^{**2}}{2}.$$

On the other hand, the right hand side in Equation (7.10) is equal to

$$\frac{X^{**}}{f(X^{**})}\big(C(C^{-1}(f(X^{**}),K),K+M) - f(X^{**})\big) + M$$

and, since $C(C^{-1}(f(X^{**}),K),K) = f(X^{**})$, for M going to 0 the above expression behaves as

$$\frac{X^{**}}{f(X^{**})}\big(\partial_K C(C^{-1}(f(X^{**}),K),K)M + \partial_K^2 C(C^{-1}(f(X^{**}),K),K)M^2\big) + M.$$

Observe that $\frac{X^{**}}{f(X^{**})} = \frac{1}{-\partial_K C(C^{-1}(f(X^{**}),K),K)}$ so that the above expression reduces to

$$\frac{X^{**}}{f(X^{**})}\partial_K^2 C(C^{-1}(f(X^{**}),K),K)M^2$$

where the term multiplying M^2 is a positive constant. Then, taking the logarithm, this behaves as $2\log M$.

Now, if $\hat{v}_K^{**} \in o\big(\sqrt{-\log \frac{M}{X^{**}}}\big)$ then

$$\log P_{\mathrm{BS}}^{**}(X^{**}, M, \hat{v}_K^{**}) \sim -\frac{1}{2\hat{v}_K^{**2}} \left(\log \frac{M}{X^{**}}\right)^2.$$

and equating this with $2 \log M$ we find $\hat{v}_K^{**} \sim \frac{\sqrt{-\log M}}{2}$, which implies $\hat{v}_K^{**} \notin o\big(\sqrt{-\log \frac{M}{X^{**}}}\big)$, so that this solution cannot be accepted. If $\hat{v}_K^{**} \sim c^{**}\sqrt{-\log \frac{M}{X^{**}}}$ with $c^{**} < \sqrt{2}$ then

$$\log P_{\mathrm{BS}}^{**}(X^{**}, M, \hat{v}_K^{**}) \sim \frac{\log \frac{M}{X^{**}}}{8c^{**2}}(2 + c^{**2})^2$$

and equating with $2 \log M$ we find that the only admissible solution is $c^{**} = 2 - \sqrt{2}$.

Regarding the limit for large strikes, the definition of $\hat{C}_K^{**}$ implies that lifted Calls on Calls vanish at ∞ iff the underlying Calls do. Then, d_1^{**} is surjective iff d_1 is. So that if d_1 has a finite limit at ∞, then $\hat{v}_K^{**}$ must explode. Otherwise, as in the relative Calls on Calls' case, the Lee right wing condition must hold: $\hat{v}_K^{**}(X^{**}, M) < \sqrt{2 \log \frac{M}{X^{**}}}$ for large enough M. Similarly to the previous section, considering the relation

$$C_{\mathrm{BS}}(X^{**}, M, \hat{v}_K^{**}) = \frac{X^{**}}{f(X^{**})} C_{\mathrm{BS}}(C^{-1}(f(X^{**}), K), K + M, v)$$

and developing for large M, we obtain that $\hat{v}_K^{**}(X^{**}, M)$ at ∞ behaves as $v(C^{-1}(f(X^{**}), K), K + M)$. In particular, if the limit exists, this is equal to the limit of $v(C^{-1}(f(X^{**}), K), K + M)$ at ∞.

□

7.5.3 Examples

Black-Scholes Calls on Calls' implied volatility

In the Black-Scholes case, the implied total volatility is constant. In Lemmas 7.19 and 7.20 we showed that both the Calls on Calls' implied total volatility and the lifted Calls on Calls' implied total volatility explode for $M = 0$ and go to the original Black-Scholes total implied volatility at $M = \infty$.

On the left of Figure 7.4 we plot the total implied volatilities $\hat{v}_K(X, M)$ and $\hat{v}_K^{**}(X, M)$ as functions of M. We take $X = \mathrm{BS}(S, K, v)$ where $S = 100$, $K = 110$ and $v = 0.2$.

SVI Calls on Calls' implied volatility

We consider now the SVI model as in Section 7.3.2. For K going to 0, the corresponding implied total volatility $\sqrt{\mathrm{SVI}(\log \frac{K}{S})}$ behaves as $\sqrt{b(1-\rho)|\log \frac{K}{S}|}$. Except for a constant, this is equivalent to the behavior of $\hat{v}_K(X, M)$ and $\hat{v}_K(X^{**}, M)$ in 0. Also, for M going to ∞, the three total implied volatilities behave similarly, exploding with a speed of $\sqrt{b(1+\rho)\log \frac{K+M}{S}}$.

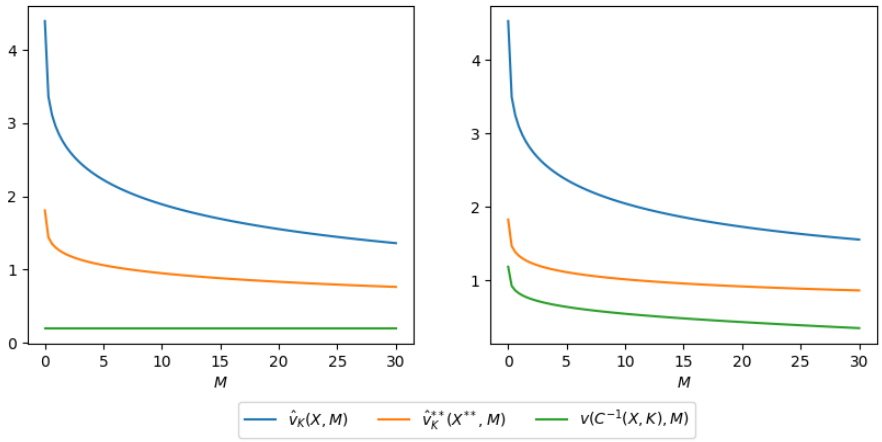

Figure 7.4: The hidden smiles $\hat{v}_k(X, M)$ and $\hat{v}_k^{**}(X^{**}, M)$ of the Black-Scholes model (left) and the smiles obtained from SVI prices (right). The current underlier has value 100*, the original strike is set at* 110*, the original Black-Scholes implied total volatility at* 0.2*, and the original SVI parameters at* $a = 0.01$*,* $b = 0.1$*,* $\rho = -0.6$*,* $m = -0.5$*,* $\sigma = 0.1$*.*

We show on the right of Figure 7.4 the total implied volatilities $\hat{v}_K(X, M)$ and $\hat{v}_K^{**}(X^{**}, M)$ as functions of M. We set $X = \mathrm{BS}(S, K, \mathrm{SVI}(\log \frac{K}{S}))$ where $S = 100$, $K = 110$ and the SVI parameters $a = 0.01$, $b = 0.1$, $\rho = -0.6$, $m = -0.5$, $\sigma = 0.1$ are taken as in Section 7.3.2 in order to guarantee arbitrage-free prices.

Part II

Risk metrics and margin computations for CCPs

Chapter 8

A closed form model-free approximation for the Initial Margin of option portfolios

Abstract

Central clearing counterparty houses (CCPs) play a fundamental role in mitigating the counterparty risk for exchange traded options. CCPs cover for possible losses during the liquidation of a defaulting member's portfolio by collecting initial margins from their members. In this chapter we analyze the current state of the art in the industry for computing initial margins for options, whose core component is generally based on a VaR or Expected Shortfall risk measure. We derive an approximation formula for the VaR at short horizons in a model-free setting. This innovating formula has promising features and behaves in a much more satisfactory way than the classical Filtered Historical Simulation-based VaR in our numerical experiments. In addition, we consider the neural-SDE model for normalized Call prices proposed by [Cohen et al., arXiv:2202.07148, 2022] and obtain a quasi-explicit formula for the VaR and a closed formula for the short term VaR in this model, due to its conditional affine structure.

From:
C. Martini and **A. Mingone**, *A closed form model-free approximation for the Initial Margin of option portfolios*, arXiv preprint https://arxiv.org/abs/2306.16346, 2023.

8.1 Structure of the chapter

In the first part of this chapter, we look at the mechanism of options' initial margin adopted by CCPs in Section 8.2. In Section 8.3 we go into detail in the practical implementations used by CCPs to calculate initial margins, followed by an assessment of their pros and cons.

In the second part of the chapter, we firstly describe in Section 8.4 the short-term model-free formula that we advance as short-term VaR formula for option portfolios, and show that it coincides with the exact one in the Stochastic Volatility model. Secondly in Section 8.5, we derive the closed margin formula in the neural-SDE model for normalized option prices proposed in [18]. In particular, we will see that the latter closed formula has the form of our short-term model-free formula for small time steps. We conclude the chapter by performing numerical experiments in Section 8.6, where plots show the more regular and promising behavior of our short-term model-free formula compared to the classic FHS VaR.

8.2 The mechanism of initial margin for options

CCPs charge clearing members, on a daily or intra-daily basis, with total risk requirements that are computed from initial margins. The initial margin aims at covering possible losses in the portfolio value when liquidating it after a default, under normal market conditions, and it is estimated considering a tail risk.

Consider a portfolio, at time t, with possibly both long $(L_i)_i$ and short $(S_j)_j$ option positions (both L_i and S_j are positive) with different strikes and expiries. In the case of default at time t, the CCP has to liquidate the portfolio in a Margin Period of Risk (MPOR) of say h days (h is usually 2 days for exchange-traded options). At date $t+h$, the portfolio could have undergone market movements, so that the CCP has to estimate its payoff after liquidation.

The initial margin (IM) is then the Value-at-Risk (VaR) or Expected-Shortfall (ES) at a confidence level of generally 0.99 of the portfolio predicted P&Ls:

$$\mathrm{IM}(t) = -\mathrm{VaR}_{0.99}\big(\mathrm{P\&L}(t+h)\big)$$

where the minus sign ensures a positive margin value.

At this point, the total risk requirement charged by the CCP does not solely include the initial margin. Indeed, the CCP eventually adds to the latter quantity some add-ons to take into account risks that are not directly related to market moves. Among these, we typically find the Wrong Way Risk add-on, the liquidity and concentration risk add-on and possibly other specific add-ons:

$$\text{Total margin}(t) = \mathrm{IM}(t) + \text{Add-ons}(t).$$

Now, the total margin is floored by the Short Option Minimum (SOM). Deep short OTM positions have very little risk since they will probably stay OTM along the MPOR. However, their extreme risk is still not 0 and the methodology should be able to capture it. This is generally not the case for strikes very far from the ATM, because of the lack of historical liquid data on these strikes. Then, to assure an enough conservative margin, a secure floor should be applied to the risk requirement. The SOM is generally the sum along all short positions of the calibrated extreme costs for these options:

$$\text{Refined total margin}(t) = \max\big(\text{Total margin}(t);\ \mathrm{SOM}(t)\big).$$

At this point, the final total risk requirement is the refined total margin adjusted by two other terms: minus the Net Option Value (NOV) on equity-style options (options for which the premium is paid in full at the settlement date, i.e. one or two days after the option trade) and the Unpaid Premium (UP), described below:

$$\text{Total risk requirement}(t) = \max\big(\text{Refined total margin}(t) - \text{NOV}(t) + \text{UP}(t);\ 0\big)$$

where the floor by zero is to avoid that the CCP pays and

$$\begin{aligned}\text{NOV}(t) &= \sum_i L_i O_i(t) - \sum_j S_j O_j(t)\\ \text{UP}(t) &= \sum_{i\ \text{unpaid}} L_i O_i(t) - \sum_{j\ \text{undelivered}} S_j O_j(t)\end{aligned}$$

with O denoting the option prices.

For the NOV, let us consider for instance the case of a short option position, where the option is of equity-style. In the case of a default, the liquidation of this position would require to buy the option in the market, which amounts to the CCP paying the option price at the time of liquidation. This means that the initial margin for a short option position should aim to cover the largest option price, up to a fixed confidence level. On the other hand, the liquidation of a long option position will always result in a positive inflow for the CCP, because the CCP will sell the long option position and receive the option price.

The reason why the CCP applies the NOV can alternatively be explained observing that the liquidation at the end of the MPOR, at time $t+h$, will result for the CCP in the monetary flow:

$$\text{Liquidation P\&L}(t+h) = \sum_i L_i O_i(t+h) - \sum_j S_j O_j(t+h).$$

The Liquidation P&L can be expressed as the sum of the NOV and the portfolio's value increment:

$$\begin{aligned}\text{Liquidation P\&L}(t+h) &= \Big(\sum_i L_i O_i(t) - \sum_j S_j O_j(t)\Big) +\\ &\quad + \Big(\sum_i L_i\big(O_i(t+h) - O_i(t)\big) - \sum_j S_j\big(O_j(t+h) - O_j(t)\big)\Big)\\ &= \text{NOV}(t) + \text{P\&L}(t+h).\end{aligned}$$

Then, the CCP has to charge to the clearing member minus the liquidation profit, i.e. the predicted losses (the initial margin appropriately adjusted by the add-ons and the SOM) minus the NOV on equity-style options.

The UP is charged by the CCP to cover from the risk of default of the counterparts before the settlement date of the option premium, and it corresponds to the net position of accrued option premiums which are still unpaid (because the settlement date has not

passed yet). In this way, the difference between the UP and the NOV can be seen as a (Contingent) Variation Margin for Options (VMO) not yet settled.

Consider a defaulting clearing member which is long an option before the settlement date. The CCP will need to pay the premium to its counterpart in the trade, and this will be done re-selling the option and collecting its new premium, and using the VMO previously required to the buyer. This latter component is needed to account for the difference between the initially established option premium and today's one. Similarly, the CCP has to liquidate defaulting short positions on not yet settled options buying the option and delivering it to the buyer counterpart. To do so, it will use the money from the buyer plus the VMO from the defaulting seller.

All in all, the final formula for the total risk requirement is

$$\text{Total risk requirement}(t) = \max\big(\max\big(\text{IM}(t)+\text{Add-ons}(t);\ \text{SOM}(t)\big)-\text{NOV}(t)+\text{UP}(t);\ 0\big).$$

In this chapter we will focus on the IM component of the total risk requirement.

8.3 Initial margin for options in the industry: a short survey

The total risk requirement mechanism and its different layers is essentially the same across all CCPs, with possible differences in wording. What really makes the difference among CCPs' requirements is the way the IMs (and the add-ons) are computed. A notorious parametric model for margining has been proposed by CME Group under the name of SPAN.[1] It consists in computing the P&Ls of the portfolio under different risk scenarios depending on the combination of underlying price changes and implied volatility changes. A similar model has been implemented by ICE with the name of IRM. These models are particularly tricky and overconservative, and for these reasons nowadays CCPs are passing to new models. In particular, both SPAN and IRM models have been upgraded to the corresponding SPAN2[2] and IRM2[3] models, which both use the Filtered Historical Simulation (FHS) techniques to create risk scenarios. Indeed, the majority of CCPs is now adopting the FHS to compute IMs for option portfolios.

8.3.1 Filtered Historical Simulation

The FHS has recently become the standard approach for VaR computations among CCPs, especially on cash equity markets. The FHS technique is indeed particularly efficient in cash equity and fixed income markets for spot instruments, but it becomes more subtle in derivatives clearing.

The FHS model is particularly appreciated since it is essentially data-driven and model-free, and it relies on few requirements to be satisfied. For a given instrument

[1] https://www.cmegroup.com/clearing/risk-management/span-overview.html

[2] https://www.cmegroup.com/clearing/risk-management/span-overview/span-2-methodology.html

[3] https://www.theice.com/clearing/margin-models/irm-2/methodology

to be cleared, firstly the CCP must choose the risk factors which drive its price; let r_s denote their returns, either logarithmic, absolute, or relative depending on the risk factor. A key property that scaled returns must satisfy is stationarity (see [7]). Indeed, the FHS model relies on the hypothesis that risk factors' returns at tomorrow's date $t+1$ behave as

$$r_{t+1} = \eta\sigma_{t+1}$$

where σ_{t+1} is the returns' simulated conditional volatility at day $t+1$ and η is drawn from the historical observations

$$\eta_s = \frac{r_s}{\sigma_s}.$$

In other words, the past historical return is *re-contextualized* to the current volatility context by the FHS devoling/revoling steps.

Generally, the industry standard is to use an Exponentially Weighted Moving Average (EWMA) variance estimator for the volatility. A EWMA volatility with decay factor λ is computed as

$$\mathrm{EWMA}_s = \sqrt{(1-\lambda)(r_s)^2 + \lambda \mathrm{EWMA}_{s-1}^2},$$

with an eventual flooring in case of too low values. Then, the historical volatility σ_s used to scale historical returns can be calculated with two possible formulations: $\sigma_s = \mathrm{EWMA}_s$ and $\sigma_s = \mathrm{EWMA}_{s-1}$ respectively. The two alternatives are discussed in [42], section 7.1, where it is acknowledged that they will lead to significantly different outcomes.

When computing the IM for portfolios of options using the FHS methodology, the CCP has to choose a set of risk factors, assessing the stationarity property in particular. Together with the underlying value (and possibly the interest/repo rate), also the Implied Volatility (IV) has to be taken as a risk factor. Since the IV is actually a surface which behaves differently depending on the strike and the maturity of the option, two alternatives can be considered in order to generate IV scenarios:

1. Identify a fixed two-dimensional grid for the IV surface and define each point as a risk factor.

2. Choose a model for option prices and take its parameters as risk factors.

For deeper insights on the VaR computation for options in a FHS approach see [40].

Implied Volatility anchor points

In the first alternative, the anchor points on the grid can be chosen with fixed time-to-maturity or fixed rolling index as first coordinate, and fixed log-forward moneyness, or fixed delta, or (equivalently) fixed ratio between log-forward moneyness and square-root of time-to-maturity as second coordinate. Since market data is more dense around the ATM point for shortest maturities and spreads out for increasing maturities, the fixed

delta grid is generally preferred. Indeed, it implies a grid in log-forward moneyness with a triangular shape, with a range that starts from the ATM point and spreads out as the time-to-maturity increases.

If choosing a dense grid guarantees a more precise fit of IVs in between the anchor points, it highly decreases computation performances and makes it difficult to identify general historical patterns in the IV surface dynamics. For this reason, Principal Component Analysis (PCA) can be performed in order to model the shifts of the surface. To cite an example, in [68], the authors test the FHS method on the (PCA) principal components in a Karhunen-Loève decomposition and find that scenarios satisfy the conditions of no Butterfly arbitrage (i.e. the requirement that for a fixed maturity, Call prices must be non-increasing and convex with respect to the strike).

Once scenarios are generated on the anchor points, the model still needs an interpolation/extrapolation criterion to predict future prices on points outside the grid. The criterion could be either a model for the implied volatility (such as SVI) or for prices (such as SABR), which has to be calibrated from the scenarios grid, or classic interpolations via b-splines. The choice can be driven by arguments of non-arbitrability of prices, or of best fit and computation efficiency of the algorithm.

Implied Volatility models

In the second alternative, the CCP chooses a pricing model for options and, once the stationarity property on the model parameters' returns is verified, performs an FHS on the model parameters.

As an example, the SABR model is an industry standard and it is driven by three parameters α, β and ρ. Generally, the β parameter is fixed a priori based on historical observations, so that only the α and ρ parameters need to be estimated. After showing the stationarity of their returns in the target market, the CCP can apply the FHS technique on the historical observations of α and ρ, and use their drawn values to simulate future prices.

Similarly, the Stochastic Volatility Inspired (SVI) model by Gatheral is largely used among CCPs, and also among crypto funds, to model the implied total variance. Its sub-model Slices SVI (SSVI) is sometimes preferred since it has more tractable arbitrage-free requirements and since it still fits data pretty well. SSVI has three parameters θ, φ and ρ per each maturity, so that if the stationarity of their returns is verified, the FHS technique can be applied to obtain simulated prices. An example of this application can be found in [40].

Lastly, a model which is sometimes considered is the Gaussian lognormal mixture model as described by Glasserman and Pirjol in [37]. It consists in a convex combination of Black-Scholes prices and the number of parameters depends on chosen number of basis prices. Even though it is easy to implement, it guarantees no arbitrage for slices and it has very good fitting ability, the model is not easy to extend to full surfaces and it hides potential issues when extrapolating in extreme events. Indeed, it has the theoretical property to have the same constant asymptotic level in the two wings of the smile (Proposition 5.1 of [37]), so that while the calibration of market smiles could suggest

a decreasing shape, the calibrated smile with a lognormal Gaussian mixtures would necessarily increase at large strikes, a pure model artefact. As a consequence, while calibration fit could be good for liquid market data (concentrated around the ATM point), in contexts such as the computation of tail risks as in margins, the extrapolation at extreme strikes would be misleading in those circumstances.

Limitations of the FHS

The FHS technique works well when a large history of risk factors is stored, which is in fact a first immediate practical limitation. Indeed, the possible number of scenarios for the FHS cannot be larger than the available history, since the normalized returns are drawn from past observations.

A second important drawback of the FHS methodology for complex products is the capture of the joint dynamics of risk factors. Indeed, in the FHS model, risk factors are re-scaled according to their own intrinsic volatility, without any reference to other risk factors and this may cause a discrepancy in the relationships between the risk factors, and in particular in their correlations, as explained in section 7.2 of [42]. Furthermore, while the property of stationarity of the normalized returns of single assets is generally historically satisfied, this is more hardly the case for the returns of IV points or model parameters, resulting in more unstable and unnatural results for the FHS methodology.

Thirdly, FHS is relatively straightforward to implement, as far as the risk factors under study do not have structural relationships which could be destroyed by the core FHS algorithm. Unfortunately, this is exactly the case for futures' curves and implied volatility surfaces.

Indeed, in the case of futures' curves, the FHS simulation considers a set of fixed pillars (i.e. futures' time-to-maturities) of the curves today and apply the re-scaled corresponding past returns. For each simulation, the simulated vector of futures values on the fixed pillars should be consistent between the spot returns and the future returns. However, this consistency is not guaranteed by the FHS simulations.

Similarly, when using the IV anchor points as risk factors, even if the calibrated volatility surfaces are perfectly calibrated and arbitrage-free, the volatility surfaces obtained by an FHS procedure have no reason to be arbitrage-free in turn (and in general will not be, because arbitrage-free surfaces do not have nice additive or multiplicative properties). Furthermore, IV anchor points returns are generally considered in absolute terms, which could cause negative simulated implied volatilities. Flooring the latter quantities to 0 is not a good choice, since: 1) prices for zero volatility are always strictly lower than the market prices for European options; 2) since Call option prices are decreasing functions of the strike, a zero volatility for an OTM option implies that all the options of same type with the same maturity and larger strikes should also have a zero volatility, so that also simulated implied volatility smiles should satisfy this property.

The possibility of generating scenarios such that each matrix of prices indexed by the moneyness and time-to-maturity grid is arbitrage-free is essentially an open question. A recent article [24] describes a weighted Monte-Carlo algorithm which penalizes arbitrageable scenarios to obtain arbitrage-free simulations with higher probability. We

explain the model in Section 8.3.3. Another alternative is to use parametric models of IV surfaces, for which no-arbitrage conditions are available, and work at the level of the parameters of such models. Yet, randomizing the model parameters may produce a lot of instability. A noteworthy attempt is the neural-SDE model of [18], that we investigate further below, which provides a consistent framework for this purpose.

Finally, the FHS methodology is known to be procyclical as shown for example in section 6 of [42]. Procyclicality has to be avoided because it implies margins which react too abruptly to market changes, and this may cause liquidity issues to the clearing members who have to post the corresponding collateral.

8.3.2 The procyclicality control by Wong and Zhang (Options Clearing Corporation)

Even though the FHS model is the most popular among CCPS, in the recent years some new models for the options clearing are born and CCPs are starting to look at these alternatives. An important feature in margin requirements that CCPs should always try to mitigate is procyclicality and we have seen that FHS does not properly satisfy this requirement. Article 28 of [21] is dictated to the procyclicality control, detailing specific actions that CCPs have to adopt for its limitation.

With this in mind, Wong and Zhang from the Options Clearing Corporation (OCC) choose a model for options initial margin (see [67]) that guarantees to control procyclicality thanks to a dynamic scaling factor that behaves as an inverted S-curve and is higher during low-volatility periods and lower during high-volatility ones.

The model specifies the log-returns of the ATM IV at expiry T_j by

$$\log \frac{\sigma_{t+h}(T_j, F_t(T_j))}{\sigma_t(T_j, F_t(T_j))} := \eta_t \Big(\frac{T_j}{T_1}\Big)^{-\alpha} \sqrt{h} z_t \tag{8.1}$$

where z_t is a normalized innovation, centered with unit variance, $F_t(T_j)$ is today's forward for maturity T_j, and T_1 is the first quoted expiry. The factor η_t in turn is a dynamic rescaling of the CBOE VVIXSM (VVIX), in particular

$$\eta_t = D(\sigma_t)\mathrm{VVIX}_t$$

where σ_t is the S&P500 ATM IV of the short-term expiry (or any reference expiry, like the one-month), and the scaling factor $D(\sigma_t)$ is a sigmoid function, which models a state transition from a risk point of view:

$$D(\sigma_t) = L + \frac{H}{1 + \exp\left(\kappa(\sigma_t - \sigma^*)\right)}.$$

Here L is the minimum of the ratio between the long-term mean of the historical vol-of-vol and the VVIX, H is the difference between the maximum and the minimum of the latter ratio, κ is the growth rate of the curve, and σ^* is the sigmoid's midpoint.

The IV surface is recovered from the ATM IV dynamics considering the second order approximation in $\log \frac{K}{F_t(T)}$:

$$\sigma_t(T,K) \approx \sigma_t(T,F_t(T)) + \Sigma_t(T) \log \frac{K}{F_t(T)} + C_t(T)\Big(\log \frac{K}{F_t(T)}\Big)^2,$$

where $\Sigma_t(T)$ and $C_t(T)$ are respectively the ATM skew and the ATM curvature.

In this way, knowing the distribution of z_t allows to perform simulations of implied volatility surfaces and to compute an empirical VaR.

Observe that the dynamics of the implied volatility in Equation (8.1) are modeled for fixed strike and expiry, i.e. for a fixed contract. This differs with the majority of other models, whose dynamics are defined for fixed time-to-maturity and log-forward moneyness.

8.3.3 Arbitrage-free simulations for options

When computing an IM, the priority of the CCP is to be conservative enough to cover for members' defaults, while not requiring too high margins to keep its competitiveness in the market and avoiding procyclicality. For this reason, arbitrage-free requirements are not necessarily taken into account as seen for the FHS methodology. However, simulating reliable scenarios (and so scenarios with no arbitrage) allows to estimate more plausible margins, and avoids the pitfall of paying for implausible scenarios.

The article [24] describes a cunning way to compute an empirical VaR tweaking the simulations from any model in favor of arbitrage-free simulations. The arbitrage considered in the article is the static arbitrage, that, in case of options, can arise in both the direction of time-to-maturity and the direction of log-forward moneyness. Arbitrage-free Call prices should:

1. lie between the discounted intrinsic value (computed with respect to the forward) and the discounted forward;
2. increase in time-to-maturity at a fixed moneyness;
3. decrease in log-forward moneyness at a fixed time-to-maturity;
4. be a convex function of the log-forward moneyness.

Note that in the article, the authors only address the last three points, but the methodology can be easily extended to include the first one.

Per each arbitrage situation, a penalization function is defined, depending only on the normalized Call prices surface on a fixed time-to-maturity and log-forward moneyness discrete grid. Penalization functions are null in case of no arbitrage and increase their value with increasing arbitrageable grid points. The target arbitrage penalty function is the sum of the three penalization functions, and it is null if and only if the discrete Call prices are free of arbitrage.

At this point, the VaR calculation algorithm is straightforward:

1. Simulate scenarios using the chosen initial model.
2. Per each simulated scenario:
   - evaluate the arbitrage penalty function;
   - compute its weight inversely proportional to the arbitrage penalty function.
3. Compute empirical VaR under the probability measure resulting from weights.

Since weights prioritize arbitrage-free scenarios, the VaR calculation hangs towards more reliable and possible values.

The methodology holds for any model that simulates scenarios. It can then be applied to both FHS and Monte-Carlo simulation models. In particular, the authors apply it to a non-parametric generative model for implied volatility surfaces called VolGAN (section 6 of [24]).

8.3.4 The neural-SDE model

In [19], Cohen at al. show very good empirical results on options' VaR estimation. The results are based on a specific model that the authors introduce in [18], which consists in a representation of normalized Call prices via non-random linear functions of some risk factors ξ_t. The articles focus on how to calibrate and consequently generate arbitrage-free Call prices surfaces via neural networks for the dynamics under consideration.

In the neural-SDE model, the normalized Call prices (i.e. Call prices divided by the forward and discount factor) are affinely decomposed into time-independent non-random surfaces G_i and time-dependent stochastic combining factors $\xi_{t,i} \in \mathbb{R}^d$:

$$\begin{aligned} c_t(\tau, k) &= G_0(\tau, k) + G(\tau, k) \cdot \xi_t \\ &= G_0(\tau, k) + \sum_{i=1}^{d} G_i(\tau, k)\xi_{t,i} \end{aligned} \tag{8.2}$$

where τ is the time-to-maturity and k is the log-forward moneyness. The underlying asset S_t and the time-dependent vector ξ_t evolve as

$$\begin{aligned} dS_t &= \alpha(\xi_t)S_t dt + \beta(\xi_t)S_t dW_{0,t} \quad & S_0 &= s_0 \in \mathbb{R}, \\ d\xi_t &= \mu(\xi_t)dt + \sigma(\xi_t) \cdot dW_t & \xi_0 &= \zeta_0 \in \mathbb{R}^d, \end{aligned} \tag{8.3}$$

where $W_0 \in \mathbb{R}$, $W = (W_1, \ldots, W_d)^T \in \mathbb{R}^d$ are independent standard Brownian motions under the real-world measure P, and the hypothesis for the existence and uniqueness of the processes hold, i.e. $\alpha(\xi_t) \in L^1_{\text{loc}}(\mathbb{R})$, $\mu(\xi_t) \in L^1_{\text{loc}}(\mathbb{R}^d)$, $\beta(\xi_t) \in L^2_{\text{loc}}(\mathbb{R})$, $\sigma(\xi_t) \in L^2_{\text{loc}}(\mathbb{R}^{d\times d})$. Observe that α, β, μ and σ are deterministic functions.

Starting from these assumptions, the factors are decoded using different PCA-based techniques to also ensure that the reconstructed prices are more likely to be arbitrage-free both in a static and in a dynamic sense. Absence of dynamic arbitrage is ensured

through Heath-Jarrow-Morton-type conditions while absence of static arbitrage is ensured by imposing that each discretized normalized Call prices' surface satisfies a set of linear conditions $Ac \geq b$ for some matrix A and vector b (see [17]). Notice that since the decomposition of the normalized Call prices is affine and the no static arbitrage conditions are linear, it is possible to rewrite the latter conditions for ξ_t as $A \cdot G \cdot \xi_t \geq b - A \cdot G_0$.

Given the history of market Call prices, the factors G_i can be calibrated for every grid point (τ_j, k_j) and factors $\xi_{s,i}$ for every past day s under the no arbitrage constraints.

After the factors decoding, Cohen at al. set up a supervised learning process to estimate $\alpha(\xi_t), \beta(\xi_t), \mu(\xi_t), \sigma(\xi_t)$ via a maximal likelihood function which ensures that the time series for ξ_t evolves inside the convex polytope generated by the no static arbitrage conditions.

Empirical VaR in the neural-SDE model

From now on we denote with α_t the process $\alpha(\xi_t)$ and similarly for β_t, μ_t and σ_t.

Suppose we want to compute the VaR of a portfolio constituted of Call options at MPOR date $t + h$, where $h = n\delta t$ and δt is the one day unit.

Having the drift and diffusion functions for the time series of ξ_t and the underlier from the model calibration, predictions can be made with an Euler scheme in a Monte-Carlo fashion. In particular (with an abuse of notation), processes values for the first step at $t + \delta t$ are

$$\begin{aligned} S_{t+\delta t} &= S_t \exp\left(\left(\alpha_t - \frac{\beta_t^2}{2}\right)\delta t + \beta_t(W_{0,t+\delta t} - W_{0,t})\right), \\ \xi_{t+\delta t} &= \xi_t + \mu_t \delta t + \sigma_t(W_{t+\delta t} - W_t), \end{aligned}$$

where

$$\begin{aligned} W_{0,t+\delta t} - W_{0,t} &= \sqrt{\delta t} X_0, \\ W_{t+\delta t} - W_t &= \sqrt{\delta t} X, \end{aligned}$$

with independent standard Gaussian random variables $X_0 \in \mathbb{R}$, $X \in \mathbb{R}^d$.

In order to guarantee more stability of simulations, a tamed Euler scheme can also be implemented.

The following steps are performed as above, using the latest values of $S_\cdot$ and $\xi_\cdot$. At each step, new parameters $\alpha_\cdot$, $\beta_\cdot$, $\mu_\cdot$ and $\sigma_\cdot$ can be estimated using the neural network algorithm implemented in [18].

Alternatively, assuming $\alpha_\cdot$, $\beta_\cdot$, $\mu_\cdot$ and $\sigma_\cdot$ to be constant between t and $t+h = t+n\delta t$, simulations for S_{t+h} and ξ_{t+h} can be faster computed as

$$\begin{aligned} S_{t+h} &= S_t \exp\left(\left(\alpha_t - \frac{\beta_t^2}{2}\right)h + \beta_t(W_{0,t+h} - W_{0,t})\right), \\ \xi_{t+h} &= \xi_t + \mu_t h + \sigma_t(W_{t+h} - W_t). \end{aligned} \tag{8.4}$$

The predicted values of ξ_{t+h} can then be used to compute predicted values of normalized Call prices using Equation (8.2), which can be de-normalized using the predicted values of S_{t+h}.

The number of simulations that can be performed is arbitrary, so that a stable value of the VaR can be computed as the empirical quantile of simulated Call prices.

Limitations of the neural-SDE model

In this calibration routine of the neural-SDE model, there is an important point which, according to us, should be taken into consideration in applications: the G parameters are calibrated on the history of market prices, but given their linear role in the normalized Call prices, there is little hope that a long history of Call prices will be well explained by the very same G factors. Indeed, normalized Call prices in this model are random linear combinations of *fixed* surfaces, so that one should probably expect the Call prices to maintain these fixed parameters for no more than a typical period of one month or so, after which they should be re-calibrated. In [18], the G parameters are calibrated on a 17-years history, which might be far from being realistic in practice. As a consequence, calibration fit is not as good as in other more dynamic models. As an example, the mean absolute percentage error (MAPE) computed by the authors in Table 2 of [19] using two components of ξ_t is around 4.61% and 5.40%, while in our tests limiting the calibration window of G to one month reduces the MAPE to about 1.5 percentage points. On the other hand, it is not possible to simply calibrate the G parameters on shorter periods of past history, since then the neural-SDE on the ξ_t factors cannot be properly trained to estimate the model parameters, given the too low amount of historical data.

To some extent, there is therefore a trade-off between the stationarity of the model and its relevance - note though that one could argue that this is a general situation for any model.

Furthermore, this stationarity of parameters is likely to be related to the low procyclicality of the obtained VaR estimations that the authors claim: because the G parameters are the same since several years, the initial margin is indeed automatically less reactive to market changes.

This being said, the neural-SDE model provides a consistent and tractable framework which seems to us very promising.

8.3.5 The market data in input of the margin computation, and Market/Model add-ons

The models described above for margin computations (FHS, arbitrage-free GANs, neural-SDE) have all in common the generation of scenarios for the risk factors. In the case of options, these scenarios can only be generated after an initial calibration of market prices using any internal model, calibration which will then be reversed to get simulated prices. Indeed, the CCP needs a model and/or an interpolation scheme to get prices at any time-to-maturity and log-forward moneyness, and this scheme is used since the beginning of the IM computation. As a result, margins are based on model prices (i.e. prices

calibrated/interpolated with the selected scheme), rather than market prices, and should then be adjusted by a term taking into consideration how the initial discrepancy between market and model prices propagates when computing the IM.

There can be two approaches to this issue in the context of a VaR-type model:

1. Apply the scenarios to the calibrated model prices, thus obtaining shocked model prices, and assume that the model P&Ls are a good representative of the market P&Ls, along each scenario. This means that the calibration error is assumed to be the same at the current date and at the future date along the shocked scenario.

2. Compute a *Market/Model add-on*, which incorporates risk coming from the fact that the model which has been used to estimate the IM does not perfectly match market prices. Since the final risk requirement is computed on model prices and captures future movements of model prices, so that it could differ from the actual requirement needed for market prices, the Market/Model add-on estimates how large the difference between the market IM and the model IM is and adds it to the final requirement.

In the second approach, market P&Ls can be decomposed in 3 terms:

- the difference between the portfolio price under the calibrated model and its market price: $P_t^{\text{mod}} - P_t^{\text{mkt}}$;
- the difference between the portfolio model prices along the scenario s: $\tilde{P}_{t+h,s}^{\text{mod}} - P_t^{\text{mod}}$;
- the difference between the portfolio price under the calibrated model and its market price at the simulated date along the scenario s: $\tilde{P}_{t+h,s}^{\text{mkt}} - \tilde{P}_{t+h,s}^{\text{mod}}$.

The first of the 3 terms above is known and can be readily computed; the second term is computed in the IM; the third term depends on each scenario and upon the assumption on the distance between the market and model prices at the future simulated date along each scenario. The Market/Model add-on aims at covering this third source of risk.

8.4 A simple short-term model-free formula

In this section we describe a new short-term model-free formula for options VaR, which only depends on market data and does not need any model-specific calibration. The short-term attribute depends on the fact that approximations are performed in the MPOR component, so that the shorter the MPOR, the more precise is the formula.

In the following we work at today's time t, so that all past quantities up to time t (included) are observed. We denote with $\text{DF}_t(\tau)$ and $F_t(\tau)$ the discount factor and the forward value for time-to-maturity τ evaluated at time t. We work under the hypothesis of known constant rates between today date t and the MPOR date $t + h$, so that for a given time-to-maturity, discount factors are constant and forward values are proportional

to the underlier S_t. In particular we write $F_t(\tau) = f(\tau)S_t$. We call δt the one day unit and consider an MPOR $h = n\delta t$ of n days. Finally, we denote respectively by p_Y and F_Y the probability density function and the cumulative density function of a generic random variable Y. The cumulative density function and the probability density function of a standard Gaussian random variable are denoted with Φ and ϕ respectively. Also, when considering the distribution of the underlier S_{t+h} at time $t+h$, we actually mean the distribution conditional to quantities at time t (i.e. S_t and other risk factors ξ_t). If not otherwise specified, the coefficients of any SDE (such as α_t, β_t, μ_t, σ_t, η_t, ρ_t) are deterministic functions of the risk factors and time.

In the following sections we will always consider a portfolio of Vanilla Calls with price at time t given by

$$\Pi_t(S_t, \xi_t) = \sum_i \pi_i C\Big(T_i - t, \log \frac{K_i}{f(T_i - t)S_t}; S_t, \xi_t\Big)$$

where $C(\tau, k; S_t, \xi_t)$ is a generic model price of a Call with time-to-maturity τ and log-forward moneyness k, depending on the current value of the underlier S_t and of the other possible risk factors ξ_t (as for example the implied volatility in the short-term model-free case). The P&Ls are defined as the finite differences of the portfolio over the MPOR:

$$\text{P\&L} := \Pi_{t+h}(S_{t+h}, \xi_{t+h}) - \Pi_t(S_t, \xi_t).$$

The h-days VaR at confidence level θ (close to 1) of the portfolio is the quantity $v(\theta, h)$ such that

$$P\big(\text{P\&L} \leq v(\theta, h)\big) = 1 - \theta.$$

Sometimes we will need to develop the above expression using conditional probabilities. In particular, it holds

$$\begin{aligned} P\big(\text{P\&L} \leq v(\theta, h)\big) &= E[\mathbb{1}_{\text{P\&L} \leq v(\theta,h)}] \\ &= E\big[E[\mathbb{1}_{\text{P\&L} \leq v(\theta,h)} | S_{t+h}]\big] \\ &= \int_0^\infty P\big(\text{P\&L} \leq v(\theta, h) | S_{t+h} = s\big)\, dF_{S_{t+h}}(s). \end{aligned} \tag{8.5}$$

In the case of existence of a probability function for S_{t+h}, the latter expression can be written as

$$P\big(\text{P\&L} \leq v(\theta, h)\big) = \int_0^\infty p_{S_{t+h}}(s) P\big(\text{P\&L} \leq v(\theta, h) | S_{t+h} = s\big)\, ds.$$

8.4.1 The Black-Scholes case and the Stochastic Volatility case

Before introducing the short-term model-free VaR formula, we firstly look at some prototypical examples such as the Black-Scholes and the Stochastic Volatility cases.

In the classic Black-Scholes case, the underlier is a geometric Brownian motion

$$dS_t = \alpha_t S_t dt + \beta_t S_t dW_t$$

under the real-world probability measure. Applying Ito's lemma, portfolio prices are processes such that

$$d\Pi_t(S_t) = \left(S_t \alpha_t \frac{d}{dS_t}\Pi_t(S_t) + \frac{1}{2} S_t^2 \beta_t^2 \frac{d^2}{dS_t^2}\Pi_t(S_t) \right) dt + S_t \beta_t \frac{d}{dS_t}\Pi_t(S_t) dW_t.$$

Writing dW_t as $\sqrt{h}X$ where X is a standard Gaussian random variable and approximating the above expression at the first order we have that the P&Ls have the form

$$\Pi_{t+h}(S_{t+h}) - \Pi_t(S_t) \approx S_t \beta_t \frac{d}{dS_t}\Pi_t(S_t)\sqrt{h}X.$$

Then, it is easy to compute the VaR with a first order approximation:

$$P\big(\text{P\&L} \leq v(\theta, h)\big) = P\left(\frac{d}{dS_t}\Pi_t(S_t) X \leq \frac{v(\theta, h)}{S_t \beta_t \sqrt{h}} \right)$$

so that

$$v(\theta, h) = \Phi^{-1}(1-\theta) S_t \beta_t \left| \frac{d}{dS_t}\Pi_t(S_t) \right| \sqrt{h}.$$

The above reasoning can actually be generalized to Stochastic Volatility models where the volatility of the underlier is a stochastic process with volatility σ_t:

$$\begin{aligned} dS_t &= \alpha_t S_t dt + \xi_t S_t dW_{0,t} \\ d\xi_t &= \mu_t dt + \sigma_t dW_t \\ dW_{0,t} dW_t &= \rho_t dt. \end{aligned}$$

In the above formulation we have dropped the dependency of volatility parameters in the volatility itself, i.e. $\mu_t = \mu_t(\xi_t)$ and $\sigma_t = \sigma_t(\xi_t)$. Indeed, in order to guarantee the positivity of the volatility there must be such a dependency. Applying Ito's lemma to the portfolio $\Pi_t(S_t, \xi_t)$ of option prices generated by the pricing version of the Stochastic Volatility model, one finds

$$d\Pi_t(S_t, \xi_t) = a_t dt + S_t \xi_t \frac{d}{dS_t}\Pi_t(S_t, \xi_t) dW_{0,t} + \sigma_t \frac{d}{d\xi_t}\Pi_t(S_t, \xi_t) dW_t$$

where

$$\begin{aligned} a_t = \alpha_t S_t \frac{d}{dS_t}\Pi_t(S_t, \xi_t) + \mu_t \frac{d}{d\xi_t}\Pi_t(S_t, \xi_t) + \frac{\xi_t^2 S_t^2}{2} \frac{d^2}{dS_t^2}\Pi_t(S_t, \xi_t) + \\ + \frac{\sigma_t^2}{2} \frac{d^2}{d\xi_t^2}\Pi_t(S_t, \xi_t) + \xi_t S_t \sigma_t \rho_t \frac{d^2}{dS_t d\xi_t}\Pi_t(S_t, \xi_t). \end{aligned}$$

Considering the finite increments of the portfolio and neglecting linear terms for h going to 0, the form of the P&Ls becomes

$$\Pi_{t+h}(S_{t+h}, \xi_{t+h}) - \Pi_t(S_t, \xi_t) \approx S_t \xi_t \frac{d}{dS_t}\Pi_t(S_t, \xi_t)\sqrt{h}X_0 + \sigma_t \frac{d}{d\xi_t}\Pi_t(S_t, \xi_t)\sqrt{h}X,$$

where X_0 and X are standard jointly normal random variables with correlation ρ_t equal to the correlation of the two Brownian motions. Then, any combination of X_0 and X is still normal and the VaR of the portfolio is

$$v(\theta,h)=\Phi^{-1}(1-\theta)\sqrt{\begin{array}{c}\Big(S_t\xi_t\frac{d}{dS_t}\Pi_t(S_t,\xi_t)\Big)^2+\Big(\sigma_t\frac{d}{d\xi_t}\Pi_t(S_t,\xi_t)\Big)^2\\ +2\rho_t S_t\xi_t\sigma_t\frac{d}{dS_t}\Pi_t(S_t,\xi_t)\frac{d}{d\xi_t}\Pi_t(S_t,\xi_t)\end{array}}\sqrt{h}. \tag{8.6}$$

8.4.2 A short-term model-free formula

Driven by the results in the Black-Scholes and the Stochastic Voaltility case, we generalize the VaR formulas to a short-term model-free formula which can be applied to any historical series of spot and option prices.

With this aim, we rather work using the implied volatility, which can always be computed from market prices using a root-finding algorithm applied to the classic Black-Scholes pricing formula

$$\mathrm{BS}_t\big(k,\tau,\omega,F_t(\tau),\mathrm{DF}_t(\tau),\sigma_t^{\mathrm{imp}}(k,\tau)\big)=\omega\mathrm{DF}_t(\tau)F_t(\tau)\big(\Phi(\omega d_1)-e^k\Phi(\omega d_2)\big)$$

where $k=\log\frac{K}{F_t(\tau)}$ is the log-forward moneyness, $\tau=T-t$ is the time-to-maturity, $\omega=+1$ if the option is a Call, -1 if it is a put, and

$$d_{1,2}=-\frac{k}{\sigma_t^{\mathrm{imp}}(k,\tau)\sqrt{\tau}}\pm\frac{\sigma_t^{\mathrm{imp}}(k,\tau)\sqrt{\tau}}{2}.$$

When computing risks, the implied volatility $\sigma_t^{\mathrm{imp}}(k,\tau)$ is generally considered as a risk factor together with the underlier S_t. For this reason, we write it as a function of a driving factor ξ_t: $\sigma_t^{\mathrm{imp}}(k,\tau)=\sigma(k,\tau,\xi_t)$, so that the dynamics of the two risk factors are

$$\begin{aligned}dS_t&=\alpha_t S_t dt+\beta_t S_t dW_{0,t}\\ d\xi_t&=\mu_t dt+\eta_t dW_t\\ dW_{0,t}dW_t&=\rho_t dt.\end{aligned}\tag{H1}$$

Observe that the implied volatility risk factor depends on the log-forward moneyness and the time-to-maturity rather than the contract strike and its maturity. Indeed, the time series of a fixed contract is available since its issue date and is then limited in time. Furthermore, we do not expect its implied volatility to have any nice statistical property of stationarity that could legitimate drawing meaningful forecasts for its historical returns between time t and $t+h$. On the contrary, we expect that the market encode the implied volatility risk rather in a log-forward moneyness, time-to-maturity map, meaning that the time series of the implied volatility at a fixed point in this relative grid will have much nicer features.

Let us consider a portfolio of Calls and Puts written as Black-Scholes functions:

$$\begin{aligned}\Pi_t(S_t,\xi_t) &= \sum_i \pi_i \mathrm{BS}_t\Big(\log\frac{K_i}{f(T_i-t)S_t}, T_i-t, \omega_i, F_t(T_i-t), \mathrm{DF}_t(T_i-t),\\ &\qquad \sigma\Big(\log\frac{K_i}{f(T_i-t)S_t}, T_i-t, \xi_t\Big)\Big)\\ &=: \sum_i \pi_i \mathrm{BS}_t^i.\end{aligned} \tag{8.7}$$

Repeating the steps in Section 8.4.1, the approximated formula for the VaR becomes

$$v(\theta,h) = \Phi^{-1}(1-\theta)\sqrt{\begin{aligned}&\Big(S_t\beta_t\frac{d}{dS_t}\Pi_t(S_t,\xi_t)\Big)^2 + \Big(\eta_t\frac{d}{d\xi_t}\Pi_t(S_t,\xi_t)\Big)^2\\ &\quad +2\rho_t S_t\beta_t\eta_t\frac{d}{dS_t}\Pi_t(S_t,\xi_t)\frac{d}{d\xi_t}\Pi_t(S_t,\xi_t)\end{aligned}}\sqrt{h}.$$

This formula is far from being model-free for two reasons:

- The term $\frac{d}{dS_t}\Pi_t(S_t,\xi_t)$ is the full derivative of the portfolio Π_t with respect to S_t, which also involves the derivative of prices with respect to the implied volatility, since it depends on $k = \log\frac{K}{f(\tau)S_t}$. As a consequence, it does not correspond to the Black-Scholes delta and its expression must be made explicit.
- The term $\frac{d}{d\xi_t}\Pi_t(S_t,\xi_t)$ is the derivative of the portfolio with respect to ξ_t, and it does not coincide with what the market indicates with vega, i.e. the sensibility of the portfolio to the option volatility.

Given the above, we shall rather develop the dynamics of the portfolio as a function of S_t and $\sigma(k,\tau,\xi_t)$ where k also depends on S_t.

Observe that for fixed k and τ, we have

$$\begin{aligned}d\sigma_t &= \partial_\xi\sigma_t\, d\xi_t\\ &= \mu_t\partial_\xi\sigma_t\, dt + \eta_t\partial_\xi\sigma_t\, dW_t\end{aligned} \tag{8.8}$$

where $\sigma_t = \sigma(k,\tau,\xi_t)$. We define $\zeta_t(k,\tau,\xi_t) := \eta_t\partial_\xi\sigma(k,\tau,\xi_t)$.

On the other hand, writing $k = \log\frac{K}{f(\tau)S_t}$ and $\tau = T-t$, we rather find

$$\begin{aligned}d\sigma_t &= a_t\, dt - \frac{\partial_k\sigma_t}{S_t}\, dS_t + \partial_\xi\sigma_t\, d\xi_t\\ &= \big(a_t - \alpha_t\partial_k\sigma_t + \mu_t\partial_\xi\sigma_t\big)\, dt - \beta_t\partial_k\sigma_t\, dW_{0,t} + \zeta_t\, dW_t\end{aligned} \tag{8.9}$$

where $a_t = -\partial_\tau\sigma_t + \partial_k\sigma_t\frac{\partial_\tau f}{f} + \frac{\beta_t^2}{2}(\partial_k^2\sigma_t + \partial_k\sigma_t) + \frac{\eta_t^2}{2}\partial_\xi^2\sigma_t - \rho_t\beta_t\eta_t\partial_\xi\partial_k\sigma_t$ and $\sigma_t = \sigma\big(\log\frac{K}{f(T-t)S_t}, T-t, \xi_t\big)$.

Using Equation (8.9) and ignoring the terms in h in the finite scheme of the portfolio increments, the P&Ls assume the form

$$\Pi_{t+h}(S_{t+h},\xi_{t+h}) - \Pi_t(S_t,\xi_t) \approx \beta_t\big(S_t\sum_i \pi_i \partial_S \mathrm{BS}_t^i - \sum_i \pi_i \partial_k \sigma_t^i \partial_\sigma \mathrm{BS}_t^i\big)\sqrt{h}X_0 + \\ + \sum_i \pi_i \zeta_t^i \partial_\sigma \mathrm{BS}_t^i \sqrt{h} X$$

where $\sigma_t^i = \sigma\big(\log\frac{K_i}{f(T_i-t)S_t}, T_i - t, \xi_t\big)$, $\zeta_t^i = \zeta_t\big(\log\frac{K_i}{f(T_i-t)S_t}, T_i - t, \xi_t\big)$, and X_0 and X are standard jointly normal random variables with correlation ρ_t.

Let us denote $\Delta_t^i := \partial_S \mathrm{BS}_t^i$ and $\mathcal{V}_t^i := \partial_\sigma \mathrm{BS}_t^i$. Using the same proof as in the Stochastic Volatility case of Section 8.4.1, we finally find a VaR formula on an MPOR horizon of h days of the form:

$$\begin{aligned} \mathrm{VaR}_{\theta,t}(h) &= \Phi^{-1}(1-\theta)\sqrt{c_t^2 + q_t^2 + 2\rho_t c_t q_t}\sqrt{h} \\ c_t &= \beta_t\big(S_t\sum_i \pi_i \Delta_t^i - \sum_i \pi_i \mathcal{V}_t^i \partial_k \sigma_t^i\big) \\ q_t &= \sum_i \pi_i \zeta_t^i \mathcal{V}_t^i. \end{aligned} \qquad \text{(Short-term formula)}$$

This expression is actually model-free. Indeed, the terms Δ_t^i and $\mathcal{V}_t^i$ are respectively the Black-Scholes delta and vega of the i-th option in the portfolio. In particular they correspond to

$$\begin{aligned} \Delta_t\big(k,\tau,\omega,\sigma_t^{\mathrm{imp}}(k,\tau)\big) &= \omega\Phi(\omega d_1), \\ \mathcal{V}_t\big(k,\tau,F_t(\tau),\mathrm{DF}_t(\tau),\sigma_t^{\mathrm{imp}}(k,\tau)\big) &= \mathrm{DF}_t(\tau)F_t(\tau)\phi(d_1)\sqrt{\tau}. \end{aligned}$$

The volatility β_t of the underlying spot S_t can be computed looking at historical moves. For example, it could be a EWMA volatility on log-returns appropriately rescaled by the square-root of the returns' distance h_r (of for example one trading day): $\beta_t = \frac{\mathrm{EWMA}(r_{S,t})}{\sqrt{h_r}}$ where $r_{S,t} = \log\frac{S_t}{S_{t-h_r}}$.

Given Equation (8.8), the quantity ζ_t^i is simply the vol-of-vol evaluated in $\big(\log\frac{K_i}{f(T_i-t)S_t}, T_i - t, \xi_t\big)$ and it could be also computed as a EWMA volatility on historical absolute returns of the implied volatility surface at the fixed log-forward moneyness and time-to-maturity grid point, rescaled by the square-root of h_r. For liquidity reasons, it is also possible to approximate the latter quantity as the vol-of-vol at the 1M ATM point, times an appropriate factor (see Section 8.6.3 for the description of a possible way to calibrate such a factor). In [67], the authors suggest to consider the historical series of the 1M ATM implied volatility point. A less procyclical alternative identified by the authors consists in rescaling the VVIX historical value with a sigmoid function which ensures a smooth vol-of-vol transition between high and low volatility regimes.

The correlation parameter ρ_t can be computed using a EWMA correlation between spot log-returns and absolute IV returns, where the IV point considered can be again the 1M ATM point.

Lastly, the derivative $\partial_k \sigma_t^i$ is the derivative of the smile with respect to the log-forward moneyness, evaluated in $\big(\log \frac{K_i}{f(T_i-t)S_t}, T_i - t, \xi_t\big)$. Since options are quoted in strike and maturity rather than log-forward moneyness and time-to-maturity, observe that

$$\partial_k \sigma_t^i = \partial_k \sigma_t \bigg(\log \frac{K_i}{f(T_i - t)S_t}, T_i - t, \xi_t \bigg) = K \partial_K \tilde{\sigma}_t^{\text{imp}}(K_i, T_i)$$

where $\tilde{\sigma}_t^{\text{imp}}(K,T) = \sigma_t^{\text{imp}}\big(\log \frac{K}{f(T-t)S_t}, T-t\big)$. The derivative of the strike smile can be recovered by simple interpolation of market data (for example, using cubic B-splines or arbitrage-free smile models), or by finite differences of market data.

t-Student short-term model-free VaR formulation

When calculating risk, a large majority of financial players considers the distribution of the returns of an underlier S_t to be t-Student. The reason is linked to the shape of the probability functions of such distribution, which are fatter, compared to a classic normal distribution. In this way the importance of extreme events is higher and this guarantees a larger conservativeness of the risk model.

Assumptions. In this section we consider then a t-Student distribution for the relative returns $\frac{S_{t+h}-S_t}{S_t}$, with ν_t degrees of freedom, location parameter $\alpha_t h$ and scale parameter $\beta_t \sqrt{h}$. In particular, we consider the model

$$S_{t+h} = S_t(1 + \alpha_t h + \beta_t T_{t+h}) \tag{H1}$$

where $T_{t+h} \in \mathbb{R}$ is a t-Student with ν_t degrees of freedom, null mean and variance equal to h. Then, the probability density function of S_{t+h} is

$$p_{S_{t+h}}(s) = \frac{\Gamma\big(\frac{\nu_t+1}{2}\big)}{\Gamma\big(\frac{\nu_t}{2}\big) S_t \beta_t \sqrt{\pi h}} \bigg(1 + \frac{1}{\nu_t} \bigg(\frac{s - S_t(1+\alpha_t h)}{S_t \beta_t \sqrt{h}} \bigg)^2 \bigg)^{-\frac{\nu_t+1}{2}}. \tag{8.10}$$

Also, for every strike K and maturity T, denoting $\tau = T - t$, $k(s) = \log \frac{K}{f(\tau)s}$, we consider the increment $\Delta\sigma_t(k(S_t), \tau) := \sigma_{t+h}(k(S_{t+h}), \tau - h) - \sigma_t(k(S_t), \tau)$ conditional to S_{t+h} to be a Gaussian random variable with mean $m_t(S_{t+h}, k(S_t), \tau)$ and variance $\zeta_t(k(S_t), \tau)^2(1 - \rho_t^2)h$, with

$$m_t(s, k(S_t), \tau) = \mu_t(k(S_t), \tau)h + \frac{s - S_t(1+\alpha_t h)}{\beta_t S_t} \zeta_t(k(S_t), \tau)\rho_t.$$

In other words, we write the conditional implied volatility increments as

$$\Delta\sigma_t(k(S_t), \tau) | (S_{t+h} = s) = m_t(s, k(S_t), \tau) + \zeta_t(k(S_t), \tau)\sqrt{1 - \rho_t^2}\sqrt{h}X \tag{H2}$$

where X is a standard Gaussian random variable.

Remark 8.1. *Since with the above hypothesis we only know the distribution of the implied volatility increments conditional to S_{t+h}, it could seem difficult to calibrate parameters μ_t, ζ_t and ρ_t on market data. However, given a conditional distribution, it is easy to recover the moments of the marginal distribution using the tower property in expectations. Indeed, moments up to the second order of $\Delta\sigma_t(k(S_t),\tau)$ (without the conditioning to S_{t+h}) are*

$$\begin{aligned} E[\Delta\sigma_t(k(S_t),\tau)] &= \mu_t(k(S_t),\tau)h \\ Var[\Delta\sigma_t(k(S_t),\tau)] &= \zeta_t(k(S_t),\tau)^2 h \\ Corr[S_{t+h},\Delta\sigma_t(k(S_t),\tau)] &= \rho_t. \end{aligned}$$

This allows to easily calibrate parameters based on the historical mean and variance of $\Delta\sigma_t(k(S_t),\tau)$.

In the next paragraphs, we justify the following formula for the h-days VaR with confidence level θ under Equations (H1) and (H2):

$$\begin{aligned} \mathrm{VaR}_{\theta,t}(h) &= F_Z^{-1}(1-\theta)\sqrt{c_t^2+q_t^2+2\rho_t c_t q_t}\sqrt{h} \\ c_t &= \beta_t\big(S_t\sum_i \pi_i\Delta_t^i - \sum_i \pi_i \mathcal{V}_t^i \partial_k\sigma_t^i\big) \\ q_t &= \sum_i \pi_i \zeta_t^i \mathcal{V}_t^i \end{aligned} \qquad \text{(Short-term t-Student)}$$

where

$$Z = \frac{q_t\sqrt{1-\rho_t^2}X + (c_t+q_t\rho_t)Y}{\sqrt{c_t^2+q_t^2+2\rho_t c_t q_t}} \tag{8.11}$$

and X is a standard Gaussian random variable and Y is a standard t-Student with ν_t degrees of freedom independent of X.

Quantities that enter Equation (Short-term t-Student) are the same as in the Gaussian case: the Black-Scholes Greeks delta Δ_t^i and vega $\mathcal{V}_t^i$, the volatility β_t of the underlying spot S_t, the vol-of-vol ζ_t^i, the correlation ρ_t, and the derivative $\partial_k\sigma_t^i$ of the smile with respect to the log-forward moneyness. See Section 8.4.2 for a description of how to compute these quantities in practice from market data.

Remark 8.2. *[9] shows in Theorem 1 that the probability density function of Z is*

$$p_Z(z) = \sum_{k=0}^{\infty} \phi_k^{(\nu_t,\gamma)} g_{k,a_1}(z)$$

where $a_1 = \frac{q\sqrt{1-\rho_t^2}}{\sqrt{c_t^2+q_t^2+2\rho_t c_t q_t}}$, $\gamma = \frac{c_t+q_t\rho_t}{q_t\sqrt{2(1-\rho_t^2)}}$ *and*

$$\begin{aligned} \phi_k^{(\nu_t,\gamma)} &= \frac{\Gamma\left(k+\frac{1}{2}\right)}{k!\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{\nu_t}{2}\right)}\int_0^\infty \exp(-f) f^{\frac{\nu_t-1}{2}}\left(f+\gamma^2\right)^{-k-\frac{1}{2}} df \\ g_{k,a_1}(z) &= \frac{\Gamma\left(\frac{1}{2}\right)}{\Gamma\left(k+\frac{1}{2}\right)a_1\sqrt{2\pi}}\left(\frac{z^2}{2a_1^2}\right)^k \exp\left(-\frac{z^2}{2a_1^2}\right). \end{aligned}$$

Quantiles of Z can also be computed empirically, simulating the distribution of the linear combination of two independent random variables distributed as a standard Gaussian and a standard t-Student with ν_t degrees of freedom. In particular, simulations of Z can be found following the steps:

1. *Simulate two independent normal random variables X and N_Y, with mean 0 and variance 1;*
2. *Turn N_Y into a uniform distribution $U_Y = \Phi(N_Y)$;*
3. *Recover the t-Student random variable via $Y = F_t^{-1}(U_Y; \nu_t)$ where $F_t(\cdot; \nu_t)$ is the cumulative density function of a t-Student with ν_t degrees of freedom;*
4. *Put $Z = \frac{q_t\sqrt{1-\rho_t^2}X + (c_t + q_t\rho_t)Y}{\sqrt{c_t^2 + q_t^2 + 2\rho_t c_t q_t}}$.*

We now explain the rationale of Equation (Short-term t-Student). Doing a first order approximation of increments of the option portfolio $\Pi_t(S_t, \sigma_t)$, taking into consideration the dependence of every options' implied volatility to the log-forward moneyness and so to the underlier, we find that the form of the P&Ls is

$$\begin{aligned}\Pi_{t+h}(S_{t+h}, \sigma_{t+h}) - \Pi_t(S_t, \sigma_t) &\approx \beta_t\big(S_t \sum_i \pi_i \partial_S \mathrm{BS}_t^i - \sum_i \pi_i \partial_k \sigma_t^i \partial_\sigma \mathrm{BS}_t^i\big)\, T_{t+h} + \\ &\quad + \sum_i \pi_i \partial_\sigma \mathrm{BS}_t^i \Delta\sigma_t^i \\ &= c_t T_{t+h} + \sum_i \pi_i \mathcal{V}_t^i \Delta\sigma_t^i,\end{aligned} \tag{8.12}$$

where we used the same notations as in Section 8.4.2. Here, we do not know the distribution of the increments of the implied volatilities, so that we cannot automatically infer the distribution of the P&Ls. However, we can still compute VaRs using the relation in Equation (8.5).

Firstly, we can write $T_{t+h} = \sqrt{h}\tilde{Y}$ where $\tilde{Y}$ is a standard t-Student with ν_t degrees of freedom, and

$$\Delta\sigma_t(k(S_t), \tau) = \mu_t(k(S_t), \tau)h + \zeta_t(k(S_t), \tau)\sqrt{h}\tilde{X}$$

for a certain random variable $\tilde{X}$ with mean 0 and variance 1. Then, given Equation (8.12), the distribution of $\frac{\text{P\&L}}{\sqrt{h}}$ tends to the distribution of $c_t\tilde{Y} + \sum_i \pi_i \mathcal{V}_t^i \zeta_t^i \tilde{X}$, which does not depend on h. In particular since

$$\begin{aligned}1 - \theta &= P\big(\text{P\&L} \leq v(\theta, h)\big) \\ &= P\bigg(\frac{\text{P\&L}}{\sqrt{h}} \leq \frac{v(\theta, h)}{\sqrt{h}}\bigg),\end{aligned}$$

and the limiting random variable has a strictly positive density, then the function $v(\theta, h)$ is asymptotic with $\sqrt{h}$, i.e. $v(\theta, h) = u(\theta)\sqrt{h} + o(\sqrt{h})$. This is consistent with Equation (Short-term formula), where the h-days VaR is proportional to the square-root of h.

Secondly, the distribution of the P&Ls conditional to S_{t+h} is Gaussian and in particular

$$P\big(\text{P\&L} \leq v(\theta,h) | S_{t+h} = s\big) = \Phi\left(\frac{v(\theta,h) - c_t t(s) - \sum_i \pi_i \mathcal{V}_t^i m_t^i(s)}{q_t\sqrt{1-\rho_t^2}\sqrt{h}}\right)$$

where $t(s) = \frac{s - S_t(1+\alpha_t h)}{\beta_t S_t}$. Then, removing the conditionality to the probability of the P&Ls, it holds

$$P\big(\text{P\&L} \leq v(\theta,h)\big) = \int_{-\infty}^{\infty} p_{S_{t+h}}(s)\Phi\left(\frac{v(\theta,h) - c_t t(s) - \sum_i \pi_i \mathcal{V}_t^i m_t^i(s)}{q_t\sqrt{1-\rho_t^2}\sqrt{h}}\right) ds$$

and this in turn is

$$\int_{-\infty}^{\infty} p_T(y)\Phi\left(\frac{v(\theta,h) - c_t y\sqrt{h} - \sum_i \pi_i \mathcal{V}_t^i m_t^i\big(S_t(1+y\beta_t\sqrt{h}+\alpha_t h)\big)}{q_t\sqrt{1-\rho_t^2}\sqrt{h}}\right) dy$$

where $p_{S_{t+h}}$ is as in Equation (8.10), p_T is the probability density function of a standard t-Student with ν degrees of freedom, and we used the transformation $y = \frac{t(s)}{\sqrt{h}}$. In this way, for the Lebesgue's dominated convergence theorem and using the fact that $v(\theta, h) = u(\theta)\sqrt{h} + o(\sqrt{h})$, the right hand side of the previous relation goes to

$$\int_{-\infty}^{\infty} p_T(y)\Phi\left(\frac{u(\theta) - (c_t + q_t\rho_t)y}{q_t\sqrt{1-\rho_t^2}}\right) dy$$

for h going to 0. Consider two independent random variables X and Y with X a standard Gaussian and Y a standard t-Student with ν_t degrees of freedom. We can write the latter expression as

$$E\left[P\left(X \leq \frac{u(\theta) - (c_t + q_t\rho_t)Y}{q_t\sqrt{1-\rho_t^2}}\right)\right] = P\left(X \leq \frac{u(\theta) - (c_t + q_t\rho_t)Y}{q_t\sqrt{1-\rho_t^2}}\right).$$

Defining the random variable Z as in Equation (8.11), we shall look at the value of $u(\theta)$ such that

$$1 - \theta = P\left(Z \leq \frac{u(\theta)}{\sqrt{c_t^2 + q_t^2 + 2\rho_t c_t q_t}}\right).$$

All in all, the short-term model-free VaR formula in the t-Student case becomes Equation (Short-term t-Student) ignoring terms in $o(\sqrt{h})$.

Remark 8.3. *In both Equation (Short-term formula) and Equation (Short-term t-Student) the vol-of-vol parameter depends on the strike and maturity of the option, while the correlation does not. This is due to the underlying hypothesis that the whole implied volatility surface is driven by one single Brownian motion, even though the magnitude of movements for each surface point depends on the point itself. The short-term model-free formulas can be generalized to the case where there is more than one Brownian motion driving the implied volatility surface (typically the target dimension is of 2 or 3).*

8.4.3 Properties and limitations

Local quantities and extreme risk: concrete practical implementation

Observe that all the above VaR estimations (the Black-Scholes formula, the Stochastic volatility formula, and the short-term model-free formula) are defined with local quantities: deltas, vegas, instantaneous volatility and correlation coefficients. The initial margin however incorporates a tail risk which looks at future moves in prices that typically correspond to large moves. Even if there is an apparent paradox here, the explanation is clear: those formulas are asymptotic formulas when the time step h goes to zero, and for sufficient small h even the tail risk will be driven by the local quantities, in so far as we deal we diffusion models.

The whole question therefore is how those asymptotic formulas will behave in practice. Obviously, the smaller the MPOR, or the less volatile the market, the better. A careful backtesting will be the clue here: it will allow to diagnose whether the coverage and procyclicality behavior of the formula are satisfactory.

In this regard, and especially from a regulatory perspective, one should keep in mind that the final IM formula will contain other components besides this core one, like a weighted Stress Historical VaR and the Short Option Minimum quantity described in Section 8.2. In general the former component will be obtained by computing price returns along stress historical scenarios with full re-evaluation (meaning, using the Black-Scholes formula for options with the shocked underlier and implied volatility) instead of the local first order Greeks. Therefore the risk of missing a convexity behavior should be largely mitigated, if not fully eliminated. Regarding the SOM, consider a portfolio of short deep OTM options. Today, this portfolio has negligible delta and vega quantities, so that the VaR estimation is around 0, even though there actually is a tail risk. This hidden risk is far to be local, but it still should be taken into consideration in the initial margin calculation. This is the rationale of the SOM, which is already implemented by CCPs and should cover the risk of those short-term portfolios, as discussed in Section 8.2.

Symmetry with respect to the portfolio

It is easy to see that all the new VaR formulas in this chapter are symmetrical with respect to the portfolio, i.e. being short or long on the same portfolio would produce the same VaR exposure. This could seem weird, especially when we suppose a log-normal distribution of the spot, which is not symmetric. The symmetricity appears when we take the limit for h going to 0. Indeed, the terms multiplying $\sqrt{h}$ are symmetrical in the portfolio position while the ones that should break the symmetricity multiply higher orders of h, so that they are canceled out in the limit.

However, this symmetry is not an issue when computing margins: as seen in Section 8.2, the final total risk requirement charged by the CCP is composed of the margin computed on P&Ls (refined by the add-ons and the SOM) minus the NOV component. In this way, neglecting the add-ons and the SOM, a long portfolio $\Pi > 0$ with initial margin IM has a total risk requirement equal to $\text{IM} - \Pi$; while the same portfolio but on a short position $-\Pi < 0$ implies a total risk requirement of $\text{IM} + \Pi$.

Comparison with FHS

In Section 8.3.1, we have seen that among its drawbacks, the FHS model is limited by the number of scenarios that it can generate, depending on the available historical data. On the other hand, the short-term model-free formula in Section 8.4.2 does not need to compute simulated scenarios and eventually requires historical data only for the calibration of volatility parameters.

Secondly, while the FHS does not capture the joint dynamics of risk factors in complex products, the short-term model-free formula in Section 8.4.2 considers both the singular margin impact of each risk factor and the joint margin impact affected by the correlation of risk factors. Furthermore, the short-term model-free formula for options is more natural than the FHS methodology, whose application to IV surface points is more subtle.

A third limitation is the difficulty of FHS to generate arbitrage-free scenarios, which is not an issue for the short-term model-free formula in Section 8.4.2 since it does not require the generation of scenarios and does not face the arbitrage issue.

Finally, regarding the procyclicality of the VaR estimation, we show in numerical experiments in Section 8.6.3 that the short-term model-free VaR is less procyclical then the FHS VaR for the tested portfolios.

We turn now to the exact computation of the VaR in the neural-SDE model.

8.5 Quasi-explicit formula for the VaR in the neural-SDE model

In this section we investigate the neural-SDE model described in Section 8.3.4 and the special specification of its parameters with the aim of applying it to an IM computation. We are not interested in the calibration of arbitrage-free Call prices surfaces via neural networks but to the affine factor model for normalized option prices itself, so that parameters can be calibrated with any algorithm of choice, which is not necessarily a neural network.

The model is particularly simple and it turns out to have a quasi-explicit formula for the VaR of option portfolios, as we show in Section 8.5.1 below. In practice, this could enable rapid computations for the IM in such models, which may prove to be highly relevant when properly calibrated.

Moreover, the VaR can be approximated by a closed formula which is proportional to the square-root of the MPOR (see Section 8.5.2). This approximated formula coincides with the VaR formula in the Stochastic Volatility model of Section 8.4.1.

We reemphasize the fact that while the model can be calibrated also in different ways as the ones described in [18], the results in this chapter are still valid and independent from the calibration setup.

We use the same notations as in Section 8.4. Furthermore, in the whole section, the notation $\|\cdot\|_2$ indicates the Euclidean 2-norm, i.e. $\|(a_1, \dots, a_d)^T\|_2^2 = \sum_{i=1}^d a_i^2$.

8.5.1 Quasi-explicit formula for the VaR

Consider a portfolio of Vanilla Calls with price at time t given by

$$\begin{aligned}\Pi_t(S_t,\xi_t) &= \sum_i \pi_i C\Big(T_i - t, \log\frac{K_i}{f(T_i-t)S_t}; S_t, \xi_t\Big)\\ &= S_t \sum_i \pi_i \mathrm{DF}_t(T_i-t) f(T_i-t) c\Big(T_i - t, \log\frac{K_i}{f(T_i-t)S_t}; \xi_t\Big)\end{aligned} \tag{8.13}$$

where

$$c\Big(T_i - t, \log\frac{K_i}{f(T_i-t)S_t}; \xi_t\Big) = G_0\Big(T_i - t, \log\frac{K_i}{f(T_i-t)S_t}\Big) + G\Big(T_i - t, \log\frac{K_i}{f(T_i-t)S_t}\Big)\cdot\xi_t.$$

From now on, we work at time t, so that quantities S_t and ξ_t are known. The P&L $:=$ $\Pi_{t+h}(S_{t+h},\xi_{t+h}) - \Pi_t(S_t,\xi_t)$ of the portfolio reads then

$$\mathrm{P\&L} = A(h, S_{t+h}) + B(h, S_{t+h})\cdot\big(\xi_{t+h} - \xi_t\big) \tag{8.14}$$

where

$$\begin{aligned}A(h,s) &= s\sum_i \pi_i \mathrm{DF}_t(T_i-(t+h)) f(T_i-(t+h))\times\\ &\quad\times\Bigg(G_0\Big(T_i-(t+h), \log\frac{K_i}{f(T_i-(t+h))s}\Big) +\\ &\quad + G\Big(T_i-(t+h), \log\frac{K_i}{f(T_i-(t+h))s}\Big)\cdot\xi_t\Bigg) - \Pi_t(S_t,\xi_t),\\ B(h,s) &= s\sum_i \pi_i \mathrm{DF}_t(T_i-(t+h)) f(T_i-(t+h)) G\Big(T_i-(t+h), \log\frac{K_i}{f(T_i-(t+h))s}\Big).\end{aligned} \tag{8.15}$$

An important consequence to the representation in Equation (8.14) is the linearity of the P&Ls in $\xi_{t+h} - \xi_t$. In terms of VaR calculations, this means that the VaR for the P&Ls' distribution conditional to S_{t+h} is linear with respect to the VaR for the ξ_{t+h} distribution.

Hypothesis on the joint increments

Since from a practical perspective market data is always related to a discrete time grid, from now on, for risk calculations we consider processes defined via their Euler scheme as in Section 8.3.4, i.e.

$$\begin{aligned}S_{t+h} &= S_t \exp\Big(\Big(\alpha_t - \frac{\beta_t^2}{2}\Big)h + \beta_t(W_{0,t+h} - W_{0,t})\Big),\\ \xi_{t+h} &= \xi_t + \mu_t h + \sigma_t\cdot(W_{t+h} - W_t)\end{aligned} \tag{8.16}$$

where $W_{0,t+h} - W_{0,t} \in \mathbb{R}$ and $W_{t+h} - W_t \in \mathbb{R}^d$ are Gaussian random variables with combined law $N(\mathbf{0}, hP_t)$ and P_t is the correlation matrix

$$P_t = \begin{pmatrix} 1 & P_{S,\xi,t}^T \\ P_{S,\xi,t} & P_{\xi,t} \end{pmatrix}.$$

We work at time t, so that quantities S_t, ξ_t, α_t, β_t, μ_t and σ_t are known. We will not need to observe $W_{0,t}$ and W_t.

Remark 8.4. *The Euler schemes with time step h for the processes S_t and ξ_t defined via the SDE Equation (8.3) are a particular case of Equation (8.16), therefore the results of this section hold also in this case.*

To develop Equation (8.5) we need a partial result regarding the distribution of the increments of ξ_t conditional to S_{t+h}.

Lemma 8.1. *For processes in Equation (8.16), the distribution of $\xi_{t+h} - \xi_t$ conditional to $S_{t+h} = s$ is a Gaussian $N(m_t(s), V_t)$ where*

$$m_t(s) = \mu_t h + \frac{1}{\beta_t}\left(\log\frac{s}{S_t} - \left(\alpha_t - \frac{\beta_t^2}{2}\right)h\right)\sigma_t \cdot P_{S,\xi,t},$$
$$V_t = h(\sigma_t \cdot b_t) \cdot (\sigma_t \cdot b_t)^T,$$

and $b_t \in \mathbb{R}^{d\times d}$ is a matrix such that $b_t \cdot b_t^T = P_{\xi,t} - P_{S,\xi,t} \cdot P_{S,\xi,t}^T$.

In the case of independent processes, $m_t(s) = \mu_t h$ and $V_t = h\sigma_t \cdot \sigma_t^T$.

Proof. Let us consider the Brownian increments $\Delta W_{0,t} = W_{0,t+h} - W_{0,t}$ and $\Delta W_t = W_{t+h} - W_t$, which have Gaussian joint distribution $N(\mathbf{0}, hP_t)$. The Gaussian random variable $Z = \Delta W_t - \Delta W_{0,t} P_{S,\xi}$ has null mean and variance equal to $h(P_{\xi,t} - P_{S,\xi,t} \cdot P_{S,\xi,t}^T)$. Since covariance matrices are symmetric and positive semi-definite, there exists a matrix $b_t \in \mathbb{R}^{d\times d}$ such that $b_t \cdot b_t^T = P_{\xi,t} - P_{S,\xi,t} \cdot P_{S,\xi,t}^T$. Also, Z and $\Delta W_{0,t}$ are independent since uncorrelated and jointly Gaussian, and it follows that the distribution of ΔW_t conditional to $\Delta W_{0,t} = w_0$ is a Gaussian with mean $w_0 P_{S,\xi,t}$ and covariance matrix $hb_t \cdot b_t^T$.

From Equation (8.16), it is immediate to recover the distribution of $\xi_{t+h} - \xi_t$ conditional to $\Delta W_{0,t} = w_0$. For the conditionality with respect to $S_{t+h} = s$, it is enough to substitute w_0 with $\frac{1}{\beta_t}\left(\log\frac{s}{S_t} - \left(\alpha_t - \frac{\beta_t^2}{2}\right)h\right)$.

If processes are independent, $P_{S,\xi,t} = \mathbf{0}$, $P_{\xi,t} = I_d$, and the conclusion follows.

□

As an immediate consequence to Lemma 8.1, we can write the increments of ξ_t conditional to S_{t+h} as

$$\xi_{t+h} - \xi_t | (S_{t+h} = s) = m_t(s) + \sqrt{h}\sigma_t \cdot b_t \cdot X \tag{8.17}$$

where $X \sim N(0, I_d)$ is a Gaussian random variable independent to S_{t+h}.

Then

$$\begin{aligned}\text{P\&L}|(S_{t+h} = s) &= A(h,s) + B(h,s) \cdot \left(m_t(s) + \sqrt{h}\sigma_t \cdot b_t \cdot X\right) \\ &= \hat{A}(h,s) + \hat{B}(h,s) \cdot X\end{aligned}$$

where

$$\begin{aligned}\hat{A}(h,s) &:= A(h,s) + B(h,s) \cdot m_t(s) \in \mathbb{R}, \\ \hat{B}(h,s) &:= B(h,s) \cdot \sqrt{h}\sigma_t \cdot b_t \in \mathbb{R}^{1\times d}.\end{aligned} \tag{8.18}$$

In particular, conditional to $S_{t+h} = s$, the P&L is a sum of jointly Gaussian variables, so it is also a Gaussian variable with law $N(\hat{A}(h,s), \|\hat{B}(h,s)\|_2^2)$. Then the quantity $P(\text{P\&L} \leq v(\theta,h)|S_{t+h} = s)$ is the cumulative function of a Gaussian variable, and in particular it is equal to

$$P(\text{P\&L} \leq v(\theta,h)|S_{t+h} = s) = \Phi\left(\frac{v(\theta,h) - \hat{A}(h,s)}{\|\hat{B}(h,s)\|_2}\right).$$

Reconsidering Equation (8.5), the VaR at risk level θ for the P&Ls can be computed as specified in the following proposition.

Proposition 8.1. *Under the model of Equations (8.2) and (8.16), the h-days VaR at confidence level θ of the portfolio Equation (8.13) is the value of $v(\theta,h)$ which solves*

$$1 - \theta = \int_0^\infty \Phi\left(\frac{v(\theta,h) - \hat{A}(h,s)}{\|\hat{B}(h,s)\|_2}\right) dF_{S_{t+h}}(s)s. \tag{8.19}$$

where Equations (8.15) and (8.18) define $\hat{A}(h,s)$ and $\hat{B}(h,s)$.

Note that Proposition 8.1 gives a semi-closed formula for the VaR in the neural-SDE model, with no need of further hypothesis. As a consequence, we can compute efficiently the VaR in this model without using any approximation.

We shall notice that since losses cannot be larger than today's position, the result $v(\theta,h)$ should always be higher than minus the current value of the portfolio, i.e. $v(\theta,h) \geq -\Pi_t(S_t,\xi_t)$. This condition holds true if and only if the P&Ls' distribution is null below $-\Pi_t(S_t,\xi_t)$, or equivalently if and only if the distribution of future prices is null below 0. In particular, it must hold

$$G \cdot \xi_{t+h} > -G_0$$

for any ξ_{t+h}. This condition does not seem to be guaranteed a priori. Indeed, it depends on how the parameters of the distribution of the ξ_t are calibrated. However, if the ξ_t are calibrated such that Call prices always satisfy no arbitrage conditions, then in particular prices will always be positive.

Calls and Puts portfolio

In the case of portfolios with both Call and Put options, it is sufficient to re-write Put options using the Put-Call parity

$$P\Big(T-t, \log\frac{K}{F_t(T-t)}\Big) = C\Big(T-t, \log\frac{K}{F_t(T-t)}\Big) - \mathrm{DF}_t(T-t)\big(F_t(T-t)-K\big).$$

In this way, the relation in Equation (8.14) still holds redefining quantities A and B with elementary steps. In particular for every Put option position $\pi P\big(T-t, \log\frac{K}{F_t(T-t)}\big)$ in the portfolio, $A(h,s)$ adds the term

$$\begin{aligned} &s\pi \mathrm{DF}_t(T-(t+h))f(T-(t+h))\Big(G_0\Big(T-(t+h), \log\frac{K}{f(T-(t+h))s}\Big)+\\ &+G\Big(T-(t+h), \log\frac{K}{f(T-(t+h))s}\Big)\cdot\xi_t - 1\Big) + \pi\mathrm{DF}_t(T-(t+h))K, \end{aligned}$$

with the term $\Pi_t(S_t,\xi_t)$ updating its value with the added Puts, while $B(h,s)$ adds

$$s\pi \mathrm{DF}_t(T-(t+h))f(T-(t+h))G\Big(T-(t+h), \log\frac{K}{f(T-(t+h))s}\Big).$$

8.5.2 Closed formula for the short term VaR

In this section we start by proving that the VaR in the neural-SDE model for option prices is of the form $u(\theta)\sqrt{h}$ asymptotically with h. This formulation reflects empirical results and standard models adopted in industry and it is consistent with Equation (Short-term formula). Then, we state the main result of this section computing the explicit form of the function $u(\theta)$.

Firstly, we look at the form of the function $v(\theta,h)$ when h is small.

Lemma 8.2. *Under the model of Equations (8.2) and (8.16), the h-days VaR at confidence level θ of the portfolio in Equation (8.13) is asymptotic to $\sqrt{h}$ for h going to 0:*

$$VaR_{\theta,t}(h) = u(\theta)\sqrt{h} + o\big(\sqrt{h}\big)$$

for a certain function $u(\theta)$ not depending on h.

We give the proof in Section 8.A.

Our next result uses the quantities

$$\begin{aligned}
c_t &:= S_t\beta_t \sum_i \pi_i \mathrm{DF}_t(T_i - t) f(T_i - t) \times \\
&\quad \times \Big[G_0\Big(T_i - t, \log \frac{K_i}{f(T_i - t)S_t}\Big) + G\Big(T_i - t, \log \frac{K_i}{f(T_i - t)S_t}\Big) \cdot \xi_t - \mathbb{1}_{\mathrm{Puts}}(i) + \\
&\quad - \partial_k \Big(G_0\Big(T_i - t, \log \frac{K_i}{f(T_i - t)S_t}\Big) + G\Big(T_i - t, \log \frac{K_i}{f(T_i - t)S_t}\Big) \cdot \xi_t \Big) \Big] + \\
&\quad + B(0, S_t) \cdot \sigma_t \cdot P_{S,\xi,t}, \\
q_t &:= \big\| B(0, S_t) \cdot \sigma_t \cdot b_t \big\|_2,
\end{aligned} \tag{8.20}$$

where $\mathbb{1}_{\mathrm{Puts}}(i)$ is 1 if the index i refers to a put, otherwise it is null, and

$$B(0, S_t) = S_t \sum_i \pi_i \mathrm{DF}_t(T_i - t) f(T_i - t) G\Big(T_i - t, \log \frac{K_i}{f(T_i - t)S_t}\Big).$$

Remark 8.5. *It is easy to prove that c_t and q_t can be written with the alternative expressions:*

$$\begin{aligned}
c_t &= S_t\beta_t \frac{d}{dS_t} \Pi_t(S_t, \xi_t) + \nabla_{\xi_t} \Pi_t(S_t, \xi_t)^T \cdot \sigma_t \cdot P_{S,\xi,t}, \\
q_t &= \| \nabla_{\xi_t} \Pi_t(S_t, \xi_t)^T \cdot \sigma_t \cdot b_t \|_2.
\end{aligned}$$

We can now state the main result of this section.

Proposition 8.2. *Under the model of Equations (8.2) and (8.16), the h-days VaR at confidence level θ of the portfolio in Equation (8.13) is*

$$VaR_{\theta,t}(h) = \Phi^{-1}(1 - \theta) \sqrt{c_t^2 + q_t^2} \sqrt{h} + o\big(\sqrt{h}\big) \tag{8.21}$$

where c_t and q_t are defined in Equation (8.20).

The proof is given in Section 8.B.

Corollary 8.1. *Under the model of Equations (8.2) and (8.16) with $d = 1$, the h-days VaR at confidence level θ of the portfolio Equation (8.13) is*

$$VaR_{\theta,t}(h) = \Phi^{-1}(1 - \theta) \sqrt{\begin{aligned} &\Big(S_t\beta_t \frac{d}{dS_t} \Pi_t(S_t, \xi_t) \Big)^2 + \Big(\sigma_t \frac{d}{d\xi_t} \Pi_t(S_t, \xi_t) \Big)^2 \\ &+ 2 P_{S,\xi,t} S_t \beta_t \sigma_t \frac{d}{dS_t} \Pi_t(S_t, \xi_t) \frac{d}{d\xi_t} \Pi_t(S_t, \xi_t) \end{aligned}} \sqrt{h} + o\big(\sqrt{h}\big).$$

Observe that this is compatible with the Stochastic Volatility model's VaR in Equation (8.6).

Normal distribution for S_{t+h}

Lemma 8.1 still holds considering normal increments for S_{t+h}:

$$S_{t+h} = S_t\Big(1 + \alpha_t h + \beta_t(W_{0,t+h} - W_{0,t})\Big),$$

appropriately redefining the quantity $m_t(s)$. Indeed, in this case $m_t(s)$ becomes $\mu_t h + \frac{s - S_t(1+\alpha_t h)}{S_t\beta_t}\sigma_t \cdot P_{S,\xi,t}$.

The results in Lemma 8.2 and Proposition 8.2 are exactly the same as for the log-normal case. Indeed, the probability density function of S_{t+h} is

$$p_{S_{t+h}}(s) = \frac{1}{S_t\beta_t\sqrt{2\pi h}} \exp\left(-\left(\frac{s - S_t(1+\alpha_t h)}{S_t\beta_t\sqrt{2h}}\right)^2\right)$$

and the conclusion can be easily attained as in the previous case, using the transformation $y = \frac{s - S_t(1+\alpha_t h)}{S_t\beta_t\sqrt{h}}$ as done in Section 8.B.

t-Student distribution for S_{t+h}

As in Section 8.4.2, we consider the case where the increments of the underlier S_t follow a t-Student distribution, while increments of the implied volatility σ_t are Gaussian conditional to S_{t+h}.

Lemma 8.1 holds true in the Gaussian case because the random variable $Z = \Delta W_t - \Delta W_{0,t} P_{S,\xi}$ is still Gaussian and its decorrelation with $\Delta W_{0,t}$ implies its independence. However, in this case $\Delta W_{0,t}$ is substituted with a t-Student and we cannot derive the same result. We need then some additional hypothesis to derive the equivalent of Proposition 8.1 when S_{t+h} is a t-Student, and in particular we shall take Equation (8.17) as granted a priori.

Lemma 8.3. *Consider the model of Equation (8.2) and the hypothesis that*

$$S_{t+h} = S_t(1 + \alpha_t h + \beta_t T_{t+h})$$

where $T_{t+h} \in \mathbb{R}$ is a t-Student with ν_t degrees of freedom, null mean and variance equal to h, and $\xi_{t+h} - \xi_t$ conditional to $S_{t+h} = s$ is a Gaussian random variable with mean $m_t(s)$ and covariance matrix V_t with

$$m_t(s) = \mu_t h + \frac{s - S_t(1+\alpha_t h)}{S_t\beta_t}\sigma_t \cdot P_{S,\xi,t},$$
$$V_t = h(\sigma_t \cdot b_t) \cdot (\sigma_t \cdot b_t)^T,$$

and $b_t \in \mathbb{R}^{d\times d}$ is a matrix such that $b_t \cdot b_t^T = P_{\xi,t} - P_{S,\xi,t} \cdot P_{S,\xi,t}^T$, for certain parameters μ_t, $P_{S,\xi,t} \in \mathbb{R}^d$, σ_t, $P_{\xi,t} \in \mathbb{R}^{d\times d}$. Then the h-days VaR at confidence level θ of the portfolio in Equation (8.13) is the value of $v(\theta, h)$ which solves

$$1 - \theta = \int_0^\infty \Phi\left(\frac{v(\theta,h) - \hat{A}(h,s)}{\|\hat{B}(h,s)\|_2}\right) dF_{S_{t+h}}(s)s.$$

where Equations (8.15) *and* (8.18) *define* $\hat{A}(h,s)$ *and* $\hat{B}(h,s)$.

The calibration of the parameters μ_t, $P_{S,\xi,t}$, σ_t and $P_{\xi,t}$ is practically difficult if performed from the distribution of $\xi_{t+h}-\xi_t$ conditional to S_{t+h}. However, as in Remark 8.1, we can recover the moments of the marginal distribution, allowing an easy calibration of the parameters based on the historical mean and covariances of $\xi_{t+h}-\xi_t$.

Remark 8.6. *Under the model*

$$S_{t+h} = S_t(1+\alpha_t h + \beta_t T_{t+h})$$

where $T_{t+h} \in \mathbb{R}$ *is a t-Student with* ν_t *degrees of freedom, null mean and variance equal to* h*, if* $\xi_{t+h}-\xi_t$ *conditional to* $S_{t+h}=s$ *is a Gaussian random variable with mean* $m_t(s)$ *and covariance matrix* V_t *as in Lemma 8.3, then*

$$\begin{aligned} E[\xi_{t+h}-\xi_t] &= \mu_t h \\ Cov[\xi_{t+h}-\xi_t] &= h\sigma_t \cdot P_{\xi,t} \cdot \sigma_t^T \\ Cov\left[\begin{pmatrix} S_{t+h} \\ \xi_{t+h}-\xi_t \end{pmatrix}\right] &= \begin{pmatrix} \beta_t^2 S_t^2 h & \beta_t S_t h(\sigma_t \cdot P_{S,\xi,t})^T \\ \beta_t S_t h \sigma_t \cdot P_{S,\xi,t} & h\sigma_t \cdot P_{\xi,t} \cdot \sigma_t^T \end{pmatrix}. \end{aligned}$$

The result in Lemma 8.2 still holds true in the t-Student case. The proof is equivalent to the one given in Section 8.A, with the adaptations regarding the distribution of S_{t+h} and $\xi_{t+h}-\xi_t$. These adaptations can be found in the proof of Equation (Short-term t-Student) in Section 8.4.2. As a consequence, the function $v(\theta,h)$ is of the form $u(\theta)\sqrt{h}+o(\sqrt{h})$ for h small.

The probability density function of S_{t+h} is as in Equation (8.10). In this way, using the transformation $y=\frac{s-S_t(1+\alpha_t h)}{S_t\beta_t\sqrt{h}}$ and repeating the calculations in Section 8.B.1, we find Equation (8.23) with exactly the same values for c_t and q_t. We then look for the value of $u(\theta)$ such that

$$1-\theta = E\left[P\left(\frac{q_t X + c_t Y}{\sqrt{c_t^2+q_t^2}} \le \frac{u(\theta)}{\sqrt{c_t^2+q_t^2}}\right)\right].$$

The random variable $Z=\frac{q_t X+c_t Y}{\sqrt{c_t^2+q_t^2}}$ is not Gaussian in this case since it is the sum of a standard Gaussian random variable and a standard t-Student random variable, which are independent. However, it still holds that the initial margin in the case of t-Student returns is

$$\mathrm{VaR}_{\theta,t}(h) = F_Z^{-1}(1-\theta)\sqrt{c_t^2+q_t^2}\sqrt{h}+o(\sqrt{h}).$$

The empirical quantile of Z can be recovered as detailed in Remark 8.2.

The above observations lead us to the following.

Corollary 8.2. *Under the framework of Lemma 8.3, the h-days VaR at confidence level* θ *of the portfolio in Equation* (8.13) *is*

$$VaR_{\theta,t}(h) = F_Z^{-1}(1-\theta)\sqrt{c_t^2+q_t^2}\sqrt{h}+o(\sqrt{h})$$

where c_t, q_t are defined in Equation (8.20) and $Z = \frac{qX+cY}{\sqrt{c_t^2+q_t^2}}$ with X and Y independent and X standard Gaussian, Y standard t-Student with ν_t degrees of freedom.

8.6 Numerical experiments

8.6.1 Backtesting option portfolios

Before going into the numerical experiments, it is worth focusing on the specific issues arising while backtesting option portfolios. Indeed, options are contracts with a fixed strike and a fixed expiry, so that considering a backtest on a real fixed contract is awkward for two reasons: at maturity the option will expire and the backtest could not continue anymore; the option could become far OTM or ITM in time, completing losing interest in the market and becoming illiquid or even not traded anymore.

Moreover, in order to focus on a given risk (like the calendar spread one), and its adequate coverage by the margin model, it is better to consider a portfolio with a constant risk profile across time, and so constant specifications in terms of moneyness and time-to-maturity.

For this reason, options are generally backtested for fixed log-forward moneyness (or delta) and fixed time-to-maturity, rather than fixed contract. Of course, these desired options are not always available among the market quoted ones, so that two possibilities arise:

1. Considering the nearest in log-forward moneyness (or delta) and time-to-maturity real quoted option.

2. Considering synthetic option prices on the chosen fixed log-forward moneyness (or delta) and fixed time-to-maturity obtained via the model pricing criteria.

In the first case, the backtested portfolios will possibly change every day, depending on how much the ATM level has moved and on the rolling maturity. In the second case, the VaR estimations are compared to model P&Ls rather than real ones, so that if the calibration model is not good enough, backtesting results could be misleading.

In general, there is no preferred way to backtest option portfolios and CCPs may adopt both methodologies. We recommend to perform the two approaches for production, because the synthetic option prices, due to the complexity of the data treatments performed, may eventually not reflect fully faithfully the effective market returns when they are available, as used directly in the first approach above.

8.6.2 VaR formula in the Heston model

The VaR formula in Equation (8.6) can be applied to any Stochastic Volatility model. In this section we test it on the classic Heston model

$$\begin{aligned} dS_t &= \alpha_t S_t dt + \sqrt{\nu_t} S_t dW_{0,t} \\ d\nu_t &= \kappa(\theta - \nu_t)dt + \xi\sqrt{\nu_t} dW_t \\ dW_{0,t}dW_t &= \rho dt. \end{aligned}$$

In this case, Equation (8.6) has the form

$$\mathrm{VaR}_{\theta,t}(h) = \Phi^{-1}(1-\theta)\sqrt{\begin{aligned} &S_t^2\nu_t\Big(\frac{d}{dS_t}\Pi_t(S_t,\nu_t)\Big)^2 + \xi^2\nu_t\Big(\frac{d}{d\nu_t}\Pi_t(S_t,\nu_t)\Big)^2 \\ &\qquad +2\rho\xi S_t\nu_t\frac{d}{dS_t}\Pi_t(S_t,\beta_t)\frac{d}{d\nu_t}\Pi_t(S_t,\nu_t)\end{aligned}}\sqrt{h}. \tag{8.22}$$

Firstly, we simulate one year (365 days) history of the process (S_t, ν_t) with a simple Euler scheme of the form

$$\begin{aligned} S_{t+\delta t} &= S_t\big(1 + \alpha_t\delta t + \sqrt{\nu_t}\sqrt{\delta t}X_0\big) \\ \nu_{t+\delta t} &= |\nu_t + \kappa(\theta - \nu_t)\delta t + \xi\sqrt{\nu_t}\sqrt{\delta t}X| \end{aligned}$$

where X_0, X are standard Gaussian random variables with correlation ρ. Differences on the final results are negligible using the log-formulation for S_t, i.e.

$$S_{t+\delta t} = S_t \exp\Big(\Big(\alpha_t - \frac{\nu_t}{2}\Big)\delta t + \sqrt{\nu_t}\sqrt{\delta t}X_0\Big),$$

or using a Milstein scheme instead of the Euler's one. Observe that more efficient simulation schemes for the Heston model could be considered, however here we are interested into the performances of the short-term model-free VaR formula, rather than efficient ways to simulate the underlier.

At this point, for different outright, calendar and butterfly portfolios, we compute daily prices using the semi-analytical formula for a Call option in the Heston model described in [47]. We take null rates, so that the forward price coincides with the spot value and the discount factor is 1. Finally we compare real PnLs with 0.99-VaR estimations as in Equation (8.22) on different MPOR horizons. In order to compute portfolio derivatives with respect to the spot and the volatility of the spot, we use the average between the corresponding backward and forward finite differences:

$$\begin{aligned} \frac{d}{dS_t}\Pi_t(S_t,\nu_t) &\approx \frac{\Pi_t(S_t+\varepsilon,\nu_t) - \Pi_t(S_t-\varepsilon,\nu_t)}{2\varepsilon}, \\ \frac{d}{d\nu_t}\Pi_t(S_t,\nu_t) &\approx \frac{\Pi_t(S_t,\nu_t+\varepsilon) - \Pi_t(S_t,\nu_t-\varepsilon)}{2\varepsilon}. \end{aligned}$$

Note in particular that these quantities are *not* the Black-Scholes ones defined through the sensitivities of the Black-Scholes formula evaluated at the implied volatility corresponding to the Heston model price.

MPOR (days)	Coverage		Size of loss	
	Mean	Median	Mean	Median
1	0.9927	0.9945	0.0485	0.0204
2	0.9902	0.9945	0.0645	0.0353
3	0.9896	0.9945	0.0682	0.0422

Table 8.1: Average coverage and size of loss of Equation (8.22) on 74 option portfolios on simulated Heston data.

We use calibrated Heston parameters on S&P500 on December 2015 (see [13]), in particular $(\kappa, \sqrt{\theta}, \xi, \rho) = (6.169, 0.16168, 0.477, -0.781)$, and we start with initial values $S_0 = 2054$ and $\sqrt{\nu_0} = 0.15562$. For the Euler scheme, we choose a step δt of 10^{-1} days. The portfolios considered are outright Calls with delta in $\{0.2, 0.35, \delta_{\text{ATM}}, 0.65, 0.8\}$ and time-to-maturity in $\{30, 90, 180, 365\}$ days, and the resulting combinations of calendar spread and butterfly spread portfolios. In particular, a calendar spread is a portfolio with one short Call at a fixed strike K and maturity T_1 and one long Call with same strike K and maturity $T_2 > T_1$. For a fixed delta, the common strike of the two options is chosen to be the one related to the shortest maturity. A butterfly spread is composed of two long Calls with deltas δ and $1 - \delta$ respectively, and two short ATM Calls. For butterfly spreads, we also test the deltas $0.1, 0.3, 0.4, 0.45$.

The number of tested portfolios is then: $5 \times 4 = 20$ outrights, $\binom{4}{2} \times 5 = 30$ calendar spreads, and $6 \times 4 = 24$ butterfly spreads; in total 74 portfolios.

For each tested portfolio, we compute the coverage ratio as the number of days where the model VaR covers the realized loss over the total number of tested days, and the size of losses as the ratio between the margin loss (difference between realized loss and model VaR) and the portfolio price. The average and the median over all portfolios is displayed in Table 8.1. The results are very satisfactory for all MPORs. Indeed, the average coverage meets the 0.99-VaR requirement and breaches are of a very small size below 7% of the portfolio value.

Similarly, we compute the initial margin using the short-term model-free formula described in Section 8.4.2. In particular, we simulate 5 years history of an Heston process with same parameters as in the previous test and compute the initial margin for the same portfolios on the last year's observations (the previous history is used to calibrate parameters). The formula used is then Equation (Short-term formula), where the delta and the vega Greeks are the Black-Scholes ones, the spot volatility, the vol-of-vol and the correlation between risk factors are computed with the EWMA specification, and the derivative of the implied volatility with respect to the log-forward moneyness in computed via finite differences. The implied volatility point used to compute the EWMA correlation is the 1M ATM point.

As in the previous test, we compute the average coverage and size of loss for each tested MPOR. Results are shown in Table 8.2. Results are less conservative than in the previous test since the coverage is around 0.983. However, the size of loss is still very small compared to the portfolio value, and actually smaller than in the previous test.

MPOR (days)	Coverage		Size of loss	
	Mean	Median	Mean	Median
1	0.9853	0.9863	0.0404	0.0194
2	0.9826	0.9822	0.0565	0.0368
3	0.9813	0.9808	0.0496	0.0255

Table 8.2: Average coverage and size of loss of Equation (Short-term formula) *on* 74 *option portfolios on simulated Heston data, under the hypothesis of log-normal distribution of spot returns.*

MPOR (days)	Coverage		Size of loss	
	Mean	Median	Mean	Median
1	0.9907	0.9918	0.0362	0.0164
2	0.9886	0.9918	0.0596	0.0364
3	0.9889	0.9945	0.0485	0.0291

Table 8.3: Average coverage and size of loss of Equation (Short-term t-Student) *on* 74 *option portfolios on simulated Heston data, under the hypothesis of t-Student with* 5 *degrees of freedom distribution of spot returns.*

Results can actually be improved using the hypothesis of a t-Student distribution for spot returns and redefining the VaR for MPOR h as in Equation (Short-term t-Student) where the distribution of Z is obtained empirically as explained in Remark 8.2. Table 8.3 shows results when considering 5 degrees of freedom in the t-Student distribution of spot returns. As expected, results are more conservative than the normal case and satisfy the 0.99 coverage requirement.

8.6.3 Coverage performances of the short-term model-free VaR

In this section we show the results of coverage of the short-term model-free 0.99 VaR formula in Section 8.4.2 compared with the classical FHS model described in Section 8.3.1.

We use a database of S&P500 data provided to Zeliade by the Clearify project[4] on end of the day option prices. Firstly, we clean the rough data removing options with 0 volume and use the Put-Call parity on mid prices to extrapolate forward and discount factors for each quoted maturity having at least two Put-Call couples. Then, we remove all Calls not satisfying the arbitrage bounds

$$\mathrm{DF}_t(T)(F_t(T)-K)_+ \leq C_t(T,K) \leq \mathrm{DF}_t(T)F_t(T),$$

and all Puts not satisfying

$$\mathrm{DF}_t(T)(K-F_t(T))_+ \leq P_t(T,K) \leq \mathrm{DF}_t(T)K.$$

[4]Clearify is a collaboration between Zeliade and the Imperial College Mathematical Finance Department funded by an Imperial Faculty of Natural Sciences Strategic Research Funding Award.

At this points, Puts are transformed into Calls and all the following computations are performed for Call prices.

In order to get normalized historical prices, we compute synthetic historical prices on a fixed grid of time-to-maturity and log-forward moneyness. At this aim, we observe that market data is generally dense around the ATM point for short maturities and it spreads out with increasing expiry. For this reason, the most cunning fixed grid should be in delta, so we identify 17 delta points from 0.015 to 0.985 and compute the corresponding log-forward moneyness for a symbolic volatility of 0.1, over the grid of time-to-maturities of 2, 5, 10, 21, 42, 63, 126 and 252 days. Observe that in this way the grid is not constant in log-forward moneyness for different time-to-maturities. At this point, we firstly compute implied volatilities based on real prices. Then, we interpolate linearly on the log-forward moneyness direction since data is dense enough. The interpolation on the time-to-maturity direction is done on the implied total variances (squared implied volatilities times the time-to-maturity) adding synthetic points for the zero maturity equal to 0 and then interpolating linearly. The interpolated prices are then corrected to avoid static arbitrage as described in [17].

For comparison, we implemented a second interpolation scheme firstly adding synthetic points for the zero maturity (setting prices equal to their intrinsic values) and for extreme moneyness (with ITM prices equal to the discounted forward and null OTM prices); then normalizing all prices by their discounted forward; finally using monotonic cubic splines on the log-forward moneyness direction and linear interpolation on the time-to-maturity direction. The final results reported in this section do not significantly change.

We consider two portfolios: the first one is an ATM calendar spread between maturities of 1M and 6M; while the second one is a butterfly spread on maturity 3M and moneyness 0.9, 1, 1.1. We compute the VaR for an MPOR horizon of 1 day and a confidence level $\theta = 0.99$.

The short-term model-free VaR is obtained computing:

1. The spot volatility β_t via a EWMA volatility algorithm with decay factor 0.97 on spot log-returns, divided by the square-root of the daily step;
2. The correlation ρ_t via a EWMA correlation on spot log-returns and 1M ATM implied volatility absolute returns;
3. The vol-of-vol $\zeta_t(k, \tau)$ as F times the 1M ATM vol-of-vol obtained as a EWMA volatility on implied volatility absolute returns, divided by the square root of the daily step. In order to be conservative enough, the F factor is 1.1 times the quantile 0.9 of 5 years history of ratios between the (k, τ) vol-of-vol and the 1M ATM vol-of-vol;
4. Delta and vega quantities as the Black-Scholes deltas and vegas evaluated at the option implied volatility;
5. The derivative of the smile with respect to the log-forward moneyness as the derivative of the interpolated smile via B-splines.

Since we consider log-normal spot returns, we use the VaR formulation with normal quantiles of Equation (Short-term formula).

The FHS risk factors are the spot prices and the 17×8 implied volatility grid points for fixed log-forward moneyness and time-to-maturity. We consider log-returns for the former risk factors and absolute returns for the latter ones. The volatility of risk factors' returns is computed via EWMA with decay factor 0.97. The future discount factors and forward values are obtained under the assumption of constant risk-free rates in the MPOR horizon.

We backtest the short-term model-free VaR and the FHS VaR against the synthetic P&Ls as explained in the second approach of Section 8.6.1.

Figure 8.1 shows the VaR patterns for the tested portfolios. We can see that the short-term model-free VaR has more breaches then the FHS VaR, however these are of small size and can be entirely removed setting a larger vol-of-vol factor F. Alternatively, one could consider the t-Student framework described in Section 8.4.2. The most noticeable feature of the short-term model-free VaR is its regularity compared to the FHS VaR. In particular, the short-term model-free VaR behaves as we expect after large moves in realized P&Ls, and it also softens its behavior, without big jumps. On the contrary, the FHS VaR is not as consistent (and sometimes seems to move without following market patterns).

Furthermore, the short-term model-free VaR looks more smooth, i.e. less procyclical. To prove this sentence, we compute the peak-to-trough ratio on the whole 2019 dates and the average n-day procyclicality measure (in percentage) for n equal to 1, 5, 10 and 20 days. In particular, the two quantities are computed as

$$\text{Peak-to-trough} = \frac{\max_t\big(-\text{VaR}_{0.99,t}(h)\big)}{\min_t\big(-\text{VaR}_{0.99,t}(h)\big)}$$
$$n\text{-day } \% = \max_t\left(\frac{-\text{VaR}_{0.99,t}(h)}{-\text{VaR}_{0.99,t-n}(h)} - 1\right) \times 100$$

where t ranges in the whole 2019 and we choose an MPOR $h = 1$. Results are displayed in Table 8.4. We see that except for the 1-day procyclicality measure in the calendar spread portfolio, all other procyclicality measures for both portfolios are largely smaller for the short-term model-free VaR, i.e. the latter model is less procyclical than the FHS VaR.

8.6.4 Practical implementation of the neural-SDE model

In this section we consider the neural-SDE model for normalized option prices, and in particular we compare VaR estimations obtained as empirical quantiles on simulations (see Section 8.3.4) and VaR values resulting from the approximated closed formula of Proposition 8.2.

We use the same data as in Section 8.6.3 and compute forward and discount factors similarly.

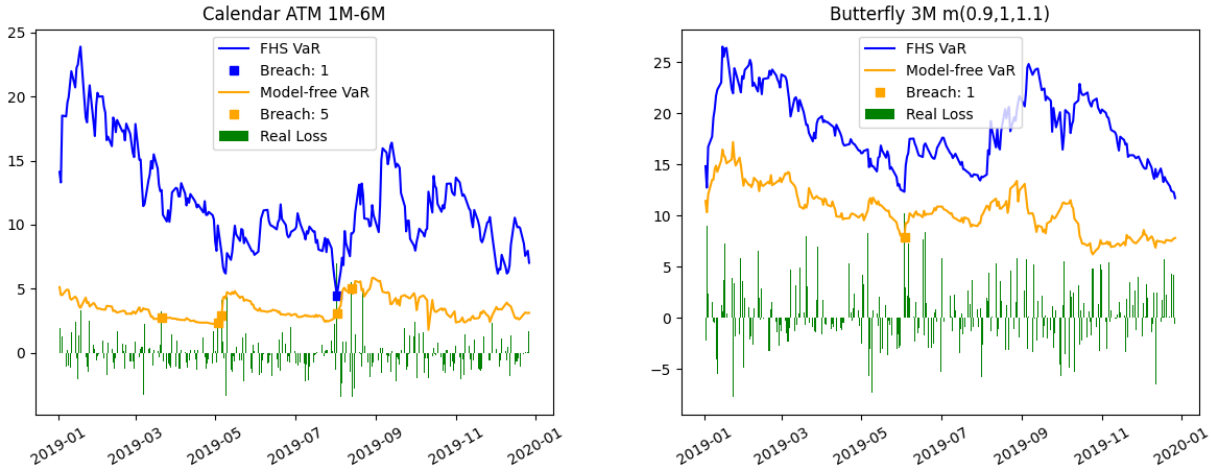

Figure 8.1: Margins obtained with the FHS algorithm (blue) and the short-term model-free VaR (orange), for a calendar spread ATM 1M-6M portfolio (left) and a butterfly spread 3M with moneyness $(0.9, 1, 1.1)$ *portfolio (right).*

	Calendar ATM 1M-6M		Butterfly 3M m$(0.9, 1, 1.1)$	
	FHS	Short-term model-free	FHS	Short-term model-free
Peak-to-trough	5.32	3.27	2.26	2.77
1-day %	42.17	45.38	31.34	17.95
5-day %	90.11	89.04	70.47	44.23
10-day %	143.88	109.28	106.69	52.17
20-day %	139.75	118.53	90.35	44.56

Table 8.4: Comparison between FHS VaR and short-term model-free VaR peak-to-trough ratio and average percentage n-day procyclicality measure for a calendar spread and a butterfly spread portfolios.

Before calibrating the G factors on the time-to-maturity, log-forward moneyness grid identified in Section 8.6.3, historical prices must be interpolated on such a grid. With this aim, we use the same interpolation/extrapolation algorithm consisting in implied volatility's linear interpolation on the space dimension and total variances' linear interpolation in the time direction.

Now, for a fixed date t, the past 5 years historical data (1260 observations) is used to calibrate the G and ξ factors. We choose to calibrate the factors ξ_s for $s \leq t$ in the most efficient way, only looking at the statistical accuracy. In particular, for the fixed time-to-maturity and log-forward moneyness grid, we choose G_0 as the average historical prices and the remaining G_i as the principal components of the residuals between prices and values of G_0. See [18] for a detailed description of the calibration algorithm. The calibration code that we use is the one implemented in the Github repository of the cited

article.[5] Based on calibration accuracy and process time, we decide to take 2 statistical accuracy factors ξ_s.

At this point we have the constant factors G on the fixed grid and an history of factors ξ_s. Furthermore, the calibration uses neural networks to estimate the distributions of S and ξ as in Equation (8.3). In particular, we have today's parameters α_t, β_t, μ_t and σ_t, and for any value of S and ξ, the neural network can predict the corresponding parameters. Observe that we take the covariance matrix $P_t = I_d$ as in [18].

For future computations, the matrix G has to be interpolated outside the fixed time-to-maturity and log-forward moneyness grid. To do so, we firstly interpolate normalized Call prices on the target couple (τ, k) for every historical past day. The interpolation is performed as in the preparation of the initial database, computing implied volatilities and interpolating them linearly on the space direction and linearly in total variance on the time direction. Once all history on (τ, k) is retrieved, the $G(\tau, k)$ factors are the intercept and the coefficients of the linear regression of prices along the history of the ξs.

At present, the two VaR calculation methodologies can be implemented. We test the same two portfolios of Section 8.6.3 consisting of an ATM calendar spread 1M-6M, and a butterfly spread on maturity 3M and moneyness 0.9, 1, 1.1. We consider an MPOR of 1 day and a VaR confidence level $\theta = 0.99$.

For both VaR methodologies, we work under the assumption of constant risk-free rates in the MPOR horizon. Then, the discount factor $\mathrm{DF}_{t+h}(\tau)$ in h days on a time-to-maturity τ is equal to today's discount factor $\mathrm{DF}_t(\tau)$, while the forward value $F_{t+h}(\tau)$ in h days on a time-to-maturity τ becomes $\frac{S_{t+h}}{S_t} F_t(\tau)$.

For the empirical VaR, simulations are performed under the hypothesis that parameters $\alpha_\cdot$, $\beta_\cdot$, $\mu_\cdot$ and $\sigma_\cdot$ are constant between t and $t+h$, following Euler's scheme in Equation (8.16). Starting with the estimation of today's parameters, values of S_{t+h} and ξ_{t+h} are simulated 10000 times. Future normalized prices are computed using the model relation in Equation (8.2) with the G factors evaluated on time-to-maturity $\tau - h$ and log-forward moneyness $k + \log \frac{F_t(\tau)}{F_{t+h}(\tau-h)}$, and the estimated values of ξ_{t+h}. Once the simulated normalized Call prices are computed, they are re-denormalized multiplying by $\mathrm{DF}_t(\tau - h)$ and $F_{t+h}(\tau - h)$. The final VaR is the $1 - \theta$ empirical quantile of P&Ls obtained as difference of simulated future prices and today price.

In the case of VaR obtained via approximated closed formula, in order to be consistent with the empirical VaR, the distribution of S_{t+h} is taken to be a log-normal distribution, so that the used closed formula coincides with Equation (8.21). Derivatives of the components of G with respect to k are computed as the average between backward and forward finite differences.

We plot the percentage ratio between the absolute difference between the two VaR estimations and the portfolio value along year 2019 in Figure 8.2. We can see that the empirical VaR and the approximated formula in Equation (8.21) for the VaR generally have a very small error (about 4% for the calendar portfolio and 1% for the butterfly

[5] https://github.com/vicaws/neuralSDE-marketmodel

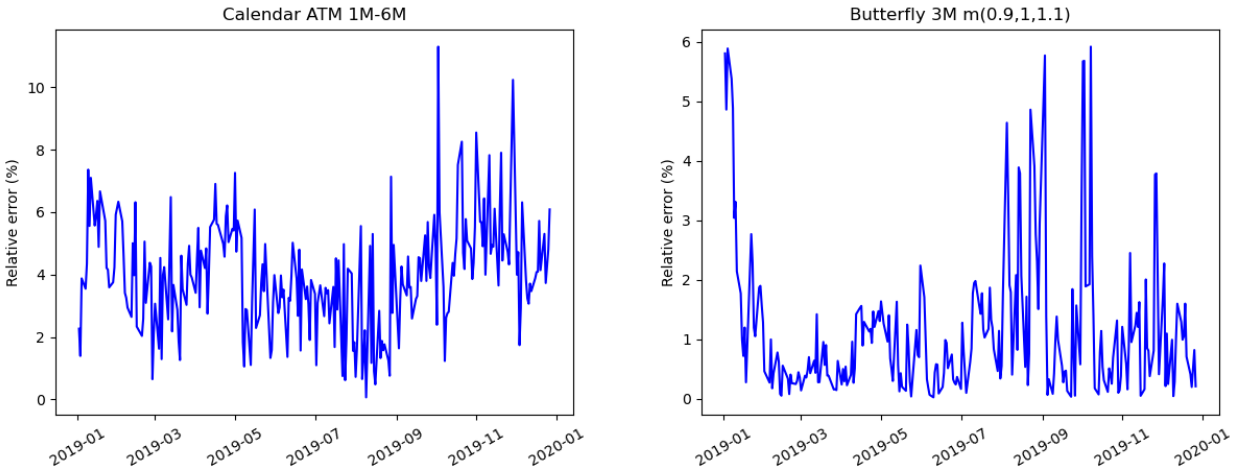

Figure 8.2: Percentage relative error between the empirical neural-SDE VaR calculated as in Section 8.3.4 and the approximated closed formula VaR in Proposition 8.2, for a calendar spread ATM $1M$*-*$6M$ *portfolio (left) and a butterfly spread* $3M$ *with moneyness* $(0.9, 1, 1.1)$ *portfolio (right).*

portfolio), with some higher picks which could reach the 10% of the portfolio. This is due to the fact that the approximated formula is less procyclical than the empirical one and reacts slower to market changes. All in all, the results confirm the consistency of hypothesis in Proposition 8.2.

8.7 Conclusion

We summarize and analyze the methodologies that CCPs currently use for the initial margin of option portfolios. In particular, we compute a quasi-explicit formula for the VaR of option portfolios in the neural-SDE model of [18], and propose a closed asymptotic short-term model-free formula for the VaR at small time horizons.

Based on the numerical experiments that we conduct, we are confident that this new short-term model-free formula could be considered as a candidate for the core component of IM methodologies for option portfolios, duly complemented by Short Option Minimum and Stress Historical VaR components.

8.A Proof of Lemma 8.2

Let us consider the distribution of $\frac{\text{P\&L}}{\sqrt{h}}$ given in Equations (8.14) and (8.16). We write $W_{0,t+h} - W_{0,t} = \sqrt{h}Y$ and $W_{t+h} - W_t = \sqrt{h}X$ where Y is a standard Gaussian random variable and X is a d-dimensional Gaussian random variable with correlation matrix P_t not depending on h.

We first consider the term $\frac{A(h,S_{t+h})}{\sqrt{h}}$ and look at its limit for h going to 0. Since S_{t+h} goes to S_t, both the numerator and the denominator go to 0. We then use L'Hôpital's rule to develop the limit. The derivative of S_{t+h} with respect to h is

$$S_{t+h}\left(\alpha_t - \frac{\beta_t^2}{2} + \frac{\beta_t}{2\sqrt{h}}Y\right)$$

so that the derivative of $A(h, S_{t+h})$ with respect to h is

$$\begin{aligned}
&S_{t+h}\Bigg[\left(\alpha_t - \frac{\beta_t^2}{2} + \frac{\beta_t}{2\sqrt{h}}Y\right)\sum_i \pi_i \mathrm{DF}_t(T_i-(t+h))f(T_i-(t+h))\times \\
&\times\left(G_0\Big(T_i-(t+h), \log\frac{K_i}{f(T_i-(t+h))S_{t+h}}\Big)+\right. \\
&\left.+\,G\Big(T_i-(t+h), \log\frac{K_i}{f(T_i-(t+h))S_{t+h}}\Big)\cdot\xi_t - \mathbb{1}_{\text{Puts}}(i)\right)+ \\
&+\sum_i \pi_i \frac{d}{dh}\Big(\mathrm{DF}_t(T_i-(t+h))f(T_i-(t+h))\Big)\times \\
&\times\left(G_0\Big(T_i-(t+h), \log\frac{K_i}{f(T_i-(t+h))S_{t+h}}\Big)+\right. \\
&\left.+\,G\Big(T_i-(t+h), \log\frac{K_i}{f(T_i-(t+h))S_{t+h}}\Big)\cdot\xi_t - \mathbb{1}_{\text{Puts}}(i)\right)+ \\
&-\sum_i \pi_i \mathrm{DF}_t(T_i-(t+h))f(T_i-(t+h))\left(\partial_\tau G_0\Big(T_i-(t+h), \log\frac{K_i}{f(T_i-(t+h))S_{t+h}}\Big)+\right. \\
&\left.+\,\partial_\tau G\Big(T_i-(t+h), \log\frac{K_i}{f(T_i-(t+h))S_{t+h}}\Big)\cdot\xi_t\right)+ \\
&-\sum_i \pi_i \mathrm{DF}_t(T_i-(t+h))f(T_i-(t+h))\left(\partial_k G_0\Big(T_i-(t+h), \log\frac{K_i}{f(T_i-(t+h))S_{t+h}}\Big)+\right. \\
&\left.+\,\partial_k G\Big(T_i-(t+h), \log\frac{K_i}{f(T_i-(t+h))S_{t+h}}\Big)\cdot\xi_t\right)\times \\
&\times\left(-\frac{\partial_\tau f(T_i-(t+h))}{f(T_i-(t+h))} + \alpha_t - \frac{\beta_t^2}{2} + \frac{\beta_t}{2\sqrt{h}}Y\right)\Bigg]
\end{aligned}$$

where $\mathbb{1}_{\text{Puts}}(i)$ is 1 if the index i refers to a put, otherwise it is null. This quantity explodes for h going to 0 with speed $\frac{\gamma Y}{2\sqrt{h}}$ where

$$\begin{aligned}\gamma = S_t\beta_t \sum_i \pi_i \text{DF}_t(T_i - t) f(T_i - t) \times \\ \times \Big[G_0\Big(T_i - t, \log \frac{K_i}{f(T_i - t)S_t}\Big) + G\Big(T_i - t, \log \frac{K_i}{f(T_i - t)S_t}\Big) \cdot \xi_t - \mathbb{1}_{\text{Puts}}(i) + \\ - \partial_k \Big(G_0\Big(T_i - t, \log \frac{K_i}{f(T_i - t)S_t}\Big) + G\Big(T_i - t, \log \frac{K_i}{f(T_i - t)S_t}\Big) \cdot \xi_t \Big) \Big].\end{aligned}$$

This means that the ratio $\frac{A(h,S_{t+h})}{\sqrt{h}}$ tends to γY, where γ does not depend on h.

We now look at the term $B(h, S_{t+h}) \cdot \frac{\xi_{t+h} - \xi_t}{\sqrt{h}}$. Firstly, the limit of $B(h, S_{t+h})$ for h going to 0 is

$$S_t \sum_i \pi_i \text{DF}_t(T_i - t) f(T_i - t) G\Big(T_i - t, \log \frac{K_i}{f(T_i - t)S_t}\Big)$$

which is simply a sum of Call surfaces and, in general, is different from 0. Secondly, the ratio $\frac{\xi_{t+h} - \xi_t}{\sqrt{h}}$ is equal to $\mu_t\sqrt{h} + \sigma_t \cdot X$ and goes to $\sigma_t \cdot X$ when h tends to 0.

All in all, $\frac{\text{P\&L}}{\sqrt{h}}$ tends to $\gamma Y + B(0, S_t) \cdot \sigma_t \cdot X$ almost surely, hence in law. Then, since the cumulative density function of the random variable $\gamma Y + B(0, S_t) \cdot \sigma_t \cdot X$ has a continuous inverse, all the quantiles of $\frac{\text{P\&L}}{\sqrt{h}}$ converge to the corresponding quantiles of $\gamma Y + B(0, S_t) \cdot \sigma_t \cdot X$ as h tends to 0.

The h-days VaR with confidence level θ is the value of $v(\theta, h)$ solving

$$\begin{aligned}1 - \theta &= P(\text{P\&L} \leq v(\theta, h)) \\ &= P\Big(\frac{\text{P\&L}}{\sqrt{h}} \leq \frac{v(\theta, h)}{\sqrt{h}}\Big).\end{aligned}$$

From the above discussion, we have $v(\theta, h) = u(\theta)\sqrt{h} + o(\sqrt{h})$, where

$$u(\theta) = F^{-1}_{\gamma Y + B(0,S_t)\cdot\sigma_t\cdot X}(1 - \theta).$$

8.B Proof of Proposition 8.2

From the definition in Equation (8.16), S_{t+h} has a log-normal distribution with density

$$p_{S_{t+h}}(s) = \frac{1}{s\beta_t\sqrt{2\pi h}} \exp\left(- \left(\frac{\log \frac{s}{S_t} - (\alpha_t - \frac{\beta_t^2}{2})h}{\beta_t\sqrt{2h}} \right)^2 \right).$$

We look at the RHS of Equation (8.19) when h goes to 0.

We can re-write the integral using a change of variable $y = \frac{\log \frac{s}{S_t} - \left(\alpha_t - \frac{\beta_t^2}{2}\right)h}{\beta_t \sqrt{h}}$ as

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) \Phi\left(\frac{v(\theta,h) - \hat{A}(h, s(h,y))}{\|\hat{B}(h, s(h,y))\|_2}\right) dy$$

where $s(h,y) = S_t \exp\big(y\beta_t\sqrt{h} + (\alpha_t - \frac{\beta_t^2}{2})h\big)$. Observe that the integrand is dominated by the integrable function $\frac{1}{\sqrt{2\pi}} \exp\big(-\frac{y^2}{2}\big)$.

If we prove that for h going to 0 the integrand converges pointwise to a function of the form $\Phi\big(\frac{-c_t y + u(\theta)}{q_t}\big)$ where c_t and q_t do not depend on y, for the Lebesgue's dominated convergence theorem the whole integral converges to

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) \Phi\left(\frac{-c_t y + u(\theta)}{q_t}\right) dy. \tag{8.23}$$

Assuming the proof of the convergence is done (see Section 8.B.1), we can pick-up a pair of independent standard Gaussian random variables X, Y and write the latter expression as

$$E\left[\mathbb{1}\left(X \leq \frac{-c_t Y + u(\theta)}{q_t}\right)\right].$$

The random variable $Z = \frac{q_t X + c_t Y}{\sqrt{c_t^2 + q_t^2}}$ has a standard normal distribution, so that the latter quantity is equal to

$$E\left[\Phi\left(\frac{u(\theta)}{\sqrt{c_t^2 + q_t^2}}\right)\right] = \Phi\left(\frac{u(\theta)}{\sqrt{c_t^2 + q_t^2}}\right).$$

Then, we can finally recover the expression of the initial margin $\mathrm{VaR}_{\theta,t}(h)$ as in Equation (8.21).

8.B.1 Proof of the pointwise convergence

The pointwise convergence of

$$\Phi\left(\frac{v(\theta,h) - \hat{A}(h, s(h,y))}{\|\hat{B}(h, s(h,y))\|_2}\right)$$

to a function of the form $\Phi\big(\frac{-c_t y + u(\theta)}{q_t}\big)$ can be proved firstly observing that we can work under hypothesis of continuous functions, given that normalized prices can be considered to be continuous in time-to-maturity and log-forward moneyness. Also, since we are looking at the limit when h is small, we can use Lemma 8.2 and substitute $v(\theta,h)$ with $u(\theta)\sqrt{h} + o(\sqrt{h})$.

Firstly observe that the matrix b_t does not depend on h, and $\|\hat{B}(h, s(h,y))\|_2 = \sqrt{h}\|B(h, s(h,y)) \cdot \sigma_t \cdot b_t\|_2$ where

$$\begin{aligned} B(h, s(h,y)) = s(h,y) \sum_i \pi_i \mathrm{DF}_t(T_i - (t+h)) f(T_i - (t+h)) \times \\ \times G\Big(T_i - (t+h), \log \frac{K_i}{f(T_i - (t+h)) s(h,y)}\Big). \end{aligned}$$

As shown in Section 8.A, the limit of $B(h, s(h,y))$ for h going to 0 is

$$S_t \sum_i \pi_i \mathrm{DF}_t(T_i - t) f(T_i - t) G\Big(T_i - t, \log \frac{K_i}{f(T_i - t) S_t}\Big)$$

which is different from 0. Then, the ratio $\frac{u(\theta)\sqrt{h} + o(\sqrt{h})}{\|\hat{B}(h,s(h,y))\|_2}$ goes to

$$\frac{u(\theta)}{\|B(0, S_t) \cdot \sigma_t \cdot b_t\|_2} = \frac{u(\theta)}{q_t}$$

in 0.

We now consider the ratio $\frac{\hat{A}(h,s(h,y))}{\sqrt{h}\|B(h,s(h,y)) \cdot \sigma_t \cdot b_t\|_2}$. The function $\hat{A}(h, s(h,y))$ is in turn the sum between $A(h, s(h,y))$ and $B(h, s(h,y)) \cdot m_t(s(h,y))$. The latter term is equal to $B(h, s(h,y)) \cdot (\mu_t h + y\sigma_t \cdot P_{S,\xi,t}\sqrt{h})$, so that its ratio with $\sqrt{h}\|B(h, s(h,y)) \cdot \sigma_t \cdot b_t\|_2$ goes to $\frac{B(0,S_t)\cdot\sigma_t \cdot P_{S,\xi,t}}{q_t} y$.

We shall now focus on the ratio $\frac{A(h,s(h,y))}{\sqrt{h}\|B(h,s(h,y)) \cdot \sigma_t \cdot b_t\|_2}$. Since both the numerator and denominator go to 0 with h, we use L'Hôpital's rule to develop the limit.

The derivative of $B(h, s(h,y))$ with respect to h is

$$\begin{aligned}
& s(h,y)\Bigg[\left(\frac{y\beta_t}{2\sqrt{h}} + \alpha_t - \frac{\beta_t^2}{2}\right) \sum_i \pi_i \mathrm{DF}_t(T_i - (t+h)) f(T_i - (t+h)) \times \\
& \times G\Big(T_i - (t+h), \log \frac{K_i}{f(T_i - (t+h)) s(h,y)}\Big) + \\
& + \sum_i \pi_i \frac{d}{dh}\Big(\mathrm{DF}_t(T_i - (t+h)) f(T_i - (t+h))\Big) \times \\
& \times G\Big(T_i - (t+h), \log \frac{K_i}{f(T_i - (t+h)) s(h,y)}\Big) + \\
& - \sum_i \pi_i \mathrm{DF}_t(T_i - (t+h)) f(T_i - (t+h)) \partial_\tau G\Big(T_i - (t+h), \log \frac{K_i}{f(T_i - (t+h)) s(h,y)}\Big) + \\
& - \sum_i \pi_i \mathrm{DF}_t(T_i - (t+h)) f(T_i - (t+h)) \partial_k G\Big(T_i - (t+h), \log \frac{K_i}{f(T_i - (t+h)) s(h,y)}\Big) \times \\
& \times \left(-\frac{\partial_\tau f(T_i - (t+h))}{f(T_i - (t+h))} + \frac{y\beta_t}{2\sqrt{h}} + \alpha_t - \frac{\beta_t^2}{2}\right)\Bigg]
\end{aligned}$$

and for h going to 0, it explodes with a speed of

$$\begin{aligned}
& \frac{1}{2\sqrt{h}} S_t y \beta_t \sum_i \pi_i \mathrm{DF}_t(T_i - t) f(T_i - t) \times \\
& \times \Big(G\Big(T_i - t, \log \frac{K_i}{f(T_i - t) S_t}\Big) - \partial_k G\Big(T_i - t, \log \frac{K_i}{f(T_i - t) S_t}\Big)\Big).
\end{aligned}$$

Similarly, the derivative of $A(h, s(h, y))$ with respect to h explodes with a speed of

$$\frac{1}{2\sqrt{h}} S_t y \beta_t \sum_i \pi_i \mathrm{DF}_t(T_i - t) f(T_i - t) \times$$
$$\times \Big[G_0\Big(T_i - t, \log \frac{K_i}{f(T_i - t)S_t}\Big) + G\Big(T_i - t, \log \frac{K_i}{f(T_i - t)S_t}\Big) \cdot \xi_t - \mathbb{1}_{\mathrm{Puts}}(i) +$$
$$- \partial_k \Big(G_0\Big(T_i - t, \log \frac{K_i}{f(T_i - t)S_t}\Big) + G\Big(T_i - t, \log \frac{K_i}{f(T_i - t)S_t}\Big) \cdot \xi_t \Big) \Big].$$

Doing the derivative of $\sqrt{h}\|B(h, s(h, y)) \cdot \sigma_t \cdot b_t\|_2$, we find

$$\frac{1}{2\sqrt{h}} \|B(h, s(h, y)) \cdot \sigma_t \cdot b_t\|_2 + \sqrt{h} \frac{\big(\frac{d}{dh} B(h, s(h, y)) \cdot \sigma_t \cdot b_t\big)^T \cdot \big(B(h, s(h, y)) \cdot \sigma_t \cdot b_t\big)}{\|B(h, s(h, y)) \cdot \sigma_t \cdot b_t\|_2},$$

and this explodes with the first term.

All in all, L'Hôpital's rule shows that the limit of $\frac{\hat{A}(h, s(h, y))}{\sqrt{h}\|B(h, s(h, y)) \cdot \sigma_t \cdot b_t\|_2}$ for h going to 0 is $\frac{c_t}{q_t} y$.

Chapter 9

Backtesting Expected Shortfall

Abstract

Since its introduction in 2001, the Expected Shortfall (ES) quickly became the standard risk measure used by financial institutions including central clearing counterparties (CCPs). Indeed, many CCPs switched from the Value at Risk (VaR) to the more conservative ES to compute their initial margins. The need of a sound backtest for the ES arose then naturally. In 2011, the proof that the Expected Shortfall (ES) lacks a property called elicitability has led to the incorrect conclusion that the ES is not backtestable. Three years later, Acerbi and Szekely designed three possible backtests for the ES and, since then, many other backtests have been proposed in the practitioner literature. In this work we study four of these test statistics from both a theoretical and practical point of view and eventually give some advice for CCPs in search of a good backtest for ES.

From:
Zeliade Systems, *Backtesting Expected Shortfall*, Zeliade Systems White Paper https://www.zeliade.com/wp-content/uploads/whitepapers/zwp-011-BacktestingExpectedShortfall.pdf, 2021.

9.1 Structure of the chapter

In this chapter we deal with one of the metrics used by CCPs to measure the counterparty risk: the Expected Shortfall. We firstly compare it to the well known Value-at-Risk in Section 9.2 and point out the pros and cons of its usage. Then, in Section 9.3 we define what a backtest is and discuss and compare different possible Expected Shortfall backtest procedures. We start in Section 9.3.1 with the analysis of the test statistic proposed by Moldenhauer and Pitera in [57]. We try to explain why this statistic is not strictly a proper backtest for the ES, but it rather tests the distribution of the P&Ls. From a theoretical point of view, one should prefer $\bar{Z}_2$ analyzed in Section 9.3.2 and the minimally biased statistic, which we denote by $\bar{Z}_{\text{MB}}$ and study in Section 9.3.3, both proposed by Acerbi and Szekely in the articles [1] and [3]. We also have a look in Section 9.3.4 at the so called $\bar{Z}_3$ statistic of [1] and see that it has the same theoretical

issues than the Moldenhauer and Pitera statistic. From a theoretical point of view, we end up by suggesting the use of $\bar{Z}_{\mathrm{MB}}$ as its own authors do. On the other hand, we will see in Section 9.4 that from a practical point of view the Moldenhauer and Pitera statistic is as good as $\bar{Z}_{\mathrm{MB}}$, at least on the tests we performed.

9.2 Expected Shortfall vs Value-at-Risk

The Value-at-Risk (VaR) has become a standard risk measure for financial risk management due to its conceptual simplicity, ease of computation, and immediate applicability. The VaR measures the maximum potential change in the value of a portfolio with a given probability over a pre-set horizon:

$$\mathrm{VaR}_\alpha\big(\mathrm{P\&L}_d\big) := -\inf\{x | P\left(\mathrm{P\&L}_d \leq x\right) > \alpha\}$$

Nevertheless, the VaR has several conceptual problems:

- It measures only a quantile of the P&Ls distribution and does not account for the losses beyond this level.
- It is not coherent, since it is not subadditive, which implies that the sum of sub-VaRs is not necessarily conservative.

The latter item means that if we split a portfolio into two sub-portfolios and compute the VaR for each sub-portfolio then the sum of the two VaRs can be smaller than the true VaR of the global portfolio.

As an alternative to the VaR risk measure, Artzner et al. (1997) [6] proposed the Expected Shortfall (ES shortly, also called "conditional VaR", "mean excess loss", "beyond VaR", or "tail VaR"). The ES is the conditional expectation of loss given that the loss is beyond the VaR level; that is

$$\mathrm{ES}_\alpha\big(\mathrm{P\&L}_d\big) := E\left[-\mathrm{P\&L}_d \mid \mathrm{P\&L}_d \leq -\mathrm{VaR}_\alpha\big(\mathrm{P\&L}_d\big)\right].$$

The ES is generally considered a more useful risk measure than VaR thanks to its robustness and to the fact that the ES verifies the subadditivity property, as opposed to the VaR. This means that the sum of two sub-ESs is not smaller than the global ES, entailing an inherent conservativeness.

The ES takes into account, by definition, the severity of the tail observations beyond the VaR. This makes the ES a more conservative risk measure than the VaR, for the same confidence level:

$$\mathrm{ES}_\alpha\big(\mathrm{P\&L}_d\big) \geq \mathrm{VaR}_\alpha\big(\mathrm{P\&L}_d\big).$$

Moreover, the ES is more robust: the fact that the ES is an average of all P&Ls beyond the VaR makes its estimation more stable, since a change in a single observation would be mitigated by the rest of the values in the average. In the VaR case, its estimation is driven by a single value (or at most two values if a linear interpolation is used), which means that the VaR would suffer from large jumps when the time window

Figure 9.1: Realized P&Ls VS VaR and ES on the Brazil Stock Market Index (BVSP) for business dates in 2018 and 2019.

moves and extreme observations are included or excluded. This robustness/stability plays an important role in diminishing the procyclicality of the margins, since when extreme market moves happen, the margins would increase slower than for VaR. This prevents, partially, from exacerbating the market stress events.

All these advantages of the ES explain its use by the majors CCPs for their margin computations, to the point where it became an industry standard.

The only weak feature of the ES was the lack of backtesting tests, while the VaR has several robust statistical tests such as the Kupiec and Christoffersen tests. This weakness was remedied by the recently proposed statistical tests, starting from the work of Acerbi and Szekely [1].

We illustrate in Figure 9.1 an example of a VaR and ES computation for the Brazil Stock Market Index (BVSP). We can see that the use of the ES allows to reduce the number of breaches from 8 to 0.

9.3 Backtesting ES

The computation of the ES needs to be backtested, which means that one should check a posteriori whether the risk prediction was correct. This check, when it leads to a negative result, is a good indicator that the ES computation method should be revised. In the past years the backtestability of the ES has been questioned: since Gneiting proved in 2011 [38] that this risk measure lacks a property called elicitability (whereas the pair (VaR, ES) is elicitable, [30]), some mathematicians concluded that the ES was not backtestable. However, in 2014 Acerbi and Szekely proved in [1] that this is not the case, by simply finding some backtest statistics for the ES. Since then, many articles regarding the backtestability of the ES have been published but a lot of them lack of

a proper definition of backtestability and, as a consequence, the tests proposed are not theoretically rigorous. For this reason we start the current section by mathematically defining what backtesting the ES means.

When we talk about backtesting a statistic, we refer to Definition 3.1 of [2]. In particular, we say that

Definition 9.1. *The statistic ES is backtestable if there exists a backtest function $Z(e, v, x)$ such that*

- $E_H[Z(e, v, X)] = 0$ *iff* $e = -E_H[X|X \leq -v]$
- $E_H[Z(e_1, v, X)] < E_H[Z(e_2, v, X)]$ *if* $e_1 < e_2$

for a fixed v.

This means that when we underestimate the value of the ES, the sign of the backtest function will be negative on average.

With these tools, we can then define the backtest test as

$$\bar{Z}(X) = \frac{1}{N} \sum_{d=1}^{N} Z(\hat{\mathrm{ES}}_d^\alpha, \hat{\mathrm{VaR}}_d^\alpha, X_d)$$

where X, $\hat{\mathrm{ES}}^\alpha$ and $\hat{\mathrm{VaR}}^\alpha$ are the vectors of respectively the valuations of a portfolio (e.g. P&Ls), the estimated ESs and the estimated VaRs. From the previous definition, we have that under the null hypothesis $H_0 : \mathrm{ES}_d^\alpha = \hat{\mathrm{ES}}_d^\alpha$, it holds $E_{H_0}[Z(\hat{\mathrm{ES}}^\alpha, \hat{\mathrm{VaR}}^\alpha, X)] = 0$ while under the alternative hypothesis of underestimation of the ES, $H_1 : \mathrm{ES}_d^\alpha \geq \hat{\mathrm{ES}}_d^\alpha \ \forall d \wedge \exists d : \mathrm{ES}_d^\alpha > \hat{\mathrm{ES}}_d^\alpha$, it holds

$$E_{H_1}[Z(\hat{\mathrm{ES}}^\alpha, \hat{\mathrm{VaR}}^\alpha, X)] < E_{H_1}[Z(\mathrm{ES}^\alpha, \mathrm{VaR}^\alpha, X)] = 0 = E_{H_0}[Z(\hat{\mathrm{ES}}^\alpha, \hat{\mathrm{VaR}}^\alpha, X)]$$

(note that here we are supposing $\hat{\mathrm{VaR}}^\alpha = \mathrm{VaR}^\alpha$ but this is not needed if we ask in the definition of backtestability that $E_H[Z(e, v_1, X)] < E_H[Z(e, v_2, X)]$ if $v_1 < v_2$ and add to the alternative hypothesis the requirement $\mathrm{VaR}_d^\alpha \geq \hat{\mathrm{VaR}}_d^\alpha \ \forall d$).

Then, in order to backtest the ES, one has to compute the value of the realized $\bar{Z}(X)$ and compare it with a threshold value, using the p-test. The p-value is the probability, under the null hypothesis, that the test statistic $\bar{Z}(X)$ is at least as extreme as the realized value of $\bar{Z}(X)$. Since we are considering a one-sided left-tail test statistic $\bar{Z}$, the p-value is $P_{H_0}(\bar{Z}(\tilde{X}) < \bar{Z}(X))$ where $\tilde{X}$ is distributed as in the null hypothesis. A test statistic $\bar{Z}(X)$ is rejected if the p-value is smaller than a level ϕ (generally about 5%). In other words, the test statistic is rejected if the realized value of $\bar{Z}(X)$ is smaller than the ϕ-quantile of $\bar{Z}(\tilde{X})$.

This threshold can be empirically obtained by repeating M times the following steps:

1. Simulate a N-vector of values $\tilde{X}$ under the distribution of the null hypothesis.
2. Calculate $\bar{Z}(\tilde{X})$ using the already computed $\hat{\mathrm{ES}}_d^\alpha$, $\hat{\mathrm{VaR}}_d^\alpha$ for $d = 1, \ldots, N$.

At this point the ϕ-quantile can be calculated as $\bar{Z}(\tilde{X})_{([x])} + (x - [x])(\bar{Z}(\tilde{X})_{([x]+1)} - \bar{Z}(\tilde{X})_{([x])})$ where $x = \frac{\phi}{100}(M-1)+1$. The realized value $\bar{Z}(X)$ is then accepted iff it is greater than the threshold.

We will refer to type I and type II errors as

1. Type I error: when $\hat{\text{ES}}^{\alpha}$ is correct but it is rejected.
2. Type II error: when $\hat{\text{ES}}^{\alpha}$ underestimates the real ES but it is accepted.

In the following subsections we will discuss and compare different possible Expected Shortfall backtest procedures. We start with the analysis of the test statistic proposed by Moldenhauer and Pitera in [57]. We try to explain why this statistic is not strictly a proper backtest for ES, in the sense of Definition 9.1, but rather tests the distribution of the P&Ls. Then, from a theoretical point of view, one should prefer $\bar{Z}_2$ and the minimally biased statistic, which we denote with $\bar{Z}_{\text{MB}}$, proposed by Acerbi and Szekely in the articles [1, 3]. We also have a look at the so called $\bar{Z}_3$ statistic in [1] and see that it has the same theoretical issues than the Moldenhauer and Pitera statistic. From a theoretical point of view, we end up by suggesting the use of $\bar{Z}_{\text{MB}}$ as its own authors do. On the other hand, we will see in the next section that from a practical point of view the Moldenhauer and Pitera statistic is as good as $\bar{Z}_{\text{MB}}$, at least on the tests we performed.

9.3.1 Moldenhauer and Pitera test statistic

This test was proposed by Moldenhauer and Pitera in [57].

Theoretical presentation

Let X_d denote the random process of the valuation of a portfolio (e.g. the P&L) and let ES_d^{α} denote the computed Expected Shortfall value for the probability α at day d. We define the random process

$$Y = X + \text{ES}^{\alpha}$$

to be the secured position. Alternatively, the definition $Y = \frac{X+\text{ES}^{\alpha}}{\text{ES}^{\alpha}}$ can be used. With this choice, the whole following discussion does not change.

The ES test statistic used by Moldenhauer and Pitera is

$$G(X, \text{ES}^{\alpha}) = \sum_{k=1}^{N} \mathbb{1}_{(Y_{[1]}+\cdots+Y_{[k]}<0)}$$

where N is the number of days in the observation window and the random process $Y_{[d]}$ denotes the ordered statistic of Y_d. Note that the authors define this statistic divided by N but for stability properties, we do not do this division (see Section 9.3.1).

In order to check whether this is a good backtest for the ES, we need to see what happens to the statistic when the ES is underestimated. Suppose that the calculated value of the ES at time d is $\hat{\text{ES}}_d^{\alpha}$. We set the null hypothesis to be $H_0 : \text{ES}_d^{\alpha} = \hat{\text{ES}}_d^{\alpha}\ \forall d$

while the alternative hypothesis will be $H_1 : \mathrm{ES}_d^\alpha \geq \hat{\mathrm{ES}}_d^\alpha \ \forall d \wedge \exists d : \mathrm{ES}_d^\alpha > \hat{\mathrm{ES}}_d^\alpha$. Note that if we use a unique method to evaluate the ES, the distributions of X are different in the two hypothesis.

Under the alternative hypothesis note that $X_d + \hat{\mathrm{ES}}_d^\alpha \leq X_d + \mathrm{ES}_d^\alpha$ for every d so that $(X_d + \hat{\mathrm{ES}}_d^\alpha)_{[k]} \leq (X_d + \mathrm{ES}_d^\alpha)_{[k]}$ for every k and $(X_d + \hat{\mathrm{ES}}_d^\alpha)_{[1]} + \cdots + (X_d + \hat{\mathrm{ES}}_d^\alpha)_{[k]} \leq (X_d + \mathrm{ES}_d^\alpha)_{[1]} + \cdots + (X_d + \mathrm{ES}_d^\alpha)_{[k]}$ for every k and the inequality is strict for some k. Then

$$\begin{aligned} E_{H_1}[G(X, \hat{\mathrm{ES}}^\alpha)] &= \sum_{k=1}^{N} P_{H_1}((X_d + \hat{\mathrm{ES}}_d^\alpha)_{[1]} + \cdots + (X_d + \hat{\mathrm{ES}}_d^\alpha)_{[k]} < 0) \\ &> \sum_{k=1}^{N} P_{H_1}((X_d + \mathrm{ES}_d^\alpha)_{[1]} + \cdots + (X_d + \mathrm{ES}_d^\alpha)_{[k]} < 0) \\ &= E_{H_1}[G(X, \mathrm{ES}^\alpha)]. \end{aligned}$$

What we found is that if we underestimate the ES value, the statistic G will have on average a value greater than its true value, which depends of course also on the distribution of X. This test statistic is then a right-tail one, so that signs in Definition 9.1 should inverted, and the realized value of the statistic is accepted iff it is smaller than ϕ. In order to find a threshold, however, we cannot use the value $E_{H_1}[G(X, \mathrm{ES}^\alpha)]$, since we do not know it. What is fundamental to prove is rather that $E_{H_1}[G(X, \hat{\mathrm{ES}}^\alpha)] > E_{H_0}[G(X, \hat{\mathrm{ES}}^\alpha)]$. This could be done if, for example, the statistic G is constructed in such a way that $G(X, \mathrm{ES}^\alpha) = 0$ iff ES^α is the true value of the ES. In this case then $E_{H_1}[G(X, \mathrm{ES}^\alpha)] = 0 = E_{H_0}[G(X, \hat{\mathrm{ES}}^\alpha)]$ and the required inequality would be automatically achieved. After, we could proceed with setting the threshold value to be the empirical ϕ-quantile obtained by simulations of $G(X, \hat{\mathrm{ES}}^\alpha)$. The requirement that a test statistic is null at the true value of the backtested quantity is also one of the requirements in Definition 9.1.

Theoretical misspecification of the backtest

In Appendix I we present two counterexamples which show that from a strict theoretical perspective, the G statistic is not the best choice for backtesting the ES, according to Definition 9.1.

In the first example, we show that $E_{H_1}[G(X, \hat{\mathrm{ES}}^\alpha)] > E_{H_0}[G(X, \hat{\mathrm{ES}}^\alpha)]$ does not imply $\hat{\mathrm{ES}}^\alpha < \mathrm{ES}^\alpha$. This means that if we calculate the threshold as the ϕ-quantile of the simulated vector of G and then accept $\hat{\mathrm{ES}}^\alpha$ iff $G(X, \hat{\mathrm{ES}}^\alpha)$ is smaller than the threshold, then we could be rejecting $\hat{\mathrm{ES}}^\alpha$ even if it correctly overestimates the real ES^α. This causes an higher probability to make type I errors.

The second example proves that the very general hypothesis $H_1 : \mathrm{ES}_d^\alpha \geq \hat{\mathrm{ES}}_d^\alpha \ \forall d \wedge \exists d : \mathrm{ES}_d^\alpha > \hat{\mathrm{ES}}_d^\alpha$ does not imply that $E_{H_1}[G(X, \hat{\mathrm{ES}}^\alpha)] > E_{H_0}[G(X, \hat{\mathrm{ES}}^\alpha)]$. This fact could cause the error of accepting $\hat{\mathrm{ES}}^\alpha$ even if it underestimates the real ES^α and so an error of type II.

These are very easy examples since we take $N = 1$ but even with only one variable X it is possible to show that the statistic G does not satisfy Definition 9.1 of a backtest function. Why then does it seem to work properly in the article of Moldenhauer and Pitera? We think that rather than being a backtest for the ES, it is a backtest for the generic distribution of X, with similar hypothesis as in Section 9.3.4. Indeed, we prove this fact in Section 9.B.

The hypothesis used does not lead to the desired result in the previous examples. Then, if we restricted ourselves to the strict theoretical aspect, the G statistic would not be considered. On the other hand, from a practical point of view, the P&Ls' distributions are generally approximated by t-Student, whose tails (in particular for negative values) can be compared by stochastic dominance. In particular, if P_{ν_1} and P_{ν_2} are the distributions of two t-Student with $\nu_1 < \nu_2$ degrees of freedom, then $P_{\nu_1} \preceq P_{\nu_2}$. This means that if we set the degrees of freedom of the distribution of X to be higher than in reality, the statistic G will correctly signal it.

How to avoid simulations

The G statistic, whatever it backtests, is rather robust with respect to the underlying distribution for N small enough and ϕ not too high and so the threshold can be simulated by standard normal distributions for X. The threshold can be calculated as follows:

1. Compute ES^α for the standard normal distribution (with closed formulas).
2. Iterate M times the following steps:
   a) Simulate a N-vector of values $\tilde{X}$ under the standard normal distribution.
   b) Calculate $G(\tilde{X}, \mathrm{ES}^\alpha)$ using the already computed ES^α.
3. Take the ϕ-quantile of the values of $G(\tilde{X}, \mathrm{ES}^\alpha)$.

In particular, for $\alpha = 0.5\%$ and $\phi = 95\%$, the threshold can be set at 6. Note that in this case, if we use a t-Student distribution for X, the result is still 6.

It must be remarked that increasing N or ϕ, the statistic G is not stable anymore and its threshold cannot be approximated in this way but it must be computed as explained in Section 9.3. Indeed, from Table 9.1 we can see that the thresholds for G under a t-Student distribution with 5 degrees of freedom or a standard normal distribution for X can drastically change (we set $\alpha = 0.5\%$).

	N	$\phi(\%)$	Normal	t-Student
0	500	95.00	6	6
1	500	99.99	12	17
2	1000	95.00	10	10
3	1000	99.99	18	24
4	2000	95.00	17	18
5	2000	99.99	27	35

Table 9.1: Thresholds of G for $\alpha = 0.5\%$.

9.3.2 Acerbi and Szekely $\bar{Z}_2$ statistic

This test was proposed by Acerbi and Szekely in their 2014 article [1].

Theoretical presentation

Define the backtest function

$$Z_2(e, v, x) = \frac{x \mathbb{1}_{\{x+v<0\}}}{\alpha e} + 1.$$

Then, under the hypothesis that $\mathrm{VaR}^\alpha(X) = v$ and $\mathrm{ES}^\alpha(X) = e$ it holds $E[Z_2(e, v, X)] = 0$. Furthermore Z_2 is strictly increasing with v and with e, meaning that when $E[Z_2(e, v, X)] < 0$, the computed VaR v and/or the computed Expected Shortfall e underestimate the real ones.

A natural test statistic for the calculated value $\hat{\mathrm{ES}}^\alpha$ can then be chosen as

$$\bar{Z}_2(X) = \frac{1}{N} \sum_{d=1}^{N} Z_2(\hat{\mathrm{ES}}_d^\alpha, \hat{\mathrm{VaR}}_d^\alpha, X_d).$$

It is easy to see that, under the null hypothesis of correctly chosen $\hat{\mathrm{ES}}_d^\alpha$, the mean value of $\bar{Z}_2(X)$ is 0. Otherwise, under the alternative hypothesis of underestimation of the risk

$$H_1 : \mathrm{ES}_d^\alpha \geq \hat{\mathrm{ES}}_d^\alpha \ \forall d \wedge \exists d : \mathrm{ES}_d^\alpha > \hat{\mathrm{ES}}_d^\alpha$$
$$\mathrm{VaR}_d^\alpha \geq \hat{\mathrm{VaR}}_d^\alpha \ \forall d,$$

it holds $E_{H_1}[\bar{Z}_2(X)] < 0 = E_{H_0}[\bar{Z}_2(X)]$.

This means that in contrast with the Moldenhauer and Pitera test, the $\bar{Z}_2$ statistic correctly backtests the ES, following our definition of backtestability.

How to avoid simulations

For fixed α and ϕ, it is possible to numerically check that the thresholds for the $\bar{Z}_2$ statistic in case of t-Student distributions are quite stable through the νs. The threshold values for $\alpha = 0.5\%$ and $\phi = 5\%$ are reported in Table 9.2 (here we do 500000 simulations).

ν	Threshold
3	-1.3
5	-1.2
10	-1.2
100	-1.1
1000	-1.1

Table 9.2: Thresholds of $\bar{Z}_2$ for $\alpha = 0.5\%$ and $\phi = 5\%$.

It follows that for this test statistic one can take as fixed threshold a value of -1.2 avoiding to calculate it.

9.3.3 Acerbi and Szekely $\bar{Z}_{\text{MB}}$ statistic

This test was proposed by Acerbi and Szekely in their 2017 article [2].

Theoretical presentation

Following the steps for $\bar{Z}_2$, the authors define a different test statistic. This time the backtest function is

$$Z_{\text{MB}}(e, v, x) = e - v + \frac{(x+v)\mathbb{1}_{\{x+v<0\}}}{\alpha}.$$

As before, if $\text{VaR}^\alpha(X) = v$ and $\text{ES}^\alpha(X) = e$, then $E[Z_{\text{MB}}(e, v, X)] = 0$ and Acerbi and Szekely show in section 4.2 of [2] that $E[Z_{\text{MB}}(e, v, X)] < 0$ when the calculated $\hat{\text{ES}}^\alpha = e$ underestimates the real one, no matter the value of v.

The corresponding test statistic for $\hat{\text{ES}}^\alpha$ is

$$\bar{Z}_{\text{MB}}(X) = \frac{1}{N}\sum_{d=1}^{N} Z_{\text{MB}}(\hat{\text{ES}}_d^\alpha, \hat{\text{VaR}}_d^\alpha, X_d).$$

Setting the less strict alternative hypothesis $H_1 : \text{ES}_d^\alpha \geq \hat{\text{ES}}_d^\alpha \ \forall d \wedge \exists d : \text{ES}_d^\alpha > \hat{\text{ES}}_d^\alpha$, it holds again $E_{H_1}[\bar{Z}_{\text{MB}}(X)] < 0 = E_{H_0}[\bar{Z}_{\text{MB}}(X)]$ and the ES can be backtested as in the previous example.

This statistic is preferred by Acerbi and Szekely since it presents a smaller sensitivity to VaR predictions. In particular, the test statistic $\bar{Z}_2$ could face type I and type II errors with more probability than the test statistic $\bar{Z}_{\text{MB}}$ if the prediction $\hat{\text{VaR}}^\alpha$ is not correct. This fact is well explained in [3]. Indeed, under the real world probability, it holds

$$E[Z_{\text{MB}}(\hat{\text{ES}}^\alpha, \hat{\text{VaR}}^\alpha, X)] = \hat{\text{ES}}^\alpha - \text{ES}^\alpha - B(\hat{\text{VaR}}^\alpha)$$

where the bias B is defined as

$$B(v) = E\left[v - \frac{(X+v)\mathbb{1}_{\{X+v<0\}}}{\alpha}\right] - \text{ES}^\alpha.$$

From [4], we have the representation

$$\text{ES}^\alpha = \min_v E\left[v + \frac{(X+v)\mathbb{1}_{\{X+v<0\}}}{\alpha}\right]$$

so that the bias is positive for any v and it vanishes iff the VaR prediction is correct. Authors show that it is small for small discrepancies of v to VaR^α, so that the $\bar{Z}_{\text{MB}}$ test statistic acquires a fundamental role: it can be used as a direct estimation of the discrepancy between the predicted and the real ES.

9.3.4 Acerbi and Szekely $\bar{Z}_3$ statistic

We now consider another statistic which does not directly backtest the computed value of the ES but which rather backtests the distribution of X used to evaluate the ES. This test was proposed by Acerbi and Szekely in their 2014 article [1].

Theoretical presentation

In particular call P_d the predicted distribution of X_d used to evaluate the VaR and the ES, and call Q_d the real unknown distribution of X_d. We put

$$H_0 : Q_d = P_d \ \forall d$$
$$H_1 : Q_d \preceq P_d \ \forall d \wedge \exists d : Q_d \prec P_d$$

where $\preceq$ denotes that the left hand side is first order stochastically dominated by the right hand side. This is equivalent to say that the CDF of Q_d is no smaller than the CDF of P_d (denoted with F_{P_d}) at every point and that for every non-decreasing function u it holds $\int u(x)\, dQ_d(x) \leq \int u(x)\, dP_d(x)$. As a consequence, both VaR and ES are underestimated under P_d.

If the test ends up to accept the null hypothesis, then it is possible to evaluate $\hat{\mathrm{ES}}_d^\alpha$ through the formula

$$\hat{\mathrm{ES}}_d^\alpha = \hat{\mathrm{ES}}_M^\alpha(Y^d) = -\frac{1}{[M\alpha]} \sum_{i=1}^{[M\alpha]} Y_{[i]}^d$$

where M is a big number (e.g. $M = N$ if an historical simulation is used) and Y^d is an M-vector of simulated variables distributed as P_d.

The test statistic used is

$$\bar{Z}_3 = -\frac{1}{N} \sum_{d=1}^{N} \frac{\hat{\mathrm{ES}}_N^\alpha(F_{P_d}^{-1}(U))}{E_V[\hat{\mathrm{ES}}_N^\alpha(F_{P_d}^{-1}(V))]} + 1$$

where U is an iid N-vector such that $U_d = F_{P_d}(X_d)$ while V is an iid N-vector of variables $U([0,1])$. Denoting a regularized incomplete beta function as $I_x(a,b)$, the denominator can be analytically computed as

$$E_V[\hat{\mathrm{ES}}_N^\alpha(F_{P_d}^{-1}(V))] = -\frac{N}{[N\alpha]} \int_0^1 I_{1-p}(N - [N\alpha], [N\alpha]) F_{P_d}^{-1}(p)\, dp.$$

This entails that $E_{H_0}[\bar{Z}_3] = 0$ and $E_{H_1}[\bar{Z}_3] < 0$.

This test statistic is very general and its alternative hypothesis does not directly involve the computed ES: this means that it is not a backtest for the ES. Furthermore, it is not as straightforward as the other statistics considered so we do not suggest its use for the precise purpose of backtesting the ES at least.

9.3.5 Conclusion

We can sum up the pros and cons of each test statistic:

- Moldenhauer and Pitera test statistic G:
  - Pros: the threshold can be calculated taking a standard normal distribution for X.
  - Cons: it does not satisfy Definition 9.1 of a backtest function.
- Acerbi and Szekely $\bar{Z}_2$:
  - Pros: extremely easy to be implemented, the threshold can be calculated taking a t-Student (with e.g. $\nu = 5$ degrees of freedom) distribution for X.
  - Cons: it could face type II errors.
- Acerbi and Szekely $\bar{Z}_{\text{MB}}$:
  - Pros: extremely easy to be implemented, it is very little influenced by the VaR predictions.
  - Cons: the threshold must be evaluated through simulations of the distribution of X.
- Acerbi and Szekely $\bar{Z}_3$:
  - Pros: not many.
  - Cons: it is the most difficult to be implemented, it is not a proper backtest for the ES.

The best theoretical choice is the $\bar{Z}_{\text{MB}}$ statistic because it correctly tests the ES, it is very easy to be implemented and its little sensitivity to VaR predictions makes it a corrective model for ES predictions. From the practical point of view, however, the fact that the threshold cannot be approximated by standard distribution requires, if using a Filtered Historical Simulation method, to store all the simulation of X used to compute the ES. The statistics G and $\bar{Z}_2$ do not face this problem, although G lacks some theoretical justifications and it is a little bit more difficult to be implemented, while $\bar{Z}_2$ is not as precise as $\bar{Z}_{\text{MB}}$ in the choice of accepting or rejecting the computed ES.

9.4 Quantitative assessment

9.4.1 Tests on simulated data

We compare here the power of the four statistics G, $\bar{Z}_2$, $\bar{Z}_{\text{MB}}$ and $\bar{Z}_3$. The simulated distribution of the P&L process are t-Student with ν degrees of freedom and the null and alternative hypothesis change for the number of degrees of freedom of the distribution. In particular, $H_0 : \nu = \nu_0$ while $H_1 : \nu = \nu_1$. The power of a statistic is the probability

to reject the null hypothesis when indeed the alternative hypothesis is correct. This means that the higher the power, the better is the test statistic in terms of avoiding type II errors. To evaluate the power of the tests under a null hypothesis H_0 which underestimates the risk, it is necessary to have $\nu_1 < \nu_0$.

We can use the function power for two purposes:

1. Evaluation of the probability to commit type I errors: this error arises when the null hypothesis is rejected even if it is true and the probability at which it arises is equal to the significance level ϕ. In order to check whether the function power is correctly written, we put $\nu_1 = \nu_0$ and see if its value is actually ϕ.

2. Evaluation of the probability to commit type II errors: this probability is the difference between 1 and the power.

We set the level of the ES at $\alpha = 99.5\%$, the significance level of the test at $\phi = 5\%$ (which means that if the p-value is less than ϕ, the statistic is rejected) and the number of days in the observation window at $N = 500$. To calculate the threshold level for the statistic we compute 250000 simulations while to calculate the power, that is the rate of rejected statistics, we do 100000 simulation.

In order to use the same input data as the Acerbi and Szekely's statistics, instead of computing the G statistic, we calculate $-G$. Furthermore, as Moldenhauer and Pitera suggest, we use the relative secured positions: $Y = \frac{X+\mathrm{ES}^\alpha}{\mathrm{ES}^\alpha}$.

We add also the power column for the $\bar{Z}_2$ statistic with a precomputed threshold equal to -1.2, calling it $\bar{Z}_2$ bis.

	ν in H_0	ν in H_1	G	$\bar{Z}_2$	$\bar{Z}_{\mathrm{MB}}$	$\bar{Z}_3$	$\bar{Z}_2$ bis
0	3	3	7.4	4.9	4.9	5.1	6.1
1	5	3	76.0	76.7	68.8	55.4	76.4
2	10	3	99.5	99.5	99.3	97.2	99.5
3	100	3	100.0	100.0	100.0	100.0	100.0
4	5	5	6.9	5.0	5.0	5.0	5.1
5	10	5	71.0	67.7	66.2	54.8	66.7
6	100	5	99.4	99.0	99.2	97.7	98.7
7	10	10	5.9	5.0	5.1	5.0	4.6
8	100	10	75.0	70.0	73.4	64.4	66.9
9	100	100	5.3	5.0	5.0	5.2	4.3

Table 9.3: Power of G, $\bar{Z}_2$, $\bar{Z}_{MB}$, $\bar{Z}_3$ and $\bar{Z}_2$ bis (%) for a value of $\phi = 5\%$.

We can see from Table 9.3 that G, $\bar{Z}_2$ and $\bar{Z}_{\mathrm{MB}}$'s powers are definitely higher than $\bar{Z}_3$'s one. Also, the evaluation of the latter statistic requires much more time than the former, whose computation times are similar.

The computation time of $\bar{Z}_2$ bis is definitely lower since it does not require the calculation of the threshold. Its power cannot actually be compared with the other

statistics' ones since it is evaluated on significance levels which differ from 5%. In particular, the significance level is 6.1% for $\nu = 3$, 4.5% for $\nu = 10$ and 4.3% for $\nu = 100$. A higher (lower) significance level leads to a higher (lower) power so setting these levels of significance for the other statistics would also increase (decrease) their power. For this reason we suggest the use of a precomputed threshold statistic only in the case of very long computation times, which in these examples do not actually arise.

This argument holds also for G, whose significance levels are somehow higher than 5% for small values of ν, which means that the corresponding powers will also be higher. Then, the real power of G is not as high as it seems. The reason why G faces a higher probability of type I error is explained in Section 9.3.1.

For a matter of completeness we report in Tables 9.4 and 9.5 the power of $\bar{Z}_2$ and $\bar{Z}_{\mathrm{MB}}$ for the actual significance levels used by $\bar{Z}_2$ bis (Table 9.4) and by G (Table 9.5). We see that their power changes as predicted.

	ϕ	$\bar{Z}_2$	$\bar{Z}_{\mathrm{MB}}$	$\bar{Z}_2$ bis
0	6.1	6.0	6.2	5.1
1	6.1	78.6	72.1	55.4
2	6.1	99.6	99.5	97.2
3	6.1	100.0	100.0	100.0
4	5.1	5.1	5.2	5.0
5	5.1	68.2	66.9	54.8
6	5.1	99.0	99.2	97.7
7	4.6	4.5	4.7	5.0
8	4.6	68.3	72.3	64.4
9	4.3	4.2	4.2	5.2

Table 9.4: Power of $\bar{Z}_2$, $\bar{Z}_{MB}$ and $\bar{Z}_2$ bis with different values of ϕ (%).

	ϕ	$\bar{Z}_2$	$\bar{Z}_{\mathrm{MB}}$	G
0	7.4	7.4	7.3	7.4
1	7.4	80.8	75.0	76.0
2	7.4	99.7	99.5	99.5
3	7.4	100.0	100.0	100.0
4	6.9	6.9	6.9	6.9
5	6.9	72.6	71.2	71.0
6	6.9	99.2	99.4	99.4
7	5.9	5.9	5.8	5.9
8	5.9	72.1	75.4	75.0
9	5.3	5.2	5.2	5.3

Table 9.5: Power of $\bar{Z}_2$, $\bar{Z}_{MB}$ and G with different values of ϕ (%).

9.4.2 Tests on historical data

The evaluation of the ES is done as described in Section 9.3 and, as before, we denote $\hat{\mathrm{ES}}_d^\alpha$ its estimated value at day d. We have $X_d = S_{d+1} - S_d$ where S_d is the value of the asset at day d. The statistic $\bar{Z}$ will be evaluated on these realized values of X:

$$\bar{Z}(X) = \frac{1}{N}\sum_{d=1}^{N} Z(\hat{\mathrm{ES}}_d^\alpha, \hat{\mathrm{VaR}}_d^\alpha, X_d).$$

In order to calculate the threshold, it is necessary to simulate the test statistic M times and then to compare the ϕ-quantile with $\bar{Z}(X)$. How can $\bar{Z}$ be simulated if we use an historical distribution? In order to calculate $\hat{\mathrm{ES}}_d^\alpha$ through an historical method as HS or FHS, we need to simulate M scenario of X_d for every d, taking into account the history of X. Since we do an historical simulation, M corresponds to the number of data available (for us, ten years so $M = 2500$). We can then use the same simulations to compute

$$\bar{Z}_k = \frac{1}{N}\sum_{d=1}^{N} Z(\hat{\mathrm{ES}}_d^\alpha, \hat{\mathrm{VaR}}_d^\alpha, X_{d,k})$$

for each $k = 1, \ldots, M$ and finally take the ϕ-quantile of the $\bar{Z}$ vector.

We now let run the G, $\bar{Z}_2$ and $\bar{Z}_{\mathrm{MB}}$ backtests on some portfolios on equity products obtained from Yahoo Finance in the period from 02/12/2005 to 02/12/2020.

	G	Threshold	Accepted
AXJO	-9	-9	No
BVSP	0	-9	Yes
FCHI	-3	-12	Yes
GDAXI	-2	-14	Yes
GSPC	-9	-11	Yes
GSPTSE	-3	-14	Yes
KS11	-7	-15	Yes
MXX	-8	-10	Yes
SSMI	-3	-11	Yes
TWII	-12	-11	No

Table 9.6: Accepted ES for G.

	$\bar{Z}_2$	Threshold	Accepted
AXJO	-2.5397	-2.76928	Yes
BVSP	-1.67839	-2.27365	Yes
FCHI	-0.527363	-3.96901	Yes
GDAXI	-0.458829	-4.07022	Yes
GSPC	-2.1588	-2.95039	Yes
GSPTSE	-0.160738	-3.80838	Yes
KS11	-1.77876	-4.18092	Yes
MXX	-1.52592	-2.83534	Yes
SSMI	-0.481356	-2.68431	Yes
TWII	-1.39444	-3.85184	Yes

Table 9.7: Accepted ES for $\bar{Z}_2$.

	$\bar{Z}_{\text{MB}}$	Threshold	Accepted
AXJO	-82.1309	-76.2563	No
BVSP	-396.507	-2533.57	Yes
FCHI	-15.1573	-132.957	Yes
GDAXI	-5.50087	-318.857	Yes
GSPC	-46.6548	-64.55	Yes
GSPTSE	4.73223	-459.015	Yes
KS11	-26.4939	-59.7472	Yes
MXX	-799.427	-904.342	Yes
SSMI	-15.3335	-228.596	Yes
TWII	-381.284	-197.07	No

Table 9.8: Accepted ES for $\bar{Z}_{MB}$.

From Tables 9.6 to 9.8 it can be seen that the backtests G and $\bar{Z}_{\text{MB}}$ lead to the same results. This is somehow surprising since from the theoretical point of view we have proven that G does not satisfy theoretical conditions of Definition 9.1 of a backtest function. However, our observations suggest that G is a backtest for the whole distribution of the P&Ls and so it could have the same results of a backtest for the ES, when the distributions employed for the evaluation of the ES are misspecified.

The backtest $\bar{Z}_2$ accepts the estimated ES for a higher number of portfolios and if G or $\bar{Z}_{\text{MB}}$ accept the ES, then also $\bar{Z}_2$ does. This could be explained by Example 4.3 and Figure 5 of [2], where it is shown that when the ES is underestimated, $\bar{Z}_2$ could fail to reject it causing a type II error, while this would never happen for $\bar{Z}_{\text{MB}}$. Which of the three statistics is then right?

Figure 9.2 shows the realized P&Ls versus the ES estimated level for the Taiwan Weighted Index (TWII). From this plot, we can see that there are only three breaches

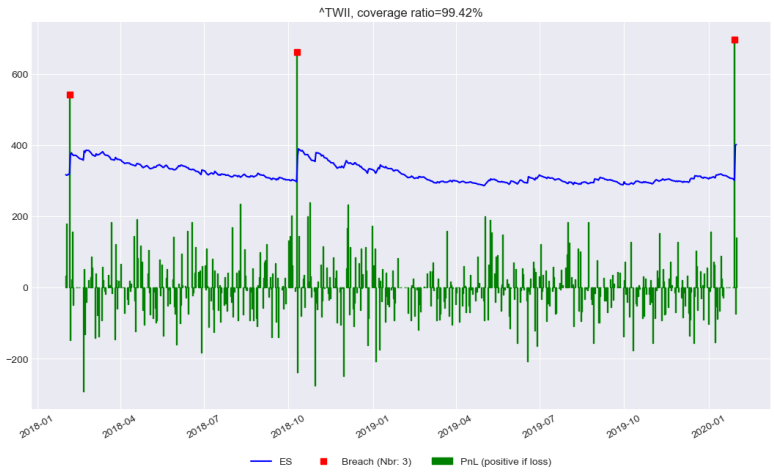

Figure 9.2: Realized P&Ls VS ES on the Taiwan Weighted Index (TWII) for business dates in 2018 and 2019.

but they are huge, so G and $\bar{Z}_{\text{MB}}$ take into account also the magnitude of the breaches while $\bar{Z}_2$ seems to slightly neglect it. Once again, we suggest the use of the $\bar{Z}_{\text{MB}}$ statistic.

9.4.3 Tests on historical data with approximated thresholds

We repeat the tests done in the previous session for G and $\bar{Z}_2$ but approximating the thresholds with pre-computed ones. This will save a lot of memory since it does not require the storage of all the P&Ls simulations but it will affect the results. This time the test cannot be done with the statistic $\bar{Z}_{\text{MB}}$ because it is not stable in the distribution of the underlying. We stress the fact that this approximation can be done on G when N is not too large and ϕ is not too small.

For G we precompute the threshold with standard normal distributions or, equivalently, with t-Student distributions for the P&Ls. Setting $\alpha = 99.5\%$ and $\phi = 5\%$, we find that the threshold value is -6. For $\bar{Z}_2$ we use a t-Student with 5 degrees of freedom. In this case the threshold is -1.2.

	G	Threshold	Accepted
AXJO	-9	-6	No
BVSP	0	-6	Yes
FCHI	-3	-6	Yes
GDAXI	-2	-6	Yes
GSPC	-9	-6	No
GSPTSE	-3	-6	Yes
KS11	-7	-6	No
MXX	-8	-6	No
SSMI	-3	-6	Yes
TWII	-12	-6	No

Table 9.9: Accepted ES for G.

	$\bar{Z}_2$	Threshold	Accepted
AXJO	-2.5397	-1.2	No
BVSP	-1.67839	-1.2	No
FCHI	-0.527363	-1.2	Yes
GDAXI	-0.458829	-1.2	Yes
GSPC	-2.1588	-1.2	No
GSPTSE	-0.160738	-1.2	Yes
KS11	-1.77876	-1.2	No
MXX	-1.52592	-1.2	No
SSMI	-0.481356	-1.2	Yes
TWII	-1.39444	-1.2	No

Table 9.10: Accepted ES for $\bar{Z}_2$.

Of course, the values of the statistics are the same as in the previous tests but the output results regarding the acceptance of the ES are different. We can see in Tables 9.6 and 9.10 that in this case, both statistics become more conservative and that $\bar{Z}_2$ seems more conservative than G, because the calculated ES for the BVSP is accepted by G and rejected by $\bar{Z}_2$. However, we can see from the Figure 9.1 that there are no breaches in the portfolio.

The reason why $\bar{Z}_2$ rejects the computed ES is that this statistic is very sensitive to VaR misspecifications. Since the number of breaches for the computed VaR amounts to 8, then there is a rejection, even if there should not be. Then G gives better results regarding the acceptance of $\hat{\text{ES}}^{\alpha}$.

9.4.4 Conclusion

To sum up, we found that the best statistic from the theoretical and numeric point of view is $\bar{Z}_{\text{MB}}$ since it is the most conservative one as it correctly accepts or rejects the computed values of ES, it is not influenced by VaR misspecifications and it provides an estimation of the real ES value. However, the evaluation of the threshold requires the storage of the historical simulations used to calculate the ES and this slows down computations. The time required for the evaluation of the threshold is in any case of some seconds so the additional storage is not computationally demanding.

If one prefers not having to deal with the storage of the P&Ls simulations, both the $\bar{Z}_2$ and G statistics can be used. From a practical point of view, G gives better results (the same results as $\bar{Z}_{\text{MB}}$) on the portfolios that we tested, although it lacks some theoretical justifications.

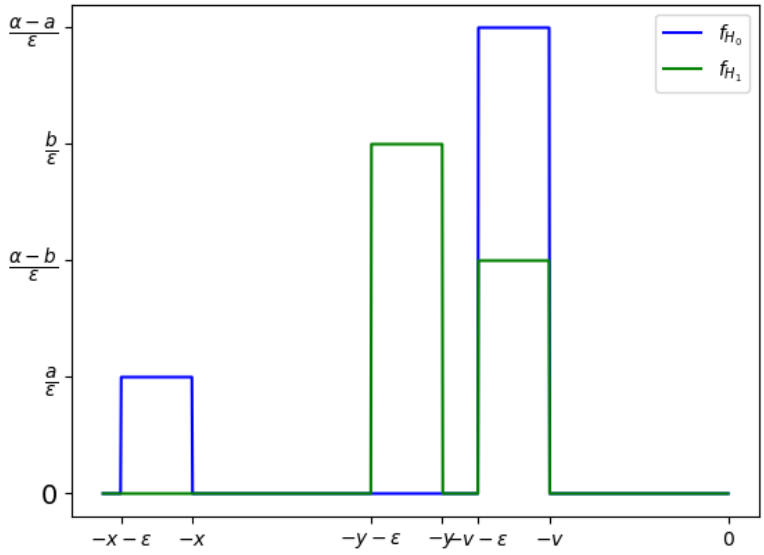

Figure 9.3: Pdf of X under H_0 (blue) and H_1 (green) for example 1.

9.A Moldenhauer and Pitera counterexamples

9.A.1 Example 1: $E_{H_1}[G(X, \hat{\mathbf{ES}}^\alpha)] > E_{H_0}[G(X, \hat{\mathbf{ES}}^\alpha)]$ does not imply $\hat{\mathbf{ES}}^\alpha < \mathbf{ES}^\alpha$

In this example we will prove that $E_{H_1}[G(X, \hat{\mathrm{ES}}^\alpha)] > E_{H_0}[G(X, \hat{\mathrm{ES}}^\alpha)]$ does not imply that we are underestimating the ES or in other words that $\hat{\mathrm{ES}}^\alpha < \mathrm{ES}^\alpha$.

Consider in particular a toy example with $N = 1$. In Figure 9.3 we plot the pdf of the unique X under the null hypothesis, denoted by f_{H_0}, and under the alternative hypothesis, denoted by f_{H_1}. We consider only the part regarding extreme losses of X, so that the distribution for $X > -v$ can be arbitrarily chosen.

The VaRs under H_0 and H_1 are both equal to v. The ES under H_0 is

$$\hat{\mathrm{ES}}^\alpha = -\frac{\frac{a}{\varepsilon}\frac{x^2}{2}\Big|_{-x-\varepsilon}^{-x} + \frac{\alpha-a}{\varepsilon}\frac{x^2}{2}\Big|_{-v-\varepsilon}^{-v}}{\alpha} = \frac{\frac{\alpha}{2}\varepsilon + (ax + (\alpha-a)v)}{\alpha}$$

and similarly the ES under H_1 is

$$\mathrm{ES}^\alpha = \frac{\frac{\alpha}{2}\varepsilon + (by + (\alpha-b)v)}{\alpha}.$$

Note that $\hat{\mathrm{ES}}^\alpha > v$ iff $\frac{\alpha}{2}\varepsilon + (ax + (\alpha-a)v) > \alpha v$ iff $\frac{\alpha}{2}\varepsilon + a(x-v) > 0$ which holds true.

Set $\varepsilon < \frac{2a(\alpha-b)}{\alpha b}(x-v)$. We have $\hat{\mathrm{ES}}^\alpha < x$ iff $\frac{\alpha}{2}\varepsilon + (ax + (\alpha-a)v) < \alpha x$ iff $\varepsilon < \frac{2(\alpha-a)}{\alpha}(x-v)$ and this is true for the chosen ε since $a < b$. We can then choose $y = \hat{\mathrm{ES}}^\alpha$. In this way $\hat{\mathrm{ES}}^\alpha > \mathrm{ES}^\alpha$, iff $ax + (\alpha-a)v > \frac{b}{\alpha}(\frac{\alpha}{2}\varepsilon + (ax + (\alpha-a)v)) + (\alpha-b)v$ iff $a(\alpha-b)(x-v) > \frac{\alpha b}{2}\varepsilon$ which is true for the chosen ε.

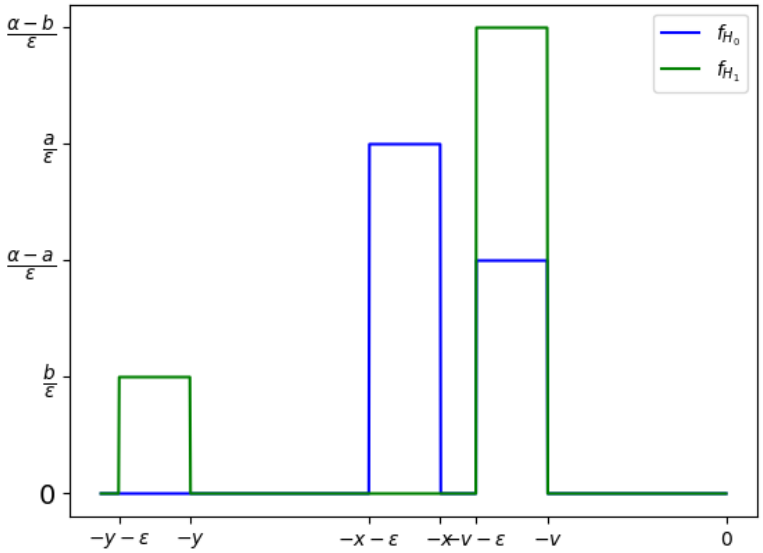

Figure 9.4: Pdf of X under H_0 (blue) and H_1 (green) for example 2.

This means that $\mathrm{ES}^\alpha < \hat{\mathrm{ES}}^\alpha$ so that we are overestimating the real ES. Let us see what happens to the statistic $G = \mathbb{1}_{X+\mathrm{ES}^\alpha<0}$. We have $E_{H_1}[G(X, \hat{\mathrm{ES}}^\alpha)] = P_{H_1}(X + \hat{\mathrm{ES}}^\alpha < 0) = b$ while $E_{H_0}[G(X, \hat{\mathrm{ES}}^\alpha)] = P_{H_0}(X + \hat{\mathrm{ES}}^\alpha < 0) = a$ so $E_{H_1}[G(X, \hat{\mathrm{ES}}^\alpha)] > E_{H_0}[G(X, \hat{\mathrm{ES}}^\alpha)]$, even if we are overestimating the ES.

9.A.2 Example 2: $\hat{\mathbf{ES}}^\alpha < \mathbf{ES}^\alpha$ does not imply $E_{H_1}[G(X, \hat{\mathbf{ES}}^\alpha)] > E_{H_0}[G(X, \hat{\mathbf{ES}}^\alpha)]$

On the other hand, we can construct an example which shows that the very general hypothesis $H_1 : \mathrm{ES}_d^\alpha \geq \hat{\mathrm{ES}}_d^\alpha \ \forall d \wedge \exists d : \mathrm{ES}_d^\alpha > \hat{\mathrm{ES}}_d^\alpha$ does not imply that $E_{H_1}[G(X, \hat{\mathrm{ES}}^\alpha)] > E_{H_0}[G(X, \hat{\mathrm{ES}}^\alpha)]$.

As in the previous example, we take $N = 1$ and we plot in Figure 9.4 the tail pdfs under the null and the alternative hypothesis.

As before, we have $\hat{\mathrm{ES}}^\alpha = \frac{\frac{\alpha}{2}\varepsilon+(ax+(\alpha-a)v)}{\alpha}$ and $\mathrm{ES}^\alpha = \frac{\frac{\alpha}{2}\varepsilon+(by+(\alpha-b)v)}{\alpha}$. For $\varepsilon < \frac{2b(\alpha-b)}{\alpha a}(y-v)$, it holds $v < \mathrm{ES}^\alpha < y$ so we can set $x = \mathrm{ES}^\alpha$. In this way it can be shown that $\hat{\mathrm{ES}}^\alpha < \mathrm{ES}^\alpha$ and $E_{H_1}[G(X, \hat{\mathrm{ES}}^\alpha)] < E_{H_0}[G(X, \hat{\mathrm{ES}}^\alpha)]$, which was our aim.

9.B Moldenhauer and Pitera alternative hypothesis

For the G statistic, let us consider the same hypothesis used for $\bar{Z}_3$, which are:

$$H_0 : Q_d = P_d \ \forall d$$
$$H_1 : Q_d \preceq P_d \ \forall d \wedge \exists d : Q_d \prec P_d$$

where P_d is the predicted distribution of X_d used to evaluate the ES and Q_d is the real unknown distribution of X_d. In particular, for every non-increasing function u, we have $E_{P_d}[u(X_d)] \leq E_{Q_d}[u(X_d)]$.

Let us consider the function $\mathbb{1}_{((X+\hat{\mathrm{ES}}^{\alpha})_{[1]}+\cdots+(X+\hat{\mathrm{ES}}^{\alpha})_{[k]}<0)}$. We prove that it is non-increasing as a function of each X_i. We have that the function $Y \to \mathbb{1}_{(Y<0)}$ is non-increasing. Call f the function $f(X_i) = (X + \hat{\mathrm{ES}}^{\alpha})_{[1]} + \cdots + (X + \hat{\mathrm{ES}}^{\alpha})_{[k]}$ where X_d is fixed for $d \neq i$. Then, it is enough to prove that f is increasing or equivalently that $f(X_i) < f(X_i + \Delta X_i)$ for every ΔX_i. Let us suppose to increase X_i to $X_i + \Delta X_i$. It follows that

- If $X_i + \hat{\mathrm{ES}}_i^{\alpha} \leq (X + \hat{\mathrm{ES}}^{\alpha})_{[k]}$ and $X_i + \Delta X_i + \hat{\mathrm{ES}}_i^{\alpha} \leq (X + \hat{\mathrm{ES}}^{\alpha})_{[k+1]}$, then $f(X_i + \Delta X_i) = f(X_i) + \Delta X_i > f(X_i)$.
- If $X_i + \hat{\mathrm{ES}}_i^{\alpha} \leq (X + \hat{\mathrm{ES}}^{\alpha})_{[k]}$ and $X_i + \Delta X_i + \hat{\mathrm{ES}}_i^{\alpha} > (X + \hat{\mathrm{ES}}^{\alpha})_{[k+1]}$, then $f(X_i + \Delta X_i) = f(X_i) - (X_i + \hat{\mathrm{ES}}_i^{\alpha}) + (X + \hat{\mathrm{ES}}^{\alpha})_{[k+1]} > f(X_i)$ since $X_i + \hat{\mathrm{ES}}_i^{\alpha} < (X + \hat{\mathrm{ES}}^{\alpha})_{[k+1]}$.
- if $X_i + \hat{\mathrm{ES}}_i^{\alpha} > (X + \hat{\mathrm{ES}}^{\alpha})_{[k]}$, then also $X_i + \Delta X_i + \hat{\mathrm{ES}}_i^{\alpha} > (X + \hat{\mathrm{ES}}^{\alpha})_{[k]}$ and $f(X_i + \Delta X_i) = f(X_i)$.

So f is an increasing function and $X_i \to \mathbb{1}_{(f(X_i)<0)}$ is decreasing for every $i = 1, \ldots, N$. We also recall that the expected value of a decreasing function is still a decreasing function.

Applying Fubini's Theorem and sequentially using the fact that $Q_d \preceq P_d$, we have

$$\begin{aligned} E_{H_0}\left[\mathbb{1}_{((X+\hat{\mathrm{ES}}^{\alpha})_{[1]}+\cdots+(X+\hat{\mathrm{ES}}^{\alpha})_{[k]}<0)}\right] &= E_{P_1}\left[E_{P_2}\left[\ldots E_{P_N}\left[\mathbb{1}_{((X+\hat{\mathrm{ES}}^{\alpha})_{[1]}+\cdots+(X+\hat{\mathrm{ES}}^{\alpha})_{[k]}<0)}\right]\right]\right] \\ &\leq E_{F_1}\left[E_{F_2}\left[\ldots E_{F_N}\left[\mathbb{1}_{((X+\hat{\mathrm{ES}}^{\alpha})_{[1]}+\cdots+(X+\hat{\mathrm{ES}}^{\alpha})_{[k]}<0)}\right]\right]\right] \\ &= E_{H_1}\left[\mathbb{1}_{((X+\hat{\mathrm{ES}}^{\alpha})_{[1]}+\cdots+(X+\hat{\mathrm{ES}}^{\alpha})_{[k]}<0)}\right]. \end{aligned}$$

From this, it follows that $E_{H_0}[G(X, \hat{\mathrm{ES}}^{\alpha})] < E_{H_1}[G(X, \hat{\mathrm{ES}}^{\alpha})]$ where the inequality is strict since $Q_d \prec P_d$ for some d.

Bibliography

[1] C. Acerbi and B. Székely. Backtesting Expected Shortfall. *RISK Magazine*, December 2014.

[2] C. Acerbi and B. Székely. General properties of backtestable statistics. *Available at SSRN 2905109*, 2017.

[3] C. Acerbi and B. Székely. The minimally biased backtest for es. *Risk. net*, 29, 2019.

[4] Carlo Acerbi and Dirk Tasche. On the coherence of expected shortfall. *Journal of banking & finance*, 26(7):1487–1503, 2002.

[5] A. Antonov, M. Konikov, and M. Spector. A New Arbitrage-Free Parametric Volatility Surface. *Available at SSRN 3403708*, 2019.

[6] P. Artzner, F. Delbaen, J. M. Eber, and D. Heath. Thinking coherently. *Risk*, pages 68–71, 1997.

[7] G. Barone-Adesi and K. Giannopoulos. Non parametric VaR techniques. Myths and realities. *Economic Notes*, 30(2):167–181, 2001.

[8] G. Barone-Adesi, K. Giannopoulos, and L. Vosper. Var without correlations for nonlinear portfolios. *Journal of Futures Markets*, 1997.

[9] C. Berg and C. Vignat. On the density of the sum of two independent Student t-random vectors. *Statistics & probability letters*, 80(13-14):1043–1055, 2010.

[10] M. Bergeron, N. Fung, J. Hull, Z. Poulos, and A. Veneris. Variational autoencoders: A hands-off approach to volatility. *The Journal of Financial Data Science*, 4(2):125–138, 2022.

[11] S. Bossu. *Advanced Equity Derivatives: Volatility and Correlation*. John Wiley & Sons, 2014.

[12] S. P. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[13] J. Cao, J. H. Kim, and W. Zhang. Pricing variance swaps under hybrid CEV and stochastic volatility. *Journal of Computational and Applied Mathematics*, 386:113220, 2021.

[14] P. Carr and G. Pelts. Duality, Deltas, and Derivatives Pricing. *Presented at the conference dedicated to Steve Shreve's 65th birthday*, June 2015.

[15] P. Carr and G. Pelts. Game of vols. *Available at SSRN 3422540*, 2015.

[16] I. J. Clark. *Foreign exchange option pricing. A Practitioner's Guide*. John Wiley & Sons, 2011.

[17] S. N. Cohen, C. Reisinger, and S. Wang. Detecting and repairing arbitrage in traded option prices. *Applied Mathematical Finance*, 27(5):345–373, 2020.

[18] S. N. Cohen, C. Reisinger, and S. Wang. Arbitrage-free neural-SDE market models. *arXiv preprint arXiv:2105.11053*, 2021.

[19] S. N. Cohen, C. Reisinger, and S. Wang. Estimating risks of option books using neural-SDE market models. *arXiv preprint arXiv:2202.07148*, 2022.

[20] P. Cohort, J. Corbetta, I. Laachir, and C. Martini. Robust calibration and arbitrage-free interpolation of SSVI slices. *Decisions in Economics and Finance*, 42(2):665–677, 2019.

[21] European Commission. Commission Delegated Regulation (EU) 153/2013 of 19 December 2012 supplementing Regulation (EU) 648/2012 of the European Parliament and Council. *Official Journal of the European Union*, 2012.

[22] European Commission. Regulation (EU) 648/2012 of 4 July 2012 of the European Parliament and Council on OTC derivatives, central counterparties and trade repositories. *Official Journal of the European Union*, 2012.

[23] European Commission. Commission Delegated Regulation (EU) No 153/2013 of 19 December 2012, supplementing Regulation (EU) No 48/2012 of the European Parliament and of the Council with regard to regulatory technical standards on requirements for central counterparties. *Official Journal of the European Union*, 2013.

[24] R. Cont and M. Vuletić. Simulation of arbitrage-free implied volatility surfaces. *Available at SSRN*, 2022.

[25] B. M. Damghani and A. Kos. De-arbitraging With a Weak Smile: Application to Skew Risk. *Wilmott*, 2013(64):40–49, 2013.

[26] S. De Marco and C. Martini. Quasi-explicit calibration of Gatheral SVI model. *Zeliade White Paper*, pages 1–15, 2009.

[27] S. De Marco and C. Martini. Moment generating functions and normalized implied volatilities: unification and extension via Fukasawa pricing formula. *Quantitative Finance*, 18(4):609–622, 2018.

[28] V. Durrleman. Implied volatility: Market models. *Encyclopedia of Quantitative Finance*, 2010.

[29] T. Ferhati. Robust Calibration For SVI Model Arbitrage Free. *Available at SSRN 3543766*, 2020.

[30] T. Fissler, J. F. Ziegel, and T. Gneiting. Expected shortfall is jointly elicitable with value at risk-implications for backtesting. *arXiv preprint arXiv:1507.00244*, 2015.

[31] M. Fukasawa. Normalization for implied volatility. *arXiv preprint arXiv:1008.5055*, 2010.

[32] M. Fukasawa. The normalizing transformation of the implied volatility smile. *Mathematical Finance*, 22(4):753–762, 2012.

[33] J. Gatheral. A Parsimonious Arbitrage-Freee Implied Volatility Parameterization with Applications to the Valuation of Volatility Derivatives. *Proceeding of the Global Derivatives and Risk Management Madrid conference*, 2004.

[34] J. Gatheral. The volatility surface: a practitioner's guide. *John Wiley & Sons*, 357, 2011.

[35] J. Gatheral and A. Jacquier. Convergence of Heston to SVI. *Quantitative Finance*, 11(8):1129–1132, 2011.

[36] J. Gatheral and A. Jacquier. Arbitrage-free SVI volatility surfaces. *Quantitative Finance*, 14(1):59–71, 2014.

[37] P. Glasserman and D. Pirjol. W-shaped implied volatility curves and the Gaussian mixture model. *Available at SSRN 3951426*, 2021.

[38] T. Gneiting. Making and evaluating point forecasts. *Journal of the American Statistical Association*, 106(494):746–762, 2011.

[39] A. Gulisashvili. *Analytically Tractable Stochastic Stock Price Models*. Springer Science & Business Media, 2012.

[40] F. Gunnarsson. *Filtered Historical Simulation Value at Risk for Options: A Dimension Reduction Approach to Model the Volatility Surface Shifts*. Master Thesis at Umeå University, 2019.

[41] G. Guo, A. Jacquier, C. Martini, and L. Neufcourt. Generalized arbitrage-free SVI volatility surfaces. *SIAM Journal on Financial Mathematics*, 7(1):619–641, 2016.

[42] P. Gurrola-Perez and D. Murphy. Filtered historical simulation Value-at-Risk models and their competitors. *Bank of England Working Paper*, 2015.

[43] P. S. Hagan, D. Kumar, A. S. Lesniewski, and D. E. Woodward. Managing smile risk. *The Best of Wilmott*, 1:249–296, 2002.

[44] S. Hendriks and C. Martini. The extended SSVI volatility surface. *Journal of Computational finance*, 22(5), 2019.

[45] A. Itkin. One more no-arbitrage parametric fit of the volatility smile. *North Am. J. Econ. Financ*, 2014.

[46] A. Itkin et al. Fitting Local Volatility: Analytic and Numerical Approaches in Black-Scholes and Local Variance Gamma Models. *World Scientific Books*, 2020.

[47] C. Kahl and P. Jäckel. Not-so-complex logarithms in the Heston model. *Wilmott magazine*, 19(9):94–103, 2005.

[48] R. W. Lee. The Moment Formula for Implied Volatility at Extreme Strikes. *Mathematical Finance*, 14(3):469–480, 2004.

[49] V. Lucic. Volatility Notes. *Available at SSRN 3211920*, 2019.

[50] V. Lucic. Lecture notes on volatility modelling. *Imperial College London*, 2020.

[51] V. Lucic. Normalizing Volatility Transforms and General Parameterization of Volatility Smile. *Available at SSRN 3835233*, 2021.

[52] C. Martini and A. Mingone. Explicit no arbitrage domain for sub-SVIs via reparametrization. *arXiv preprint arXiv:2106.02418*, 2021.

[53] C. Martini and A. Mingone. No arbitrage SVI. *SIAM Journal on Financial Mathematics*, 13(1):227–261, 2022.

[54] C. Martini and A. Mingone. Refined analysis of the no-butterfly-arbitrage domain for SSVI slices. *Journal of Computational Finance*, 27(2):1–32, 2023.

[55] R. C. Merton. On the pricing of corporate debt: the risk structure of interest rates. *The Journal of finance*, 29(2):449–470, 1974.

[56] A. Mingone. Smiles in delta. *Quantitative Finance doi:10.1080/14697688.2023.2258932*, 2023.

[57] F. Moldenhauer and M. Pitera. Backtesting expected shortfall: a simple recipe? *Journal of Risk*, 22(1), 2019.

[58] L. Nagy and M. Ormos. Volatility Surface Calibration to Illiquid Options. *The Journal of Derivatives*, 26(3):87–96, 2019.

[59] B. Ning, S. Jaimungal, X. Zhang, and M. Bergeron. Arbitrage-free implied volatility surface generation with variational autoencoders. *arXiv preprint arXiv:2108.04941*, 2021.

[60] D. Reiswich and W. Uwe. FX volatility smile construction. *Wilmott*, 2012(60):58–69, 2012.

[61] E. Renault and N. Touzi. Option hedging and implied volatilities in a stochastic volatility model. *Mathematical Finance*, 6(3):279–302, 1996.

[62] M. Roper. Arbitrage Free Implied Volatility Surfaces. *preprint*, 2010.

[63] M. R. Sampford. Some inequalities on Mill's ratio and related functions. *The Annals of Mathematical Statistics*, 24(1):130–132, 1953.

[64] S. Schlüter and M. J. Fischer. A tail quantile approximation formula for the student t and the symmetric generalized hyperbolic distribution. Technical report, IWQW Discussion Papers, 2009.

[65] M. R. Tehranchi. Symmetric martingales and symmetric smiles. *Stochastic Processes and Their Applications*, 119(10):3785–3797, 2009.

[66] M. R. Tehranchi. A Black-Scholes inequality: applications and generalisations. *Finance and Stochastics*, 24(1):1–38, 2020.

[67] L. W. Wong and Y. Zhang. Procyclicality control in risk-based margin models. *Journal of Risk*, 23(5), 2021.

[68] J. Yao, S. Laurent, and B. Bénaben. Managing Volatility Risk: An Application of Karhunen-Loève Decomposition and Filtered Historical Simulation. *arXiv preprint arXiv:1710.00859*, 2017.

ECOLE
DOCTORALE
DE MATHEMATIQUES
HADAMARD

Titre: Modèles avancés de volatilité implicite pour la gestion des risques et les Chambres de compensation.

Mots clés: volatilité implicite, paramétrisation sans arbitrage, calibration, Chambres de compensation, marge initiale

Résumé: Dans la première partie de cette thèse, nous abordons la tâche de construire des surfaces de volatilité implicite sans arbitrage qui puissent être utilisées par les opérateurs de marché. Nous étudions les contraintes d'arbitrage statique pour les portefeuilles d'options et nous les appliquons à des modèles de volatilité implicite connus. Tout d'abord, nous caractérisons l'absence d'arbitrage Butterfly dans le modèle SVI de Gatheral, et nous étudions le cas de certains modèles sous-SVI à 3 paramètres, tels que le SVI symétrique, le SVI Vanishing Upward/Downward et le SSVI. Nous reconsidérons ensuite ce dernier modèle, étendu à plusieurs maturités, et nous combinons les conditions d'absence d'arbitrage Butterfly avec celles d'absence d'arbitrage Calendar Spread. Nous identifions donc un algorithme de calibration globale garantissant l'absence d'arbitrage pour le modèle eSSVI. Ensuite, nous étudions la caractérisation d'une notion plus faible d'absence d'arbitrage Butterfly, c'est-à-dire les deux conditions de monotonie des fonctions d_1 et d_2 de la formule de Black-Scholes, dans le cadre des smiles paramétrés en delta. Enfin, en nous basant sur la remarque que les options Call peuvent être vues comme des Calls écrits sur d'autres Calls, nous étudions les propriétés dynamiques de ces contrats.

Dans la deuxième partie, nous étudions le problème de la quantification du risque de contrepartie auquel les Chambres de compensation sont confrontées quotidiennement. Nous identifions une nouvelle formule model-free pour la VaR à court terme des portefeuilles d'options qui montre d'avoir des meilleurs performances que celles de l'approche classique de la Filtered Historical Simulation dans nos tests numériques. Enfin, nous considérons la notion d'Expected Shortfall, dont nous comparons différents types de mesures de backtesting.

Title: Advanced implied volatility modeling for risk management and central clearing

Keywords: implied volatility, arbitrage-free parametrization, calibration, Central Clearing Counterparties, initial margin

Abstract: In the first part of this thesis we address the non trivial task of building arbitrage-free implied volatility surfaces which could be used by market operators for practical purposes. We study in depth static arbitrage constraints for option portfolios and apply them to notorious implied volatility models. We firstly fully characterize the absence of Butterfly arbitrage in the SVI model by Gatheral, and study the case of some 3-parameter sub-SVIs models, such as the Symmetric SVI, the Vanishing Upward/Downward SVI, and SSVI. We then reconsider the latter model, extended to multiple maturity slices, and combine the so identified conditions of no Butterfly arbitrage with the already known conditions of no Calendar Spread arbitrage. As a result, we identify a global calibration algorithm for the eSSVI model ensuring the absence of arbitrage. Secondly, we study the characterization of a weaker notion of absence of Butterfly arbitrage, i.e. the two monotonicity requirements of the functions d_1 and d_2 in the Black-Scholes formula, in the framework of smiles parameterized in delta. Finally, based on the remark that Call options can be seen as Calls written on other Calls, we study the dynamic properties of these contracts.

In the second part, we consider the problem of quantifying the counterparty risk for option portfolios that Central Clearing Counterparties face daily. We identify a new model-free formula for the short-term VaR of option portfolios which performs better than the classical approach of Filtered Historical Simulation in our numerical tests. Finally, we look at the notion of Expected Shortfall, and compare different types of backtesting measures.

Institut Polytechnique de Paris
91120 Palaiseau, France