universität wien

DISSERTATION / DOCTORAL THESIS

Titel der Dissertation / Title of the Doctoral Thesis

„Signature-based models in finance and robust risk measures“

verfasst von / submitted by

Dott. Dott.mag. Guido Gazzani

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Doctor of Philosophy (PhD)

Wien, 2023 / Vienna, 2023

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 794 370 136
Dissertationsgebiet lt. Studienblatt / field of study as it appears on the student record sheet:	PhD program in Statistics and Operations Research
Betreut von / Supervisor:	Univ.-Prof. Dipl.-Ing. Dr. Christa Cuchiero Univ.-Prof. Mag. Dr. Irene Klein

Abstract

This thesis is devoted to the study of signature-based models in finance and robust risk measures.

The former is a class of asset price models whose dynamics are described by linear functions of the (time extended) signature of a primary underlying process, which can range from a (market-inferred) Brownian motion to a general multidimensional continuous semimartingale. The framework is universal in the sense that classical models can be approximated arbitrarily well and that the model's parameters can be learned from all sources of available data by simple methods. We provide conditions guaranteeing absence of arbitrage as well as tractable option pricing formulas for so-called sig-payoffs, exploiting the polynomial nature of generic primary processes. One of our main focus lies on calibration, where we consider both time-series and implied volatility surface data, generated from classical stochastic volatility models and also from S&P 500 index market data. For both tasks the linearity of the model turns out to be the crucial tractability feature which allows to get fast and accurate calibrations results. We propose additionally a modification of the previous asset price models, by parameterizing the volatility process as a linear functional of the signature of a primary process. We then obtain both closed formulas for VIX index futures and a tractable expression for the corresponding SPX index dynamics under a pricing measure. This in turn allows to tackle the complex problem of joint calibration to SPX and VIX options which we solve by employing correlated Ornstein-Uhlenbeck processes as primary processes.

The second line of research consists in first introducing two kinds of risk measures with respect to some reference probability measure, which both allow for a certain order structure and domination property. Analyzing their relation to each other leads to the question when a certain minimax inequality is actually an equality. We then provide conditions under which the corresponding robust risk measures, being defined as the supremum over all risk measures induced by a set of probability measures, can be represented classically in terms of one single probability measure. We focus in particular on the mixture probability measure obtained via mixing over a set of probability measures using some prior, which represents for instance the regulator's beliefs. The classical representation in terms of the mixture probability measure can then be interpreted as a Bayesian approach to robust risk measures. The last part of the thesis is devoted to an outlook in the direction of dynamic robust risk measures under model uncertainty and their properties in light of the aforementioned findings in the static setup.

Kurzfassung

Diese Arbeit ist der Untersuchung von Signatur-basierten Modellen und robusten Risikomaßen in Finanzmathematik gewidmet.

Bei ersteren handelt es sich um eine Klasse von Aktienpreismodellen, deren Dynamik durch lineare Funktionen der Signatur eines zugrunde liegenden Prozesses beschrieben wird, der von einer aus dem Markt abgeleiteten Brownschen Bewegung bis hin zu einem allgemeinen mehrdimensionalen stetigen Semimartingal reichen kann. Der Modellrahmen ist universell in dem Sinne, dass klassische Modelle approximiert werden können und dass die Parameter des Modells durch einfache Methoden aus allen verfügbaren Datenquellen gelernt werden können. Wir untersuchen weiters Bedingungen, die die Abwesenheit von Arbitrage garantieren, und leiten Optionspreisformeln für sogenannte Sig-Payoffs her, die die polynomielle Eigenschaft generischer Prozesse ausnutzen. Einer unserer Schwerpunkte liegt auf der Kalibrierung, bei der wir sowohl Zeitreihen- als auch implizite Volatilitätsdaten berücksichtigen, die einerseits aus klassischen stochastischen Volatilitätsmodellen und andererseits auch aus S&P 500-Indexmarktdaten generiert werden. Für beide Aufgaben erweist sich die Linearität des Modells als entscheidendes Merkmal, das es ermöglicht, schnelle und genaue Kalibrierungsergebnisse zu erhalten. Darüber hinaus schlagen wir eine andere Version des bisherigen Aktienpreismodells vor, indem wir den Volatilitätsprozess als lineares Funktional der Signatur eines zugrundeliegenden Prozesses parametrisieren. Auf diese Weise erhalten wir sowohl geschlossene Formeln für VIX-Index-Futures als auch eine gut handhabare Dynamik des SPX-Index unter einem risiko-neutralen Maß. Dies liefert dann den Ausgangspunkt, um das komplexe Problem der gemeinsamen Kalibrierung an SPX- und VIX-Optionen anzugehen, das wir unter Verwendung korrelierter Ornstein-Uhlenbeck-Prozesse lösen.

Der zweite Forschungsstrang besteht darin, zunächst zwei Arten von Risikomaßen in Bezug auf ein Referenzwahrscheinlichkeitsmaß einzuführen, die beide eine bestimmte Ordnungsstruktur und Dominationseigenschaft zulassen. Die Analyse in welchem Verhältnis diese zwei Arten von Risikomaßen zueinander stehen führt zu der Frage, wann eine bestimmte Minimax-Ungleichung tatsächlich eine Gleichheit ist. Wir stellen dann Bedingungen auf, unter denen die entsprechenden robusten Risikomaße, die als Supremum über alle durch eine Menge von Wahrscheinlichkeitsmaßen induzierten Risikomaße definiert sind, klassisch durch ein einziges Wahrscheinlichkeitsmaß dargestellt werden können. Wir konzentrieren uns insbesondere auf die Mischverteilung, die man durch Mischen über eine Menge von Wahrscheinlichkeitsmaßen unter Verwendung eines Priors erhält, der z.B. die Einschätzungen des Regulators darstellt. Diese klassische Darstellung in Form der Mischverteilung kann dann als Bayesianischer Ansatz für robuste Risikomaße interpretiert werden. Der letzte Teil der Arbeit widmet sich einem Ausblick in Richtung dynamischer robuster Risikomaße unter Modellunsicherheit und deren Eigenschaften im Lichte der

oben genannten Erkenntnisse im statischen Setup.

Acknowledgements

First of all, I would like to thank my advisors Christa Cuchiero and Irene Klein for their expertise and for creating a stimulating, serene environment which fostered the interesting lines of research in the present thesis and thus, made it possible. Not only I benefited of their rigorous mathematical approaches and of our meetings, but I also appreciated their kindness and their always sincere viewpoint. In particular I would like to thank Christa for her capacity in transmitting me the passion and patience of research, for teaching me always to see the *big picture*, and Irene for her precious availability and our talks in Servitengasse.
I also want to thank Christian Bayer and Giulia Di Nunno for accepting the task of reviewing my thesis.
A self-explanatory acknowledgment goes to Sara Svaluto-Ferro, for always pushing me not to give up the challenges I faced and for enhancing my curiosity and attention to details.
My gratitude goes also to my colleagues and friends of the QUARIMAFI group and of the Institute of Statistics at WU for always sharing a coffee, their experiences or just a mathematical thought. I feel especially in debt with Camilla Damian, Luca Gonzato and Janka Möller both for our stimulating discussions and for their friendship.
A special thank goes clearly to my friends from the university and those from my hometown for keeping our bonds tight disregarding the distance and in particular for sharing together all our achievements in these three years.
A big thank to my family, for supporting me in every step of my life. I take the chance to send them my wishes to find the most deserved serenity in these periods of change in our lives.
Finally, my sincere and deepest gratitude goes to Ilaria for her patience and love: these years assume a unique meaning in my path particularly thanks to you.

Contents

List of Figures

Introduction

In the past few years data-driven models have successfully entered the area of stochastic modeling and mathematical finance. The paradigm of calibrating a few well interpretable parameters has changed to learning the model's characteristics as a whole, thereby exploiting all available sources of data. Thus highly parametric and overparametrized models have gained more and more importance. On the one hand this has opened the door to robust and more data-driven model selection mechanisms, while on the other hand model classes still have to be chosen in a way to guarantee first principles from finance like "no arbitrage". Relying on different universal approximation theorems leads to different well-suited classes of dynamic processes that can serve both purposes.

One class of such financial models are so-called *neural stochastic differential equations* (SDEs) which are defined as Itô-diffusions where the drift and the volatility function are parameterized via neural networks (see e.g. Gierjatowicz et al. (2020); Cuchiero et al. (2020); Cohen et al. (2021)). Another class of models, considered in Perez Arribas et al. (2020) and inspiring the current work, are so-called *Sig-SDEs*. These are again Itô-diffusions, however in this case the characteristics are linear functions of the *signature* (more precisely introduced below) of some (properly extended) driving Brownian motion.

In the first part of the present thesis we consider a related approach, where the asset price model itself is parameterized as a linear function of the signature of a primary underlying process. This primary process can either be a classical driving signal, e.g. a Brownian motion, but also a more general tractable stochastic model describing well observable quantities.
The notion of *signature* goes back to Chen (1977, 1957) and plays a crucial role in the context of rough path theory initiated by Lyons (1998). Indeed, we consider in Chapter 1-2 the *time extended signature* of an $\mathbb{R}^d$-valued path which serves as linear regression basis for continuous path functionals, since

- it is *point-separating*, as its final value uniquely determines the underlying path;
- linear functions on the signature form an algebra of continuous functions (with respect to a certain variation distance) that contains 1. More precisely, every polynomial on the signature may be realized as a linear function via the so-called shuffle product $\sqcup\!\sqcup$.

The Stone-Weierstrass theorem therefore yields a Universal Approximation Theorem (UAT), telling that continuous path functionals on compact sets can be uniformly approximated by a linear function of the time extended signature.

Therefore signature-based methods provide a non-parametric way to extract characteristic features (linearly) from time series data, which is essential in machine learning

tasks in finance. This explains why these techniques become more and more popular in econometrics and mathematical finance, see e.g., Buehler et al. (2020); Kalsi et al. (2020); Perez Arribas et al. (2020); Lyons et al. (2020); Ni et al. (2021); Bayer et al. (2021); Min and Hu (2021); Akyildirim et al. (2022) and the references therein.

In the first part of the thesis we consider signature-based models with the goal to provide a data-driven, universal, tractable and easy to calibrate model for a set of traded assets S. For the sake of exposition we shall assume throughout that S is one-dimensional. As already mentioned above the main ingredient of our modeling framework is a d-dimensional *primary process* $X = (X_t^1, \dots, X_t^d)_{t\geq 0}$, in most case augmented with time t, where X is supposed to be a continuous semimartingale. We shall denote the time augmented process via $(\widehat{X}_t)_{t\geq 0} = (t, X_t^1, \dots, X_t^d)_{t\geq 0}$ and assume that its signature denoted by $\widehat{\mathbb{X}}$ serves as linear regression basis for S. Indeed, S is modeled/approximated via a process $S_n(\ell)$ defined as

$$S_n(\ell)_t := \ell(\widehat{\mathbb{X}}_t), \tag{$*$}$$

where ℓ is a linear map of the signature of $\widehat{X}$ up to degree $n \in \mathbb{N}$ that has to be inferred from data (see Definition 1.2.3 for further details). The attractiveness of this model class arises from several important features that we summarize in the sequel.

No arbitrage: As shown in Section 1.2.1, the model of form ($*$) can also be expressed in terms of stochastic integrals. From this representation conditions guaranteeing no-arbitrage can in turn be easily deduced.

Universality: As we argue in Example 1.2.15 classical stochastic volatility models driven by Brownian motions with sufficiently regular coefficients can be arbitrarily well approximated by models of form ($*$), when choosing the driving Brownian motions (possibly modified with some drift) as primary process X. Locally in time this can be inferred from the stochastic Taylor expansion that we state in Proposition 1.1.15, which gives quantitative estimates. It can also seen as a consequence of the Stone-Weierstrass theorem in the way formulated e.g. in Lemma 5.2 in Bayer et al. (2021) (see also Cuchiero and Möller (2023)), however without a convergence rate.

Tractable option pricing formulas for sig-payoffs: By relying on the above UAT and in turn on approximations via so-called sig-payoffs of the form $L(\widehat{\mathbb{S}}_T(\ell))$ (see also Lyons et al. (2020)), where $\widehat{\mathbb{S}}$ denotes the signature of $(\widehat{S}_t)_t := (t, S_t)_t$ and L a linear map, this kind of approximate options pricing reduces to the computation of the expected signature of $\widehat{X}$. Indeed, in Section 1.2.2 we first show how to express $\widehat{\mathbb{S}}_T(\ell)$ in terms of the signature of $\widehat{X}$ and then translate this to the computation of $\mathbb{E}_{\mathbb{Q}}[L(\widehat{\mathbb{S}}_T(\ell)\rangle]$ under some pricing measure $\mathbb{Q}$. These formulas can then be applied whenever $\mathbb{E}_{\mathbb{Q}}[\widehat{\mathbb{X}}_T]$ can be easily computed. This is the case for highly generic primary processes of the form

$$\mathrm{d}\widehat{X}_t = \mathbf{b}(\widehat{\mathbb{X}}_t)\mathrm{d}t + \sqrt{\mathbf{a}(\widehat{\mathbb{X}}_t)}\mathrm{d}B_t,$$

where $\mathbf{b}$ and $\mathbf{a}$ are linear maps. Indeed, as shown in Cuchiero et al. (2023b) these processes can be seen as projections of extended tensor algebra valued polynomial processes (introduced in Cuchiero et al. (2012); Filipović and Larsson (2016)), which implies that the expected signature can be computed by solving a linear ODE. This ODE is usually infinite dimensional, but if $\widehat{X}$ is itself a polynomial process it becomes finite dimensional. We exemplify this polynomial process point of view by means of time extended correlated Brownian motions, a particular simple polynomial process, in Section A.1.1 and for time extended classical polynomial processes in Section 2.2.

Calibration to time series data: The tractability of the model class given by $(*)$ becomes particularly clear in view of calibration tasks. Indeed, when the goal is to calibrate to times series data of prices $(S_{t_i})_{i=1}^N$ or spot volatility/spot variance $(V_{t_i})_{i=1}^N$, this task reduces to a simple linear regression. We illustrate this well working procedure on out-of-sample data generated from Heston, SABR and multivariate Black-Scholes models in Section 1.4.1.

Calibration to options: When calibrating to the market's volatility surface (see Section 1.4.2), we exploit again the linearity of the model by precomputing Monte Carlo samples of $\widehat{\mathbb{X}}$ and then performing a standard optimization to find the parameters of the linear map ℓ. By initializing the parameters of ℓ appropriately this optimization task actually becomes a convex problem, which makes it particularly tractable. On simulated and real market data (S&P 500 index) we show that a full calibration to the volatility surface, in particular when using time dependent parameters, is highly accurate and very fast. Note that *sampling from the calibrated model* is also particularly easy since it just means computing a scalar product between the parameters and the trajectory of the signature.

Closely related to the work presented in Chapter 1 is the inspiring paper by Perez Arribas et al. (2020) (see also Perez Arribas (2020)). Indeed, the class of Sig-SDEs considered there can be embedded in our framework by choosing a properly extended one-dimensional Brownian motion (namely a lead-lag transformation) as primary process. This is due to the fact that $(*)$ can also be expressed in terms of stochastic integrals as stated in Proposition 1.2.9. While in Perez Arribas et al. (2020) calibration was considered for certain options within the Black-Scholes model, we here provide calibration results to both, time-series data and volatility surfaces generated from classical stochastic volatility models and also from real market data. A crucial difference is that the model in Perez Arribas et al. (2020) is calibrated to sig-payoffs that approximate vanilla call and puts. As we encountered some problems with this procedure (as outlined in Appendix A.2), we rely on Monte Carlo pricing which is fast due to the possibility to precompute all samples of $\widehat{\mathbb{X}}$ in advance.
Chapter 2 treats a sligthly different class of signature-based models and constitutes a novel contribution for joint VIX and S&P 500 index modeling and calibration to the

corresponding options. The joint calibration to SPX[1] and VIX options is a problem that has gained a lot of attention in quantitative finance since several years. It is sometimes still regarded as the *holy grail of volatility modeling,* even though significant progress has been made recently (see below in the literature overview).

One main reason for the increased interest, is that the VIX index is no longer only used as indicator of volatility, but rather as important underlying for many derivatives. In fact, futures and options written on it are extensively used to hedge the volatility exposure of option portfolios, see e.g. Rhoads (2011). Suppose that an investor has a long position on the S&P 500 index. Although she might believe it has long-term prospects, it would be desirable to reduce her exposure to short-term volatility. Buying VIX derivatives with the belief that volatility is going to increase she might balance out these positions, while she was wrong, the losses to her VIX position could be mitigated by gains to the existing trade.

On February 24, 2006, European options written on the VIX index were also launched on the CBOE in the form of options on VIX futures. We address the reader to the official website of CBOE[2] for details on how the VIX is computed and traded in practice.

The higher liquidity on the market of VIX index products has marked the need to jointly calibrate to both, the prices of options on the underlying S&P 500 index and to prices of VIX derivatives.

From a data perspective the challenge, especially for short maturities, is to reconcile the large negative skew of SPX options' implied volatilities with relatively lower implied volatilities arising from the VIX options. This has been discussed rigorously in Guyon (2020a), where the author shows a necessary condition to solve the joint calibration problem. As observed in the paper this translates, in the context of models with continuous trajectories, to a volatility process with high mean-reversion speed and large negative correlation to the S&P 500 index. In addition to that a high vol-of-vol is desirable in order to reproduce the aforementioned negative skew of at-the-money SPX options, but can yield too high implied volatilities for the VIX options.

Remark 0.0.1. It is also important to notice that given a type of option just a fixed number of maturities are available. This in particular implies that on different days options with different time to maturities are traded. For example, options with exactly one year time to maturity are not available on a daily basis. For what concerns the options we are interested in, typically, VIX options expire on the Wednesday 30 days (or closest to 30 days) prior to the third Friday of the next calendar month. On the other hand a monthly SPX option typically expires on the third Friday of every month.

Inspired by Perez Arribas et al. (2020) and the work in Chapter 1, we consider here a new type of stochastic volatility model for the discounted price process $S = (S_t)_{t\geq 0}$ with continuous trajectories. It is given by

$$\mathrm{d}S_t(\ell) = S_t(\ell)\sigma_t^S(\ell)\mathrm{d}B_t,$$

[1]The SPX is a theoretical index, in the sense that it is pegged to the value of the S&P 500 without being based on a held portfolio of stock shares. In the present we use SPX and S&P 500 interchangeably.
[2]www.cboe.com/tradable_products/vix/

for an initial condition $S_0 \in \mathbb{R}_+$, a standard Brownian motion B, and a volatility process σ^S satisfying

$$\sigma_t^S(\ell) := \ell(\widehat{\mathbb{X}}_t), \tag{0.0.1}$$

where ℓ is a linear map of the signature $\widehat{\mathbb{X}}_t$ of a time extended d-dimensional continuous semimartingale X. By modeling σ^S via (0.0.1) we assume that the signature of $\widehat{X}$, denoted by $\widehat{\mathbb{X}}$ (and rigorously introduced in Section 1.1), serves as a linear regression basis for the volatility process, while the parameters of the linear map ℓ have to be learned from (option price) data. Note that the parameters of X are prespecified beforehand and can thus be seen – in analogy to machine learning terminology – as *hyperparameters* (that of course can be optimized over some training or validation set). As outlined below this is one of the crucial features that allows for the split of the calibration task into precomputable samples and parameters to optimize.

Let us now highlight the implications of this modeling framework and the novelty of the present work.

- The modeling framework can be seen as *universal* in the class of continuous non-rough stochastic volatility models, which is a consequence of the universal approximation result stated in Theorem 1.1.12.
- It is not only universal in an approximate sense, but truly nests several classical models (see Remark 2.1.5) and for instance also the 'quintic Ornstein-Uhlenbeck volatility model', recently proposed by Abi Jaber et al. (2022b), which – with an additional input curve – fits SPX and VIX smiles.
- By choosing the parameters of ℓ appropriately, the modeling framework incorporates both, purely Markovian (in (S, X)) and path-dependent models.
- Up to our knowledge, it is the first signature-based model that is employed for pricing and calibration of VIX options as well as joint calibration, together with SPX options.
- We illustrate that the joint calibration problem can be solved in this framework without jumps and rough volatility (compare also Guyon and Lekeufack (2022); Rømer (2022); Abi Jaber et al. (2022b)).
- By using time-varying parameters we can go beyond short maturities both for SPX and VIX options (as classically tackled in the literature) and achieve a joint calibration also for longer maturities.

In order to achieve the highly accurate calibration results, illustrated in Section 2.3.4 and Section 2.5, we exploit the following mathematical and numerical properties.

- Defining $Z := (X, B)$, then not only $\sigma^S(\ell)$ but also the log-price $\log(S(\ell))$ can be expressed as a linear function of the signature of $\widehat{Z}$. The computational benefit is immediate, since no (Euler) simulation scheme is needed to sample from the

marginals of the price process. In terms of the parameters ℓ, $\log(S(\ell))$ is the sum of a quadratic function and a linear one, see Proposition 2.4.5.

- If X is additionally a polynomial process (see Cuchiero et al. (2012); Filipović and Larsson (2016)), then the VIX under our model can be computed analytically via matrix exponential. Indeed, in this case the forward variance can be represented by a quadratic form in the parameters ℓ and the corresponding matrix can be computed by polynomial technology, i.e. via matrix exponentials, see Theorem 2.3.1. This efficient tractability property is a consequence of the fact that the truncated signature of a polynomial process is again a polynomial process (see Section 2.2).

- We can apply a Monte Carlo approach (potentially with variance reduction) for option pricing *and* calibration, since the signature samples of $\widehat{Z}$ can be computed offline. Indeed, due to the representations of VIX and $\log(S(\ell))$ described above, the same samples can be used for *every* linear map ℓ. Therefore, the calibration task can be split into an offline sampling and a standard optimization, as no simulation is needed during the latter.

To compare our results we also give an extensive literature review on the joint calibration problem in Chapter 2.

The second line of research concerns robust risk measures and their representation. In quantitative risk management, risk measures are used to determine the minimal capital requirement so that a financial position, described by a random variable, becomes acceptable. Proceeding from the pioneering works of Artzner et al. (1999), Föllmer and Schied (2002) and Frittelli and Gianin (2002), several ideas have been proposed to extend the framework based on a single probability measure to a robust setup where the risk evaluation is not only based on one probability measure, but rather on a set of possibly non-dominated ones. A common approach among these is to *robustify* a particular risk measure by considering its worst-case counterpart, meaning to take the supremum over all risk measures induced by the specified set of probability measures. We shall call this *robust (version of a) risk measure.*

This approach has been pursued e.g. for Value at Risk (VaR) by Peng et al. (2020) and Pei et al. (2021), where the risk measure is robustified via the concept of sublinear expectations, and also by Embrechts et al. (2013), where the supremum of VaR over a set of possible joint distributions with prespecified marginals is computed. In the articles by Pei et al. (2021), Ghaoui et al. (2003) and Zymler et al. (2013) the supremum over all distributions within a certain class of models is considered to compute a 'robust' quantile. Similar approaches have been adopted for two particular coherent risk measures, namely Expected Shortfall (ES) and the superhedging price (see e.g., Zhu and Fukushima (2009), Fertis et al. (2012), Arias-Serna et al. (2020), Obłój and Wiesel (2021)).

Another possible perspective to account for model uncertainty is to modify the functional form of a risk measure by distorting the probability measure or a quantile, compare for instance Belles-Sampera et al. (2014); Wang and Ziegel (2018); Embrechts et al. (2018);

Wang et al. (2020); Liu et al. (2021, 2022) for distorted versions of VaR and ES or more recently Jaimungal et al. (2022) for a distortion of a utility based risk measure. In Wang and Ziegel (2018) also different notions of law-invariance for robust versions of VaR and ES are discussed.

These viewpoints have been extensively employed not only in risk evaluation but also in stochastic optimization, see e.g. Wang and Xu (2020), and in insurance, see for instance Escobar and Pflug (2018) and Birghila and Pflug (2019).

In Chapter 4 we introduce a different approach to robustness which is inspired by the following reasoning. Suppose that there is some probability measure $\mathbb{P}$ and an associated risk measure which yields for all financial positions a maximal value, defined e.g. by the regulator of the financial market as worst case. We shall thus call such a 'maximal' measure '*worst-case probability measure*' and introduce a certain order structure to incorporate it into a family of risk measures. Indeed, we require that

(i) the risk measured under any absolutely continuous measure is smaller or equal than the risk computed under $\mathbb{P}$;

(ii) for any equivalent measure we get the same risk.

Indeed, this worst-case point of view should capture the risk coming from all possible scenarios rather than necessarily taking into account how probable they are, which is of course reminiscent of option pricing under no-arbitrage assumptions. Let us motivate this by a simple but extreme example with two equivalent probability measures $P \sim P'$. Suppose that two investors start with zero initial capital, one believes in P the other one in P'. Under P one can win 1 Million of dollars with probability 99% and lose 1 Million with 1%, while under P' it is the opposite. How should financial regulators judge such an investment? They could apply for instance the superhedging price which in this simple complete market situation is just 0 and balances out the heterogeneous believes. In contrast to that the Value at Risk at level λ (defined as in (Föllmer and Schied, 2011, Definition 4.40)) under P would lead to -1 Million of risk capital for all $\lambda \geq 1\%$ (meaning one can throw 1 Million away and is still safe), whereas under P' it would be 1 Million of risk capital for all $0 \leq \lambda < 99\%$. So if $\lambda = 5\%$, we would have either -1 Million or 1 Million of risk capital depending on the probability measure, showing that the Value of Risk (like many other risk measures mentioned below) does not satisfy the above condition (ii). From a heterogeneous beliefs perspective it is of course valid to associate different risks to the equivalent measures P and P' and the regulator could for instance always choose the VaR under P' to adapt a worst case point of view. However, as argued above the superhedging price allows for a more balanced risk assessment, taking the different heterogeneous beliefs seriously, and satisfies the conditions (i) and (ii). Another example from the literature that satisfies these conditions is the worst case risk measure given by $-\operatorname{ess\,inf} X$ as discussed in Section 4.3.3. Our goal is thus to introduce appropriate (families of) risk measures so that the requirements (i) and (ii) from above can be fulfilled. With this approach we then aim to find a classical representation of the robust risk measure in terms of a *single* probability measure, which can then be interpreted as such a 'worst-case probability measure'.

To make this program work we need to introduce appropriate (families of) risk measures so that the requirements (i) and (ii) from above can be fulfilled. In this respect we consider two different ways of defining risk measures with respect to some reference probability measure P (not necessarily equal to $\mathbb{P}$ here). In both approaches we fix a monetary risk measure ρ (in the sense of Section 4.1 or Föllmer and Schied (2002)) on the space $\mathcal{X}$ of bounded measurable random variables.

In the *first* approach described in Section 4.2.1 we then define for a given probability measure P and for all $X \in L^\infty(P)$,

$$\rho^P(X) := \inf_{\{\widetilde{X}\in\mathcal{X}:P(\widetilde{X}=X)=1\}} \rho(\widetilde{X}).$$

With this definition it can be easily verified that whenever $P' \ll P$, we have $\rho^{P'}(X) \leq \rho^P(X)$ for all $X \in L^\infty(P)$ and thus clearly also $\rho^{P'}(X) = \rho^P(X)$ if $P' \sim P$ (see Remark 4.2.2). Hence, whenever $\mathbb{P}$ takes the role of P, the desired properties (i) and (ii) are satisfied.

In the case where the fixed risk measure ρ from which we start is convex and continuous from below (see Section 4.1.2 for details) we can pursue a *second* approach, outlined in Section 4.2.2. Indeed, fix some penalty function α on the space of probability measures $\mathcal{M}_1$ and define $\rho : \mathcal{X} \to \mathbb{R}$ via

$$\rho(X) := \sup_{Q\in\mathcal{M}_1} \{\mathbb{E}_Q[-X] - \alpha(Q)\}, \qquad \forall X \in \mathcal{X}, \tag{‡}$$

such that it coincides with the given ρ, or just use (‡) as a way to fix a monetary risk measure via a fixed penalty function α. Given a probability measure P, the second approach then consists in defining another risk measure $\widehat{\rho}^P$ via

$$\widehat{\rho}^P(X) := \sup_{Q\in\mathcal{P}^P} \{\mathbb{E}_Q[-X] - \alpha(Q)\}, \quad X \in L^\infty(P),$$

where $\mathcal{P}^P$ denotes the set of probability measures dominated by P. Also this definition yields easily the properties $\widehat{\rho}^{P'}(X) \leq \widehat{\rho}^P(X)$ or any $P' \ll P$ and thus also $\widehat{\rho}^{P'}(X) = \widehat{\rho}^P(X)$ if $P' \sim P$ (see Remark 4.2.4).

Note that the families of risk measures considered in the literature cited above usually do not satisfy these requirements. This holds especially for the approach relying on distortions, which includes as particular cases Expected Shortfall and Value at Risk. Indeed, even for the most basic premium calculation principle (i.e. risk measure) in life insurance mathematics, which is just the expected value of the present value of a payment, the order condition is clearly not satisfied. Similarly, for an arbitrary distortion risk measure, as e.g. Value at Risk, simple examples with $P \sim P'$ and random variables X_1, X_2 where P gives a smaller value to X_1 while P' gives a smaller value to X_2 can be constructed to see that the order condition does not hold either.

Coming back to the above defined risk measures ρ^P and $\widehat{\rho}^P$, note that for ρ given by (‡) both ρ^P and $\widehat{\rho}^P$ can be defined, whence we aim for a comparison of them (see Section 4.2.3). While $\rho^P(X)$ is by definition given as the infimum of (‡), taken over P-almost

surely equal claims $\widetilde{X}$, it turns out that $\widehat{\rho}^P$ is equal to the corresponding expression where the infimum and the supremum are interchanged (see Lemma 4.2.5). Thus for $X \in L^\infty(P)$, $\rho^P(X) \geq \widehat{\rho}^P(X)$ in general. By means of a classical minimax theorem, we then derive sufficient conditions under which they coincide, see Proposition 4.2.6, but also provide an example where the strict inequality holds (see Example 4.2.8).

This comparison is continued in Section 4.2.4, where we specialize the previous setting further, assuming that the initial risk measure ρ is law-invariant with respect to some fixed probability measure, denoted here by $\mathbb{P}$, whose significance will become clear below. Starting from this law-invariant risk measure we proceed as above and define ρ^P and $\widehat{\rho}^P$, for some probability measure P and show that under certain assumptions, depending on the relation between $\mathbb{P}$ and P, ρ^P and $\widehat{\rho}^P$ coincide (see Proposition 4.2.10, Proposition 4.2.12 and Corollary 4.2.14). The economically most interesting aspect of this setup is that the starting probability measure $\mathbb{P}$ with respect to which law-invariance is considered turns out to qualify as 'worst case probability measure'. Indeed, the risk computed under $\mathbb{P}$ dominates the risk computed under any other probability measure P, no matter in which relation $\mathbb{P}$ and P are (see Proposition 4.2.10 for details).

Having introduced the above setup that allows to consider families of risk measures, i.e., either $(\rho^P)_{P\in\mathcal{M}}$ or $(\widehat{\rho}^P)_{P\in\mathcal{M}}$, induced by some family $\mathcal{M}$ of probability measures, we consider in Section 4.4 and Section 4.5 the corresponding robust versions, i.e.

$$\sup_{P\in\mathcal{M}} \rho^P \quad \text{or} \quad \sup_{P\in\mathcal{M}} \widehat{\rho}^P. \tag{$\dagger$}$$

Observe that these can be seen as particular *generalized risk measures*, introduced recently in Fadina et al. (2021). As mentioned above our main goal is to link the robust risk measures given by (†) with a *single* 'worst case probability measure' $\mathbb{P}$ and to express (†) via $\rho^{\mathbb{P}}$ or $\widehat{\rho}^{\mathbb{P}}$ respectively. Indeed, as we prove in Theorem 4.4.6 and Theorem 4.4.8 such a classical representation is possible if $\mathcal{M}$ is closed under countable convex combinations and if there is a measure $\mathbb{P}$ that satisfies the condition of *generalized equivalence* (see Definition 4.4.5). The latter means in particular that every measure in $\mathcal{M}$ can be dominated by $\mathbb{P}$. Even though any countable set of probability measures can be dominated, this can be considered a strong condition, as it excludes uncountable families of mutually singular measures. Note, however, that this rather strong domination property *would still not* be enough to get a similar result for other robust versions of risk measures, induced e.g. via distortions, since they do not satisfy the required order structure. Indeed, this generalized equivalence property only works in combination with the ordering requirements from above, which led to the specification of the considered family of risk measures being crucial for this result.

In the setup of Section 4.2.4, where the starting measure $\mathbb{P}$ with respect to which law-invariance is considered qualifies as 'worst case probability measure' we get a classical representation without any assumption except that $\mathbb{P}$ has to lie in $\mathcal{M}$. In particular, the *fully non-dominated case* can be considered in this setup. The reason why this works here is because the ordering is achieved on the level of the risk measures rather than the probability measures.

In the last part of Chapter 4, namely Section 4.5, we analyze a particular candidate for a 'worst case probability measure', giving rise to a *Bayesian point of view* to robust risk measures. Indeed, let $\mathcal{M} = \{P^\theta : \theta \in \Theta\}$ be a family of probability measures defined via a parameter space Θ, on which some (prior) distribution ν is specified to model for instance the regulator's beliefs. Consider then as candidate for the 'worst case probability measure', the so-called *mixture probability measure,* given by

$$\mathbb{P}(\cdot) = \int_\Theta P^\theta(\cdot)\nu(\mathrm{d}\theta).$$

We then show that under the (rather strong) assumption of continuity in total variation of $\theta \mapsto P^\theta$, this mixture probability measure actually qualifies as 'worst case probability measure' in the sense that it is a generalized equivalent measure for a subset of $\mathcal{M}$, induced from a subset of Θ that has ν-measure 1. Even though this continuity assumption implies the existence of a dominating measure (see Lemma 4.5.11), this measure is not necessarily equivalent to the mixture probability measure $\mathbb{P}$ (see Remark 4.5.14) and can thus potentially yield higher values for the associated risk. Taking instead the mixture measure as the 'worst case probability measure' means to take the (regulator's) beliefs expressed via ν seriously. As above we obtain a certain classical representation for the robust risk measures, now via the mixture probability measure (see Theorem 4.5.13). In the setup of Section 4.2.4, we can again consider a non-dominated situation to get, under mild conditions, a classical representation with respect to the mixture probability measure $\mathbb{P}$, where in this case $\mathbb{P}$ does not even have to lie in $\mathcal{M}$ (see Corollary 4.5.17). Summarizing our main contributions are twofold:

(1) we introduce two families of risk measures allowing for a certain domination property and analyze their relation to each other leading to the question when a certain minimax inequality is actually an equality. In the case of law-invariance our approaches naturally lead to a 'worst case probability measure' under which the risk is always maximal.

(2) we show under which conditions a classical representation of the robust risk measures in terms of a single probability holds, where a particular focus is on the mixture probability measure giving rise to a Bayesian point of view on robust risk measures.

Bayesian updating of the prior ν and computing the risk measure with respect to the new mixture measure is then equivalent to computing a robust risk measure where the set over which the supremum is taken has to be in accordance with the posterior distribution. It thus gives a way how to dynamically update robust risk measures. First attempts towards a generalization of the results obtained in Chapter 4 are outlined in Chapter 5, where we consider conditional risk measure as building block of dynamic risk measures (see e.g. Acciaio and Penner (2011)). There particular attention has to be taken, if we would like to mimick the approach in Chapter 4. Indeed to start from a pointwise definition of the risk measure, in contrast to the static case, we have to take care about the measurability of the risk measure.

Part I.

Signature-based models in finance

1. Signature-based models: theory and calibration

The present chapter is based on Cuchiero et al. (2022b), a joint work with Christa Cuchiero and Sara Svaluto-Ferro.

1.1. An introduction to signatures

We start by introducing basic notions related to the definition of the signature of an $\mathbb{R}^d$-valued continuous semimartingale. For similar introductions to the concept of signature we refer e.g. to Section 2.2. in Bayer et al. (2021).

1.1.1. Basic notions

For each $m \in \mathbb{N}_0$ consider the m-fold tensor product of $\mathbb{R}^d$ given by

$$(\mathbb{R}^d)^{\otimes 0} := \mathbb{R}, \qquad (\mathbb{R}^d)^{\otimes m} := \underbrace{\mathbb{R}^d \otimes \cdots \otimes \mathbb{R}^d}_{m}.$$

For $d \in \mathbb{N}$, we define the extended tensor algebra on $\mathbb{R}^d$ as

$$T((\mathbb{R}^d)) := \{\mathbf{a} := (a_0, \dots, a_m, \dots) : a_n \in (\mathbb{R}^d)^{\otimes m}\}.$$

Similarly we introduce the truncated tensor algebra of order $n \in \mathbb{N}$

$$T^{(n)}(\mathbb{R}^d) := \{\mathbf{a} \in T((\mathbb{R}^d)) : a_m = 0, \forall m > n\},$$

and the tensor algebra $T(\mathbb{R}^d) := \bigcup_{n\in\mathbb{N}} T^{(n)}(\mathbb{R}^d)$. Note that $T^{(n)}(\mathbb{R}^d)$ has dimension $d_n := \sum_{i=0}^n d^i = (d^{n+1} - 1)/(d - 1)$.

For each $\mathbf{a}, \mathbf{b} \in T((\mathbb{R}^d))$ and $\lambda \in \mathbb{R}$ we set

$$\begin{aligned} \mathbf{a} + \mathbf{b} &:= (a_0 + b_0, \dots, a_n + b_n, \dots), \\ \lambda \cdot \mathbf{a} &:= (\lambda a_0, \dots, \lambda a_n, \dots), \\ \mathbf{a} \otimes \mathbf{b} &:= (c_0, \dots, c_n, \dots), \end{aligned}$$

where $c_n := \sum_{k=0}^n a_k \otimes b_{n-k}$, for some $n > 0$. Observe that $(T((\mathbb{R}^d)), +, \cdot, \otimes)$ is a real non-commutative algebra. For a multi-index $I := (i_1, \dots, i_n)$ we set $|I| := n$. We also consider the empty index $I := \emptyset$ and set $|I| := 0$. For each index I we then define

$$I' := \begin{cases} (i_1, \dots, i_{|I|-1}) & \text{if } |I| \geq 2, \\ \emptyset & \text{if } |I| = 1, \\ 0 & \text{if } |I| = 0. \end{cases} \tag{1.1.1}$$

Similarly we set $I'' := (I')'$ if $|I| \neq 0$ and $I'' = 0$ if $|I| = 0$. We also use the notation

$$\{I : |I| = n\} := \{1, \dots, d\}^n,$$

omitting the parameter d whenever this does not introduce ambiguity.

Next, for each $|I| \geq 1$ we set

$$e_I := e_{i_1} \otimes \cdots \otimes e_{i_n},$$

where $e_1, \dots, e_d$ denote the canonical basis vectors of $\mathbb{R}^d$. Observe that the set $\{e_I : |I| = n\}$ is an orthonormal basis of $(\mathbb{R}^d)^{\otimes n}$. Observe that multi-indices can be identified with words, i.e., the following equivalence holds true for any couple of multi-indices I, J,

$$I \otimes J \Longleftrightarrow e_I \otimes e_J.$$

We refer the reader to Perez Arribas (2020); Lyons et al. (2020); Bayer et al. (2021), for additional details concerning the previous isomorphism.

Denoting by $e_\emptyset$ the basis element corresponding to $(\mathbb{R}^d)^{\otimes 0}$, each element of $\mathbf{a} \in T((\mathbb{R}^d))$ can thus be written as

$$\mathbf{a} = \sum_{|I| \geq 0} \mathbf{a}_I e_I,$$

for some $\mathbf{a}_I \in \mathbb{R}$. Note that if $a_n \in (\mathbb{R}^d)^{\otimes n}$ we use non-bold notation whereas for the components $\mathbf{a}_I \in \mathbb{R}$ we write them bold. Finally, for each $\mathbf{a} \in T(\mathbb{R}^d)$ and each $\mathbf{b} \in T((\mathbb{R}^d))$ we set

$$\langle \mathbf{a}, \mathbf{b} \rangle := \sum_{|I| \geq 0} \langle \mathbf{a}_I, \mathbf{b}_I \rangle.$$

Observe in particular that $\mathbf{b}_I = \langle e_I, \mathbf{b} \rangle$. In the present work it will be useful to enumerate the elements of the truncated tensor algebra. To this extent we introduce the isomorphism $\mathbf{vec} : T^{(n)}(\mathbb{R}^d) \to \mathbb{R}^{d_n}$ and an injective labeling function $\mathscr{L} : \{I : |I| \leq n\} \longrightarrow \{1, \dots, d_n\}$, such that

$$\mathbf{vec}(\mathbf{u}) := \sum_{|I| \leq n} e_{\mathscr{L}(I)} \mathbf{u}_I. \tag{1.1.2}$$

Finally, we denote by $(\,\cdot\,)^{(1)} : T((\mathbb{R}^d)) \to (T((\mathbb{R}^d)))^d$ and $(\,\cdot\,)^{(2)} : T((\mathbb{R}^d)) \to (T((\mathbb{R}^d)))^{d \times d}$ the shifts given by

$$\begin{aligned} \mathbf{u}^{(1)} &:= \Big(\sum_{|I| \geq 0} \mathbf{u}_I e_{I'} \Big) (1_{\{i_{|I|}=1\}}, \dots, 1_{\{i_{|I|}=d\}})^\top, \\ \mathbf{u}^{(2)} &:= \Big(\sum_{|I| \geq 0} \mathbf{u}_I e_{I''} \Big) \begin{pmatrix} 1_{\{i_{|I|-1}=i_{|I|}=1\}} & \cdots & 1_{\{i_{|I|-1}=1, i_{|I|}=d\}} \\ \vdots & \dots & \vdots \\ 1_{\{i_{|I|-1}=d, i_{|I|}=1\}} & \cdots & 1_{\{i_{|I|-1}=i_{|I|}=d\}} \end{pmatrix}. \end{aligned} \tag{1.1.3}$$

Observe in particular that

$$\langle \mathbf{a}, \mathbf{u}^{(1)} \rangle = \sum_{|I| \geq 0} \big(\mathbf{a}_I \mathbf{u}_{(I1)}, \dots, \mathbf{a}_I \mathbf{u}_{(Id)} \big)^\top \quad \text{and} \quad \langle \mathbf{a}, \mathbf{u}^{(2)} \rangle = \begin{pmatrix} \mathbf{a}_I \mathbf{u}_{(I11)} & \cdots & \mathbf{a}_I \mathbf{u}_{(I1d)} \\ \vdots & \dots & \vdots \\ \mathbf{a}_I \mathbf{u}_{(Id1)} & \cdots & \mathbf{a}_I \mathbf{u}_{(Idd)} \end{pmatrix}$$

for each $\mathbf{a} \in T(\mathbb{R}^d)$.

Throughout the chapter we consider a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$. Whenever not specified, stochastic processes are always supposed to be defined there. We are now ready to define the signature of an $\mathbb{R}^d$-valued continuous semimartingale.

Definition 1.1.1 (Signature)**.** Let $(X_t)_{t\in[0,T]}$ be a continuous $\mathbb{R}^d$-valued semimartingale. The *signature of* X is the $T((\mathbb{R}^d))$-valued process $(s,t) \mapsto \mathbb{X}_{s,t}$ whose components are recursively defined as

$$\langle e_\emptyset, \mathbb{X}_{s,t}\rangle := 1, \qquad \langle e_I, \mathbb{X}_{s,t}\rangle := \int_s^t \langle e_{I'}, \mathbb{X}_{s,r}\rangle \circ \mathrm{d}X_r^{i_n},$$

for each $I = (i_1, \dots, i_n)$, $I' = (i_1, \dots, i_{n-1})$ and $0 \leq s \leq t \leq T$, where $\circ$ denotes the Stratonovich integral. Its projection $\mathbb{X}^n$ on $T^{(n)}(\mathbb{R}^d)$ is given by

$$\mathbb{X}^n_{s,t} = \sum_{|I|\leq n} \langle e_I, \mathbb{X}_{s,t}\rangle e_I$$

and is called *signature of* X *truncated at level* n. If $s = 0$, we use the notation $\mathbb{X}_t$ and $\mathbb{X}^n_t$, respectively.

Observe that the signature of X and the signature of $X - c$ coincides for each $c \in \mathbb{R}$. Moreover, with an equivalent notation we can write

$$\begin{aligned}\mathbb{X}_t = \Bigg(1, \int_0^t 1 \circ \mathrm{d}X_s^1, \dots, \int_0^t 1 \circ \mathrm{d}X_s^d, \int_0^t \left(\int_0^s 1 \circ \mathrm{d}X_r^1\right) \circ \mathrm{d}X_s^1, \\ \int_0^t \left(\int_0^s 1 \circ \mathrm{d}X_r^1\right) \circ \mathrm{d}X_s^2, \dots, \int_0^t \left(\int_0^s 1 \circ \mathrm{d}X_r^d\right) \circ \mathrm{d}X_s^d, \dots\Bigg).\end{aligned}$$

Using Itô integrals this can thus be rewritten as

$$\begin{aligned}\mathbb{X}_t = \Bigg(1, X_t^1 - X_0^1, \dots, X_t^d - X_0^d, \int_0^t (X_s^1 - X_0^1)\mathrm{d}X_s^1 + \frac{1}{2}[X^1]_t, \\ \int_0^t (X_s^1 - X_0^1)\mathrm{d}X_s^2 + \frac{1}{2}[X^1, X^2]_t, \dots, \int_0^t (X_s^d - X_0^d)\mathrm{d}X_s^d + \frac{1}{2}[X^d]_t, \dots\Bigg),\end{aligned}$$

where the square brackets denote the quadratic variation and covariation processes. In particular, by the definition of the Stratonovich integral and Itô's formula we can write

$$\langle e_{(i,i)}, \mathbb{X}_t\rangle = \int_0^t \left(\int_0^s 1 \circ \mathrm{d}X_r^i\right) \circ \mathrm{d}X_s^i = \frac{1}{2}(X_t^i - X_0^i)^2 = \langle e_i, \mathbb{X}_t\rangle^2,$$

showing that the quadratic expression on the right hand side has a linear representation. This property generalizes to every polynomial function. For the precise statement we first need to introduce the shuffle product.

Definition 1.1.2 (Shuffle product)**.** For every two multi-indices $I := (i_1, \dots, i_n)$ and $J := (j_1, \dots, j_m)$ the *shuffle product* is defined recursively as

$$e_I \sqcup\!\sqcup e_J := (e_{I'} \sqcup\!\sqcup e_J) \otimes e_{i_n} + (e_I \sqcup\!\sqcup e_{J'}) \otimes e_{j_m},$$

with $e_I \sqcup\!\sqcup e_\emptyset := e_\emptyset \sqcup\!\sqcup e_I = e_I$. It extends to $\mathbf{a}, \mathbf{b} \in T(\mathbb{R}^d)$ as

$$\mathbf{a} \sqcup\!\sqcup \mathbf{b} = \sum_{|I|,|J| \geq 0} a_I b_J (e_I \sqcup\!\sqcup e_J).$$

For later convenience, let us denote by $\iota^K_{(I \sqcup\!\sqcup J)}$ the number of times, the term e_K appears in $e_{I \sqcup\!\sqcup J}$. Note that $\iota^K_{(I \sqcup\!\sqcup J)} = 0$ whenever $|K| \neq |I| + |J|$. Moreover, we can write

$$e_{I \sqcup\!\sqcup J} = \sum_{|K|=|I|+|J|} \iota^K_{(I \sqcup\!\sqcup J)} \cdot e_K.$$

Observe that $(T(\mathbb{R}^d), +, \sqcup\!\sqcup)$ is a commutative algebra, which in particular means that the shuffle product is associative and commutative. Recall that,

$$|\{e_I \sqcup\!\sqcup e_J : |I| = n, |J| = m\}| = \frac{(m+n)!}{m!n!}.$$

The proof of the rough paths version of the next result for can be found for instance in Ree (1958) or Lyons et al. (2007).

Proposition 1.1.3 (Shuffle property)**.** Let $(X_t)_{t \in [0,T]}$ be a continuous $\mathbb{R}^d$-valued semimartingale and I, J be two multi-indices. Then,

$$\langle e_I, \mathbb{X} \rangle \langle e_J, \mathbb{X} \rangle = \langle e_I \sqcup\!\sqcup e_J, \mathbb{X} \rangle. \tag{1.1.4}$$

Proof. The result follows by induction using the chain rule for Stratonovich integrals. □

Example 1.1.4. Let $(X_t)_{t \in [0,T]}$ be a continuous $\mathbb{R}^d$-valued semimartingale with $X_0 = 0$. Then the (Stratonovich) integration by parts formula yields, for any $i, j \in \{1, \dots, d\}$

$$\begin{aligned} \langle e_i, \mathbb{X}_T \rangle \langle e_j, \mathbb{X}_T \rangle = X^i_T X^j_T &= \int_0^T X^i_t \circ \mathrm{d}X^j_t + \int_0^T X^j_t \circ \mathrm{d}X^i_t, \\ &= \langle e_i \otimes e_j, \mathbb{X}_T \rangle + \langle e_j \otimes e_i, \mathbb{X}_T \rangle \\ &= \langle e_i \sqcup\!\sqcup e_j, \mathbb{X}_T \rangle. \end{aligned}$$

Example 1.1.5. Let $(X_t)_{t \in [0,T]}$ be a continuous $\mathbb{R}$-valued semimartingale with $X_0 = 0$. Then $\langle e_1, \mathbb{X}_T \rangle = X_T - X_0$. Since

$$\underbrace{e_1 \sqcup\!\sqcup \ldots \sqcup\!\sqcup e_1}_{k\text{-times}} = k! \underbrace{e_1 \otimes \ldots \otimes e_1}_{k\text{-times}},$$

Proposition 1.1.3 yields

$$\mathbb{X}_T = (1, X_T - X_0, \frac{1}{2}(X_T - X_0)^2, \frac{1}{3!}(X_T - X_0)^3, \dots, \frac{1}{k!}(X_T - X_0)^k, \dots).$$

We recall now an important property of the signature. The result is known in the rough paths literature (see for instance Boedihardjo et al. (2016)), but can also be proved directly in the simpler situation of a continuous semimartingale that contains time as strictly monotone component.

Lemma 1.1.6 (Uniqueness of the signature)**.** Let $(X_t)_{t\in[0,T]}$ and $(Y_t)_{t\in[0,T]}$ be two continuous $\mathbb{R}^d$-valued semimartingales with $X_0 = Y_0 = 0$. Set $\widehat{X}_t := (t, X_t)$, $\widehat{Y}_t := (t, Y_t)$ and let $\widehat{\mathbb{X}}$ and $\widehat{\mathbb{Y}}$ be the corresponding signature processes. Then $\widehat{\mathbb{X}}_T = \widehat{\mathbb{Y}}_T$ if and only if $X_t = Y_t$ for each $t \in [0, T]$.

Proof. Denote by $0, \dots, d$ the indices of $\widehat{X}$ and $\widehat{Y}$, where 0 corresponds to the time component. The claim follows by noticing that $\langle (e_i \sqcup\!\sqcup (e_0^{\otimes k})) \otimes e_0, \widehat{\mathbb{X}}_T\rangle = \int_0^T X_t^i \frac{t^k}{k!}\mathrm{d}t$, which implies that $\widehat{\mathbb{X}}_T$ uniquely determines $\widehat{X}_t$ for every $t \in [0, T]$. □

Remark 1.1.7. Observe that $(t^k/k!)_{k\in\mathbb{N}_0}$ is a basis of $L^2([0, T], \mathrm{d}t)$ and

$$\langle (e_i \sqcup\!\sqcup (e_0^{\otimes k})) \otimes e_0, \widehat{\mathbb{X}}_T\rangle = \int_0^T X_t^i \frac{t^k}{k!}\mathrm{d}t$$

is the corresponding projection of X^i on $t^k/k!$. This property can be used to explicitly construct a sequence of polynomials with random coefficients on $[0, T]$ converging almost surely to X^i on $L^2([0, T], \mathrm{d}t)$. It also establishes a link between signature-based methods and quantization methods (see for instance Luschgy and Pagès (2002); Pagès and Printems (2005) and Bonesini et al. (2021); Tissot-Daguette (2022) for recent advances).

Let us now introduce the definition of the half-shuffle product (see also Ebrahimi-Fard and Patras (2015), Diehl et al. (2020), where it is usually denoted with $\prec$). Given an $\mathbb{R}^d$-valued continuous semimartingale, this operation permits to write the signature of $\mathbb{X}^n$ as a linear map of $\mathbb{X}$. This result will be useful later when we will need to compute the signature of a model given by a linear combination of terms of $\widehat{\mathbb{X}}^n$ (see Section 1.3).

Definition 1.1.8 (Half-shuffle)**.** Set $I = (i_1, \dots, i_n)$ and $J = (j_1, \dots, j_m)$ for some $n \in \mathbb{N}$ and $m > 0$. We define the half-shuffle $\widetilde{\sqcup\!\sqcup}$ as $e_I \widetilde{\sqcup\!\sqcup} e_\emptyset := 0$ and

$$e_I \widetilde{\sqcup\!\sqcup} e_J := (e_I \sqcup\!\sqcup e_{J'}) \otimes e_{j_m}.$$

Lemma 1.1.9. For each multi-indices $I = (i_1, \dots, i_n)$ and $J = (j_1, \dots, j_m)$ it holds

$$\int_0^t \langle e_I, \mathbb{X}_s\rangle \circ \mathrm{d}\langle e_J, \mathbb{X}_s\rangle = \langle e_I \widetilde{\sqcup\!\sqcup} e_J, \mathbb{X}_t\rangle, \tag{1.1.5}$$

for each $t \geq 0$.

Proof. By definition of the Stratonovich integral, Proposition 1.1.3, and by the definition of the signature we know that

$$\int_0^t \langle e_I, \mathbb{X}_s\rangle \circ \mathrm{d}\langle e_J, \mathbb{X}_s\rangle = \int_0^t \langle e_I, \mathbb{X}_s\rangle\langle e_{J'}, \mathbb{X}_s\rangle \circ \mathrm{d}\langle e_{j_m}, X_s\rangle = \langle (e_I \sqcup\!\sqcup e_{J'}) \otimes e_{j_m}, \mathbb{X}_t\rangle.$$

□

In order to combine the value of the signature on different time intervals Chen's identity going back to Chen (1957, 1977) turns out to be fundamental. For the reader convenience, we propose here a direct proof using the definition of Stratonovich integrals.

Lemma 1.1.10 (Chen's identity)**.** Let $(X_t)_{t\in[0,T]}$ be an $\mathbb{R}^d$-valued semimartingale. Then

$$\mathbb{X}_{s,t} = \mathbb{X}_{s,u} \otimes \mathbb{X}_{u,t}$$

for each $s \leq u \leq t \leq T$. This can equivalently be written as

$$\langle e_I, \mathbb{X}_{s,t}\rangle = \sum_{e_{I_1}\otimes e_{I_2}=e_I} \langle e_{I_1}, \mathbb{X}_{s,u}\rangle\langle e_{I_2}, \mathbb{X}_{u,t}\rangle,$$

for each multi-index I.

Proof. We proceed by induction on the length of the multi-index I. For $|I| = 0$ we have $I = \emptyset$ and the statement is clear. Suppose now that the claim holds for each $|J| < n$ and set $I = (i_1, \dots, i_n)$. Then, applying Chen's identity to $\langle e_{I'}, \mathbb{X}_{s,r}\rangle$ yields

$$\begin{aligned}
\langle e_I, \mathbb{X}_{s,t}\rangle &= \int_s^t \langle e_{I'}, \mathbb{X}_{s,r}\rangle \circ \mathrm{d}X_r^{i_n} \\
&= \int_s^u \langle e_{I'}, \mathbb{X}_{s,r}\rangle \circ \mathrm{d}X_r^{i_n} + \int_u^t \sum_{e_{I'_1}\otimes e_{I'_2}=e_{I'}} \langle e_{I'_1}, \mathbb{X}_{s,u}\rangle\langle e_{I'_2}, \mathbb{X}_{u,r}\rangle \circ \mathrm{d}X_r^{i_n} \\
&= \langle e_I, \mathbb{X}_{s,u}\rangle + \sum_{e_{I'_1}\otimes e_{I'_2}=e_{I'}} \langle e_{I'_1}, \mathbb{X}_{s,u}\rangle\langle e_{I'_2} \otimes e_{i_n}, \mathbb{X}_{u,t}\rangle \\
&= \sum_{e_{I_1}\otimes e_{I_2}=e_I} \langle e_{I_1}, \mathbb{X}_{s,u}\rangle\langle e_{I_2}, \mathbb{X}_{u,t}\rangle,
\end{aligned}$$

whence the claim follows. □

1.1.2. Universal approximation theorem

This section is devoted to the universal approximation result mentioned in the introduction. Loosely speaking, it states that every quantity of the form

$$f\Big((\widehat{\mathbb{X}}_t^2)_{t\in[0,T]}\Big)$$

for some continuous map f and some $T > 0$ can be approximated arbitrarily well on compact sets by linear functions of the signature of the form $\langle \ell, \widehat{\mathbb{X}}_T\rangle$ where $\ell \in T(\mathbb{R}^d)$. Observe that the latter just involves the *final value* $\widehat{\mathbb{X}}_T$ of $\widehat{\mathbb{X}}$, instead of its whole trajectory. We refer to Cuchiero and Möller (2023) or Lemma 5.2 in Bayer et al. (2021), in the same context, for the situation involving not just the approximation of a functional up to a fixed *final value* T, but a uniform approximation on the whole time interval $[0, T]$. Different versions of this result for a fixed final time, e.g. for finite variation paths or for

continuous functions depending on the whole signature (instead of level 2), are available in the literature (see for instance Theorem 3.1 in Levin et al. (2013), Theorem 1 in Király and Oberhauser (2019), Proposition 4.5 in Lyons et al. (2020) and Section 3 in Cuchiero et al. (2022c)). For completeness and to keep the paper self-contained we prove it here in the current context of continuous semimartingales and continuous functions of the respective paths lifted up to order 2.

Fix $T > 0$, consider a continuous $\mathbb{R}^d$-valued semimartingale $(X_t)_{t\in[0,T]}$, and let $\widehat{X}_t := (t, X_t)$. Denote by $0, \dots, d$ the indices of $\widehat{X}$, where 0 corresponds to the time component. For each $N \in \mathbb{N}$ define the set

$$\mathcal{S}^{(N)} := \{(\widehat{\mathbb{X}}^N_t)_{t\in[0,T]}(\omega) \colon \omega \in \Omega\},$$

which, without loss of generality (passing to a subset of Ω of measure 1), corresponds to a set of signature paths of $\widehat{X}$ up to time T.

The next lemma states that the signature of a continuous semimartingale coincides with the so-called Lyons lift (see for instance Theorem 2.2.1 in Lyons (1998)). This in particular implies that higher order terms of the signature can be defined pathwise starting from the trajectories of $\widehat{\mathbb{X}}^2$.

Lemma 1.1.11. For each $n \in \mathbb{N}$ there exists a map $S^{(n)} : \mathcal{S}^{(2)} \to \mathcal{S}^{(n)}$ such that

$$\widehat{\mathbb{X}}^N = S^{(N)}(\widehat{\mathbb{X}}^2) \tag{1.1.6}$$

almost surely.

Proof. The claim follows from Exercise 17.2 and Theorem 9.5 in Friz and Victoir (2010). □

Without loss of generality (passing again to a subset of Ω of measure 1) assume that conditions (1.1.4), (1.1.5), (1.1.6) and

$$\langle e_I \otimes e_0, \widehat{\mathbb{X}}_t(\omega)\rangle = \int_0^t \langle e_I, \widehat{\mathbb{X}}_s(\omega)\rangle \mathrm{d}s$$

holds for each $\omega \in \Omega$ and each multi-index I. Consider a generic distance $d_{\mathcal{S}^{(2)}}$ on the set of trajectories given by $\mathcal{S}^{(2)}$, with respect to which the map from $\mathcal{S}^{(2)}$ to $\mathbb{R}$ given by

$$\hat{\mathbf{x}}^2 \mapsto \langle e_I, S^{(|I|)}(\hat{\mathbf{x}}^2)_t\rangle$$

is continuous for each multi-index I and every $t \in [0, T]$.

Theorem 1.1.12 (Universal approximation theorem)**.** *Let K be a compact subset of $\mathcal{S}^{(2)}$ and consider a continuous map $f : K \to \mathbb{R}$.*[1] *Then for every $\varepsilon > 0$ there exists some $\ell \in T(\mathbb{R}^d)$ such that*

$$\sup_{(\widehat{\mathbb{X}}^2_t)_{t\in[0,T]}\in K} |f((\widehat{\mathbb{X}}^2_t)_{t\in[0,T]}) - \langle \ell, \widehat{\mathbb{X}}_T\rangle| < \varepsilon,$$

almost surely.

[1] Compactness and continuity are defined with respect to $d_{\mathcal{S}^{(2)}}$.

Proof. The result follows by an application of the Stone-Weierstrass theorem. To check the needed properties, observe that by Proposition 1.1.3 the set

$$A := \operatorname{span}\{\hat{\mathbf{x}}^2 \mapsto \langle e_I, S^{(|I|)}(\hat{\mathbf{x}}^2)_T\rangle \colon I \in \{0,\dots,d\}^{|I|}\}$$

is an algebra of continuous maps from $(K, d_{\mathcal{S}^{(2)}})$ to $\mathbb{R}$. Choosing $I = \emptyset$ we get that A is vanishing nowhere and we just need to check that it is point separating in $\mathcal{S}^{(2)}$. Lemma 1.1.6 yields the result for $\langle e_1, \hat{\mathbf{x}}^2\rangle, \dots, \langle e_d, \hat{\mathbf{x}}^2\rangle$. Due to Lemma 1.1.9, this then applies also to the remaining components of $\hat{\mathbf{x}}^2$. □

Remark 1.1.13. (i) Observe that the assumptions on $d_{\mathcal{S}^{(2)}}$ guarantee that every map of the form

$$f\Big((\widehat{\mathbb{X}}^2_t)_{t\in[0,T]}\Big) := \tilde{f}(\langle e_{I_1}, \widehat{\mathbb{X}}_{t_1}\rangle, \dots, \langle e_{I_n}, \widehat{\mathbb{X}}_{t_n}\rangle),$$

for some $\tilde{f} \in C(\mathbb{R}^n)$, is continuous with respect to $d_{\mathcal{S}^{(2)}}$.

(ii) The rough paths literature provides several metrics $d_{\mathcal{S}^{(n)}}$ with respect to which the maps $S^{(n)} : (\mathcal{S}^{(2)}, d_{\mathcal{S}^{(2)}}) \to (\mathcal{S}^{(n)}, d_{\mathcal{S}^{(n)}})$ are almost surely continuous on bounded sets. Among them one can for instance consider

$$d_{\mathcal{S}^{(n)}}(\hat{\mathbf{x}}, \hat{\mathbf{y}}) := \max_{k\in\{1,\dots,n\}} \sup_{\mathcal{D}} \left(\sum_{t_i\in\mathcal{D}} |\pi_k(\hat{\mathbf{x}}_{t_{i-1},t_i} - \hat{\mathbf{y}}_{t_{i-1},t_i})|^{p/k} \right)^{k/p}.$$

where $p \in (2,3)$, $\pi_k(\hat{\mathbf{x}}) := \sum_{|I|=k}\langle e_I, \hat{\mathbf{x}}\rangle e_I$, and $\mathcal{D}$ denotes the set of all partitions of $[0,T]$, see Corollary 9.11, Definition 8.6 and Theorem 8.10 in Friz and Victoir (2010) for the result or Cuchiero et al. (2022c) for a more detailed explanation. Note that

$$\pi_1(\hat{\mathbf{x}}_{t_{i-1},t_i}) = \sum_{i=1}^{d} \langle e_i, \hat{\mathbf{x}}_{t_i}\rangle - \langle e_i, \hat{\mathbf{x}}_{t_{i-1}}\rangle$$

and

$$\pi_2(\hat{\mathbf{x}}_{t_{i-1},t_i}) = \sum_{i,j=1}^{d} \langle e_{ij}, \hat{\mathbf{x}}_{t_i}\rangle - \langle e_{ij}, \hat{\mathbf{x}}_{t_{i-1}}\rangle - \langle e_i, \hat{\mathbf{x}}_{t_{i-1}}\rangle(\langle e_j, \hat{\mathbf{x}}_{t_i}\rangle - \langle e_j, \hat{\mathbf{x}}_{t_{i-1}}\rangle).$$

(iii) The universal approximation theorem can also be stated without constructing the involved spaces based on a fixed semimartingale. Since this would require a more advanced knowledge of rough paths theory, we address the interested reader to Lyons (1998) or Chevyrev and Friz (2019) for uniqueness and continuity of $S^{(n)} : (\mathcal{S}^{(2)}, d_{\mathcal{S}^{(2)}}) \to (\mathcal{S}^{(n)}, d_{\mathcal{S}^{(n)}})$ and to Cuchiero et al. (2022c) for a formulation and proof of the resulting statement in the case of càdlàg semimartingales.

(iv) In other versions of the universal approximation theorem the map f is defined on a subset of $T((\mathbb{R}^d))$-valued paths, which includes the realisations of $\mathbb{X}$, see for instance Theorem 3.1 in Levin et al. (2013) and Proposition 4.5 in Lyons et al. (2020). Here, the critical issues are to define the involved metric explicitly and to identify compact subsets of $T((\mathbb{R}^d))$-valued paths.

Finally, we introduce the concept of polynomial process which will play a key role for the computation of expected signatures. Here we denote by $\sqrt{\cdot}$ the matrix square root.

Definition 1.1.14. Suppose that an $\mathbb{R}^d$-valued process $X = (X_t)_{t\geq 0}$ satisfies

$$\mathrm{d}X_t = b(X_t)\mathrm{d}t + \sqrt{a(X_t)}\mathrm{d}W_t, \qquad X_0 = x_0 \tag{1.1.7}$$

for some d-dimensional Brownian motion W and some maps $a : \mathbb{R}^d \to \mathbb{S}_+^d$ and $b : \mathbb{R}^d \to \mathbb{R}^d$ such that a_{ij} is a polynomial of degree at most 2 and b_j is a polynomial of degree at most 1 for each $i, j \in \{1, \dots, d\}$.

1.1.3. Stochastic Stratonovich Taylor expansion

In addition to the (global) universal approximation theorem formulated above, we now also state a well-known quantitative approximation result for solutions of SDEs based on the (stochastic) Stratonovich Taylor expansion, see Kloeden and Platen (1992). We also refer to Litterer and Oberhauser (2014) for approximations of non-anticipative path functionals of SDEs via Chen-Fliess series of iterated stochastic integrals. In contrast to Kloeden and Platen (1992) and Litterer and Oberhauser (2014), where the L^2-error is considered (see Remark 1.1.16 below), we here directly provide a 'convergence rate in probability' without requiring second moment estimates on the coefficients of the stochastic Taylor expansion. To this end we rely on the deterministic estimates provided for rough differential equations as in Section 10 in Friz and Victoir (2010). We additionally refer the reader to Dupire and Tissot-Daguette (2022) for a recent contribution concerning functional expansions.

Proposition 1.1.15. Let $\widehat{X}$ be a time extended d-dimensional Brownian motion and $(Y_t)_{t\in[0,T]}$ be an D-dimensional strong solution of the SDE

$$\mathrm{d}Y_t = \sum_{j=0}^{d} \Psi_j(Y_t)^\top \circ \mathrm{d}\widehat{X}_t^j,$$

for some smooth $\Psi_j : \mathbb{R}^D \to \mathbb{R}^D$. For each I let $\Psi_I : \mathbb{R}^D \to \mathbb{R}^D$ be the map whose k-th component is given by

$$\Psi_I^k(y) = \Psi_{i_1}^\top \nabla(\Psi_{i_2}^\top \nabla \cdots (\Psi_{i_n}^\top e_k))(y).$$

Then, for each $m \geq 2$ and $\varepsilon > 0$ there is a constant $C(m, \varepsilon)$ such that

$$\mathbb{P}\left(\left|Y_t - Y_0 - \sum_{0<|I|\leq m} \Psi_I(Y_0)\langle e_I, \widehat{\mathbb{X}}_t\rangle\right| > C(m,\varepsilon)t^{(m+1)/2}\right) < \varepsilon.$$

Proof. For each $K > 0$ consider the stopping times $\tau_K := \inf\{t \geq 0 \colon |Y_t| \geq K\}$. For each j and K fix smooth $\Psi_j^K : \mathbb{R}^D \to \mathbb{R}^D$ such that $\Psi_j^K(y) = \Psi_j(y)$ for each $|y| \leq K$ and

$\operatorname{supp}(\Psi_j^K) \subseteq \{y \in \mathbb{R}^D \colon |y| \leq K+1\}$. Choose $\gamma := m+1$ and $2 < p \leq \gamma$. Exercise 17.2, Theorem 17.3 and Corollary 10.15 in Friz and Victoir (2010)[2] applied to

$$\mathrm{d}Y_t = \sum_{j=0}^{d} \Psi_j^K(Y_t) \circ \mathrm{d}\widehat{X}^j_{t\wedge\tau_k}$$

yields the existence of a constant C such that

$$\left| Y_t - Y_0 - \sum_{0<|I|\leq m} \Psi_I(y_0)\langle e_I, \widehat{\mathbb{X}}_t\rangle \right| 1_{\{\tau_K \geq T\}} \leq |\Psi|^\gamma_{\mathrm{Lip}^{\gamma-1}} C \|\widehat{\mathbb{X}}^2\|^\gamma_{1/p-\mathrm{Höl};[0,T]} t^{\gamma/p},$$

almost surely, where $|\Psi|^\gamma_{\mathrm{Lip}^{\gamma-1}}$ and $\|\cdot\|^\gamma_{1/p-\mathrm{Höl};[0,T]}$ are defined in Definition 10.2 and Equation (8.2) of Friz and Victoir (2010), respectively. Observe that $\gamma/p \leq (m+1)/2$ and

$$|\Psi|^\gamma_{\mathrm{Lip}^{\gamma-1}} \leq \sup_{k\leq m+1, |y|\leq K+1} |\Psi^{(k)}(y)|^\gamma \leq \left(1 + \sup_{k\leq m+1, |y|\leq K+1} |\Psi^{(k)}(y)|\right)^{m+1}.$$

By the discussion at the beginning of Section 3 in Friz and Victoir (2010) we also get that $\|\widehat{\mathbb{X}}^2\|^\gamma_{1/p-\mathrm{Höl};[0,T]}$ is finite almost surely and the claim follows. □

Remark 1.1.16. Alternatively, an L^2-error for the stochastic Taylor expansion is given in Proposition 5.10.1 in Kloeden and Platen (1992). It provides (under some technical conditions, in particular second moments of Y_t) the following estimate

$$\mathbb{E}\left[\left| Y_t - Y_0 - \sum_{0<|I|\leq m} \Psi_I(Y_0)\langle e_I, \widehat{\mathbb{X}}_t\rangle \right|^2\right] \leq C t^{m+1},$$

for some $C \geq 0$. An application of the Markov inequality then yields the previous estimate

$$\mathbb{P}\left(\left| Y_t - Y_0 - \sum_{0<|I|\leq m} \Psi_I(Y_0)\langle e_I, \widehat{\mathbb{X}}_t\rangle \right| > \widetilde{C}^{1/2} t^{(m+1)/2}\right) \leq \varepsilon,$$

where $\widetilde{C} := C/\varepsilon$, for each $\varepsilon > 0$.

We now introduce a framework for signature-based asset price models. To this end we fix a time horizon $T > 0$, consider a d-dimensional continuous semimartingale $X := (X^1, \ldots, X^d)$ and its matrix-valued quadratic covariation $[X]$. We suppose that X encodes all the information to represent the market's assets S. For notational convenience we only consider a single asset, i.e. assume that S is one-dimensional.

We shall furthermore assume that X has some tractability properties which are made precise below and which are for instance satisfied by a d-dimensional Brownian motion (see Section A.1.1 and Example 1.4.1). Alternatively, one can choose X to be a collection of liquid financial products, whose time series are easily accessible and can be interpreted as realizations of continuous semimartingales.

[2] Observe that in Friz and Victoir (2010) a vector field Ψ_j is identified with the first order operator $f \mapsto \Psi_j \cdot \nabla f$ (see for instance the discussion on page 125). The Euler scheme $\mathcal{E}_\Psi$ is defined in Definition 10.1.

1.2. Definition and first properties

As the continuous semimartingale X shall serve as main modeling building block, we call it *primary process* and extend it appropriately as made precise in the next definition.

Definition 1.2.1 (Primary process)**.** We refer to X as *primary process*. Depending on the context, we then consider one of the following two extensions of X.

(i) $\widehat{X}_t := (t, X_t, [X]_t)$, where $[X]$ denotes the d^2-dimensional process given by the quadratic covariation of X.

(ii) $\widehat{X}_t := (t, X_t)$. This extension is always be paired with the assumption that X is an Itô-semimartingale with absolutely continuous characteristics such that each element of its diffusion matrix can be written as linear combination of elements of its time extended signature.

Remark 1.2.2. Note that (i) of Definition 1.2.1 is more general than (ii) in two respects: first the components of the diffusion matrix do not need to be linear functions of the time extended signature, but could for instance be more general path-dependent functionals; second $t \mapsto [X]_t$ does not need to be absolutely continuous with respect to the Lebesgue measure, hence X does not need to be an Itô-semimartingale.

Throughout the paper we will always denote with $\widehat{\mathbb{X}}_t$ the signature of the previous extensions of X_t. Our goal consists in describing/approximating the dynamics of S with a *signature model*.

Definition 1.2.3 (Signature model)**.** A *signature model* is a stochastic process of the form

$$S_n(\ell)_t := \ell_\emptyset + \sum_{0<|I|\leq n} \ell_I \langle e_I, \widehat{\mathbb{X}}_t \rangle, \tag{1.2.1}$$

where $n \in \mathbb{N}$ and $\ell := \{\ell_\emptyset, \ell_I \colon 0 < |I| \leq n\}$.

Remark 1.2.4. By Proposition 1.2.9 below the class of Sig-SDEs models considered in Perez Arribas et al. (2020) can be embedded in our framework by choosing a properly extended one-dimensional Brownian motion as primary process.

In the following we list several important properties which make signature models a tractable framework for stochastic finance.

- For each $t \in [0, T]$, $S_n(\ell)_t$ is linear in $\widehat{\mathbb{X}}_t$. This in particular implies that having pre-computed the signature of $\widehat{X}$ an update of the parameters ℓ boils down to computing (1.2.1), which is nothing else than a scalar product.
- The quadratic variation of processes of form (1.2.1) is again of the form (1.2.1).
- Form (1.2.1) remains invariant under polynomial transformations.

- Itô-integrals of processes of form (1.2.1) with respect to processes of form (1.2.1) are again processes of form (1.2.1). This includes in particular the signature $\widehat{\mathbb{S}}_n(\ell)$ of $\widehat{S}_n(\ell)_t := (t, S_n(\ell)_t)$ or expressions of the form

$$\int_0^{\cdot} S_n(\ell)_s \mathrm{d}X_s^i.$$

- The latter implies that the expected signature of $\widehat{S}_n(\ell)$ is given by

$$\mathbb{E}[\langle e_J, \widehat{\mathbb{S}}_n(\ell)_t \rangle] = P_J(\ell, \mathbb{E}[\widehat{\mathbb{X}}_t]),$$

  for some P_J such that $P_J(\,\cdot\,, \mathbb{E}[\widehat{\mathbb{X}}_t])$ is a polynomial of degree $|J|$ and $P_J(\ell, \,\cdot\,)$ is a linear map for each ℓ (see Theorem 1.2.16 and Remark 1.2.17 for more details).

- Due to Theorem 1.1.12 this provides approximations

$$\mathbb{E}\Big[f\Big((\widehat{\mathbb{S}}_n^2(\ell)_t)_{t\in[0,T]}\Big)\Big] \approx P_f(\ell, \mathbb{E}[\widehat{\mathbb{X}}_T])$$

  for each map f, which is continuous with respect to $d_{\mathcal{S}^{(2)}}$, where P_f is given by a finite linear combination of maps P_J as above. This includes representations for

$$\mathbb{E}\Big[\tilde{f}(\widehat{S}_n(\ell)_T)\Big] \qquad \text{and} \qquad \mathbb{E}\Big[\tilde{f}\Big(\int_0^T \widehat{S}_n(\ell)_t \mathrm{d}t\Big)\Big]$$

  for maps $\tilde{f}$ being payoff functions and where the expectation is taken with respect to a pricing measure.

We now prove several representation results for the signature model of form (1.2.1). Throughout we shall apply the following notation.

Notation 1.2.5. We use e_0 for the component of $\widehat{X}$ corresponding to time, e_k for its component corresponding to X^k, and (if needed) ε_{ij} for the component of $\widehat{X}$ corresponding to $[X^i, X^j]$.

Consider the following assumption, which includes in particular the case where X is a d-dimensional Brownian motion.

Assumption 1.2.6. *For all $i, j \in \{1, \dots, d\}$ it holds*

$$\mathrm{d}[X^i, X^j]_t = \sum_{|I|\le m} a_{ij}^I \langle e_I, \widehat{\mathbb{X}}_t \rangle \mathrm{d}t$$

for some $m \in \mathbb{N}$ where $\widehat{X}_t = (t, X_t)$.

Remark 1.2.7. Note that for $\widehat{X}$ as of Definition 1.2.1(ii), Assumption 1.2.6 is satisfied by definition.

Proposition 1.2.8. Suppose that $S_n(\ell)$ has a representation of form (1.2.1) with $\widehat{X}$ as in Definition 1.2.1(i) and that Assumption 1.2.6 holds. Then $S_n(\ell)$ has a representation of the same form but with $\widehat{X}$ as in Definition 1.2.1(ii).

Proof. Observe that by Assumption 1.2.6 it holds

$$[X^i, X^j]_t = \int_0^t \sum_{|I|\le m} a_{ij}^I \langle e_I, \widehat{\mathbb{X}}_s\rangle \circ \mathrm{d}s = \sum_{|I|\le m} a_{ij}^I \langle e_I \otimes e_0, \widehat{\mathbb{X}}_t\rangle.$$

The claim follows by Lemma 1.1.9. □

Proposition 1.2.9. Fix $n \in \mathbb{N}$ and suppose that $\widehat{X}$ is given by Definition 1.2.1(i). Then there is a one to one correspondence between representations of the form (1.2.1) and representations of the form

$$S_n(\ell)_t = \ell_\emptyset + \int_0^t \Big(\ell_\emptyset^0 + \sum_{0<|I|\le n-1} \ell_I^0 \langle e_I, \widehat{\mathbb{X}}_s\rangle\Big)\mathrm{d}s + \sum_{k=1}^d \int_0^t \Big(\ell_\emptyset^k + \sum_{0<|I|\le n-1} \ell_I^k \langle e_I, \widehat{\mathbb{X}}_s\rangle\Big)\mathrm{d}X_s^k$$
$$+ \sum_{k_1,k_2=1}^d \int_0^t \Big(\ell_\emptyset^{k_1,k_2} + \sum_{0<|I|\le n-1} \ell_I^{k_1,k_2} \langle e_I, \widehat{\mathbb{X}}_s\rangle\Big)\mathrm{d}[X^{k_1}, X^{k_2}]_s,$$

for $\ell := \{\ell_\emptyset, \ell_I^0, \ell_I^{k_1}, \ell_I^{k_1,k_2} \colon |I| \le n-1 \text{ and } k_1, k_2 \in \{1,\dots,d\}\}$.

If $\widehat{X}$ satisfies Assumption 1.2.6, then the same is true with $\ell_I^{k_1,k_2} = 0$ for each $|I| \le n-1$ and $k_1, k_2 \in \{1,\dots,d\}$.

Before presenting the proof of this proposition let us formulate the following technical lemma.

Lemma 1.2.10. Suppose that $\widehat{X}$ is given by Definition 1.2.1(i). Then

$$\int_0^t \langle e_I, \widehat{\mathbb{X}}_s\rangle \mathrm{d}s = \langle e_I \otimes e_0, \widehat{\mathbb{X}}_t\rangle, \qquad \int_0^t \langle e_I, \widehat{\mathbb{X}}_s\rangle \mathrm{d}X_s^k = \langle \tilde{e}_I^k, \widehat{\mathbb{X}}_s\rangle, \quad \text{and}$$
$$\int_0^t \langle e_I, \widehat{\mathbb{X}}_s\rangle \mathrm{d}[X^{k_1}, X^{k_2}]_s = \langle e_I \otimes \varepsilon_{k_1 k_2}, \widehat{\mathbb{X}}_t\rangle,$$

where $\tilde{e}_\emptyset^k = e_k$ and $\tilde{e}_I^k = e_I \otimes e_k - \frac{1}{2} e_{I'} \otimes \varepsilon_{i_{|I|}k}$ for each $|I| > 0$. If Assumption 1.2.6 is in force, then the same holds true with

$$\tilde{e}_I^k = e_I \otimes e_k - \sum_{|J|\le m} \frac{a_{i_{|I|}k}^J}{2} (e_{I'} \sqcup\!\sqcup e_J) \otimes e_0$$

for each $|I| > 0$.

Proof. The representations of $\int_0^t \langle e_I, \widehat{\mathbb{X}}_s\rangle \mathrm{d}s$ and $\int_0^t \langle e_I, \widehat{\mathbb{X}}_s\rangle \mathrm{d}[X^{k_1}, X^{k_2}]_s$ follow by definition of the signature. We proceed with the proof of the representation of $\int_0^t \langle e_I, \widehat{\mathbb{X}}_s\rangle \mathrm{d}X_s^k$. For $I = \emptyset$ the claim is clear. By definition of the Stratonovich integral we have that

$$\begin{aligned}\int_0^t \langle e_I, \widehat{\mathbb{X}}_s\rangle \mathrm{d}X_s^k &= \int_0^t \langle e_I, \widehat{\mathbb{X}}_s\rangle \circ \mathrm{d}X_s^k - \frac{1}{2}[\langle e_I, \widehat{\mathbb{X}}\rangle, X^k]_t \\ &= \langle e_I \otimes e_k, \widehat{\mathbb{X}}_t\rangle - \frac{1}{2}\int_0^t \langle e_{I'}, \widehat{\mathbb{X}}_s\rangle \mathrm{d}[X^{i_{|I|}}, X^k]_s \\ &= \langle e_I \otimes e_k, \widehat{\mathbb{X}}_t\rangle - \frac{1}{2}\langle e_{I'} \otimes \varepsilon_{i_{|I|}k}, \widehat{\mathbb{X}}_t\rangle \\ &= \langle \tilde{e}_I^k, \widehat{\mathbb{X}}_t\rangle.\end{aligned}$$

Suppose now that Assumption 1.2.6 is in force. Then Proposition 1.1.3 and the definition of the signature yield

$$\begin{aligned}\int_0^t \langle e_I, \widehat{\mathbb{X}}_s\rangle \mathrm{d}[X^{k_1}, X^{k_2}]_s &= \int_0^t \sum_{|J|\le m} a_{k_1k_2}^J \langle e_I, \widehat{\mathbb{X}}_s\rangle \langle e_J, \widehat{\mathbb{X}}_s\rangle \mathrm{d}s \\ &= \sum_{|J|\le m} a_{k_1k_2}^J \langle (e_I \sqcup\!\sqcup e_J) \otimes e_0, \widehat{\mathbb{X}}_t\rangle,\end{aligned}$$

and the claim follows. □

We are now ready to provide the proof of Proposition 1.2.9.

Proof of Proposition 1.2.9. Let $S_n(\ell)$ be as in the statement of the proposition. By Lemma 1.2.10 it holds

$$\begin{aligned}S_n(\ell)_t = \ell_\emptyset + \Big(\ell_\emptyset^0 \langle e_0, \widehat{\mathbb{X}}_t\rangle + \sum_{0<|I|\le n-1} \ell_I^0 \langle e_I \otimes e_0, \widehat{\mathbb{X}}_t\rangle\Big) + \sum_{k=1}^d \Big(\ell_\emptyset^k \langle \tilde{e}_\emptyset^k, \widehat{\mathbb{X}}_t\rangle + \sum_{0<|I|\le n-1} \ell_I^k \langle \tilde{e}_I^k, \widehat{\mathbb{X}}_t\rangle\Big) \\ + \sum_{k_1,k_2=1}^d \Big(\ell_\emptyset^{k_1,k_2} \langle \varepsilon_{k_1k_2}, \widehat{\mathbb{X}}_t\rangle + \sum_{0<|I|\le n-1} \ell_I^{k_1,k_2} \langle e_I \otimes \varepsilon_{k_1k_2}, \widehat{\mathbb{X}}_s\rangle\Big),\end{aligned}$$

which is of the form given by (1.2.1). Conversely, by Lemma 1.2.10 we also have

$$\langle e_I \otimes e_0, \widehat{\mathbb{X}}_t\rangle = \int_0^t \langle e_I, \widehat{\mathbb{X}}_s\rangle \mathrm{d}s, \qquad \langle e_I \otimes \varepsilon_{k_1k_2}, \widehat{\mathbb{X}}_t\rangle = \int_0^t \langle e_I, \widehat{\mathbb{X}}_s\rangle \mathrm{d}[X^{k_1}, X^{k_2}]_s, \quad \text{and}$$
$$\langle e_I^k \otimes e_k, \widehat{\mathbb{X}}_s\rangle = \int_0^t \langle e_I, \widehat{\mathbb{X}}_s\rangle \mathrm{d}X_s^k + \frac{1}{2}\int_0^t \langle e_{I'}, \widehat{\mathbb{X}}_s\rangle \mathrm{d}[X^{i_{|I|}}, X^k]_s,$$

and the claim follows. □

In the case of Brownian motion the formulas of Lemma 1.2.10 simplify as follows.

Example 1.2.11. Suppose that X is a vector of correlated Brownian motions with correlation matrix ρ and $\widehat{X}$ is given by Definition 1.2.1(ii). Then it holds

$$\tilde{e}_\emptyset^k = e_k \qquad \text{and} \qquad \tilde{e}_I^k = e_I \otimes e_k - \frac{\rho_{i_{|I|},k}}{2} 1_{\{i_{|I|}\neq 0\}} e_{I'} \otimes e_0.$$

1.2.1. Absence of arbitrage and universality properties

The above representation results will allow us to establish conditions that guarantee absence of arbitrage. We shall therefore assume that *no free lunch with vanishing risk* (Delbaen and Schachermayer (1994)) holds, which is – due to the continuity of sample paths of $S_n(\ell)$ – equivalent to the existence of an equivalent local martingale measure $\mathbb{Q}$, i.e. a measure $\mathbb{Q} \sim \mathbb{P}$ under which the (discounted) asset price model $S_n(\ell)$ is a local martingale. We shall always assume here that interest rates are zero and that the asset is already discounted. To formulate a precise no-arbitrage condition, the following corollary which is a direct consequence of Proposition 1.2.9 and Lemma 1.2.10 is essential.

Corollary 1.2.12. *Suppose that X is a local martingale. Then $S_n(\ell)$ is a local martingale if and only if it admits a representation of the form*

$$S_n(\ell)_t = \ell_\emptyset + \sum_{k=1}^{D} \left(\ell_\emptyset^k \langle e_k, \widehat{\mathbb{X}}_t \rangle + \sum_{0<|I|\leq n-1} \ell_I^k \langle \tilde{e}_I^k, \widehat{\mathbb{X}}_t \rangle \right), \tag{1.2.2}$$

for some $D \in \{1, \dots, d\}$ and $\ell := \{\ell_\emptyset, \ell_I^k \colon 0 \leq |I| \leq n-1 \text{ and } k \in \{1, \dots, d\}\}$.

Proof. Due to the local martingale property of X, $S_n(\ell)$ in the representation of Proposition 1.2.9 is a local martingale if and only if all integrals with respect to time and the quadratic variation process vanish. This means that $S_n(\ell)$ is of form

$$S_n(\ell) = \ell_\emptyset + \sum_{k=1}^{d} \int_0^t \left(\ell_\emptyset^k + \sum_{0<|I|\leq n-1} \ell_I^k \langle e_I, \widehat{\mathbb{X}}_s \rangle \right) \mathrm{d}X_s^k \tag{1.2.3}$$

and Lemma 1.2.10 yields the assertion. □

We are now ready to formulate sufficient no-arbitrage conditions.

Corollary 1.2.13. *Suppose that there is an equivalent measure $\mathbb{Q} \sim \mathbb{P}$ such that X is a local $\mathbb{Q}$-martingale. Then the following holds.*

(i) The model $S_n(\ell)$ is free of arbitrage if it admits a representation as of (1.2.2).

(ii) If $\mathbb{Q}$ is an equivalent local martingale measure for $S_n(\ell)$, then $S_n(\ell)$ is necessarily of form (1.2.2).

Proof. The first assertion is a direct consequence of Corollary 1.2.12, as form (1.2.2) implies that $S_n(\ell)$ is a local $\mathbb{Q}$-martingale and thus $\mathbb{Q}$ is an equivalent local martingale measure. The assumption of the second assertion implies that $S_n(\ell)$ is a local martingale under $\mathbb{Q}$, whence by Corollary 1.2.12 it has to be of form (1.2.2). □

Remark 1.2.14. (i) Note that $S_n(\ell)$ could be free of arbitrage without being of form (1.2.2). This is the case if the primary process X is not a local martingale under any of the equivalent local martingale measures.

(ii) Observe also that if $S_n(\ell)$ is form (1.2.2) and a local martingale under $\mathbb{Q}$, then X does not necessarily need to be a local $\mathbb{Q}$-martingale. This happens if drift terms in (1.2.3) cancel out. Note however that in dimension $d = 1$ this cannot occur and X is thus necessarily a local $\mathbb{Q}$-martingale. In particular, in this case $\mathbb{Q}$ is unique and the model thus complete.

In the following example we show how classical stochastic volatilty models can be approximated by arbitrage-free signature models, thus making the announced universality properties precise.

Example 1.2.15. We consider a generic stochastic volatility model driven by two correlated Brownian motions $(B_t)_{t\in[0,1]}$ and $(W_t)_{t\in[0,1]}$ with correlation coefficient $\rho \in [-1, 1]$. More precisely, we assume that under some equivalent local martingale measure $\mathbb{Q}$ the dynamics of S are given by

$$\begin{aligned} \mathrm{d}S_t &= g(S_t, V_t)\mathrm{d}B_t, \\ \mathrm{d}V_t &= h(S_t, V_t)\mathrm{d}t + \sigma(S_t, V_t)\mathrm{d}W_t, \end{aligned}$$

for some functions $g, h, \sigma : \mathbb{R}^2 \to \mathbb{R}$. As a possible choice for the parameters we can for instance consider

- $g(s, v) = s\sqrt{v}$, $h(s, v) = \kappa(\theta - v)$ and $\sigma(s, v) = \sigma_0\sqrt{v}$, for some parameters $\kappa, \theta \geq 0$ and $\sigma_0 \in \mathbb{R}$ to retrieve the Heston model (see Heston (1993));
- $g(s, v) = s^\beta v$, $h(s, v) = 0$ and $\sigma(s, v) = \alpha v$ for $0 < \beta \leq 1$ and $\alpha \in \mathbb{R}$ to retrieve the SABR model (see Hagan et al. (2002)).

Set now $X_t := (B_t, W_t)$ and let $\widehat{X}_t$ be given by the representation of Definition 1.2.1(ii), i.e., $\widehat{X}_t := (t, B_t, W_t)$. For some fixed $n \in \mathbb{N}$, consider then the following model

$$S_n(\ell)_t := \ell_\emptyset + \int_0^t \Big(\ell^B_\emptyset + \sum_{0<|I|\leq n-1} \ell^B_I \langle e^B_I, \widehat{\mathbb{X}}_s\rangle\Big)\mathrm{d}B_s,$$

which is a local martingale under $\mathbb{Q}$ and thus absence of arbitrage is guaranteed. Moreover, by Proposition 1.2.9, we know that this model has a representation of the form

$$S_n(\ell)_t = \ell_\emptyset + \ell^B_\emptyset B_t + \sum_{0<|I|\leq n-1} \ell^B_I \langle \tilde{e}^B_I, \widehat{\mathbb{X}}_t\rangle,$$

for each $t \in [0, 1]$, where by Lemma 1.2.10

$$\tilde{e}^B_I := e_I \otimes e_1 - \frac{1}{2}(1_{\{i_{|I|}=1\}} + \rho 1_{\{i_{|I|}=2\}})e_{I'} \otimes e_0, \tag{1.2.4}$$

for each I with $0 < |I| \leq n - 1$, implying that it is a signature model as of Definition 1.2.3.

Concerning universality note that if the coefficients ℓ^B_I are chosen such that (1.2.4) matches the signature approximation of S up to order $n-1$ as specified in Proposition 1.1.15

(assuming appropriate regularity conditions on the coefficients g, h, σ), then at least locally in t, $S_n(\ell)$ is close to S in probability.

Let us exemplify this by means of the SABR model for $\beta = 1$ and, for simplicity, $\rho = 0$. Set $\ell_\emptyset := S_0$, $\ell^B_\emptyset := y_1 y_2$, $\ell^B_{(0)} := -\frac{1}{2} y_1 y_2^3 - \frac{1}{2}\alpha^2 y_1 y_2$, $\ell^B_{(1)} := y_1 y_2^2$, and $\ell^B_{(2)} := \alpha y_1 y_2$. By an application of Proposition 1.1.15 for $X_t = (t, W_t, B_t)$, $Y^1 = S$, $Y^2 = V$, $m = 2$,

$$\Psi_0(y) = (1, -\frac{1}{2} y_1 y_2^2, -\frac{1}{2}\alpha^2 y_2)^\top,$$

$\Psi_1(y) = (0, y_1 y_2, 0)^\top$, and $\Psi_2(y) = (0, 0, \alpha y_2)^\top$ we can conclude that for each $\varepsilon > 0$ there is a constant $C(\varepsilon)$ such that

$$\mathbb{Q}\Big(|S_t - S_2(\ell)_t| > C(\varepsilon) t^{3/2}\Big) < \varepsilon,$$

where $S_2(\ell)$ is given by Definition 1.2.3 for $\widehat{X}_t = (t, B_t, W_t)$. The same procedure for general n leads to a similar result for an arbitrary speed of convergence.

Another possibility to establish universality is to use the fact that the solution map of a stochastic differential equation (with sufficiently regular coefficients) is a continuous map of the signature of the driving signal (see e.g. Corollary 10.40 Friz and Victoir (2010)). The universal approximation theorem (Theorem 1.1.12, see also the formulation of Lemma 5.2 in Bayer et al. (2021) or Cuchiero and Möller (2023)) then yields the result, however without quantitative estimates.

1.2.2. The expected signature of $S_n(\ell)$

For pricing purposes of so-called *sig-payoffs* treated in Section 1.3 it will be important to be able to compute the expected signature of $\widehat{S}_n(\ell)_t = (t, S_n(\ell)_t)$. In this section we thus provide formulas which trace this computation back to the calculation of the expected signature of $\widehat{X}_t$, which in many cases is well-known (see e.g. Fawcett (2003) for Brownian motion) and can often be computed by techniques of polynomial processes (see Section A.1 for the explicit computation if X is a Brownian motion and Section 2.2 if X being a polynomial processes) or by solving an infinite dimensional system of linear PDEs corresponding to the Kolmogorov forward equation of the signature process (see Ni (2012)). For a unified treatment of signature cumulants, i.e. the logarithm of expected signature, we refer to Friz et al. (2022).

Here, we first suppose that $\widehat{X}$ is given by Definition 1.2.1(ii).

Theorem 1.2.16. *Fix $n \in \mathbb{N}$, a multi-index J, and $D \in \{1, \dots, d\}$ and denote by $\widehat{\mathbb{S}}_n(\ell)_t$ the signature of $\widehat{S}_n(\ell)_t$. Let e_0 be the component of $\widehat{S}_n(\ell)$ corresponding to time and e_1 its component corresponding to $S_n(\ell)$. Define $e(\emptyset, \ell) := \tilde{e}(\emptyset, \ell) := e_\emptyset$ and*

$$e(J, \ell) = \widetilde{\sqcup\!\sqcup}_{i=1}^{|J|}\Big(e_0 1_{\{j_i=0\}} + \Big(\sum_{0<|I|\le n} \ell_I e_I\Big) 1_{\{j_i=1\}}\Big),$$

$$\tilde{e}(J, \ell) = \widetilde{\sqcup\!\sqcup}_{i=1}^{|J|}\Big(e_0 1_{\{j_i=0\}} + \Big(\sum_{k=1}^{D}\Big(\ell^k_\emptyset e_k + \sum_{0<|I|\le n-1} \ell^k_I \tilde{e}^k_I\Big)\Big) 1_{\{j_i=1\}}\Big),$$

for $|J| > 0$ with $\widetilde{\sqcup\!\sqcup}$ the half-shuffle being introduced in Definition 1.1.8. Then the following representations hold.

- $\langle e_J, \widehat{\mathbb{S}}_n(\ell)_t\rangle = \langle e(J,\ell), \widehat{\mathbb{X}}_t\rangle$, *if $S_n(\ell)$ is given by (1.2.1), and*
- $\langle e_J, \widehat{\mathbb{S}}_n(\ell)_t\rangle = \langle \tilde{e}(J,\ell), \widehat{\mathbb{X}}_t\rangle$, *if $S_n(\ell)$ is given as in Corollary 1.2.12.*

Proof. In order to prove the claim we proceed by induction. Fix $S_n(\ell)$ as in (1.2.1). For $J = \emptyset$ the claim follows by the definition of signature. Suppose the claim holds true for each J such that $|J| = m-1$, and fix J with $|J| \le m$. Then

$$\begin{aligned}
\langle e_J, \widehat{\mathbb{S}}_n(\ell)_t\rangle &= \int_0^t \langle e_{J'}, \widehat{\mathbb{S}}_n(\ell)_s\rangle \circ \mathrm{d}\langle e_{j_m}, \widehat{S}_n(\ell)_s\rangle \\
&= \int_0^t \langle e_{J'}, \widehat{\mathbb{S}}_n(\ell)_s\rangle \circ \mathrm{d}\langle e_0 1_{\{j_m=0\}} + \Big(\sum_{0<|I|\le n} \ell_I e_I\Big) 1_{\{j_m=1\}}, \widehat{\mathbb{X}}_s\rangle \\
&= 1_{\{j_m=0\}} \int_0^t \langle e_{J'}, \widehat{\mathbb{S}}_n(\ell)_s\rangle \circ \mathrm{d}\langle e_0, \widehat{\mathbb{X}}_s\rangle \\
&\qquad + 1_{\{j_m=1\}} \sum_{0<|I|\le n} \ell_I \int_0^t \langle e_{J'}, \widehat{\mathbb{S}}_n(\ell)_s\rangle \circ \mathrm{d}\langle e_I, \widehat{\mathbb{X}}_s\rangle.
\end{aligned}$$

The induction hypothesis and Lemma 1.1.9 yield the first claim and the second one is analogous. □

Remark 1.2.17. (i) Let now $S_n(\ell)$ be as in (1.2.1) and observe that

$$e(J,\ell) = \widetilde{\sqcup\!\sqcup}_{i=1}^{|J|} \sum_{0<|I|\le n} \Big(e_I 1_{\{j_i=0\}} 1_{\{I=(0)\}} + \ell_I e_I 1_{\{j_i=1\}} \Big).$$

Setting $c(j, I, \ell) := 1_{\{j=0\}} 1_{\{I=(0)\}} + \ell_I 1_{\{j=1\}}$ we thus obtain that

$$\mathbb{E}[\langle e_J, \widehat{\mathbb{S}}_n(\ell)_t\rangle] = \mathbb{E}[\langle e(J,\ell), \widehat{\mathbb{X}}_t\rangle] = \sum_{|I_1|,\dots,|I_{|J|}|=1}^{n} \mathbb{E}[\langle \widetilde{\sqcup\!\sqcup}_{i=1}^{|J|} e_{I_i}, \widehat{\mathbb{X}}_t\rangle] \prod_{i=1}^{|J|} c(j_i, I_i, \ell).$$

Even if at a first sight this representation appears involved, it is in fact very handy. Indeed the expectations $\mathbb{E}[\langle \widetilde{\sqcup\!\sqcup}_{i=1}^{|J|} e_{I_i}, \widehat{\mathbb{X}}_t\rangle]$ can be computed just once in advance. Since $c(j, I, \cdot\,)$ is affine, we also immediately obtain that the map

$$(\ell, \mathbb{E}[\widehat{\mathbb{X}}]) \mapsto P_J(\ell, \mathbb{E}[\widehat{\mathbb{X}}_t]) := \mathbb{E}[\langle e_J, \widehat{\mathbb{S}}_n(\ell)_t\rangle]$$

is polynomial of degree $|J|$ in its first argument and linear in the second one.

(ii) Similarly, let $S_n(\ell)$ be as in Corollary 1.2.12, set $\tilde{e}_\emptyset^k := e_k$, and observe that

$$\tilde{e}(J,\ell) = \widetilde{\sqcup\!\sqcup}_{i=1}^{|J|} \sum_{k=1}^{D} \sum_{0\le|I|\le n-1} \Big(\tilde{e}_I^k 1_{\{j_i=0\}} 1_{\{I=(0)\}} 1_{\{k=1\}} + \ell_I^k \tilde{e}_I^k 1_{\{j_i=1\}} \Big).$$

Setting $\tilde{c}(j,I,k,\ell) := 1_{\{j=0\}}1_{\{I=(0)\}}1_{\{k=1\}} + \ell_I^k 1_{\{j=1\}}$ yields

$$\mathbb{E}[\langle e_J, \widehat{\mathbb{S}}_n(\ell)_t\rangle] = \sum_{k_1,\dots,k_{|J|}=1}^{D} \sum_{|I_1|,\dots,|I_{|J|}|=0}^{n-1} \mathbb{E}[\langle \widetilde{\sqcup\!\sqcup}_{i=1}^{|J|} \tilde{e}_{I_i}^{k_i}, \widehat{\mathbb{X}}_t\rangle] \prod_{i=1}^{|J|} \tilde{c}(j_i, I_i, k_i, \ell).$$

We again obtain that the map

$$(\ell, \mathbb{E}[\widehat{\mathbb{X}}]) \mapsto \widetilde{P}_J(\ell, \mathbb{E}[\widehat{\mathbb{X}}_t]) := \mathbb{E}[\langle e_J, \widehat{\mathbb{S}}_n(\ell)_t\rangle]$$

is polynomial of degree $|J|$ in its first argument and linear in the second one.

The expressions above clearly also provide a formula of the variance of $S_n(\ell)$.

Corollary 1.2.18. *Let $S_n(\ell)$ be as in Corollary 1.2.12 and assume that it is a true martingale. Then*

$$Var(S_n(\ell)_t) = 2 \sum_{k_1,k_2=1}^{D} \sum_{|I_1|,|I_2|=1}^{n-1} \mathbb{E}[\langle \tilde{e}_{I_1}^{k_1} \sqcup\!\sqcup \tilde{e}_{I_2}^{k_2}, \widehat{\mathbb{X}}_t\rangle]\ell_{I_1}\ell_{I_2}. \tag{1.2.5}$$

Proof. The martingale property guarantees that $\mathbb{E}[S_n(\ell)_t] = S_n(\ell)_0$ and hence, by Proposition 1.1.3 (see also Example 1.1.5), $\mathrm{Var}(S_n(\ell)_t) = 2\mathbb{E}[\langle e_1 \otimes e_1, \widehat{\mathbb{S}}_n(\ell)_t\rangle]$. By Remark 1.2.17 we can conclude that

$$\mathrm{Var}(S_n(\ell)_t) = 2 \sum_{k_1,k_2=1}^{D} \sum_{|I_1|,|I_2|=1}^{n-1} \mathbb{E}[\langle \tilde{e}_{I_1}^{k_1} \sqcup\!\sqcup \tilde{e}_{I_2}^{k_2}, \widehat{\mathbb{X}}_t\rangle]\tilde{c}(1,I_1,k_1,\ell)\tilde{c}(1,I_2,k_2,\ell)$$

and the claim follows. □

Let us now consider the case where X is additionally extended by its quadratic variation. In that case we need the following representation.

Lemma 1.2.19. Suppose that $\widehat{X}$ is given by Definition 1.2.1(i) and fix two multi-indices I and J such that $|I| > 0$ and $|J| > 0$. Using the notation of Remark 1.2.5 set

$$e_I[\sqcup\!\sqcup]e_J := \sum_{k_1,k_2=1}^{d} 1_{\{i_{|I|}=k_1, j_{|J|}=k_2\}}(e_{I'} \sqcup\!\sqcup e_{J'}) \otimes \varepsilon_{k_1k_2}.$$

Then $[\langle e_\emptyset, \widehat{\mathbb{X}}\rangle, \langle e_J, \widehat{\mathbb{X}}\rangle]_t = [\langle e_\emptyset, \widehat{\mathbb{X}}\rangle]_t = 0$, and $[\langle e_I, \widehat{\mathbb{X}}\rangle, \langle e_J, \widehat{\mathbb{X}}\rangle]_t = \langle e_I[\sqcup\!\sqcup]e_J, \widehat{\mathbb{X}}_t\rangle$. If Assumption 1.2.6 is in force the same is true with

$$e_I[\sqcup\!\sqcup]e_J := \sum_{k_1,k_2=1}^{d} \sum_{|H|\le m} a_{k_1k_2}^H 1_{\{i_{|I|}=k_1, j_{|J|}=k_2\}}(e_{I'} \sqcup\!\sqcup e_{J'} \sqcup\!\sqcup e_H) \otimes e_0.$$

Proof. By definition of the signature and Proposition 1.1.3 we can compute

$$\begin{aligned}[\langle e_I, \widehat{\mathbb{X}}\rangle, \langle e_J, \widehat{\mathbb{X}}\rangle]_t &= \int_0^t \langle e_{I'}, \widehat{\mathbb{X}}_s\rangle\langle e_{J'}, \widehat{\mathbb{X}}_s\rangle \mathrm{d}[\langle e_{i_{|I|}}, \widehat{\mathbb{X}}\rangle, \langle e_{j_{|J|}}, \widehat{\mathbb{X}}\rangle]_s \\ &= \sum_{k_1,k_2=1}^{d} 1_{\{i_{|I|}=k_1, j_{|J|}=k_2\}} \langle e_{I'}, \widehat{\mathbb{X}}_s\rangle\langle e_{J'}, \widehat{\mathbb{X}}_s\rangle \mathrm{d}\langle \varepsilon_{k_1k_2}, \widehat{\mathbb{X}}_s\rangle \\ &= \langle e_I[\sqcup\!\sqcup]e_J, \widehat{\mathbb{X}}_t\rangle.\end{aligned}$$

Since under Assumption 1.2.6 it holds $\mathrm{d}[\langle e_{i_{|I|}}, \widehat{\mathbb{X}}\rangle, \langle e_{j_{|J|}}, \widehat{\mathbb{X}}\rangle]_t = \sum_{|H|\le m} a^H_{k_1k_2}\langle e_H, \widehat{\mathbb{X}}_t\rangle \mathrm{d}\langle e_0, \widehat{\mathbb{X}}_t\rangle$, the second claim follows again by Proposition 1.1.3. □

Using the above lemma we can now express the signature components of $\widehat{S}_n(\ell)_t$ here defined as $\widehat{S}_n(\ell)_t := (t, S_n(\ell)_t, [S_n(\ell)]_t)$ also via $\widehat{\mathbb{X}}_t$.

Theorem 1.2.20. *Fix $n \in \mathbb{N}$, a multi-index J, and suppose that $\widehat{X}$ is given by Definition 1.2.1(i). Set $\widehat{S}_n(\ell)_t := (t, S_n(\ell)_t, [S_n(\ell)]_t)$ and let $\widehat{\mathbb{S}}_n(\ell)_t$ denote the corresponding signature. Let e_0 be the component of $\widehat{S}_n(\ell)_t$ corresponding to time, e_1 its component corresponding to $S_n(\ell)_t$, and e_2 its component corresponding to $[S_n(\ell)]_t$. Define $e(J,\ell) := e_\emptyset$ for $J = \emptyset$ and*

$$e(J,\ell) = \widetilde{\sqcup\!\sqcup}_{k=1}^{|J|}\Big(e_0 1_{\{j_k=0\}} + \Big(\sum_{0<|I|\le n} \ell_I e_I\Big)1_{\{j_k=1\}} + \Big(\sum_{0<|I|,|H|\le n} \ell_I\ell_H e_I[\sqcup\!\sqcup]e_H\Big)1_{\{j_k=2\}}\Big).$$

Then $\langle e_J, \widehat{\mathbb{S}}_n(\ell)_t\rangle = \langle e(J,\ell), \widehat{\mathbb{X}}_t\rangle$.

Proof. First observe that by Lemma 1.2.19 it holds

$$[S_n(\ell)] = \Big[\Big(\ell_\emptyset + \sum_{0<|I|\le n} \ell_I\langle e_I, \widehat{\mathbb{X}}\rangle\Big)\Big] = \sum_{0<|I|,|H|\le n} \ell_I\ell_H\langle e_I[\sqcup\!\sqcup]e_H, \widehat{\mathbb{X}}\rangle.$$

Next, note that

$$\begin{aligned}\langle e_j, \widehat{S}_n(\ell)_t\rangle &= t1_{\{j=0\}} + (S_n(\ell)_t - S_n(\ell)_0)1_{\{j=1\}} + [S_n(\ell)]_t 1_{\{j=2\}} \\ &= \langle e_0 1_{\{j=0\}} + \Big(\sum_{0<|I|\le n} \ell_I e_I\Big)1_{\{j=1\}} + \Big(\sum_{0<|I|,|H|\le n} \ell_I\ell_H e_I[\sqcup\!\sqcup]e_H\Big)1_{\{j=2\}}, \widehat{\mathbb{X}}_t\rangle.\end{aligned}$$

The proof now follows the proof of Theorem 1.2.16. □

1.3. Pricing of sig-payoffs

We recall here the notion of *sig-payoffs* as introduced in Lyons et al. (2020).

Definition 1.3.1 (Sig-payoff)**.** Suppose that the price process S is given by a continuous semimartingale. A payoff $F : \Omega \to \mathbb{R}$ is said to be a sig-payoff if there exists $m \in \mathbb{N}$, and $f := \{f_\emptyset, f_J \colon 0 < |J| \leq m\}$, such that

$$F := f_\emptyset + \sum_{0<|J|\leq m} f_J \langle e_J, \widehat{\mathbb{S}}_T \rangle,$$

where $\widehat{\mathbb{S}}$ denotes the signature of $\widehat{S}_t = (t, S_t)$.

Example 1.3.2. Let $K > 0$ be a strike price and $T > 0$ a maturity time. Then, Asian forwards written on a stock S are payoffs of the form

$$\frac{1}{T}\int_0^T S_t \mathrm{d}t - K = \frac{1}{T}\int_0^T (S_t - S_0)\mathrm{d}t - K + S_0 = \frac{1}{T}\langle e_1 \otimes e_0, \widehat{\mathbb{S}}_T \rangle + (K - S_0)\langle e_\emptyset, \widehat{\mathbb{S}}_T \rangle,$$

and are thus sig-payoffs.

Even though the standard vanilla derivatives such as call and put options are *not* sig-payoffs, approximate sig-payoffs can be used as efficient control variates in Monte Carlo pricing, which we outline in Section 1.4.2.

Note that in this section we shall use $\widehat{S}_t$ always for $\widehat{S}_t = (t, S_t)$ and consider pricing of sig-payoffs when S is given by a signature model, as made precise in the following corollary.

Corollary 1.3.3. *Let the dynamics of $S_n(\ell)$ under a local martingale measure $\mathbb{Q}$ be specified as in Corollary 1.2.12 for $\widehat{X}$ given by Definition 1.2.1(ii). Consider a sig-payoff*

$$F = f_\emptyset + \sum_{0<|J|\leq m} f_J \langle e_J, \widehat{\mathbb{S}}_n(\ell)_T \rangle.$$

Then, using the notation of Remark 1.2.17 we can write the corresponding price as

$$\begin{aligned}\mathbb{E}_\mathbb{Q}[F] &= f_\emptyset + \sum_{0<|J|\leq m} f_J \widetilde{P}_J(\ell, \mathbb{E}_\mathbb{Q}[\widehat{\mathbb{X}}_T]) \\ &= f_\emptyset + \sum_{|J|=1}^{m} \sum_{k_1,\ldots,k_{|J|}=1}^{D} \sum_{|I_1|,\ldots,|I_{|J|}|=0}^{n} \mathbb{E}[\langle \widetilde{\sqcup\!\sqcup}_{i=1}^{|J|} \tilde{e}_{I_i}^{k_i}, \widehat{\mathbb{X}}_T \rangle] f_J \prod_{i=1}^{|J|} \tilde{c}(j_i, I_i, k_i, \ell), \qquad (1.3.1)\end{aligned}$$

for $\tilde{c}(j, I, k, \ell) := 1_{\{j=0\}} 1_{\{I=(0)\}} 1_{\{k=1\}} + \ell_I^k 1_{\{j=1\}}$.

Proof. This is a direct consequence of Theorem 1.2.16 and Remark 1.2.17. □

Remark 1.3.4. (i) Expression (1.3.1) admits also a second representation that turns out to be useful for coding:

$$\begin{aligned}\mathbb{E}_\mathbb{Q}[F] = f_\emptyset + \sum_{|J|=1}^{m} \sum_{k_1,\ldots,k_{|J|}=1}^{D} \sum_{I_1,\ldots,I_{|J|}\in\mathcal{I}} &\mathbb{E}[\langle \widetilde{\sqcup\!\sqcup}_{i=1}^{|J|} \tilde{e}_{I_i}^{k_i}, \widehat{\mathbb{X}}_T \rangle] \\ &\times f_{(1_{\{I_1 \neq I^t, k_1 \neq 0\}}, \ldots, 1_{\{I_{|J|} \neq I^t, k_{|J|} \neq 0\}})} \prod_{i=1}^{|J|} \Big(1_{\{I_i = I^t, k_i = 0\}} + \ell_{I_i}^k 1_{\{I_i \neq I^t, k_i \neq 0\}} \Big).\end{aligned}$$

(ii) With a similar procedure it is also possible to provide a representation of $\mathrm{Var}_{\mathbb{Q}}(F)$. By Proposition 1.1.3 we know that

$$(F - f_\emptyset)^2 = \sum_{|J_1|,|J_2|=1}^{m} f_{J_1} f_{J_2} \langle e_{J_1} \sqcup\!\sqcup e_{J_2}, \widehat{\mathbb{S}}_n(\ell)_T \rangle.$$

Setting $\widetilde{P}_{I_1 \sqcup\!\sqcup I_2} := \sum_{i=1}^{K} \widetilde{P}_{J_i}$ for J_i and K satisfying $e_{I_1} \sqcup\!\sqcup e_{I_2} = \sum_{i=1}^{K} e_{J_i}$ we thus obtain

$$\begin{aligned}\mathrm{Var}_{\mathbb{Q}}(F) &= \mathbb{E}_{\mathbb{Q}}[(F - f_\emptyset)^2] - \mathbb{E}_{\mathbb{Q}}[F - f_\emptyset]^2 \\ &= \sum_{|J_1|,|J_2|=1}^{m} f_{J_1} f_{J_2} \Big(\widetilde{P}_{J_1 \sqcup\!\sqcup J_2}(\ell, \mathbb{E}_{\mathbb{Q}}[\widehat{\mathbb{X}}_T]) - \widetilde{P}_{J_1}(\ell, \mathbb{E}_{\mathbb{Q}}[\widehat{\mathbb{X}}_T]) \widetilde{P}_{J_2}(\ell, \mathbb{E}_{\mathbb{Q}}[\widehat{\mathbb{X}}_T]) \Big).\end{aligned}$$

(iii) From this representations we can see that the maps $P_F(f, \ell, \mathbb{E}[\widehat{\mathbb{X}}_T]) := \mathbb{E}[F]$ and $P_{\mathrm{Var}(F)}(f, \ell, \mathbb{E}[\widehat{\mathbb{X}}_T]) := \mathrm{Var}(F)$ inherit good properties from $\widetilde{P}_J$. In particular, P_F is linear in f and polynomial of degree m in ℓ, and $P_{\mathrm{Var}(F)}$ is quadratic in f and polynomial of degree $2m$ in ℓ. Both maps are linear in $\mathbb{E}[\widehat{\mathbb{X}}_T]$.

1.4. Calibration

This section is dedicated to illustrate how the signature model can be calibrated to market data. We here consider three different tasks: first, calibration to time series data, second calibration to option prices, and third a combination of these two tasks.

Throughout the section we fix $n \geq 1$ and $d \geq 1$. Consider a model given as in Corollary 1.2.12 with $\widehat{X}$ as in Definition 1.2.1(ii). In order to simplify the notation we set $\ell_\emptyset := S_0$ and we drop the index 1 from ℓ_I^1 and $\tilde{e}_I^1$.

1.4.1. Calibration to time-series data

For the time-series calibration we investigate two different methods. In both cases we suppose that time series data for $\widehat{X}$ and S are available on a time grid $t_1, \dots, t_N$, with $N > 1$.

Regression using price data

Our first method consists in performing a *linear* regression where we directly regress – according to the signature model as in Corollary 1.2.12 for $D = 1$ – the trajectories of S on the signature of $\widehat{X}$. For the computation of $(\widehat{\mathbb{X}}_{t_i})_{i=1}^N$ based on the observations $(\widehat{X}_{t_i})_{i=1}^N$ we can resort to several available packages, e.g. the package `iisignature` developed by Reizenstein and Graham (2020) or `signatory` by Kidger and Lyons (2020) in Python.

Let $d^* := (d+1)_n = \frac{(d+1)^n - 1}{d}$ denote the dimension of the signature vector of $\widehat{X}$ truncated at level $n-1$. Then the current calibration problem consists in finding $\ell^* \in \mathbb{R}^{d^*}$ such that

$$\ell^* \in \operatorname{argmin}_\ell \, L_{\text{price},\alpha}(\ell),$$

where the loss function $L_{\text{price},\alpha}(\ell)$ is given by

$$\begin{aligned} L_{\text{price},\alpha}(\ell) &:= \sum_{i=1}^N \Big(S_n(\ell)_{t_i} - S_{t_i} \Big)^2 + \alpha(\ell) \\ &= \sum_{i=1}^N \Big(S_0 + \ell_\emptyset \langle e_1, \widehat{\mathbb{X}}_{t_i} \rangle + \sum_{0 < |I| \leq n-1} \ell_I \langle \tilde{e}_I, \widehat{\mathbb{X}}_{t_i} \rangle - S_{t_i} \Big)^2 + \alpha(\ell), \end{aligned} \tag{1.4.1}$$

where α denotes a fixed penalization function. An example for α is given by a convex combination of L^1 and L^2 penalizations, i.e., $\alpha(\ell) = \lambda \|\ell\|_1 + (1-\lambda)\|\ell\|_2$, where $\lambda \in [0,1]$. Note here that the number of parameters corresponds to d^*, i.e. the dimension of the signature truncated at level $n-1$. This is due to the representation of Corollary 1.2.12, as also seen from the second equality in (1.4.1).

Example 1.4.1. Depending on the choice of the primary process, the time series for X may not be directly readable from the market. This is for instance the case if we assume X to be a vector of correlated Brownian motions, whose trajectories have to be extracted

from the observations of S. We discuss such a procedure in the following within the class of classical stochastic volatility models.

Suppose that the dynamics of the asset price process are described by a stochastic volatility model under the physical measure $\mathbb{P}$, e.g.,

$$\begin{aligned} \mathrm{d}S_t &= \mu(S_t, V_t)\mathrm{d}t + g(S_t, V_t)\mathrm{d}B_t^{\mathbb{P}} \\ \mathrm{d}V_t &= h(S_t, V_t)\mathrm{d}t + \sigma(S_t, V_t)\mathrm{d}W_t^{\mathbb{P}}, \end{aligned}$$

with $\mathrm{d}[B^{\mathbb{P}}, W^{\mathbb{P}}]_t = \rho \mathrm{d}t$ where $\rho \in [-1, 1]$ and μ, h, g, σ functions from $\mathbb{R}^2$ to $\mathbb{R}$ such that g and σ are strictly positive. Choosing $d = 2$, we aim to estimate the trajectories of a primary process $X_t := (B_t^{\mathbb{Q}}, W_t^{\mathbb{Q}})$ specified by

$$B_t^{\mathbb{Q}} = \int_0^t \frac{\mu(S_s, V_s)}{g(S_s, V_s)}\mathrm{d}s + B_t^{\mathbb{P}}, \qquad W_t^{\mathbb{Q}} = \int_0^t \frac{h(S_s, V_s)}{\sigma(S_s, V_s)}\mathrm{d}s + W_t^{\mathbb{P}}. \tag{1.4.2}$$

Note that under sufficient integrability conditions on the market prices of risk $\mathbb{Q}$ is an equivalent local martingale measure, under which $(B^{\mathbb{Q}}, W^{\mathbb{Q}})$ are correlated $\mathbb{Q}$-Brownian motions and under which the dynamics of (S, V) are of the following form

$$\begin{aligned} \mathrm{d}S_t &= g(S_t, V_t)\mathrm{d}B_t^{\mathbb{Q}}, \\ \mathrm{d}V_t &= \sigma(S_t, V_t)\mathrm{d}W_t^{\mathbb{Q}}. \end{aligned}$$

Note that other choices of equivalent local martingale measure $\widetilde{\mathbb{Q}}$ are of course possible. This then amounts to specify the second Brownian motion $W_t^{\widetilde{\mathbb{Q}}}$ (under the assumption that we stay in a Markovian framework) as

$$W_t^{\widetilde{\mathbb{Q}}} = \int_0^t \frac{h(S_s, V_s) - h^{\widetilde{\mathbb{Q}}}(S_s, V_s)}{\sigma(S_s, V_s)}\mathrm{d}s + W_t^{\mathbb{P}}.$$

Under this measure the dynamics of V, then read as

$$\mathrm{d}V_t = h^{\widetilde{\mathbb{Q}}}(S_t, V_t)\mathrm{d}t + \sigma(S_t, V_t)\mathrm{d}W_t^{\widetilde{\mathbb{Q}}}.$$

A simple choice is of course to choose $h^{\widetilde{\mathbb{Q}}} = h$ so that $W^{\widetilde{\mathbb{Q}}} = W^{\mathbb{P}}$. We consider in the applications below always the case $h^{\widetilde{\mathbb{Q}}} = 0$, i.e. $W^{\mathbb{Q}}$ as defined in (1.4.2).

In order to extrapolate the trajectories of $(B^{\mathbb{Q}}, W^{\mathbb{Q}})$, we estimate the spot quadratic variation of the price and of the volatility process (see e.g. Jacod and Protter (2011) for different types of pathwise spot covariance estimators in general semimartingale contexts), to get an estimator of $\Sigma_{11} := g(S_t, V_t)^2$ and $\Sigma_{22} := \sigma(S_t, V_t)^2$, which in turn allows to compute

$$\mathrm{d}B_t^{\mathbb{Q}} = \frac{\mathrm{d}S_t}{\sqrt{\Sigma_{11,t}}}, \qquad \mathrm{d}W_t^{\mathbb{Q}} = \frac{\mathrm{d}V_t}{\sqrt{\Sigma_{22,t}}}. \tag{1.4.3}$$

Observe that assuming g and σ to be strictly positive we get that $\sqrt{\Sigma_{11,t}} = g(S_t, V_t)$ and $\sqrt{\Sigma_{22,t}} = h(S_t, V_t)$. This is important to guarantee that $\mathrm{d}S_t = g(S_t, V_t)\mathrm{d}B_t^{\mathbb{Q}}$ (as requested in Corollary 1.2.12) and $\mathrm{d}V_t = h(S_t, V_t)\mathrm{d}W_t^{\mathbb{Q}}$.

Having extracted the trajectories of $B^{\mathbb{Q}}$ and $W^{\mathbb{Q}}$ we can readily compute the trajectories of their time extended signature. If one wants to work with a different equivalent local martingale measure $\widetilde{\mathbb{Q}}$ parameterized via $h^{\widetilde{\mathbb{Q}}}(S_s, V_s)$ as above, then $W_t^{\widetilde{\mathbb{Q}}}$ can obtained via

$$\mathrm{d}W_t^{\widetilde{\mathbb{Q}}} = \frac{\mathrm{d}V_t}{\sqrt{\Sigma_{22,t}}} - \frac{h^{\widetilde{\mathbb{Q}}}(S_t, V_t)}{\sqrt{\Sigma_{22,t}}}\mathrm{d}t.$$

Regression using spot volatility data

Let us now discuss the second method, which works when the primary process X is given by a vector of correlated Brownian motions similarly as in Example 1.4.1. Indeed, we then can also calibrate to the time-series data of the spot volatility of the trajectories of S. The same task was tackled in Perez Arribas et al. (2020) for the one-dimensional case with one driving Brownian motion as primary process and with a lead-lag extension instead of a time extension. In our framework the calibration to the spot volatility consists in finding $\ell^* \in \mathbb{R}^{d^*}$ such that

$$\ell^* \in \operatorname{argmin}_\ell \, L_{\mathrm{vol},\alpha}(\ell),$$

where the loss function $L_{\mathrm{vol},\alpha}(\ell)$ is now given by

$$L_{\mathrm{vol},\alpha}(\ell) := \sum_{i=1}^{N} \Big(\ell_\emptyset + \sum_{0<|I|\le n} \ell_I \langle e_I, \widehat{\mathbb{X}}_{t_i} \rangle - \sqrt{\frac{\mathrm{d}}{\mathrm{d}t}[S]_{t_i}} \Big)^2 + \alpha(\ell), \tag{1.4.4}$$

where $\frac{\mathrm{d}}{\mathrm{d}t}[S]$ denotes the spot quadratic variation process and α a penalty function as in (1.4.1). Note that we still work with the signature model as of Corollary 1.2.12 (for $D = 1$) exploiting its representation in form of (1.2.3).

Remark 1.4.2. (i) As the spot volatility is not directly available but has to be estimated, the penalty function plays an important role in order to avoid overfitting to noisy data. Note also that not all coefficients of the signature might be statistically significant. This includes in particular the case where d is large and the components of X are strongly correlated, and the case where the number of parameters to be calibrated is much higher than the number of points we aim to fit.

(ii) Let us stress that in the context of Example 1.4.1 the calibration to the price path and to the volatility process $g(S, V)$ are equivalent in the following sense. Once we have found ℓ^* solving (1.4.4), we obtain

$$[S]_t \approx \int_0^t \Big(\ell^*_\emptyset + \sum_{0<|I|\le n} \ell^*_I \langle e_I, \widehat{\mathbb{X}}_s \rangle \Big)^2 \mathrm{d}s.$$

Since by definition $S_t = \int_0^t g(S_s, V_s)\mathrm{d}B_s^{\mathbb{Q}}$, $g(S_t, V_t) > 0$, and $[B^{\mathbb{Q}}]_t = t$, this yields

$$S_t \approx S_0 + \int_0^t \Big(\ell^*_\emptyset + \sum_{0<|I|\le n} \ell^*_I \langle e_I, \widehat{\mathbb{X}}_s \rangle \Big) \mathrm{d}B_s^{\mathbb{Q}},$$

which by Lemma 1.2.10 can be rewritten as

$$S_t \approx S_0 + \ell^*_\emptyset B^{\mathbb{Q}}_t + \sum_{0<|I|\leq n} \ell^*_I \langle \tilde{e}_I, \widehat{\mathbb{X}}_t \rangle.$$

As we have the same expression for the parameters ℓ^* obtained via (1.4.1), the claim follows.

Numerical examples

We now study the above regression tasks by means of several examples based on classical models. We start with a SABR-type and a Heston model.

Example 1.4.3. Consider two classical stochastic volatility models, namely a SABR-type model

$$\begin{aligned} \mathrm{d}S_t &= \mu S_t \mathrm{d}t + S_t V_t \mathrm{d}B^{\mathbb{P}}_t \\ \mathrm{d}V_t &= \kappa(\theta - V_t)\mathrm{d}t + \sigma V_t \mathrm{d}W^{\mathbb{P}}_t, \end{aligned}$$

and a Heston model

$$\begin{aligned} \mathrm{d}S_t &= \mu S_t \mathrm{d}t + S_t \sqrt{V_t} \mathrm{d}B^{\mathbb{P}}_t \\ \mathrm{d}V_t &= \kappa(\theta - V_t)\mathrm{d}t + \sigma \sqrt{V_t} \mathrm{d}W^{\mathbb{P}}_t, \end{aligned} \tag{1.4.5}$$

where in both cases $\mathrm{d}[B^{\mathbb{P}}, W^{\mathbb{P}}]_t = \rho \mathrm{d}t$ where $\rho \in [-1, 1]$.

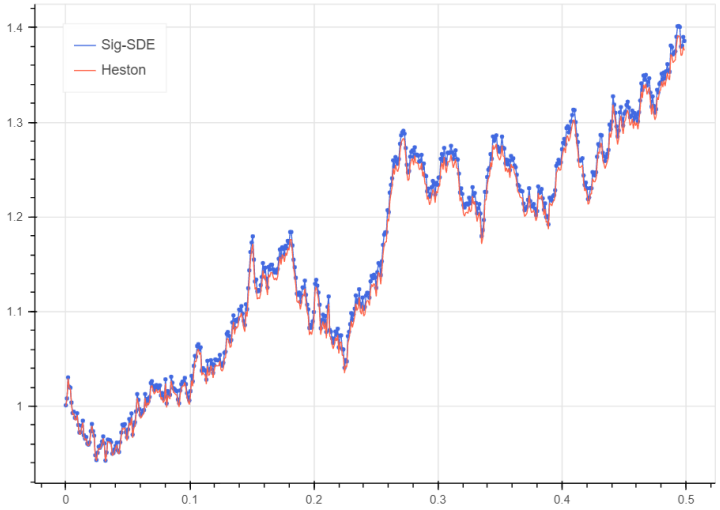
We now aim to approximate these models with signature models whose primary processes are $\mathbb{Q}$-Brownian motions extracted from the time series data of S and V as in Example 1.4.1. The calibration is performed applying both procedures described above, namely the calibration to the price's trajectory and the calibration to the spot volatility trajectory. The corresponding loss functions are given by (1.4.1) and (1.4.4), respectively. In both cases the truncation parameter is chosen to be $n = 2$ and a Lasso's penalty function $\alpha(\ell) = 10^{-5}\|\ell\|_1$ has been employed (a Ridge regression has led to similar results). For the experiment, we select the models parameters

$$\{S_0, V_0, \mu, \kappa, \theta, \sigma, \rho\} := \{1, 0.08, 0.001, 0.5, 0.15, 0.25, -0.5\}.$$

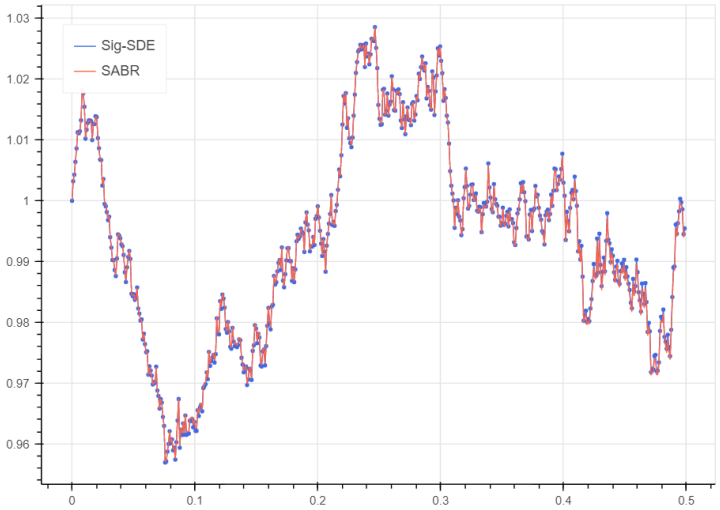
Observe that in the two models the process V has a different meaning. Indeed, in the SABR it corresponds to the spot volatility while in the Heston it is the spot variance. As we choose the same parameters, this means that the volatility V in the SABR model is smaller that the volatility $\sqrt{V}$ in the Heston model.

For both fit we extract 3 ticks of Brownian motions per day from trajectories that are assumed to be observable with a frequency of 4 seconds for 8 hours during 1 calendar year.

The calibrations have thus been performed over 1 calendar year of observations with 3 ticks per day ($T = 1$ and $N = 1095$). The obtained parameters ℓ^* are then tested simulating 0.5 years of new trajectories of the stochastic volatility models, extracting the corresponding $\mathbb{Q}$-Brownian motions, and computing the trajectories of $S_3(\ell^*)$ as in Corollary 1.2.12. Results are illustrated in Figure 1.1.



(a) Regression on the price of a Heston model as described in (1.4.1).



(b) Regression on the volatility of a SABR-type model as described in (1.4.4).

Figure 1.1.: Out-of-sample of half calendar year with training sample of one calendar year.

To assess the accuracy of the above calibration to time-series data we consider the mean squared error (MSE) between the predicted path and the observed path, i.e.

$$\frac{1}{K}\sum_{i=1}^{K}(S_n(\ell^*)_{t_i} - S_{t_i})^2.$$

In the following tables we state the in-sample MSE denoted by $\mathrm{MSE}_{\mathrm{ex}}^{\mathrm{in}}$ as well as the out-of-sample MSE denoted by $\mathrm{MSE}_{\mathrm{ex}}^{\mathrm{out}}$. To justify the robustness of the estimated

parameters the out-of-sample MSEs are calculated as an average of the mean squared errors on 1000 different realizations of the simulated model.
Regression on the price as in (1.4.1) under a Heston model:

$\mathrm{MSE}_{\mathrm{ex}}^{\mathrm{in}}$	$\mathrm{MSE}_{\mathrm{ex}}^{\mathrm{out}}$
$2.574 \cdot 10^{-6}$	$2.841 \cdot 10^{-4}$

Regression on the volatility as in (1.4.4) under the SABR-type model:

$\mathrm{MSE}_{\mathrm{ex}}^{\mathrm{in}}$	$\mathrm{MSE}_{\mathrm{ex}}^{\mathrm{out}}$
$3.204 \cdot 10^{-6}$	$1.060 \cdot 10^{-6}$

It is worth mentioning that learning the volatility as in (1.4.4) requires an estimation of the paths of $(\sqrt{\frac{\mathrm{d}}{\mathrm{d}t}[S]_t})_{t\in[0,T]}$ from the price's trajectories. If such estimation is too noisy the procedure can yield unsatisfactory results. One reason why this is here not visible in the out-of-sample MSEs is that in the SABR model the volatility enters linearly in the price dynamics, and hence is easier to be learned via (1.4.4), which can be seen from the above tables. Note also that the smaller volatility in the SABR models plays role in the order of magnitude of $\mathrm{MSE}_{\mathrm{ex}}^{\mathrm{out}}$.

In the next example we consider a multivariate asset price process whose dynamics are given by a multi-dimensional Black-Scholes model.

Example 1.4.4 (The multi-dimensional case)**.** Consider a d-dimensional Black-Scholes model under the physical probability measure $\mathbb{P}$, i.e.,

$$\mathrm{d}S_t = \mu^\top \operatorname{diag}(S_t)\mathrm{d}t + \operatorname{diag}(S_t)\sqrt{\Sigma}\mathrm{d}B_t^{\mathbb{P}},$$

where $B^{\mathbb{P}}$ is a d-dimensional standard Brownian motion (with independent components), $\mu \in \mathbb{R}^d$, $\Sigma \in \mathbb{S}_+^d$ and $\sqrt{\cdot}$ denotes the matrix square root. Similarly as in Example 1.4.1 we extract independent $\mathbb{Q}$-Brownian motions (i.e. $\mathbb{P}$ with the corresponding market price of risk) which serve as primary process X in the signature model, now however of dimension $d > 1$.

In order to calibrate to the simulated price path S, we perform a Ridge regression using the prices' trajectories with penalty function $\alpha(\ell) = 5 \cdot 10^{-5}\|\ell\|_2$. The truncation parameter is chosen to be $n = 5$, corresponding to 46655 parameters to be calibrated. We considered the case of five positively correlated components, i.e. $\Sigma_{ij} = \sigma_i\sigma_j\rho_{i,j}$ with $\sigma_i \in (0.15, 0.34)$ and $\rho_{i,j} \in (0.4, 0.9)$ for $i, j = 1, \ldots, 5$, with drifts $\mu_i \in (-0.003, 0.001)$ for $i = 1, \ldots, 5$. The calibration has been performed over 1 calendar year of observations with 3 ticks per day ($T = 1$ and $N = 1095$). The obtained parameters ℓ^* are then tested simulating 4 months of new trajectories of the Black-Scholes model, extracting

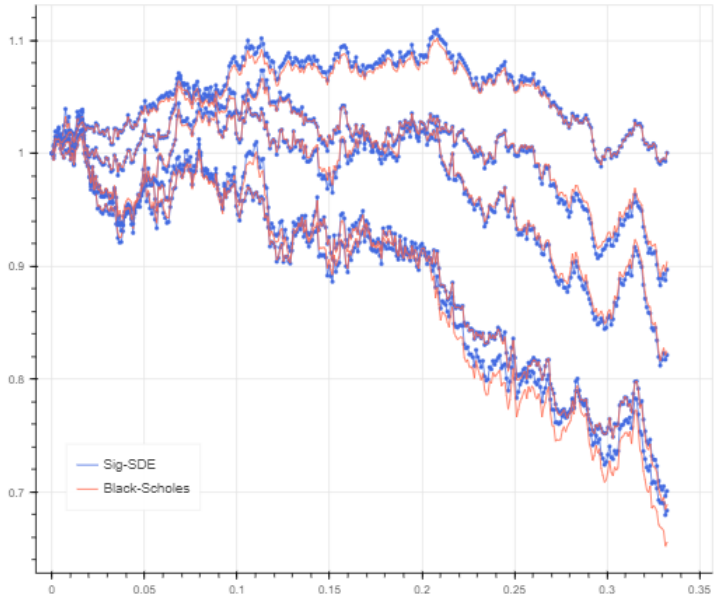


Figure 1.2.: Regression on the price of a 5-dimensional Black-Scholes model as described in (1.4.1). Out-of-sample of 4 months with training sample of 1 calendar year.

the corresponding $\mathbb{Q}$-Brownian motions, and computing the trajectories of $S_6(\ell^*)$ as in Corollary 1.2.12. Results are illustrated in Figure 1.2. The $\mathrm{MSE}_{\mathrm{ex}}^{\mathrm{out}}$ on the single trajectories are of order 10^{-5}.

1.4.2. Calibration to option prices

We now pass to the more common way of calibrating models in mathematical finance, namely to calibration to call (and put) option prices, with the goal to match their implied volatilities as well as possible.

Throughout we here assume that there exists a pricing measure $\mathbb{Q}$ and that the primary process X is a vector of correlated $\mathbb{Q}$-Brownian motions. The goal of this approach is to choose the parameters ℓ in order to fit the prices of options on S available on the market. In the context of Sig-SDEs, a similar approach has been considered in Perez Arribas et al. (2020). There the authors however only consider option prices generated from a Black-Scholes model and maturities longer than approximately 5 months. To our knowledge, in the context of signature based models for asset prices, we are the first ones dealing with market data, thus including shorter maturities and realistic volatility smiles. Suppose we are given prices of N call options $\pi^*(T_1, K_1), \dots, \pi^*(T_N, K_N)$ with maturities T_i and strikes K_i. Then in spirit of Cont and Ben Hamida (2005) (see also Cuchiero et al. (2020)) we translate the current calibration task to finding ℓ such that

$$\ell \in \operatorname{argmin}_\ell \; L_{\text{option}}(\ell),$$

where

$$L_{\text{options}}(\ell) = \sum_{i=1}^{N} \gamma_i \Big(\pi^*(T_i, K_i) - \pi^{\text{model}}(\ell, T_i, K_i) \Big)^2, \tag{1.4.6}$$

with γ_i Vega weights and $\pi^{\text{model}}(\ell, T_i, K_i)$ denoting the price of the option under the signature model with parameters ℓ, maturity T_i and strike K_i.

Remark 1.4.5. The loss function described in (1.4.6) can be adapted depending on the available data. A common approach is for instance to weigh the options by their bid-ask spreads, see for instance Cont and Ben Hamida (2005). This reflects the relative importance of reproducing different option prices precisely. In the present work we shall employ Vega weights since bid-ask spreads are not always provided in the data and since they are well-suited to match the implied volatility surface.

The same calibration procedure would apply to any other type of option, e.g. path-dependent exotic ones: as we shall see below, the pricing of the latter with a signature model does not require any additional effort. For liquidity reasons we however decided to stick to vanilla options and calibrate to the corresponding implied volatility surface.

To fix notation we briefly recall the definition of implied volatility, which is our goodness of fit criterium.

Definition 1.4.6. Let $\pi(T, K)$ be the price of a call option with maturity T and strike price K written on the asset S. The implied volatility of $\pi(T, K)$ is defined as the volatility $\sigma(T, K)$ that solves the equation

$$\pi^{BS}(T, K, \sigma(T, K)) = \pi(T, K), \tag{1.4.7}$$

where π^{BS} Black-Scholes option price. Then $(\sigma(T, K))_{T,K}$ is called (implied) volatility surface, and $(\sigma(T, K))_K$ is called (implied) volatility smile for each fixed maturity T.

Once the optimal parameters ℓ^* are found via (1.4.6) we then need to solve (1.4.7) numerically for $\pi(T,K) = \pi^{\text{model}}(\ell^*, T, K)$ to assess the goodness of fit. Indeed we measure it in terms of the absolute error (in basis points) between the implied volatilities from the market and the model. In Chapter 2 we will on the other hand test our calibration results by checking weather the calibrated implied volatilities lie in the market's bid-ask spread.

Monte Carlo pricing

We now address the problem of computing for a fixed strike $K \in \mathbb{R}$ and maturity $T > 0$, the price $\pi^{\text{model}}(\ell, T, K)$ of the corresponding option under the signature model.

Before presenting our method based on Monte Carlo, recall that Section 1.3 provides a closed-form formula for the computation of sig-payoffs without the need to resort to Monte Carlo techniques. By the universal approximation theorem (Theorem 1.1.12), we also know that call options payoffs can be approximated arbitrarily well by sig-payoffs, or in this case even just by polynomials, on compacts. The combination of these two properties leads to the approach exploited by Perez Arribas et al. (2020) of splitting the computation of $\pi^{\text{model}}(\ell, T, K)$ in two parts: an approximation of the call (put) payoffs using sig-payoffs and pricing the latter with the closed formula proved in Section 1.3. However, in order to achieve an acceptable error between the original call payoff and the sig-payoff one needs to choose a sig-payoff involving signature terms of quite high order. Indeed, such an approximation needs to be valid for every set of coefficients ℓ involved in the optimization procedure and thus on a very large compact set K. From a computational point of view we observed that this sig-payoff or polynomial approximation is not tenable and we thus opted for abandoning this approach. Note however that sig-payoffs can still be used as control variates in Monte Carlo pricing techniques as we shall outline in Section 1.4.2. For more details on the issues that can occur in view of sig-payoff or polynomial approximations we refer to Section A.2.

For these reasons we therefore adopted a Monte Carlo approach, which is due to the linearity of the model particularly tractable, making the whole calibration computationally feasible within a reasonable amount of time while providing highly accurate results. For the Monte Carlo price we thus fix a number of samples $N_{MC} > 0$ and approximate $\pi^{\text{model}}(\ell, T, K)$ via

$$\pi^{\text{model}}(\ell, T, K) \approx \frac{1}{N_{MC}} \sum_{i=1}^{N_{MC}} (S_n(\ell)_T(\omega_i) - K)^+.$$

We stress again that this can be computed fast. Indeed, by the linearity of the model simulating $(S_n(\ell)_T(\omega_i))_{i=1}^{N_{MC}}$ boils down to the following steps:

- simulate $(X_t(\omega_i))_{t\in[0,T]}$, which in the current setting are just trajectories for correlated Brownian motions, for each $i \in \{1, \dots, N_{MC}\}$;
- compute $\langle e_I, \widehat{\mathbb{X}}_T(\omega_i)\rangle$ for all $i = 1, \dots, N_{MC}$ and for all multi-indices I such that $|I| \leq n$;

- take linear combinations to compute $\langle \tilde{e}_I, \widehat{\mathbb{X}}_T(\omega_i)\rangle$ for all $i = 1, \dots, N_{MC}$ and for all multi-indices $|I| \leq n-1$ as described in Lemma 1.2.10;
- retrieve $(S_n(\ell)_T(\omega_i))_{i=1}^{N_{MC}}$ via (1.2.2).

Observe that the parameters ℓ only enter in the very last step which allows to precompute and store all other quantities. This is in contrast to other models where the parameters that need to be calibrated enter at each time point in the simulation steps of e.g. an Euler scheme or more complicated schemes used for instance for rough volatility models.

Remark 1.4.7. Since the calibration problem is usually ill-posed as inverse problem, there might exist more than one local minimum of the loss function which is able to reproduce the prices on the market, and their implied volatility. We refer to Cont and Tankov (2004) where the sensitivity, e.g. in terms of initial starting points in gradient descent optimization algorithms, of the nonlinear least squares calibration problem in the case of exponential Lévy models is discussed. In our setting it is worth mentioning that since $S_n(\ell)_T$ is linear and the call payoffs are convex we have a convex optimization problem whenever for all maturities and strikes

$$\pi^*(\ell, T, K) \leq \frac{1}{N_{MC}} \sum_{i=1}^{N_{MC}} (S_n(\ell)_T(\omega_i) - K)^+.$$

Therefore the initial random parameter ℓ for the optimization can for instance be initialized according to this condition.

Variance reduction with sig-payoffs

Even though Monte Carlo pricing is fast since all essential quantities can be precomputed as explained above, we here discuss variance reduction techniques (see e.g. Glasserman (2004)) that can speed up the procedure even further. The idea is to introduce a control variate, i.e., a random variable Φ^{cv} such that:

$$\mathbb{E}_{\mathbb{Q}}[\Phi^{cv}] = 0, \qquad \mathrm{Var}\big((S_n(\ell)_T - K)^+ - \Phi^{cv}\big) < \mathrm{Var}\big((S_n(\ell)_T - K)^+\big).$$

An example of control variates used for pricing and calibrating neural SDE models can be found in Cuchiero et al. (2020); Gierjatowicz et al. (2020), where Φ^{cv} is constructed from hedging strategies. A possible other choice of control variates for signature models are sig-payoffs. Indeed, one can use the pricing formula derived in Section 1.3 to define:

$$\Phi^{cv}(\ell, T, K) := f_\emptyset + \sum_{0<|J|\leq m} f_J \langle e_J, \widehat{\mathbb{S}}_n(\ell)_T\rangle - \widetilde{P}_f(\ell, \mathbb{E}_{\mathbb{Q}}[\widehat{\mathbb{X}}_T]),$$

for some fixed $m > 0$ where

$$(S_n(\ell)_T - K)^+ \approx f_\emptyset + \sum_{0<|J|\leq m} f_J \langle e_J, \widehat{\mathbb{S}}_n(\ell)_T\rangle,$$

for a wide range of ℓ and with high probability. This can be done by performing a linear regression to obtain the coefficients f. Alternatively, a polynomial approximation of the payoff's function can also be employed.

The properties of Φ^{cv} then guarantee the accuracy of the approximation

$$\pi^{\text{model}}(\ell, T, K) \approx \frac{1}{N_{MC}} \Big(\sum_{i=1}^{N_{MC}} (S_n(\ell)_T(\omega_i) - K)^+ - \Phi^{cv}(\ell, T, K)(\omega_i) \Big),$$

already for smaller values of N_{MC}. We will not use the previous variance reduction techniques within the present chapter but we will discuss it again in Section 2.3.3, in application to price VIX options and to Remark 2.4.9 for SPX options.

Model's performance

In the following we discuss the problem of minimizing the functional (1.4.6) using the Monte Carlo method as described in Section 1.4.2 to compute the model prices. We consider the model described in Corollary 1.2.12 for $\widehat{X}_t = (t, B_t, W_t)$ as in Definition 1.2.11.2.1 for two correlated Brownian motions B and W with correlation coefficient $\rho = -0,5$, and $D = 1$.

As a first example we consider synthetic data, where the implied volatility surface we aim to fit is generated by a Heston model given by (1.4.5). We consider 7 maturities $(T_k)_{k=1}^7$ ranging from 30 days to 2 years and 13 strikes $(K_j)_{j=1}^{13}$ ranging from 80% to 120% of the spot price. The truncation parameter is fixed to $n = 3$ and the number of Monte Carlo samples to $N_{MC} = 10^6$. The results for the following two sets of parameters under the risk neutral measure

κ	θ	σ	ρ	V_0
0.1	0.1	0.4	-0,5	0.08
0.2	0.3	0.5	-0,5	0.08

are displayed in the first and in the second row of Figure 1.3, respectively.

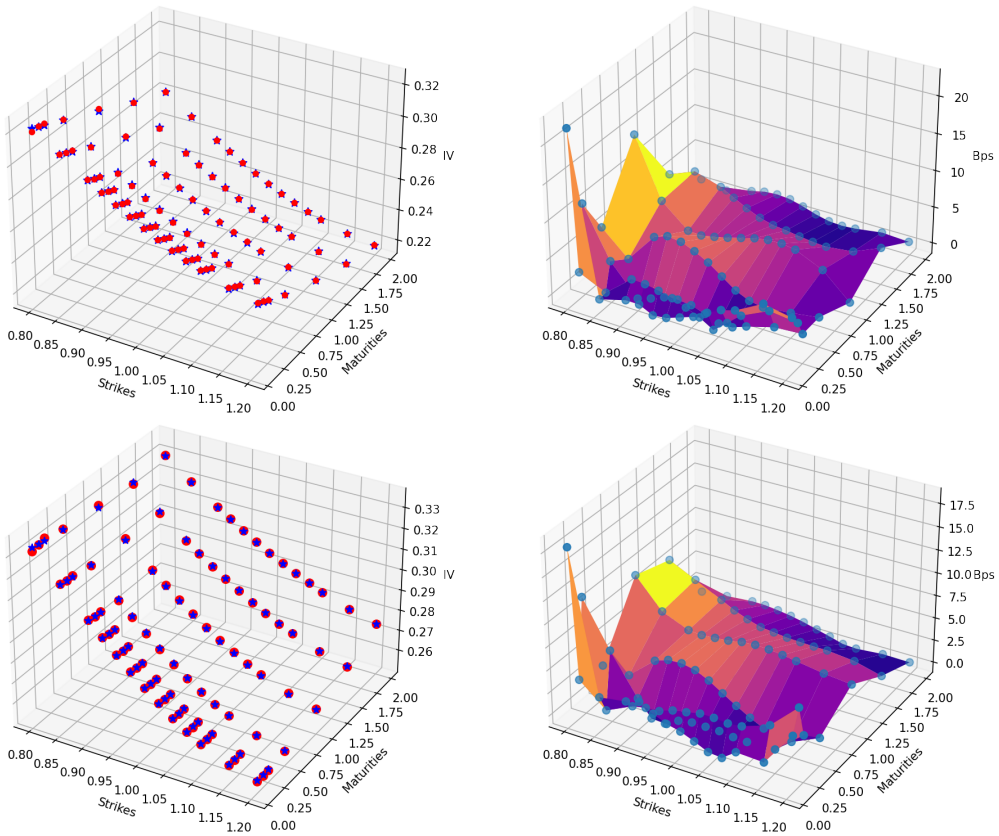


Figure 1.3.: On the left: blue stars correspond to the implied volatilities of the Heston models, red dots denote the calibrated implied volatilities of $S_n(\ell)$ with $n = 3$ (13 estimated parameters). On the right: absolute errors between the two surfaces expressed in basis points (Bps).

We stress that these calibrations to Heston generated implied volatility surfaces can take between 4 to 15 minutes on a standard machine.

Let us now turn to market data. We consider the trading day 17/03/2021 for call options written on the S&P 500 index. Our dataset provided by Bloomberg consists of 7 maturities $(T_k)_{k=1}^7$, ranging from 30 days to 2 years, and 9 strike prices $(K_j)_{j=1}^9$ for each maturity which vary between 80% and 120% of the spot price. Again, the truncation parameter is fixed to $n = 3$ and the Monte Carlo's parameter to $N_{MC} = 10^6$. The results are displayed in Figure 1.4.

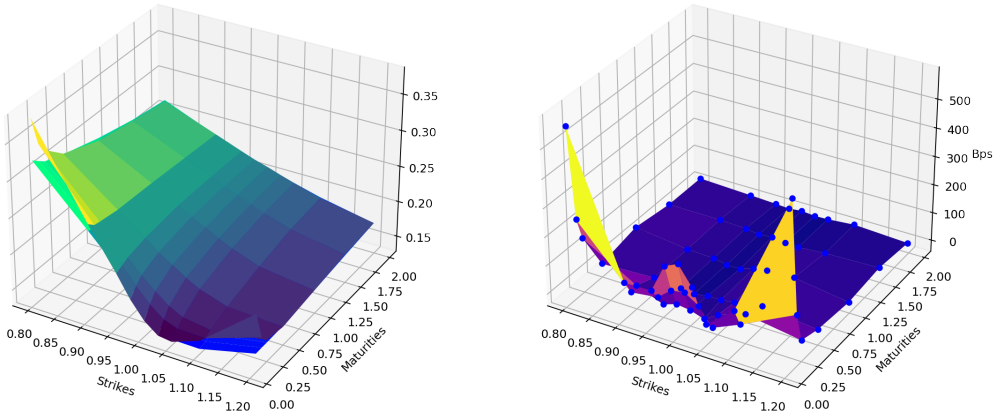


Figure 1.4.: On the left: the upper surface represents the implied volatility of the S&P 500 index as of 17-03-2021, the lower one is the calibrated implied volatility of $S_n(\ell^*)$ with $n = 3$ (13 parameters). On the right: absolute error between the two surfaces (Bps).

From Figure 1.4 we notice that volatility smiles for short maturities have not been captured. In particular, as for many continuous models, the most problematic part consists in fitting the shortest maturities (see Chapter 3 and Chapter 7 of Gatheral (2011)).

To investigate the ability of the model to reproduce short maturity smiles we calibrate it using different loss functions penalizing outliers more severely. Precisely, we first calibrate the model using the procedure described above and we denote by w_i the absolute error between the target implied volatility and the approximated one for maturity T_i and strike K_i. Then, in spirit of generative-adversarial distances as for instance considered in Cuchiero et al. (2020), we define a new loss function

$$L_{p,\alpha}(\ell) = \sum_{i=1}^{N} (\gamma_i + \alpha w_i) |\pi^*(T_i, K_i) - \pi^{\text{model}}(\ell, T_i, K_i)|^p, \tag{1.4.8}$$

depending on parameters p and α that need to be chosen. By taking high values for p and α we can approximate the sup-distance between the two price surfaces, i.e.,

$$L_\infty(\ell) := \sup_{i=1,\dots,N} |\pi^*(T_i, K_i) - \pi^{\text{model}}(\ell, T_i, K_i)|,$$

without compromising differentiability with respect to ℓ. The result for different choices of the parameters α and p but also the truncation level n is displayed in Figure 1.5. As can be guessed from the figure, although the maximal absolute error for maturities larger than 60 days is (almost) acceptable (96 Bps for $n = 3$, $p = 1000$, and $\alpha = 500$ and 34 Bps for $n = 4$ and $p = 1000$, and $\alpha = 500$), the absolute error for the shortest maturity and

the far in and out of the money strikes are still above 270 Bps and 395 Bps respectively. Observe that the performance for $n = 4$ is the best for every maturity larger than 60 days as well as for the at-the-money region of the shortest maturity.

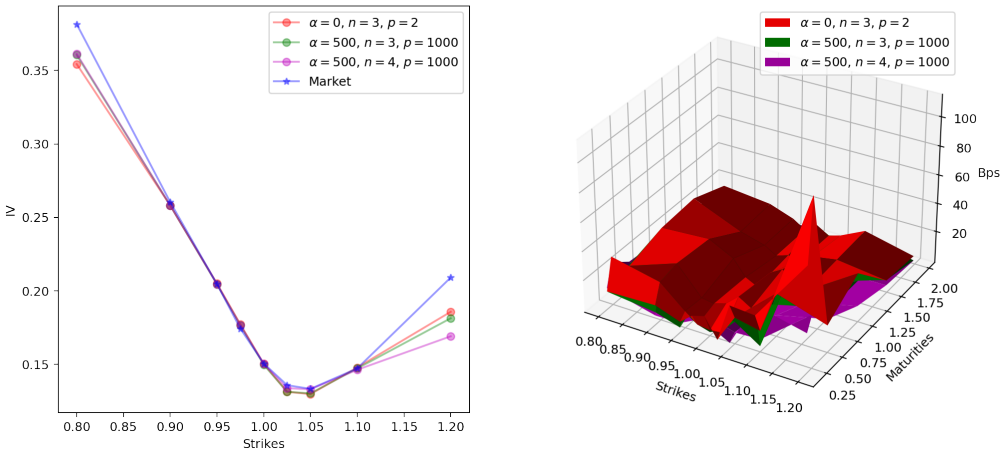


Figure 1.5.: Comparison between the calibrated implied volatility smiles for different parameters and the S&P 500 index as of 17-03-2021 smile (in blue) at maturity $T_1 = 30$ days (on the left) and for maturities ranging from 60 days to 2 years (on the right). Calibration has been performed using (1.4.8).

Finally, as a further experiment, we calibrate the model to the shortest maturity alone and to every other maturity together. The first calibration is thus performed to $T_1 = 30$ days and the second one to 6 maturities ranging from 60 days to 2 years. In both cases 9 strikes ranging from 80% to 120% of the spot price are considered. The parameters are fixed to $n = 2$, $p = 2$, and $\alpha = 0$ for the calibration to the first maturity and $n = 4$, $p = 300$, $\alpha = 500$ for the remaining maturities. The result is displayed in Figure 1.6. We can see that the absolute error is below 12 Bps for the first maturity and 45 for every other maturity, thus yielding a high accuracy. The computational time needed for the calibration to the first smile is the range of a few minutes, the calibration to the remaining surface can take longer, which is due to the higher value of n and the joint calibration to all maturities.

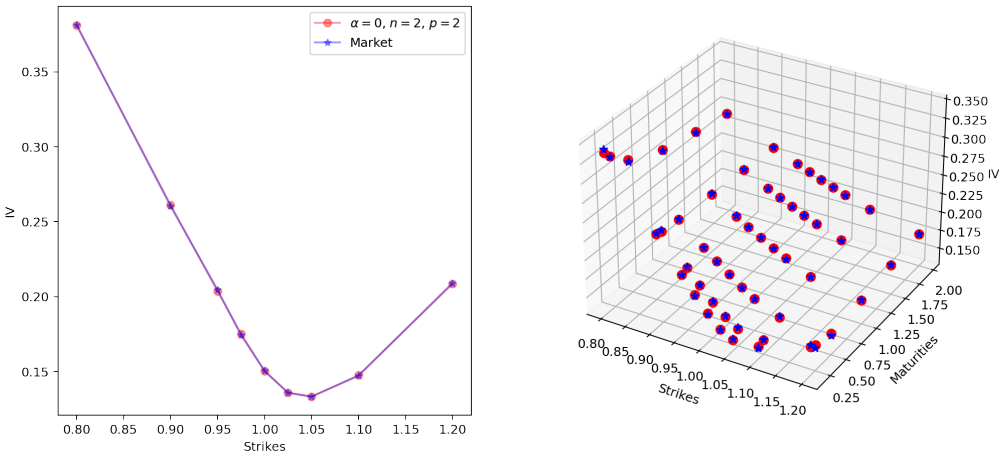


Figure 1.6.: Comparison between the calibrated implied volatility smiles and the SPX-500 Index as of 17-03-2021 smile (in blue) at maturity $T_1 = 30$ days (on the left) and for maturities ranging from 60 days to 2 years (on the right). Two different calibrations have been performed using (1.4.8).

These results suggest that introducing maturity-dependent parameters and performing a slice-wise calibration to the individual smiles as for instance in Cuchiero et al. (2020) or Gierjatowicz et al. (2020) can be of interest to obtain both, an excellent accuracy and a low computational time. We investigate this further in Section 1.4.2.

Remark 1.4.8. Note that a signature model with a two dimensional primary process and $n = 4$ as in the Figure 1.6 (right) has 121 parameters. This model is calibrated to 54 options prices. To investigate possible over-fitting, we computed values of the implied volatility surface for out-of-sample strikes and maturities. More precisely, we added in total 76 new points. The result is displayed in Figure 1.7 where the out-of-sample points are displayed in red and the surface corresponds to the one of Figure 1.6 (right) for maturities from 60 days to 2 years. This shows that the model exhibits highly promising generalization features and is thus well suited for fast arbitrage-free smile inter- and extrapolation.

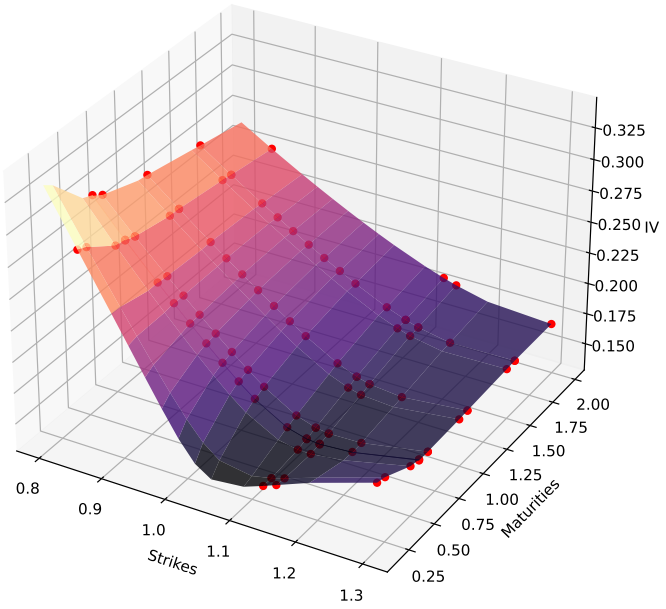


Figure 1.7.: Implied volatility surface generated by $S_4(\ell^*)$ with ℓ^* as in Figure 1.6 (right). Red dots denote out of sample points.

Time-dependent parameters

Motivated by the above results, we propose in the following a variant of the initial model where the parameters ℓ are allowed to depend on time such that we can perform a slice-wise calibration. Set $\widehat{X}$ as in Definition 1.2.11.2.1 and let $\{T_1, \dots, T_N\}$ with $N > 1$ be a set of fixed times, usually corresponding to the available maturities, and set

$$\overline{S}_n(\ell, \ell^{\mathrm{corr}})_t := S_n(\ell)_t + R_n(\ell^{\mathrm{corr}})_t, \tag{1.4.9}$$

where $S_n(\ell)$ is given as in Corollary 1.2.12 for $D = 1$ and

$$R_n(\ell^{\mathrm{corr}})_t = \sum_{j=1}^{N-1} \sum_{|I| \leq n-1} \ell_I^{\mathrm{corr},j} 1_{\{t \geq T_j\}} \langle \tilde{e}_I, \widehat{\mathbb{X}}_t - \widehat{\mathbb{X}}_{T_j} \rangle.$$

Observe that $R_n(\ell^{\mathrm{corr}})_t = 0$ for each $t \leq T_1$. This choice is motivated by the observation that $S_n(\ell)$ performs very well if calibrated on the first maturity only (see Figure 1.6). On every other T_k we have

$$R_n(\ell^{\mathrm{corr}})_{T_k} = \sum_{j=1}^{k-1} \sum_{|I| \leq n-1} \ell_I^{\mathrm{corr},j} \langle \tilde{e}_I, \widehat{\mathbb{X}}_{T_k} - \widehat{\mathbb{X}}_{T_j} \rangle.$$

Moreover, for any $t \geq 0$ Lemma 1.2.10 yields

$$\overline{S}_n(\ell, \ell^{\text{corr}})_t = S_0 + \int_0^t \sum_{|I| \leq n-1} \left(\ell_I + \sum_{j=1}^{N-1} \ell_I^{\text{corr},j} 1_{\{s \geq T_j\}} \right) \langle e_I, \widehat{\mathbb{X}}_s \rangle \mathrm{d}X_s^1,$$

showing that $\overline{S}_n(\ell, \ell^{\text{corr}})$ is still continuous. We can also see that if X^1 is a local martingale the same is true for $\overline{S}_n(\ell, \ell^{\text{corr}})$.

Remark 1.4.9. The added correction term can be interpreted to take a similar role as the so-called leverage function in local stochastic volatility models going back to Lipton (2002); Ren et al. (2007). Indeed, the leverage function is an adjustment to stochastic volatility models that depends on time (and state) and has the purpose to perfectly fit the observed smiles, which by the underlying stochastic volatility model can usually not be achieved at the desired accuracy.

Let us now describe the calibration procedure for the adjusted model with time-dependent parameters. Suppose (for simplicity) that for each maturity $T_1, \dots, T_N$ the price of call options with strikes $K_1, \dots, K_M$ is known. The calibration is then performed recursively, adapting the loss function to the already calibrated parameters. Precisely, the loss function used to calibrate ℓ is given by

$$L_{\text{options}}(\ell) = \sum_{j=1}^{M} \gamma_{1,j} \Big(\pi^*(T_1, K_j) - \pi^{\text{model}}(\ell, T_1, K_j) \Big)^2, \tag{1.4.10}$$

which is a version of (1.4.6). Next, fix $k \in \{1, \dots, N-1\}$ and let $\ell^{<k} := (\ell, \ell^{\text{corr},1}, \dots, \ell^{\text{corr},k-1})$ be the already calibrated coefficients. The loss function used to calibrate $\ell^{\text{corr},k}$ is given by

$$L_{\text{options},k}(\ell^{\text{corr},k}) = \sum_{j=1}^{M} \gamma_{k,j} \Big(\pi^*(T_k, K_j) - \pi^{\text{model}}(\ell^{<k}, \ell^{\text{corr},k}, T_k, K_j) \Big)^2. \tag{1.4.11}$$

In both cases $\gamma_{k,j}$ denotes the Vega weight corresponding to maturity T_k and strike K_j and π^{model} is computed using Monte Carlo as outlined in Section 1.4.2, namely

$$\pi^{\text{model}}(\ell^{<k}, \ell^{\text{corr},k}, T_k, K_j) \approx \frac{1}{N_{MC}} \sum_{i=1}^{N_{MC}} (\overline{S}_n(\ell^{<k}, \ell^{\text{corr},k})_{T_k}(\omega_i) - K_j)^+.$$

To test this alternative approach we consider again the trading day 17/03/2021 for call options written on the S&P 500 index, 7 maturities ranging from 30 days to 2 years, and 9 strike prices for each maturity which vary between 80% and 120% of the spot price. The underlying process is chosen to be $X = (B, W)$ for two Brownian motions B and W with correlation $\rho = -0.5$. The truncation parameter is fixed to $n = 2$ and the number of Monte Carlo samples to $N_{MC} = 10^6$.

In Figure 1.8 we report the results after minimizing (1.4.10) and (1.4.11) for all $k = 1, \dots, 6$. This shows that the signature model with time-dependent parameters perfectly fits each volatility smile. As illustrated in Figure 1.9 the model also manages to replicate the observed term structure of the at-the-money (ATM) volatility skew, defined via $\psi(T) := \left|\frac{\partial \sigma(T,K)}{\partial K}\right|_{K=S_0}$, for any maturity $T > 0$, very accurately. This is a feature which can also be reproduced by rough volatility models, but classical stochastic volatility models rather generate a term structure of the ATM skew that is constant for small maturities (see Gatheral et al. (2018)).

Finally, we stress the fact that the calibration procedure is also very fast, as for each smile it only took about approximately 1 minute and 30 seconds. These results suggest that the signature model with time-dependent coefficients qualifies for very fast and highly accurate calibration fits.

In Figure 1.10 we plot the values of the 91 calibrated parameters $\ell, \ell^{\text{corr},1}, \dots, \ell^{\text{corr},6}$. One can see that the corrections parameters get smaller and smaller. Note that we could have worked with a smaller number of corrections leading to a similar performance in terms of accuracy. As shown at the beginning of the section one set of parameters is indeed enough to capture long maturities (e.g. maturities longer than one year).

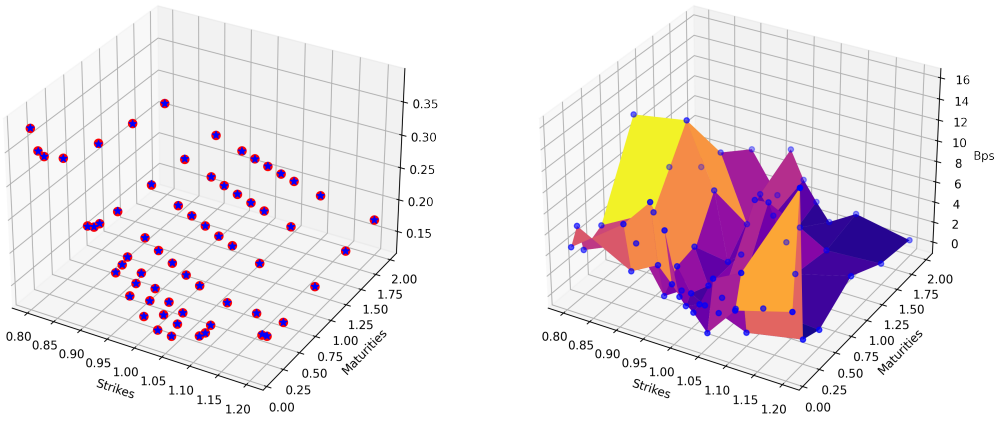


Figure 1.8.: On the left: blue stars denotes the implied volatilities on the market as of 17-03-21 on S&P 500 index and the red dots denote the calibrated implied volatilities with a signature model as in (1.4.9) with $n = 2$. On the right: absolute error between the two surfaces (Bps).

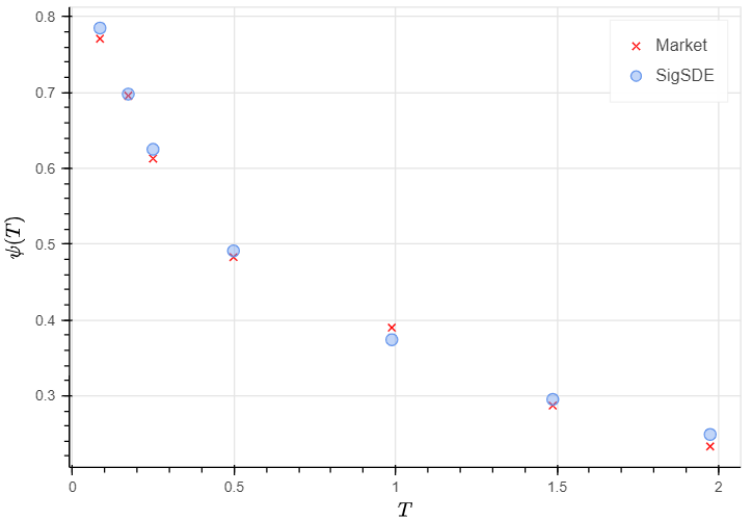


Figure 1.9.: Comparison between market term structure of the at-the-money (red crosses) implied volatility skew ψ and the calibrated one from the SigSDE (azure circles) minimizing (1.4.10) and (1.4.11).

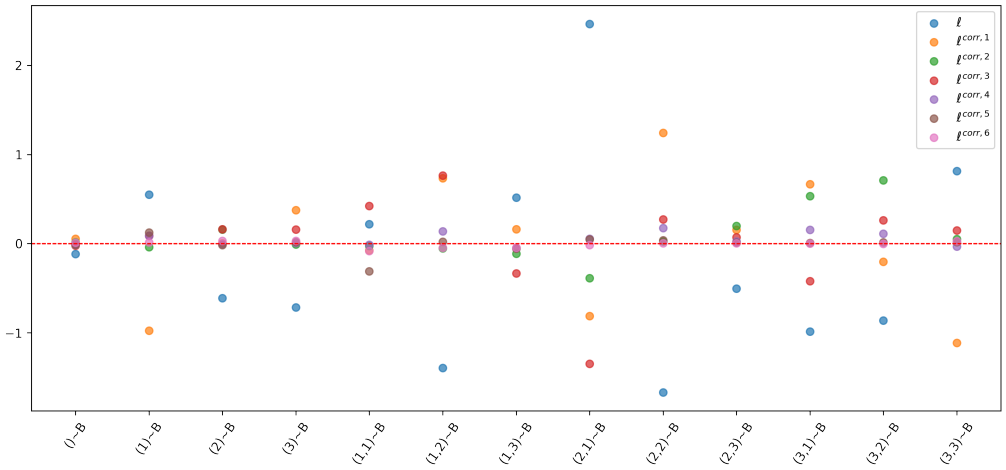


Figure 1.10.: Optimal parameters $\ell, \ell^{\text{corr},1}, \ldots, \ell^{\text{corr},6}$ (consisting of 13 parameters each) found calibrating to the volatility surface of the market as of 17-03-21 on S&P 500 index.

1.4.3. Joint calibration to time-series data and option prices

In this section we combine the promising results obtained in Section 1.4.1 and 1.4.2 to solve the following joint calibration problem. The main focus is to find a signature model that fits at the same time both, a path realization coming from time series data as well as a volatility smile implied from option prices. Let $\widehat{X}$ be as in Definition 1.2.11.2.1, let $\{T_1, \dots, T_N\}$ with $N > 1$ be fixed times corresponding to the options' maturities and consider a signature model $S_n(\ell)$ as in Corollary 1.2.12. Suppose that for each maturity $T_1, \dots, T_N$ the prices of call options are available for several strikes, and that we have time series data for the primary process and for the target model on a given time grid Π at our disposal. Consider now the loss function used to calibrate ℓ given by

$$L_{\text{joint}}(\ell) := \lambda L_{\text{options}}(\ell) + (1-\lambda) L_{\text{price},0}(\ell)$$

where $L_{\text{price},0}$ is given by (1.4.1) with $\alpha = 0$ and times $t_i \in \Pi$, L_{options} is given by (1.4.10), and $\lambda \in [0,1]$.

Remark 1.4.10. When performing this joint calibration we implicitly assume to have a common primary process X for both tasks. We here suppose that X is given by (1.4.3), i.e. we work with Brownian motions under the local martingale measure $\mathbb{Q}$ specified in Example 1.4.1. Note that under this measure $\mathbb{Q}$ the volatility/variance process V is necessarily a local martingale. Below we shall use a Heston model to generate both the price trajectory (under $\mathbb{P}$) and the option prices (under $\mathbb{Q}$). For the latter we specify the parameters to guarantee that V is a (local) martingale under $\mathbb{Q}$. One can of course also work with different local martingale measures as outlined in Example 1.4.1.

We consider an example where the traded asset S follows Heston dynamics as described in (1.4.5), with the following parameters.

μ	κ	θ	σ	ρ	V_0	S_0
0.001	0.8	0.1	0.55	-0.5	0.12	1

under the physical measure $\mathbb{P}$. We fix $N = 2$ and generate option prices at maturity T_1 =3 months and T_2 =1 year, under the set of parameters

μ	κ	θ	σ	ρ	V_0	S_0
0	0	0	0.55	-0.5	0.12	1

which ensures S and V to be local martingales under $\mathbb{Q}$. We consider additionally equi-spaced strike prices $\{K_1, \dots, K_9\}$ with moneyness ranging between $70 - 130\%$ of the spot price. In Figure 1.11 we report the fit of the implied volatility smiles for T_1, T_2 respectively, with less than 25 bps of absolute error. For the path calibration we extract daily Brownian motions from a Heston trajectory which is assumed to be observable with a frequency of 3 seconds for 8 hours during 9 months. The training sample consists of 3 months, namely 91 points and the out-of-sample performance in Figure 1.12, is tested on 6 months. As hyper-parameters we choose $n = 2$, $\lambda = 0.9$, and $N_{MC} = 10^6$ trajectories

for the Monte Carlo pricing. These plots indicate that the model is also able to jointly fit time-series and option price data.

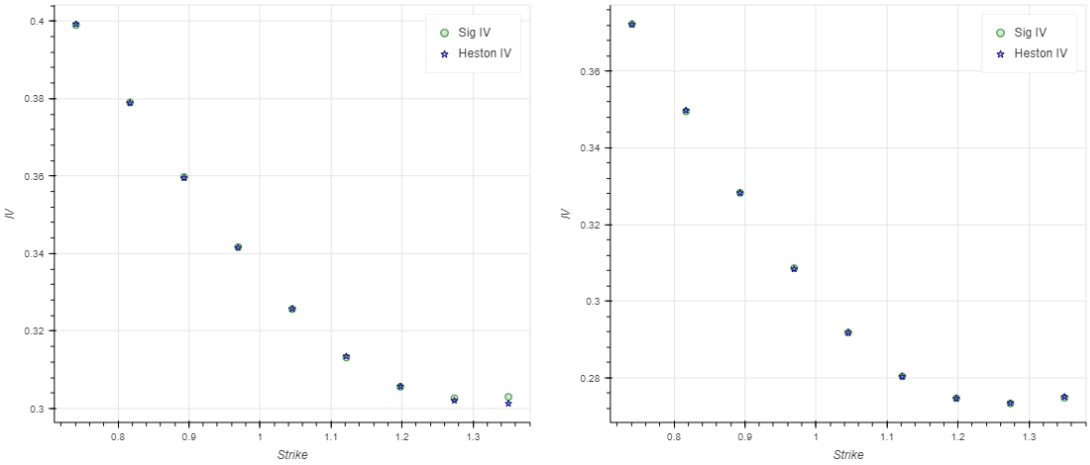


Figure 1.11.: Comparison of implied volatilities for T_1 =3 months and T_2 =1 year.

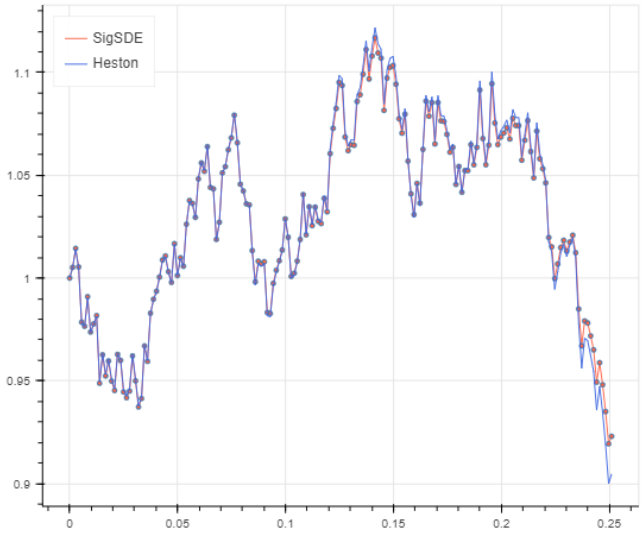


Figure 1.12.: Out-of-sample path calibration performance, with MSE of $1.497 \cdot 10^{-5}$.

2. Joint calibration of SPX and VIX options with signature-based models

The present chapter is based on Cuchiero et al. (2023a), which can be seen as a continuation of the work presented in Chapter 1 and is a joint work with Christa Cuchiero, Janka Möller and Sara Svaluto-Ferro. Before stating our modeling choice we discuss the current state of the art concerning joint calibration to SPX and VIX options.

2.0.1. State of the art

First attempts to solve the joint calibration problem appear in Gatheral (2008), with a double constant elasticity of variance model (CEV), which despite being rather flexible cannot fit accurately the implied volatilities of SPX and VIX options jointly. Later on, the belief that jumps in the SPX (or additionally also in the volatility) are necessary has lead to the following contributions, Sepp (2012); Papanicolaou and Sircar (2014); Baldeaux and Badran (2014); Kokholm and Stisen (2015); Pacati et al. (2018) and more recently Grzelak (2022) with a new perspective on randomization.

Continuous stochastic volatility models based on Markovian semimartingales have also been employed to solve the joint calibration problem. For instance, in Fouque and Saporito (2018) a Heston model with stochastic vol-of-vol has been calibrated, however only for maturities above 4 months where VIX options are less liquid. More recently, Rømer (2022) considered a model where the volatility is driven by two Ornstein-Uhlenbeck (OU) processes using a non-standard transformation function. The well-working choice of two OU-processes illustrated there has been an inspiration for our concrete numerical implementations. We also point out that the (non-rough) model introduced in Abi Jaber et al. (2022a,b), where the volatility is described by a polynomial of order five in one single OU-process, falls (apart from the additional input curve) into this class of continuous Markovian models and is a particular instance of our framework. Let us also refer to the paper by Guyon and Mustapha (2022), where a neural SDE model has been successfully jointly calibrated.

Within the class of continuous, however not necessarily Markovian models, Guyon and Lekeufack (2022) conduct an empirical and statistical analysis as well as a joint calibration for a family of models where the volatility depends on the paths of the asset. These models can be turned into Markovian ones by using exponential kernels instead of general ones.

Two further distinct and rather new lines of research are worth being mentioned as well: first, martingale optimal transport and second rough volatility.

The martingale optimal transport approach (see e.g. Guyon and Henry-Labordere

(2013)) is used to calibrate discrete-time models as proposed in Guyon (2020b, 2021). These models are closely related to Schrödinger bridge problems, where the idea is to calibrate only the drift of the volatility while keeping the volatility of volatility unchanged, see e.g. Henry-Labordere (2019) or Guo et al. (2021, 2022) as well as the references therein regarding an optimal transport approach. Although the calibration within that setting is accurate, it is also computationally rather expensive and not amenable to calibrate to several maturities jointly. These computational challenges have been tackled recently in Guyon and Bourgey (2022) both in discrete and continuous time extending the contributions of Guyon (2020b, 2021).

In the area of rough volatility modeling, initiated by the seminal paper Gatheral et al. (2018), the main idea is to replace the standard Brownian motion in the volatility process by a fractional Brownian motion. Even though the roughness of the trajectories found in Gatheral et al. (2018), can also be explained by the estimation procedure as discussed e.g. in Cont and Das (2023), the non-Markovianity given by the fractional Brownian motion with Hurst parameter $H < 0.5$, manages to reproduce many stylized facts arising in financial time-series. Several classical models have been enhanced with rougher noise, but for simplicity we mention those employed in the SPX/VIX calibration. One example is the quadratic rough Heston model introduced in Gatheral et al. (2020), which was in turn calibrated in Rosenbaum and Zhang (2021) by relying on neural networks approaches, also exploited in e.g. Bayer et al. (2019); Cuchiero et al. (2020). For a large class of rough volatility models Jacquier et al. (2021) give new insights on the joint calibration problem by providing small-time formulas of the at-the-money implied volatility, skew and curvature for both European options on SPX and VIX options. In Rømer (2022) an exhaustive study of the flexibility of different rough volatility models to joint SPX/VIX calibration is carried out, including the rough Bergomi, the rough Heston and an extended rough Bergomi model. Some of these, for instance the rough Heston model, have an affine structure i.e., can be embedded in the class of affine Volterra processes, considered in Abi Jaber et al. (2019); Cuchiero and Teichmann (2019, 2020), which allow for Fourier pricing after solving the related Riccati equations. This underlying structure is the building block of an extension with jumps investigated in Bondi et al. (2022a) and recently employed in the context of the joint calibration in Bondi et al. (2022b), where a rough Heston model with Hawkes-type jumps with Fourier pricing is employed.

2.1. The model

We introduce now the model $(S_t)_{t\geq 0}$ for the discounted dynamics of the S&P 500 index already outlined in the introduction. Its dynamics under a risk-neutral probability measure $\mathbb{Q}$ are given by

$$\mathrm{d}S_t = S_t \sigma_t^S \mathrm{d}B_t, \tag{2.1.1}$$

where $S_0 \in \mathbb{R}^+$, $\sigma^S = (\sigma_t^S)_{t\geq 0}$ is the volatility process to be specified and $B = (B_t)_{t\geq 0}$ is a one-dimensional Brownian motion, correlated with σ^S. We define additionally the instantaneous variance via $V_t := (\sigma_t^S)^2$ for every $t \geq 0$. Our modeling choice is to

parametrize the volatility process σ^S as a linear function of the time-extended signature of a primary underlying process X, namely

$$\sigma_t^S(\ell) := \ell_\emptyset + \sum_{0<|I|\leq n} \ell_I \langle e_I, \widehat{\mathbb{X}}_t \rangle, \tag{2.1.2}$$

where

- $(X_t)_{t\geq 0}$ is a d-dimensional continuous semimartingale with positive quadratic variation and $(\widehat{\mathbb{X}}_t)_{t\geq 0}$ denotes the signature of its time extension $(\widehat{X}_t)_{t\geq 0}$ given by $\widehat{X}_t := (t, X_t)$;
- $\ell := \{\ell_I \in \mathbb{R} : |I| \leq n\}$ denotes the collection of parameters of the model, i.e., $\ell \in \mathbb{R}^{(d+1)_n}$.

Furthermore, we denote by $(Z_t)_{t\geq 0}$ the process given by $Z_t = (X_t, B_t)$, by $(\widehat{Z}_t)_{t\geq 0}$ its time extension, and by $(\widehat{\mathbb{Z}}_t)_{t\geq 0}$ the signature of $(\widehat{Z}_t)_{t\geq 0}$. The correlation matrix process between the components of Z is denoted by

$$\rho_{ij} = \frac{[Z^i, Z^j]}{\sqrt{[Z^i]}\sqrt{[Z^j]}} \in [-1, 1],$$

for all $i, j = 1, \ldots, d+1$, where $[\cdot, \cdot]$ denotes the quadratic variation. Observe that ρ encodes in particular the correlation between X and B. In order to simplify the notation we will drop the dependence on ℓ for the processes $S = (S_t)_{t\geq 0}$ and $(\sigma_t^S)_{t\geq 0}$ as in (2.1.1), whenever this does not cause any confusion.

Remark 2.1.1. As an alternative definition for the volatility process $(\sigma_t^S)_{t\geq 0}$ one can set

$$\sigma_t^S(\ell) := \ell_\emptyset + \sum_{0<|I|\leq n} \ell_I \langle e_I, \widehat{\mathbb{X}}_{t-\varepsilon,t} \rangle,$$

for some fixed $\varepsilon > 0$. In this case the value of the volatility process σ^S at time t does not depend on the whole trajectory of the primary process X, but just on its evolution from $t - \varepsilon$ to t. For an economically reasonable choice for ε the lags used in Section 3.4 of Guyon and Lekeufack (2022) can be adapted to the current setting.

Remark 2.1.2 (Interest rates and dividends)**.** In the model given by (2.1.1) we describe the discounted prices and construct the VIX from them, in line with the definition of the CBOE for the computation of the VIX. However, contingent claims are often expressed in terms of undiscounted prices. If the dynamics of the discounted price process are given by (2.1.1), the undiscounted one fulfills

$$\mathrm{d}\tilde{S}_t = (r - q)\tilde{S}_t \mathrm{d}t + \tilde{S}_t \sigma_t^S(\ell) \mathrm{d}B_t,$$

where here $r, q \in \mathbb{R}$ denote the interest rate and the dividend, respectively. Therefore $\tilde{S}_t(\ell) = e^{(r-q)t} S_t(\ell)$ and the price of a call option on the S&P 500 index under our model, reads

$$C(T, K) = \mathbb{E}\left[e^{-rT}(\tilde{S}_T(\ell) - K)^+\right] = \mathbb{E}[e^{-rT}(e^{(r-q)T} S_T(\ell) - K)^+]$$

where $T > 0$ denotes the maturity time and $K \in \mathbb{R}$ the undiscounted strike price.

We will refer to X as the primary underlying process of the model (2.1.1) as in the previous chapter. Possible choices for the primary process X can be *tractable* processes such as an Ornstein-Uhlenbeck (OU) process or a Brownian motion, where with tractable we mean that their expected signature can be computed easily. To be more precise we state the following assumption.

Assumption 2.1.3. *$\widehat{X} = (t, X_t)_{t\geq 0}$ is a polynomial process in the sense of Definition 1.1.14.*

Remark 2.1.4. If Assumption 2.1.3 is in force then $\widehat{\mathbb{X}}^n$ is a finite-dimensional polynomial process in sense of Filipović and Larsson (2016) and Cuchiero et al. (2012). Hence expected signature of $\widehat{X}$ and conditional expected signature can be found by solving a finite-dimensional ODE, i.e., can be written as a finite-dimensional matrix exponential.

It is worth mentioning that the pool of primary processes that satisfy Assumption 2.1.3 is rather wide, including for example correlated Brownian motions, geometric Brownian motions, OU processes, Cox-Ingersoll-Ross (CIR) processes, Jacobi processes, and all continuous affine processes. If Assumption 2.1.3 is in force then $\widehat{\mathbb{X}}^n$ is a finite-dimensional polynomial process in sense of Filipović and Larsson (2016) and Cuchiero et al. (2012). Hence, the (conditional) expected signature of $\widehat{X}$ can be found by solving a finite-dimensional ODE, i.e., can be written in terms of a matrix exponential. For further details see Section 2.2 below.

Remark 2.1.5. We illustrate here that several classical and also recently considered stochastic volatility models are nested within our modeling choice (2.1.2) under Assumption 2.1.3.

- Suppose that $(X_t)_{t\geq 0}$ is a 1-dimensional OU process and let the order of the signature be $n = 1$, with $\ell_\emptyset = \ell_{(0)} = 0$ and $\ell_{(1)} \neq 0$. Then the process $S = (S_t)_{t\geq 0}$ coincides with the Stein-Stein model, as introduced in Stein and Stein (1991).

- Suppose that $(X_t)_{t\geq 0}$ is a 1-dimensional geometric Brownian motion without drift and let the order of the signature be $n = 1$, with $\ell_\emptyset = \ell_{(0)} = 0$ and $\ell_{(1)} \neq 0$. Then the process $S = (S_t)_{t\geq 0}$ coincides with the SABR model, as introduced in initially in Hagan et al. (2002) with $\beta = 1$.

- Suppose that $(X_t)_{t\geq 0}$ is a 1-dimensional OU process and let the order of the signature be $n = 5$, with $\ell_\emptyset, \ell_{(1)}, \ell_{(1,1,1)}, \ell_{(1,1,1,1,1)}$ non-zero and $\ell_I = 0$ otherwise. Then the process $S = (S_t)_{t\geq 0}$ coincides with the model considered in Abi Jaber et al. (2022a,b) with an exponential kernel (a part from the deterministic input curve considered there additionally). If we allow for $(X_t)_{t\geq 0}$ not to be a semimartingale and we do not consider the time augmentation, we can also include fractional kernels and therefore the whole class of Gaussian polynomial volatility models introduced in Abi Jaber et al. (2022a) within our framework.

2.2. Expected signature of polynomial processes

Let Y be a polynomial process in sense of Definition 1.1.14 whose dynamics are given by

$$\mathrm{d}Y_t = b(Y_t)\mathrm{d}t + \sigma(Y_t)\mathrm{d}W_t, \qquad Y_0 = y_0 \tag{2.2.1}$$

where $\sigma(Y_t)$ denotes the matrix square root of $a(Y_t)$. Recall that in this case the components of $a : \mathbb{R}^d \to \mathbb{S}_+^d$ are polynomials of degree at most 2, the components of $b : \mathbb{R}^d \to \mathbb{R}^d$ are polynomials of degree at most 1, and $W = (W_t)_{t\geq 0}$ is a d-dimensional Brownian motion. Denote then by $\mathbb{Y}$ the corresponding signature.

We now explain how to employ the polynomial technology to compute the conditional expected signature of $(Y_t)_{t\geq 0}$. Several representations of related quantities in particular for the Brownian case can be found in the literature, see for instance Fawcett (2003), Lyons and Victoir (2004), Lyons and Ni (2015), Boedihardjo et al. (2021), Rossi Ferrucci and Cass (2022). Our approach follows Cuchiero et al. (2023b) and is based on the classical theory of polynomial processes (see Cuchiero et al. (2012) and Filipović and Larsson (2016)). Even though results for the corresponding infinite dimensional stochastic processes (see for instance Cuchiero and Svaluto-Ferro (2021); Cuchiero et al. (2021c)) are needed in the case of general signature SDEs considered in Cuchiero et al. (2023b), the polynomial property of $(Y_t)_{t\geq 0}$ here permits to stay in the finite dimensional setting.

Lemma 2.2.1. Let $(Y_t)_{t\geq 0}$ be the polynomial process given by (2.2.1) and b and a be the corresponding drift and diffusion coefficients. Then,

$$b_j(y) = b_j^c + \sum_{k=1}^d b_j^k y_k \qquad \text{and} \qquad a_{ij}(y) = a_{ij}^c + \sum_{k=1}^d a_{ij}^k y_k + \sum_{k,h=1}^d a_{ij}^{kh} y_k y_h,$$

for some $b_j^c, b_j^k, a_{ij}^c, a_{ij}^k, a_{ij}^{kh} = a_{ij}^{hk} \in \mathbb{R}$. Moreover,

$$b_j(Y_t) = \langle \mathbf{b}_j, \mathbb{Y}_t^1 \rangle \qquad \text{and} \qquad a_{ij}(Y_t) = \langle \mathbf{a}_{ij}, \mathbb{Y}_t^2 \rangle$$

for

$$\mathbf{b}_j = \Big(b_j^c + \sum_{k=1}^d b_j^k Y_0^k\Big) e_\emptyset + \sum_{k=1}^d b_j^k e_k \qquad \text{and}$$

$$\mathbf{a}_{ij} = \Big(a_{ij}^c + \sum_{k=1}^d a_{ij}^k Y_0^k + \sum_{k,h=1}^d a_{ij}^{kh} Y_0^k Y_0^h\Big) e_\emptyset + \sum_{k=1}^d \Big(a_{ij}^k + 2\sum_{h=1}^d a_{ij}^{kh} Y_0^h\Big) e_k + \sum_{k,h=1}^d a_{ij}^{kh} e_k \sqcup\!\sqcup e_h.$$

Observe that the upper index on Y_0^k and Y_0^h refers to Y's components and not to powers.

Proof. The first part follows by the observation that by definition of polynomial processes b and a are polynomials of degree at most 1 and 2, respectively. For the second part it then suffices to note that $\langle e_\emptyset, \mathbb{Y}_t^1 \rangle = \langle e_\emptyset, \mathbb{Y}_t^2 \rangle = 1$, $\langle e_k, \mathbb{Y}_t^1 \rangle = \langle e_k, \mathbb{Y}_t^2 \rangle = (Y_t^k - Y_0^k)$, and $\langle e_k \sqcup\!\sqcup e_h, \mathbb{Y}_t^2 \rangle = (Y_t^k - Y_0^k)(Y_t^h - Y_0^h)$. □

Lemma 2.2.2. Let $(Y_t)_{t\geq 0}$ be the polynomial process given by (2.2.1) and let $\mathbf{b}$ and $\mathbf{a}$ as in Lemma 2.2.1. The truncated signature $(\mathbb{Y}^n_t)_{t\geq 0}$ is a polynomial process in the sense of Definition 1.1.14 and for each $|I| \leq n$ it holds that

$$\langle e_I, \mathbb{Y}^n_t\rangle = \int_0^t \langle Le_I, \mathbb{Y}^n_s\rangle \mathrm{d}s + \int_0^t \langle e_{I'}, \mathbb{Y}^n_s\rangle \sigma_{i_{|I|}}(Y_s)\mathrm{d}W_s,$$

where the operator $L : T((\mathbb{R}^d)) \to T((\mathbb{R}^d))$ satisfies $L(T^{(n)}(\mathbb{R}^d)) \subseteq T^{(n)}(\mathbb{R}^d)$ and is given by

$$Le_I = e_{I'} \sqcup\!\sqcup \mathbf{b}_{i_{|I|}} + \frac{1}{2} e_{I''} \sqcup\!\sqcup \mathbf{a}_{i_{|I|-1}i_{|I|}}. \tag{2.2.2}$$

Proof. Let $\sigma_j(Y_t)$ denote the j-th row of $\sigma(Y_t)$. By definition of the signature, Stratonovich integral and by the shuffle property we can compute

$$\begin{aligned}
\langle e_I, \mathbb{Y}_t\rangle &= \int_0^t \langle e_{I'}, \mathbb{Y}_s\rangle \circ \mathrm{d}\langle e_{i_{|I|}}, Y_s\rangle \\
&= \int_0^t \langle e_{I'}, \mathbb{Y}_s\rangle \mathrm{d}\langle e_{i_{|I|}}, Y_s\rangle + \frac{1}{2}\int_0^t \langle e_{I''}, \mathbb{Y}_s\rangle \mathrm{d}[\langle e_{i_{|I|-1}}, Y_s\rangle, \langle e_{i_{|I|}}, Y_s\rangle]_s \\
&= \int_0^t \langle e_{I'}, \mathbb{Y}_s\rangle \langle \mathbf{b}_{i_{|I|}}, \mathbb{Y}_s\rangle \mathrm{d}s + \int_0^t \langle e_{I'}, \mathbb{Y}_s\rangle \sigma_{i_{|I|}}(Y_s)\mathrm{d}W_s \\
&\quad + \frac{1}{2}\int_0^t \langle e_{I''}, \mathbb{Y}_s\rangle \langle \mathbf{a}_{i_{|I|-1}}, \mathbb{Y}_s\rangle \mathrm{d}s \\
&= \int_0^t \langle e_{I'} \sqcup\!\sqcup \mathbf{b}_{i_{|I|}}, \mathbb{Y}_s\rangle \mathrm{d}s + \int_0^t \langle e_{I'}, \mathbb{Y}_s\rangle \sigma_{i_{|I|}}(Y_s)\mathrm{d}W_s + \frac{1}{2}\int_0^t \langle e_{I''} \sqcup\!\sqcup \mathbf{a}_{i_{|I|-1}i_{|I|}}, \mathbb{Y}_s\rangle \mathrm{d}s \\
&= \int_0^t \langle Le_I, \mathbb{Y}_s\rangle \mathrm{d}s + \int_0^t \langle e_{I'}, \mathbb{Y}_s\rangle \sigma_{i_{|I|}}(Y_s)\mathrm{d}W_s,
\end{aligned}$$

for each $|I| \geq 0$. Since $|I \sqcup\!\sqcup J| = |I| + |J|$ it holds that $L(T^{(n)}(\mathbb{R}^d)) \subseteq T^{(n)}(\mathbb{R}^d)$. For $|I| \leq n$ we thus get that the corresponding drift's components are linear maps in $\mathbb{Y}^n$. Similarly, since $\mathbf{a}_{ij} = \sum_{|I|,|J|\leq 1} \lambda^{IJ}_{ij} e_I \sqcup\!\sqcup e_J$ for some $\lambda^{IJ}_{ij} \in \mathbb{R}$ and for $|I| \leq n$ we can compute

$$\begin{aligned}
\langle e_{I'}, \mathbb{Y}_s\rangle \sigma_{i_{|I|}}(Y_t)\big(\langle e_{J'}, \mathbb{Y}_s\rangle \sigma_{i_{|J|}}(Y_t)\big)^\top &= \langle e_{I'}, \mathbb{Y}_s\rangle \langle e_{J'}, \mathbb{Y}_s\rangle \langle \mathbf{a}_{i_{|I|}j_{|J|}}, \mathbb{Y}_s\rangle \\
&= \sum_{|H_1|,|H_2|\leq 1} \lambda^{H_1H_2}_{i_{|I|}j_{|J|}} \langle e_{I'} \sqcup\!\sqcup e_{H_1}, \mathbb{Y}_s\rangle \langle e_{J'} \sqcup\!\sqcup e_{H_2}, \mathbb{Y}_s\rangle,
\end{aligned}$$

we also have that the components of the corresponding diffusion matrix are polynomials of degree 2 in $\mathbb{Y}^n$. Lemma 2.2 in Filipović and Larsson (2016) yields the polynomial property. □

Since the linear operator L maps the finite dimensional vector space $T^{(n)}(\mathbb{R}^d)$ to itself, it admits a matrix representation.

Definition 2.2.3. We call the operator L defined in (2.2.2) *dual operator corresponding to* $\mathbb{Y}$. For each $|I| \leq n$ set then $\eta_{IJ} \in \mathbb{R}$ such that

$$Le_I = \sum_{|J| \leq n} \eta_{IJ} e_J,$$

and fix a labelling injective function $\mathscr{L} : \{I \colon |I| \leq n\} \to \{1, \dots, d_n\}$ as introduced before (1.1.6). We then call the matrix $G \in \mathbb{R}^{d_n \times d_n}$ given by

$$G_{\mathscr{L}(I)\mathscr{L}(J)} := \eta_{IJ}, \tag{2.2.3}$$

the d_n*-dimensional matrix representative of* L.

Observe that using the notation of (1.1.6), for each $\mathbf{u} \in T^{(n)}(\mathbb{R}^d)$ the matrix representative G of L satisfies

$$\mathbf{vec}(L\mathbf{u}) = G\mathbf{vec}(\mathbf{u}).$$

Theorem 2.2.4. *Let $(Y_t)_{t\geq 0}$ be the polynomial process given by (2.2.1), $(\mathcal{F}_t)_{t\geq 0}$ be the filtration generated by $(Y_t)_{t\geq 0}$ and let G be the d_n-dimensional matrix representative of the dual operator corresponding to $\mathbb{Y}$. Then for each $T, t \geq 0$ and each $|I| \leq n$ it holds*

$$\mathbb{E}[\boldsymbol{vec}(\mathbb{Y}^n_{T+t})|\mathcal{F}_T] = e^{tG^\top} \boldsymbol{vec}(\mathbb{Y}^n_T),$$

or equivalently,

$$\mathbb{E}[\langle e_I, \mathbb{Y}^n_{T+t}\rangle|\mathcal{F}_T] = \sum_{|J|\leq n} (e^{tG^\top})_{\mathscr{L}(I)\mathscr{L}(J)} \langle e_J, \mathbb{Y}^n_T\rangle,$$

where $e^{(\cdot)}$ denotes the matrix exponential.

Proof. By Lemma 2.2.2 we know that $\mathbf{vec}(\mathbb{Y}^n)$ is a polynomial process and Theorem 3.1 in Filipović and Larsson (2016) for polynomials of degree 1 yields the claim. □

Example 2.2.5.

$$\mathrm{d}X^j_t = \kappa^j(\theta^j - X^j_t)\mathrm{d}t + \sqrt{a(X_t)}\mathrm{d}W_t, \qquad X_0 = x_0,$$

for $a_{ij}(X_t) = \sigma^i\sigma^j\rho_{ij}$, and W being a d-dimensional Brownian motion. We denote by $\rho_{j(d+1)}$ the correlation between X^j and B. Setting $\kappa^{d+1} := 0$ and $\sigma^{d+1} := 1$ we can see that $\widehat{Z}$ satisfies (2.2.1) in $d+2$ dimensions for

$$b_j(\widehat{Z}_t) = 1_{\{j=0\}} + \kappa^j(\theta^j - \widehat{Z}^j_t)1_{\{j\neq 0\}} \qquad \text{and} \qquad a_{ij}(\widehat{Z}_t) = \sigma^i\sigma^j\rho_{ij}1_{\{i,j\neq 0\}}.$$

The corresponding $\mathbf{b}$ and $\mathbf{a}$ are given by $\mathbf{b}_j = e_\emptyset(1_{\{j=0\}} + \kappa^j(\theta^j - \widehat{Z}^j_0)1_{\{j\neq 0\}}) - e_j\kappa^j 1_{\{j\neq 0\}}$ and $\mathbf{a}_{ij} = e_\emptyset\sigma^i\sigma^j\rho_{ij}1_{\{i,j\neq 0\}}$ and we thus get

$$\begin{aligned} Le_I = {} & e_{I'}(1_{\{i_{|I|}=0\}} + \kappa^{i_{|I|}}(\theta^{i_{|I|}} - \widehat{Z}^{i_{|I|}}_0)1_{\{i_{|I|}\neq 0\}}) - (e_{I'} \sqcup\!\sqcup e_{i_{|I|}})\kappa^{i_{|I|}}1_{\{i_{|I|}\neq 0\}} \\ & + \frac{1}{2}e_{I''}\sigma^{i_{|I|-1}}\sigma^{i_{|I|}}\rho_{i_{|I|-1}i_{|I|}}1_{\{i_{|I|-1},i_{|I|}\neq 0\}}. \end{aligned}$$

An application of L to the first basis elements yields the following results:

- $L(e_1) = e_\emptyset \kappa^1(\theta^1 - X_0^1) - e_1\kappa^1$;
- $L(e_I \otimes e_0) = e_I \sqcup\!\sqcup \mathbf{b}_0 + e_{I'} \sqcup\!\sqcup \mathbf{a}_{i_{|I|}0} = e_I$;
- $L(e_0 \otimes e_1 \otimes e_2) = e_0 \otimes e_1\kappa^2(\theta^2 - X_0^2) - (e_0 \otimes e_1) \sqcup\!\sqcup e_2\kappa^2 + \frac{1}{2}e_0\sigma^1\sigma^2\rho_{12}$.

Letting $(\mathcal{F}_t)_{t\geq 0}$ be the filtration generated by $(\widehat{Z}_t)_{t\geq 0}$ by Theorem 2.2.4 we can conclude that

$$\mathbb{E}[\mathbf{vec}(\widehat{\mathbb{Z}}^n_{T+t})|\mathcal{F}_T] = e^{tG^\top}\mathbf{vec}(\widehat{\mathbb{Z}}^n_T), \tag{2.2.4}$$

or equivalently,

$$\mathbb{E}[\langle e_I, \widehat{\mathbb{Z}}^n_{T+t}\rangle|\mathcal{F}_T] = \sum_{|J|\leq n} (e^{tG^\top})_{\mathscr{L}(I)\mathscr{L}(J)}\langle e_J, \widehat{\mathbb{Z}}^n_T\rangle, \tag{2.2.5}$$

where G denotes the $(d+2)_n$-dimensional matrix representative of L. In order to work with the VIX it will be convenient to restrict our attention to the signature components of $(\widehat{\mathbb{Z}}_t)_{t\geq 0}$ not involving B. The following remark will be useful.

Remark 2.2.6. Observe that given a subset $E \subseteq \{0, \dots, d+1\}$, setting $\mathcal{I}_E := \{I \colon i_j \in E\}$ it holds $L(\mathcal{I}_E) \subseteq \mathcal{I}_E$. This in particular implies that

$$Le_I = \sum_{I\in\mathcal{I}_E} \eta_{IJ}e_J$$

for each $I \in \mathcal{I}_E$. Choosing $E = \{0, \dots, d\}$, letting $\mathscr{L}_E : \mathcal{I}_E \to \{1, \dots, (d+1)_n\}$ be a labelling injective function, and setting $G^E_{\mathscr{L}_E(I)\mathscr{L}_E(J)} := \eta_{IJ}$ we can see that (2.2.5) reduces to

$$\mathbb{E}[\langle e_I, \widehat{\mathbb{X}}^n_{T+t}\rangle|\mathcal{F}_T] = \sum_{|J|\leq n} (e^{t(G^E)^\top})_{\mathscr{L}_E(I)\mathscr{L}_E(J)}\langle e_J, \widehat{\mathbb{X}}^n_T\rangle.$$

To simplify the notation we often drop the E from G^E whenever this does not introduce any confusion.

In the next section we discuss the implication on pricing VIX options under (2.1.2) and Assumption (2.1.3) while the implications of these on the log-price will be investigated in Section 2.4.

Remark 2.2.7. Let $(Y_t)_{t\geq 0}$ be a polynomial process and fix $\mathbb{Y}^{-1}$ such that $e_\emptyset = \mathbb{Y}^{-1}_s \otimes \mathbb{Y}^{-1}_s$, i.e. $\langle e_\emptyset, \mathbb{Y}^{-1}_s\rangle = 1$ and

$$\sum_{e_{I_1}\otimes e_{I_2} = e_I} \langle e_{I_1}, \mathbb{Y}^{-1}_s\rangle\langle e_{I_2}, \mathbb{Y}_s\rangle = 0,$$

for each $|I| > 0$. Observe that it can be defined recursively on $|I|$ and each component of $\mathbb{Y}^{-1}_s$ corresponds to a linear combination of components of $\mathbb{Y}_s$ of the same length or shorter. Since by Chen's identity (see Lemma 1.1.10) we have $\mathbb{Y}_s \otimes \mathbb{Y}_{s,t} = \mathbb{Y}_t$, for each $s \leq u \leq t$ and $|I| \leq n$ we then get

$$\begin{aligned}\mathbb{E}[\langle e_I, \mathbb{Y}_{s,t}\rangle|\mathcal{F}_u] = \mathbb{E}[\langle e_I, \mathbb{Y}^{-1}_s \otimes \mathbb{Y}_t\rangle|\mathcal{F}_u] &= \sum_{e_{I_1}\otimes e_{I_2} = e_I} \langle e_{I_1}, \mathbb{Y}^{-1}_s\rangle\mathbb{E}[\langle e_{I_2}, \mathbb{Y}_t\rangle|\mathcal{F}_u] \\ &= \sum_{e_{I_1}\otimes e_{I_2} = e_I} \langle e_{I_1}, \mathbb{Y}^{-1}_s\rangle\mathbf{vec}(e_{I_2})^\top e^{(t-u)G^\top}\mathbf{vec}(\mathbb{Y}^n_u),\end{aligned}$$

where G denotes the d_n-dimensional matrix representative of the dual operator of $\mathbb{Y}$.

2.3. VIX options with signatures

In this section we discuss the implication on pricing VIX options under the model (2.1.1)-(2.1.2) when Assumption 2.1.3 is in force. The implications of these on the log-price will be investigated in Section 2.4.

The CBOE Volatility Index, known by its ticker symbol VIX, is a popular measure of the market's expected volatility of the S&P 500 index, calculated and published by the Chicago Board Options Exchange (CBOE). The current VIX index value quotes the expected annualized change in the S&P 500 index over the following 30 days, based on options-based theory and current options-market data, more precisely

$$\mathrm{VIX}_T := \sqrt{\mathbb{E}\left[-\frac{2}{\Delta}\log\left(\frac{S_{T+\Delta}}{S_T}\right)|\mathcal{F}_T\right]}, \tag{2.3.1}$$

where $\Delta = 30$ days and S_T denotes the price process at time $T > 0$. With the term VIX options we here usually refer to either put or calls written on VIX. In the present work we will take into account without loss of generality only call options.

2.3.1. Explicit formulas for the VIX

This section is dedicated to one of the main implication of our modeling framework, namely an explicit formula for the VIX expression (2.3.1) for S following (2.1.1)-(2.1.2) under Assumption 2.1.3. In particular we show in the next theorem that the computation of the VIX squared reduces to a quadratic form in the parameters ℓ. The entries of the corresponding positive semidefinite matrix can be computed by polynomial technology, i.e. by matrix exponential as proved in Section 2.2.

Theorem 2.3.1. *Let $S = (S_t)_{t\geq 0}$ be a price process described by*

$$\mathrm{d}S_t = S_t\sigma_t^S \mathrm{d}B_t,$$

where $\sigma^S = (\sigma_t^S)_{t\geq 0}$ denotes the volatility process, $B = (B_t)_{t\geq 0}$ a one-dimensional Brownian motion.

(i) Assume that $V = (\sigma^S)^2$ satisfies

$$\mathbb{E}\left[\int_0^T V_s \mathrm{d}s\right] < \infty. \tag{2.3.2}$$

Then the VIX index at $T > 0$ is given by

$$VIX_T = \sqrt{\frac{1}{\Delta}\mathbb{E}\left[\int_T^{T+\Delta} V_t \mathrm{d}t|\mathcal{F}_T\right]}. \tag{2.3.3}$$

(ii) Assume that σ^S and X satisfy (2.1.2) and Assumption 2.1.3, respectively. Fix an injective labeling function $\mathscr{L} : \{I : |I| \leq n\} \to \{1, \dots, (d+1)_{2n+1}\}$ and let G be the $(d+1)_{(2n+1)}$-dimensional matrix representative of the dual operator corresponding to $\widehat{\mathbb{X}}$. Then (2.3.2) is satisfied and

$$VIX_T(\ell) = \sqrt{\frac{1}{\Delta}\ell^\top Q(T,\Delta)\ell}, \tag{2.3.4}$$

where

$$Q_{\mathscr{L}(I)\mathscr{L}(J)}(T,\Delta) = \boldsymbol{vec}((e_I \sqcup\!\sqcup e_J) \otimes e_0)^\top (e^{\Delta G^\top} - \mathrm{Id})\boldsymbol{vec}(\widehat{\mathbb{X}}_T^{2n+1}), \tag{2.3.5}$$

and $\mathrm{Id} \in \mathbb{R}^{(d+1)_{2n+1} \times (d+1)_{2n+1}}$ denotes the identity matrix. More explicitly without the vectorisation this reads

$$Q_{\mathscr{L}(I)\mathscr{L}(J)}(T,\Delta) = \sum_{e_K = (e_I \sqcup\!\sqcup e_J)\otimes e_0} \sum_{|H| \leq 2n+1} (e^{\Delta G^\top} - \mathrm{Id})_{\mathscr{L}(K)\mathscr{L}(H)} \langle e_H, \widehat{\mathbb{X}}_T \rangle.$$

Proof. Part (i) follows directly from an application of Itô's formula. Indeed, for any $t, T \geq 0$

$$\log(S_t) = \log(S_0) - \frac{1}{2}\int_0^t V_s \mathrm{d}s + \int_0^t \sigma_s^S \mathrm{d}B_s,$$

and hence

$$-\frac{2}{\Delta}\log\left(\frac{S_{T+\Delta}}{S_T}\right) = -\frac{2}{\Delta}(\log(S_{T+\Delta}) - \log(S_T)) = \frac{1}{\Delta}\int_T^{T+\Delta} V_s \mathrm{d}s - \frac{2}{\Delta}\int_T^{T+\Delta} \sigma_s^S \mathrm{d}B_s,$$

where the integral with respect to the Brownian motion $B = (B_t)_{t\geq 0}$ vanishes once we take the risk neutral conditional expectation, due to (2.3.2). This yields the alternative expression for VIX_T^2 and thus for VIX_T.

For part (ii), observe that

$$V_t(\ell) = \left(\sum_{|I|\leq n} \ell_I \langle e_I, \widehat{\mathbb{X}}_t \rangle\right)^2 = \sum_{|I|,|J|\leq n} \ell_I \ell_J \langle e_I \sqcup\!\sqcup e_J, \widehat{\mathbb{X}}_t \rangle.$$

Since continuous polynomials processes have finite moments of every degree and (2.3.2) is satisfied due to Lemma 1.1.9. The expression for $V_t(\ell)$ yields then

$$\begin{aligned}\mathrm{VIX}_T^2(\ell) &= \frac{1}{\Delta}\sum_{|I|,|J|\leq n} \ell_I \ell_J \mathbb{E}\left[\int_T^{T+\Delta} \langle e_I \sqcup\!\sqcup e_I, \widehat{\mathbb{X}}_t \rangle \mathrm{d}t | \mathcal{F}_T\right] \\ &= \frac{1}{\Delta}\ell^\top Q(T,\Delta)\ell,\end{aligned}$$

where for each $T > 0$ the matrix Q is given by

$$\begin{aligned} Q_{\mathscr{L}(I)\mathscr{L}(J)}(T,\Delta) :&= \mathbb{E}\left[\int_T^{T+\Delta} \langle e_I \sqcup\!\sqcup e_J, \widehat{\mathbb{X}}_t\rangle \mathrm{d}t | \mathcal{F}_T\right] \\ &= \mathbb{E}\left[\int_0^{T+\Delta} \langle e_I \sqcup\!\sqcup e_J, \widehat{\mathbb{X}}_t\rangle \mathrm{d}t - \int_0^{T} \langle e_I \sqcup\!\sqcup e_J, \widehat{\mathbb{X}}_t\rangle \mathrm{d}t | \mathcal{F}_T\right] \\ &= \mathbb{E}\big[\langle (e_I \sqcup\!\sqcup e_J)\otimes e_0, \widehat{\mathbb{X}}_{T+\Delta}\rangle - \langle (e_I \sqcup\!\sqcup e_J)\otimes e_0, \widehat{\mathbb{X}}_{T}\rangle | \mathcal{F}_T\big] \\ &= \mathbb{E}\big[\langle (e_I \sqcup\!\sqcup e_J)\otimes e_0, \widehat{\mathbb{X}}_{T+\Delta}\rangle | \mathcal{F}_T\big] - \langle (e_I \sqcup\!\sqcup e_J)\otimes e_0, \widehat{\mathbb{X}}_{T}\rangle. \end{aligned}$$

By Theorem 2.2.4 we can rewrite the matrix Q as

$$\begin{aligned} Q_{\mathscr{L}(I)\mathscr{L}(J)}(T,\Delta) &= \mathbf{vec}((e_I \sqcup\!\sqcup e_J)\otimes e_0)^\top e^{\Delta G^\top} \mathbf{vec}(\widehat{\mathbb{X}}_T^{2n+1}) \\ &\quad - \mathbf{vec}((e_I \sqcup\!\sqcup e_J)\otimes e_0)^\top \mathbf{vec}(\widehat{\mathbb{X}}_T^{2n+1}) \\ &= \mathbf{vec}((e_I \sqcup\!\sqcup e_J)\otimes e_0)^\top (e^{\Delta G^\top} - \mathrm{Id}) \mathbf{vec}(\widehat{\mathbb{X}}_T^{2n+1}), \end{aligned}$$

and the claim follows. □

Remark 2.3.2. Consider now the model described in Remark 2.1.1 and set for simplicity $\varepsilon \geq \Delta$. Then the results of Theorem 2.3.1(ii) still hold however with

$$Q_{\mathscr{L}(I)\mathscr{L}(J)}(T,\Delta) = \sum_{e_{I_1}\otimes e_{I_2} = e_I \sqcup\!\sqcup e_J} \int_T^{T+\Delta} \langle e_{I_1}, \widehat{\mathbb{X}}_{t-\varepsilon}^{-1}\rangle \mathbf{vec}(e_{I_2})^\top e^{(t-T)G^\top} \mathbf{vec}(\widehat{\mathbb{X}}_T) \mathrm{d}t,$$

where G denotes the $(d+1)_{2n+1}$-dimensional matrix representative of the dual operator corresponding to $\widehat{\mathbb{X}}$. To adapt the proof we just need to note that for each $t \in [T, T+\Delta]$ Remark 2.2.7 yields

$$\mathbb{E}[\langle e_I \sqcup\!\sqcup e_J, \widehat{\mathbb{X}}_{t-\varepsilon,t}\rangle | \mathcal{F}_T] = \sum_{e_{I_1}\otimes e_{I_2} = e_I \sqcup\!\sqcup e_J} \langle e_{I_1}, \widehat{\mathbb{X}}_{t-\varepsilon}^{-1}\rangle \mathbf{vec}(e_{I_2})^\top e^{(t-T)G^\top} \mathbf{vec}(\widehat{\mathbb{X}}_T).$$

Note that since the integration's variable t appears twice in this expression the time integral cannot be incorporated in the signature.

Remark 2.3.3. Observe that accounting for the scaling factor of 100, conventionally introduced by CBOE, the VIX index squared can equivalently be redefined (see e.g., Rosenbaum and Zhang (2021); Rømer (2022)) as

$$\mathrm{VIX}_T^2 := \frac{100^2}{\Delta} \mathbb{E}\left[\int_T^{T+\Delta} V_t \mathrm{d}t | \mathcal{F}_T\right], \tag{2.3.6}$$

where $T, t > 0$ and $\Delta = \frac{1}{12}$, i.e., approximately 30 days. Notice that since the expressions (2.3.3) and (2.3.6) differ only by a scaling factor, all the theoretical results of the present work hold true disregarding this scaling. For sake of simplicity we will always use (2.3.3). The motivation of the scaling is that, if on one side volatility is quoted usually in percent,

the VIX is not. For instance if the VIX index is 25, it means that the hypothetical SPX option with 30 days to expiration has annualized implied volatility of 25%. For this reason the scaling of 100 is also called annualization factor. We address the reader to Chapter 11 in Gatheral (2011) for further details about the conventions of CBOE and its link with (2.3.1).

We observe that the expression (2.3.5) is computationally appealing as we can unpack the computation in three parts: compute the coordinate vector $\mathbf{vec}((e_I \sqcup\!\sqcup e_J) \otimes e_0)$, which depends just on $d > 0$ and $n > 0$, calculate the matrix exponential of $G^\top$ which depends on the choice of the primary process X, and finally sample $\widehat{\mathbb{X}}_T^{2n+1}$ which is the only part that depends on the chosen maturity time T.

Remark 2.3.4. In general the computation of $G \in \mathbb{R}^{(d+1)_{2n+1} \times (d+1)_{2n+1}}$, even if done only once, can be costly. For this reason it can sometimes be interesting to avoid the last time integral and to consider the following equivalent expression of the matrix Q, for $|I|, |J| \leq n$:

$$Q_{\mathscr{L}(I)\mathscr{L}(J)}(T, \Delta) = \left(\int_T^{T+\Delta} \mathbf{vec}(e_I \sqcup\!\sqcup e_J)^\top e^{(t-T)G^\top} \mathrm{d}t \right) \mathbf{vec}(\widehat{\mathbb{X}}_T^{2n}), \tag{2.3.7}$$

where now $G \in \mathbb{R}^{(d+1)_{2n} \times (d+1)_{2n}}$ and where we use the fact that we can interchange the conditional expectation with the time-integral by dominated convergence. As G is singular, this time integral has to be computed numerically, in general. We propose here two possible methods that can be used in order to compute it efficiently.

(i) **Approximation of the time integral**: e.g., via the trapezoidal rule also applied for VIX^2 in Bourgey and De Marco (2021). Hence if we consider the shuffled coordinates $\mathbf{vec}(e_I \sqcup\!\sqcup e_J)$ of the exponential matrix we can use the symmetry of the shuffle to reduce the number of integrals to be solved from $((d+1)_{2n})^2$ to $\frac{(d+1)_n((d+1)_n+1)}{2} \cdot (d+1)_{2n}$, instead of $(d_{2n})^2$. Observe that for our integral the error of such an approximation is given by

$$\mathrm{Err}(N) = -\frac{\Delta^2}{12N^2} G^\top (e^{G^\top \Delta} - I) + \mathcal{O}(N^{-3}),$$

as $N \to +\infty$. As a further dimension reduction one can exploit the polynomial nature of $\widehat{\mathbb{X}}^n$ to obtain a matrix representation of its second order moments. Without entering into details, the matrix G would then be the matrix corresponding to the linear operator acting on coefficients polynomials of degree 2 in $\widehat{\mathbb{X}}^n$. Its dimension would thus be $\frac{(d+1)_n((d+1)_n+1)}{2}$..

(ii) **Approximation of the matrix exponential**: we can avoid to approximate the integral by approximating the matrix exponential. Assuming that

$$\lim_{N \to +\infty} (G^\top \Delta)^N = 0, \tag{2.3.8}$$

this can for instance be done via its Taylor expansion:

$$\int_T^{T+\Delta} e^{tG^\top} \mathrm{d}t = \Delta \left(I + \frac{G^\top \Delta}{2!} + \cdots + \frac{(G^\top \Delta)^N}{N+1!} + \mathcal{O}((G^\top \Delta)^{N+1}) \right).$$

Observe that (2.3.8) holds true whenever the spectral radius, i.e., the maximal eigenvalue in absolute value, of the matrix $G^\top \Delta$ is less than 1 (see for instance Theorem 1.5 in Quarteroni et al. (2010)). This requirement suggests that for numerical purposes the parameters of the primary underlying process have to be chosen accordingly.

An interesting example is given by the case where X is a d-dimensional correlated Brownian motion, as considered for instance in Cuchiero et al. (2022b). In this case the process has no linear drift and the corresponding matrix G is nilpotent, meaning that $G^n = 0$, for each n big enough.

In general, this Taylor approach permits to avoid a numerical integration and produces an accurate approximation, allocating as few memory as possible.

Remark 2.3.5. A further step in the direction of a fast evaluation of $\mathrm{VIX}_T(\ell)$ can be taken by noticing that the matrix Q in (2.3.5) admits a Cholesky decomposition. Indeed since Q is positive semidefinite and symmetric by the shuffle property, we know that there exists an upper triangular matrix $U_T \in \mathbb{R}^{(d+1)_n \times (d+1)_n}$, with possible zero elements on the diagonal, such that

$$Q(T, \Delta) = U_T U_T^\top,$$

where for sake of simplicity we drop the dependence on Δ of U_T. Hence the evaluation of the $\mathrm{VIX}_T(\ell)$ reduces to

$$\mathrm{VIX}_T(\ell) = \sqrt{\frac{1}{\Delta} \ell^\top U_T U_T^\top \ell} = \frac{1}{\sqrt{\Delta}} \sqrt{(U_T^\top \ell)^2} = \frac{1}{\sqrt{\Delta}} \| U_T^\top \ell \|,$$

where here $\| \cdot \|$ denotes the Euclidean norm. We stress the fact that the Cholesky decomposition can be carried out offline, and the computational benefit is immediate if several samples of the signature are considered.

In the following remark we discuss a possible dimension reduction technique from which one can benefit computationally. We follow the approach of Cuchiero et al. (2021a,b), where by employing the Johnson-Lindenstrauss Lemma a random projection of the signature is considered, to which we refer as a *randomized signature*. A first way to use this tool is the following.

Remark 2.3.6. Let $d_< \in \mathbb{N}$ be the dimension of the space to which we would like to project the signature of order $n > 0$, such that $d_< \ll (d+1)_n$. Consider $A = (\alpha_{ij}) \in \mathbb{R}^{d_< \times (d+1)_n}$, such that $\alpha_{ij} \sim \mathcal{N}(0, 1/d_<)$. Then a possible way to employ the randomised signature is to parametrize the volatility process as follows,

$$\sigma_t^S(\ell) := \tilde{\ell}^\top A \cdot \mathbf{vec}(\widehat{\mathbb{X}}_t^n)$$

where with $\tilde{\ell} = \ell \cdot A^\top \in \mathbb{R}^{d_<}$ we denote the randomised parameters. Due to the linearity of integral and conditional expectation in (2.3.3) this modeling choice is equivalent to consider the randomised matrix $\widetilde{Q} \in \mathbb{R}^{d_< \times d_<}$ given by

$$\widetilde{Q}_{\mathscr{L}(I)\mathscr{L}(J)}(T, \Delta) := AQ_{\mathscr{L}(I)\mathscr{L}(J)}(T, \Delta)A^\top,$$

which leads to the following representation of $\mathrm{VIX}_T(\ell)$:

$$\mathrm{VIX}_T(\ell) = \sqrt{\frac{1}{\Delta}\tilde{\ell}^\top \tilde{Q}(T, \Delta)\tilde{\ell}}.$$

Observe that even if this procedure does not reduce the number iterated integrals to be computed offline, it reduces the number of parameters to calibrate, yielding in general to a faster evaluation of $\mathrm{VIX}_T(\ell)$.

2.3.2. Options on VIX

We here briefly describe some important aspect of options on VIX. First of all, note that VIX options are written on VIX-futures. The price process of a VIX-future contract with maturity $T > 0$, is given by

$$F_t(T) := \mathbb{E}\left[\mathrm{VIX}_T | \mathcal{F}_t\right]. \tag{2.3.9}$$

This has the following implications:

- The price of a VIX-option depends on the maturity of the corresponding VIX-future.
- In calibration procedures one should therefore not normalize, i.e. not use $\overline{\mathrm{VIX}}_t := \frac{\mathrm{VIX}_t}{\mathbb{E}[\mathrm{VIX}_T]}$, as $\mathbb{E}[\overline{\mathrm{VIX}}_T] = 1$.
- When calibrating to VIX options, we stress that we additionally calibrate to VIX futures' prices, see Section 2.3.4. This is important since future prices under the calibrated model are employed to compute its implied volatility surface. Including VIX futures in the calibration leads to a consistent model, both for VIX options and VIX futures, see e.g. Pacati et al. (2018); Guyon (2020a, 2021); Guo et al. (2022). Using market prices of the VIX futures to invert the implied volatility surface could lead to inconsistencies if one would like to price further derivatives with the calibrated model.

In this respect, let us here also comment on the computation of implied volatilities for VIX call options. Two equivalent approaches can be used to compute implied volatilities for a given maturity $T > 0$:

- use $F_0(T)$ in the so-called Black formula for options on futures (see Section 2.2 of Papanicolaou and Sircar (2014)).
- Consider $e^{-rT}F_0(T)$, with $r > 0$ the interest rate, in the Black-Scholes formula.

Recall that the Black formula coincides with Black-Scholes' one when choosing the underlying in the Black-Scholes formula to be $e^{-r(T-t)}F_t(T)$. This implies that the two approaches are equivalent, i.e. lead to the same implied volatilities. In Section 2.3.4 we will only consider futures' prices at time $t = 0$ and maturity time $T > 0$, hence for sake of simplicity we use the notation $F(T)$ instead of $F_0(T)$.

2.3.3. Variance reduction for pricing VIX options

We here discuss variance reduction techniques (see e.g. Glasserman (2004)) that can speed up the calibration in the subsequently applied Monte Carlo approach further. The key idea is to introduce a control variate, namely an easy to evaluate random variable Φ^{cv} such that given $T > 0$ and $K > 0$,

$$\mathbb{E}[\Phi^{cv}] = 0, \qquad \mathrm{Var}\big((\mathrm{VIX}_T(\ell) - K)^+ - \Phi^{cv}\big) < \mathrm{Var}\big((\mathrm{VIX}_T(\ell) - K)^+\big).$$

An example of control variates used for pricing and calibrating neural SDE models can be found in Cuchiero et al. (2020); Gierjatowicz et al. (2020), where Φ^{cv} is constructed from hedging strategies.

In the following we describe two possible choices of control variates, which consist of polynomials on VIX futures. We stress the fact that these can be seen as linear functions of the signature of the primary process $\widehat{X}$, hence they belong to the class of sig-payoffs, see Lyons et al. (2020); Perez Arribas et al. (2020) and Section 1.4.2.

- The first example is to employ the VIX squared as main ingredient, see for instance Bourgey and De Marco (2021); Guerreiro and Guerra (2022) for a similar choice within a rough Bergomi model for pricing VIX options. This is particularly easy to treat in our set up, as for any given maturity $T > 0$ we have

$$\mathbb{E}[\mathrm{VIX}_T^2(\ell)] = \frac{1}{\Delta}\ell^\top Q^{cv}(T,\Delta)\ell,$$

  with $Q^{cv}(T,\Delta) := \mathbb{E}[Q_{\mathscr{L}(I)\mathscr{L}(J)}(T,\Delta)]$. By Theorem 2.3.1 and Theorem 2.2.4 we indeed have

$$\begin{aligned} Q^{cv}_{\mathscr{L}(I)\mathscr{L}(J)}(T,\Delta) &= \mathbf{vec}((e_I \sqcup\!\sqcup e_J)\otimes e_0)^\top (e^{\Delta G^\top} - \mathrm{Id})\,\mathbb{E}[\mathbf{vec}(\widehat{\mathbb{X}}_T^{2n+1})] \\ &= \mathbf{vec}((e_I \sqcup\!\sqcup e_J)\otimes e_0)^\top (e^{\Delta G^\top} - \mathrm{Id}) e^{TG^\top}\mathbf{vec}(\widehat{\mathbb{X}}_0^{2n+1}) \\ &= \mathbf{vec}((e_I \sqcup\!\sqcup e_J)\otimes e_0)^\top (e^{(T+\Delta)G^\top} - e^{TG^\top})\mathbf{vec}(\widehat{\mathbb{X}}_0^{2n+1}) \end{aligned}$$

  where G denotes the $(d+1)_{2n+1}$-dimensional matrix representative of the dual operator corresponding to $\widehat{\mathbb{X}}$ and $\mathbf{vec}(\widehat{\mathbb{X}}_0^{2n+1}) = e_\emptyset \in \mathbb{R}^{(d+1)_{2n+1}}$.

  Observe that Q^{cv} can again be computed offline similarly to the matrix Q. Thus to compute the expectation of $\mathrm{VIX}_T^2(\ell)$ we only have to evaluate the previous quadratic form. To apply this now for pricing a call option with maturity $T > 0$ and strike $K > 0$, we set

$$\begin{aligned} \Phi^{cv}(\ell,T,K) :&= c_{T,K}(\Delta\mathrm{VIX}_T^2(\ell) - \ell^\top Q^{cv}(T,\Delta)\ell), \\ &= c_{T,K}(\ell^\top(Q(T,\Delta) - Q^{cv}(T,\Delta))\ell), \end{aligned}$$

where the constant $c_{T,K}$ maximizing the variance reduction is given by:

$$c^*_{T,K} = \frac{\mathrm{Cov}((\mathrm{VIX}_T(\ell) - K)^+, \ell^\top Q(T,\Delta)\ell)}{\mathrm{Var}(\ell^\top Q(T,\Delta)\ell)}.$$

Notice that also in this case both Q and Q^{cv} satisfy the condition for applying the Cholesky decomposition, leading to a faster evaluation of the control variate as discussed in Remark 2.3.5. Note that the Cholesky decomposition cannot be applied to $Q - Q^{cv}$, as this is in general an indefinite matrix.

- As a second example we consider a generic polynomial in VIX^2 as control variate by defining

$$Y^{cv}_m(\ell,T,K) = \sum_{i=0}^{m} \alpha_i(T,K)(\mathrm{VIX}^2_T(\ell))^i \tag{2.3.10}$$

where $\alpha_i(T,K)$ are chosen to approximate the payoff $(\mathrm{VIX}_T - K)^+$ with strike price K for some $m \geq 1$. The corresponding control-variate is then defined as $\Phi^{cv}(\ell,T,K) := c_{T,K}\left(Y^{cv}_m(\ell,T,K) - \mathbb{E}[Y^{cv}_m(\ell,T,K)]\right)$. Regarding the computational effort, let us remark the following.

  (i) VIX^2_T is computed anyway for every realisation and is hence already available, therefore the computation of $Y^{cv}_m(\ell,T,K)$ is not expensive.

  (ii) It is possible to calculate $\mathbb{E}[Y^{cv}_m(\ell,T,K)]$ analytically relying on the moment formula, see Theorem 2.2.4.

  (iii) The choice of $c_{T,K} \in \mathbb{R}$ is important and the optimal one, i.e., the one leading the highest variance reduction, is given by the following expression

$$c^*_{T,K} = \frac{\mathrm{Cov}((\mathrm{VIX}_T(\ell) - K)^+, Y^{cv}_m(\ell,T,K))}{\mathrm{Var}(Y^{cv}_m(\ell,T,K))},$$

  see for instance Section 4.1.1 in Glasserman (2004).

We stress the fact that for $m = 1$ the two control variates introduced coincide.

2.3.4. Calibration on VIX options

In this section we focus on the calibration to VIX options only. Let $\mathcal{T}$ be a set of maturities and $\mathcal{K}$ a collection of strikes. Consider the model given by (2.1.1) and (2.1.2) and assume that Assumption 2.1.3 is in force.

Using Monte Carlo compute an approximation of option and futures' prices with $N_{MC} > 0$ samples, i.e.

$$\pi^{\mathrm{model}}_{\mathrm{VIX}}(\ell,T,K) \approx \frac{e^{-rT}}{N_{MC}} \sum_{i=1}^{N_{MC}} (\mathrm{VIX}_T(\ell,\omega_i) - K)^+, \qquad F^{\mathrm{model}}_{\mathrm{VIX}}(\ell,T) \approx \frac{1}{N_{MC}} \sum_{i=1}^{N_{MC}} \mathrm{VIX}_T(\ell,\omega_i), \tag{2.3.11}$$

where

$$\mathrm{VIX}_T(\ell,\omega) = \sqrt{\frac{1}{\Delta}\ell^\top Q(T,\Delta)(\omega)\ell} = \frac{1}{\sqrt{\Delta}}\|U_T^\top(\omega)\ell\|.$$

It is crucial to note that in this framework a Monte Carlo approach is tractable since for *every* ℓ the same samples can be used. This means that we do not need to carry out any simulation during the optimization task. Indeed, the matrix Q can be simulated offline while only the products with $\ell \in \mathbb{R}^{(d+1)_n}$ enter in the calibration step.

Observe that an auxiliary randomization can be employed in every optimisation step as discussed in Remark 2.3.6. Moreover, if we want to use control variates to reduce the variance of the Monte Carlo estimator as described in the previous section, we would consider

$$\pi^{\mathrm{model}}_{\mathrm{VIX}}(\ell,T,K) \approx \frac{e^{-rT}}{N_{VR}}\sum_{i=1}^{N_{VR}}(\mathrm{VIX}_T(\ell,\omega_i)-K)^+ - \Phi^{cv}(\ell,T,K)(\omega_i).$$

Due to the variance reduction the number of samples needed is $N_{VR} \ll N_{MC}$ and Φ^{cv} is as in Section 2.3.3.

The calibration to VIX call options and the corresponding futures on $\mathcal{T}$ and $\mathcal{K}$ consists in minimizing the functional

$$L_{\mathrm{VIX}}(\ell) := \sum_{T\in\mathcal{T},K\in\mathcal{K}} \mathcal{L}\left(\pi^{\mathrm{model}}_{\mathrm{VIX}}(\ell,T,K), \pi^{b,a}_{\mathrm{VIX}}(T,K), \sigma^{b,a}_{\mathrm{VIX}}(T,K), F^{\mathrm{model}}_{\mathrm{VIX}}(\ell,T), F^{mkt}_{\mathrm{VIX}}(T)\right), \tag{2.3.12}$$

where $\mathcal{L}$ denotes a real-valued loss function, $F^{mkt}_{\mathrm{VIX}}(T)$ the market's futures' prices and

$$\pi^{b,a}_{\mathrm{VIX}}(T,K) := \{\pi^{mkt,b}_{\mathrm{VIX}}(T,K), \pi^{mkt,a}_{\mathrm{VIX}}(T,K)\}, \quad \sigma^{b,a}_{\mathrm{VIX}}(T,K) := \{\sigma^{mkt,b}_{\mathrm{VIX}}(T,K), \sigma^{mkt,a}_{\mathrm{VIX}}(T,K)\},$$

the market's option bid/ask prices $\pi^{mkt,b}_{\mathrm{VIX}}(T,K), \pi^{mkt,a}_{\mathrm{VIX}}(T,K)$, and bid/ask implied volatilities $\sigma^{mkt,b}_{\mathrm{VIX}}(T,K), \sigma^{mkt,a}_{\mathrm{VIX}}(T,K)$, respectively. We will specify the choice of the function $\mathcal{L}$ in Section 2.3.4 and Section 2.5.1.

Remark 2.3.7 (Initial guess search)**.** Since within our model choice we are given a quadratic function in ℓ to be minimized, a stochastic optimization with an initial guess is employed. In order to achieve faster convergence we consider an hyperparameter search to choose the starting parameters. The steps are outlined as follows.

- Find the magnitude of the coefficients returning Monte Carlo prices of the VIX options *close* to the one observable on the market. To this extent we sample $N_\ell > 0$ times parameters $\ell \in J_i = [-10^{-i}, 10^{-i}]^{(d+1)_n}$, for $i = 1,\dots,m$ with $m > 0$.
- Select $J^* \in (J_i)_{i=1}^m$ such that

$$J^* \in \mathrm{argmin}_{i:\ \ell\in J_i} L_{\mathrm{VIX}}(\ell).$$

- Choose the initial guess to be

$$\ell_{\mathrm{initial}} \in \mathrm{argmin}_{\ell\in J^*} L_{\mathrm{VIX}}(\ell).$$

Numerical results

In the present section we report the results of the calibration to VIX options only. Here we consider call options written on the VIX on the trading day 02/06/2021, the same as in Guyon and Lekeufack (2022). We stress that for such recent dates the bid-ask spreads for VIX options are rather tight with respect to older dated options as considered for instance in Gatheral et al. (2020); Bondi et al. (2022b). The maturities are reported in the following table with the corresponding range of strikes (in percentage) with respect to the market's futures prices.

$T_1 = 0.0383$	$T_2 = 0.0767$	$T_3 = 0.1342$	$T_4 = 0.2108$	$T_5 = 0.2875$	$T_6 = 0.3833$
(90%,250%)	(90%,250%)	(80%,310%)	(80%,300%)	(75%,395%)	(80%,405%)

We underline that the shortest maturity considered is 14 days. Regarding our modeling choice we fix $d = 2$, $n = 3$, which means to calibrate 40 parameters. For X we choose a 2-dimensional Ornstein-Uhlenbeck processes, see Example 2.2.5, with the following (hyperparameter) configuration:

$$\kappa = (0.1, 25)^\top, \qquad \theta = (0.1, 4)^\top, \qquad \sigma = (0.7, 10)^\top, \qquad \rho = \begin{pmatrix} 1 & -0.577 & 0.3 \\ \cdot & 1 & -0.6 \\ \cdot & \cdot & 1 \end{pmatrix},$$

where the last column of ρ are the correlations with the Brownian motion B driving the price process S. The motivation of this parameters choice is to mimic a rough or strong mean-reverting model as suggested in Rogers; Rømer (2022). We refer to the Appendix A.3, for the numerical results where as primary process we consider simply a correlated 2-dimensional Brownian motion. Before stating the loss function $\mathcal{L}$ that we employed in the calibration task, let us make the following remark.

Remark 2.3.8. Let $f : \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}^+$ be the call pricing functional in the Black-Scholes model, depending on the volatility σ^{BS} and the spot price ξ, i.e., $f : (\sigma^{\mathrm{BS}}, \xi) \mapsto f(\sigma^{\mathrm{BS}}, \xi)$. By Taylor expansion in an appropriate neighbourhood of $(\sigma^{mkt}, \xi^{mkt})$ we obtain

$$f(\sigma^{\mathrm{BS}}, \xi) \approx f(\sigma^{mkt}, \xi^{mkt}) + \frac{\partial f}{\partial \sigma}(\sigma^{mkt}, \xi^{mkt})(\sigma^{\mathrm{BS}} - \sigma^{mkt}) + \frac{\partial f}{\partial \xi}(\sigma^{mkt}, \xi^{mkt})(\xi - \xi^{mkt}),$$

which equivalently gives

$$(\sigma^{\mathrm{BS}} - \sigma^{mkt}) \approx \frac{1}{\frac{\partial f}{\partial \sigma}(\sigma^{mkt}, \xi^{mkt})} \big(f(\sigma^{\mathrm{BS}}, \xi) - f(\sigma^{mkt}, \xi^{mkt})\big) - \frac{\frac{\partial f}{\partial \xi}(\sigma^{mkt}, \xi^{mkt})}{\frac{\partial f}{\partial \sigma}(\sigma^{mkt}, \xi^{mkt})}(\xi - \xi^{mkt}), \tag{2.3.13}$$

where we recognize for the derivatives with respect to σ and ξ, the Greeks Vega and Delta, respectively.

Motivated by Remark 2.3.8 we propose, for a fixed maturity and strike price, the following loss-function for $\beta \in \{0, 1\}$

$$\mathcal{L}^{\beta}(\pi, \pi^{mkt,b,a}, \sigma^{mkt,b,a}, F, F^{mkt}) = \tag{2.3.14}$$
$$\left(\frac{\left(\beta \tilde{1}_{\{\pi \notin [\pi^{mkt,b}, \pi^{mkt,a}]\}} + (1-\beta)\right) \left| \pi - (\pi^{mkt,a} + \pi^{mkt,b})/2 \right| + \left| \delta^{mkt} e^{-rT} (F - F^{mkt}) \right|}{\upsilon^{mkt} (\sigma^{mkt,a} - \sigma^{mkt,b})} \right)^2,$$

where

- υ^{mkt} and δ^{mkt} denote the Vega and Delta of the option under the Black-Scholes model which depend on the maturity and on the strike price;
- F and F^{mkt} denote futures with maturity T such that $\xi = e^{-rT} F$ and $\xi^{mkt} = e^{-rT} F^{mkt}$, respectively;
- $\tilde{1}_{\{x \notin [y^b, y^a]\}} := s(y^b - x) + s(x - y^a)$ for $s(x) := \frac{1}{2} \tanh(100x) + \frac{1}{2}$ a smooth version of the indicator function.

Remark 2.3.9. (i) We observe that by Remark 2.3.8 minimizing $\mathcal{L}^0$ is equivalent to minimizing an upper bound of the square of the right-hand side of (2.3.13) normalized by the bid-ask spread of the implied volatilities. Note that we slightly abused notation, since υ^{mkt} and δ^{mkt} of course depend on the strike and the maturity.

(ii) Note that as $\ell \mapsto \mathrm{VIX}_T(\ell, \omega) = \frac{1}{\sqrt{\Delta}} \|U_T^\top \ell\|$ is convex and the call payoff is convex and increasing, the model option and future prices are convex in ℓ. If $\beta = 0$ and the initialization of ℓ is such that both the model and future prices are higher than the market ones, then we actually deal with a convex optimization problem.

(iii) If our aim does not consist in calibrating to the mid-price or mid-implied-volatility precisely, but we merely want to be within the bid-ask spreads we can set $\beta = 1$

For the next calibration result we minimize $\mathcal{L}^1$ as introduced above with $N_{MC} = 80000$ Monte Carlo samples for the previous maturities and strikes.

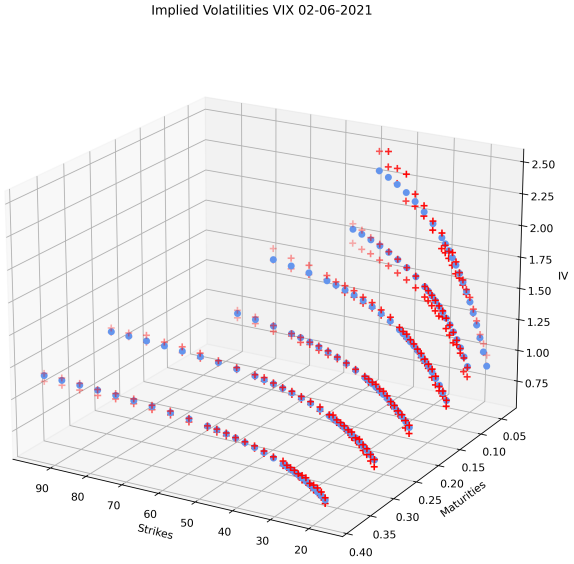


Figure 2.1.: The red crosses denote the bid-ask spread (of the implied volatilities) for each maturity, while the azure dots denote the calibrated implied volatilities of the model. On the x-axis we find the strikes and on the y-axis we find the maturities.

We observe that the calibrated VIX smiles fall systematically in the bid-ask corridor for all the maturities considered. We report additionally in the next tables the relative absolute error between the market future prices and the calibrated ones for each maturity, i.e.,

$$\varepsilon_T := \frac{|F^{mkt}(T) - F^{\text{model}}_{\text{VIX}}(\ell^*, T)|}{F^{mkt}(T)}, \tag{2.3.15}$$

where $\ell^* \in \mathbb{R}^{40}$ denotes the calibrated parameters and here $F^{\text{model}}_{\text{VIX}}(\ell^*, T)$ stands for the calibrated future model price. In Figure 2.2 we can find an illustration of the calibrated and the market futures' term structure.

$T_1 = 0.0383$	$T_2 = 0.0767$	$T_3 = 0.1342$
$\varepsilon_{T_1} = 7.0 \times 10^{-6}$	$\varepsilon_{T_2} = 2.1 \times 10^{-3}$	$\varepsilon_{T_3} = 1.3 \times 10^{-5}$

$T_4 = 0.2108$	$T_5 = 0.2875$	$T_6 = 0.3833$
$\varepsilon_{T_4} = 1.5 \times 10^{-4}$	$\varepsilon_{T_5} = 1.9 \times 10^{-6}$	$\varepsilon_{T_6} = 1.3 \times 10^{-6}$

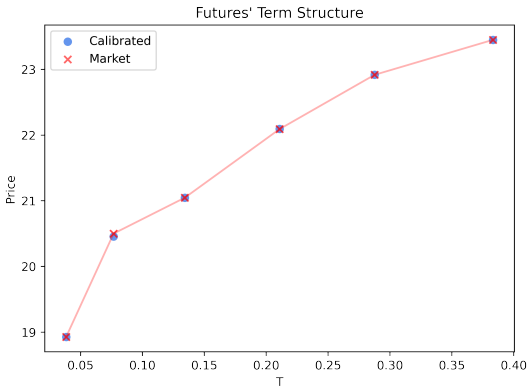


Figure 2.2.: The blue circles denote the calibrated futures prices and the red crosses the futures prices on the market, in between a linear interpolation is reported.

2.3.5. The case of time-varying parameters

We now consider the case of maturity dependent parameters as for instance employed in Gierjatowicz et al. (2020); Cuchiero et al. (2022b). Since it will be important later on to distinguish maturities of options written on the VIX index from maturities of options written on the SPX index, we introduce the two sets $\mathcal{T}^{\text{VIX}}$ and $\mathcal{T}^{\text{SPX}}$. Let us fix here $\mathcal{T}^{\text{VIX}} = \{T_1, \dots, T_N\}$, where $T_i < T_{i+1}$ for any in $i = 1, \dots, N-1$ and denote by $\ell(T_i) \in \mathbb{R}^{d_n}$ the parameters depending on the maturity $T_i > 0$. We set $T_0 = 0$ and $T_{N+1} = +\infty$. Then, we consider for any $t \geq 0$ the volatility process to be

$$\sigma_t^S(\ell) = \sum_{i=0}^{N} \sum_{|I| \leq n} \ell_I(T_i) 1_{[T_i, T_{i+1})}(t) \langle e_I, \widehat{\mathbb{X}}_t \rangle. \tag{2.3.16}$$

Therefore the variance process reads as follows,

$$V_t(\ell) = \sum_{i=0}^{N} \sum_{|J|,|I| \leq n} \ell_I(T_i) \ell_J(T_i) 1_{[T_i, T_{i+1})}(t) \langle e_I \sqcup\!\sqcup e_J, \widehat{\mathbb{X}}_t \rangle. \tag{2.3.17}$$

Assumption 2.3.10. *Assume that for a set of maturities $\mathcal{T}^{VIX}$ it holds that $|T_i - T_j| \geq \Delta$ for all $i \neq j$.*

Proposition 2.3.11. Let $\mathcal{T}^{\text{VIX}}$ be a set of maturities on the VIX index and let $Q(T, \tau)$ be the matrix as defined in (2.3.5) (here for general $\tau > 0$ instead of Δ). Then, under (2.3.17) the VIX squared at time $T_i \in \mathcal{T}^{\text{VIX}}$ is given by

$$\text{VIX}^2_{T_i}(\ell) = \frac{1}{\Delta} \Big(\sum_{j=i}^{N} \ell(T_j)^\top \left(Q(T_i, (T_{j+1} - T_i) \wedge \Delta) - Q(T_i, (T_j - T_i) \wedge \Delta) \right) \ell(T_j) \Big).$$

Note that, if $T_{i+1} - T_i > \Delta$ (which is in particular holds under Assumption 2.3.10) then,

$$\mathrm{VIX}^2_{T_i}(\ell) = \frac{1}{\Delta}\ell(T_i)^\top Q(T_i, \Delta)\ell(T_i).$$

Proof. By the definition of the VIX, it holds that

$$\begin{aligned}
\mathrm{VIX}^2_{T_i}(\ell) =& \frac{1}{\Delta}\mathbb{E}\left[\int_{T_i}^{T_i+\Delta} \sum_{j=i}^{N} \sum_{|J|,|I|\leq n} \ell_I(T_j)\ell_J(T_j) 1_{[T_j,T_{j+1})}(t)\langle e_I \sqcup\!\sqcup e_J, \widehat{\mathbb{X}}_t\rangle \mathrm{d}t \middle| \mathcal{F}_{T_i}\right] \\
=& \frac{1}{\Delta}\sum_{j=i}^{N} \sum_{|J|,|I|\leq n} \ell_I(T_j)\ell_J(T_j)\mathbb{E}\left[\int_{T_j\wedge(T_i+\Delta)}^{T_{j+1}\wedge(T_i+\Delta)} \langle e_I \sqcup\!\sqcup e_J, \widehat{\mathbb{X}}_t\rangle \mathrm{d}t \middle| \mathcal{F}_{T_i}\right] \\
=& \frac{1}{\Delta}\sum_{j=i}^{N} \sum_{|J|,|I|\leq n} \ell_I(T_j)\ell_J(T_j)\left(\mathbb{E}\left[\int_{T_i}^{T_{j+1}\wedge(T_i+\Delta)} \langle e_I \sqcup\!\sqcup e_J, \widehat{\mathbb{X}}_t\rangle \mathrm{d}t \middle| \mathcal{F}_{T_i}\right]\right. \\
& \left. - \mathbb{E}\left[\int_{T_i}^{T_j\wedge(T_i+\Delta)} \langle e_I \sqcup\!\sqcup e_J, \widehat{\mathbb{X}}_t\rangle \mathrm{d}t \middle| \mathcal{F}_{T_i}\right]\right)
\end{aligned}$$

and hence the first statement follows by the definition of Q in (2.3.5). □

Notice that also in the case of Proposition 2.3.11, Remark 2.3.5 applies.

2.4. SPX as a signature-based model

The goal of this section is to express the discounted price of the SPX, modeled via (2.1.1)-(2.1.2)

$$\mathrm{d}S_t(\ell) = S_t(\ell)\sigma^S_t(\ell)\mathrm{d}B_t,$$

in terms of the signature of $(\widehat{Z}_t)_{t\geq 0} = (t, X_t, B_t)_{t\geq 0}$, allowing again to precompute its samples and use the same ones for every ℓ. This is in the same spirit as in Chapter 1, even though there the asset price was directly modeled as linear function of the signature of some primary process. Recall that by (2.1.2) σ^S is parametrized as follows

$$\sigma^S_t(\ell) := \ell_\emptyset + \sum_{0<|I|\leq n} \ell_I\langle e_I, \widehat{\mathbb{X}}_t\rangle,$$

where $\widehat{X}_t = (t, X_t)$ with X a d-dimensional continuous semimartingale. Before addressing a more tractable expression for S, that allows to avoid (Euler) simulation schemes, we recall the following well-known integrability result.

Lemma 2.4.1. Assume that $\mathbb{E}[S_0] < \infty$. Then, the process $(S_t)_{t\geq 0}$ is a (non-negative) supermartingale and in particular $\mathbb{E}[S_t] < \infty$ for each $t \geq 0$.

Proof. Note that $S_t = S_0\mathcal{E}\left(\int_0^\cdot \sigma^S_s \mathrm{d}B_s\right)_t$ for all $t \geq 0$. Moreover $(\int_0^t \sigma^S_s \mathrm{d}B_s)_{t\geq 0}$ is a local martingale and hence, by the properties of the stochastic exponential, S_t is a non-negative local martingale. It follows from Fatou's Lemma that non-negative local martingales are supermartingales. □

In the following we suppose without loss of generality that $S_0 = 1$.

Remark 2.4.2. Recall that if Novikov's condition is satisfied, then a stochastic exponential of the form $S_t = \mathcal{E}\left(\int_0^\cdot \sigma_s^S \mathrm{d}B_s\right)_t$ for $t \in [0,T]$ is a true martingale. For σ_s^S as in (2.1.2), such condition reads

$$\mathbb{E}\left[\exp\left\{\frac{1}{2}\int_0^T V_t(\ell)\mathrm{d}t\right\}\right] < +\infty.$$

Observe that

$$\begin{aligned}\mathbb{E}\left[\exp\left\{\frac{1}{2}\int_0^T V_t(\ell)\mathrm{d}t\right\}\right] &= \mathbb{E}\left[\exp\left\{\frac{1}{2}\sum_{|I|,|J|\le n}\ell_I\ell_J\int_0^T\langle e_I \sqcup\!\sqcup e_J, \widehat{\mathbb{X}}_t\rangle \mathrm{d}t\right\}\right] \\ &= \mathbb{E}\left[\exp\left\{\frac{1}{2}\ell^\top Q^0(T)\ell\right\}\right], \end{aligned} \tag{2.4.1}$$

where for $\mathscr{L} : \{I : |I| \le n\} \to \{1, \dots, (d+1)_n\}$,

$$Q^0_{\mathscr{L}(I)\mathscr{L}(J)}(T) := \langle (e_I \sqcup\!\sqcup e_J) \otimes e_0, \widehat{\mathbb{X}}_T\rangle.$$

We point out that the previous condition is not necessarily satisfied for all $\ell \in \mathbb{R}^{(d+1)_n}$. Indeed, let us consider X to be a one-dimensional Brownian motion and choose ℓ such that the only non trivial component is the last one, i.e.,

$$\ell_I := \begin{cases} c \in \mathbb{R}, & \text{if } I = (1,\dots,1),\ |I| = n, \\ 0, & \text{otherwise.}\end{cases} \tag{2.4.2}$$

Then, (2.4.1) translates into

$$\mathbb{E}\left[\exp\left\{\frac{c^2}{2}\int_0^T \frac{2n!}{n!n!}\frac{X_t^{2n}}{2n!}\mathrm{d}t\right\}\right] = \mathbb{E}\left[\exp\left\{\frac{c^2}{2(n!)^2}\int_0^T X_t^{2n}\mathrm{d}t\right\}\right],$$

which is not finite in general, e.g. if $n = 2$, $c = \sqrt{2}(n!)$ then by Jensen's inequality it follows

$$\mathbb{E}\left[\exp\left\{\int_0^T X_t^4\mathrm{d}t\right\}\right] \ge \mathbb{E}\left[\frac{1}{T}\int_0^T e^{TX_t^4}\mathrm{d}t\right] = \frac{1}{T}\int_0^T \mathbb{E}[e^{TX_t^4}]\mathrm{d}t = +\infty.$$

Remark 2.4.3. Let us underline the fact that, even if the process $S_t(\ell)$ is a non-negative local martingale, we can still price options via risk neutral expectations without introducing arbitrage. The reason is that, although $\mathbb{E}[S_T(\ell)] < S_0$, it is impossible to take a short position on $S(\ell)$ and a long position on $V_t := \mathbb{E}[S_T(\ell)|\mathcal{F}_t]$ for all $t \le T$ because of credit constraints, therefore, the condition of No Free Lunch with Vanishing Risk (NFLVR) as introduced in Delbaen and Schachermayer (1994) is not violated. We address the reader to Kardaras et al. (2015) for further details.

The key idea is to rewrite (2.1.1) as a type of signature-based model in sense of Cuchiero et al. (2022b) including $B = (B_t)_{t\geq 0}$ as part of the primary process. This is possible since Itô integrals with respect to primary process' components can be rewritten as linear functions of the signature of the primary process itself. To do so, we introduce Assumption 1.2.6 on $Z = (X, B)$ in order to describe the correlation structure between B and X.

Assumption 2.4.4. *For all $i, j \in \{1, \dots, d+1\}$ it holds*

$$\mathrm{d}[Z^i, Z^j]_t = \sum_{|J|\leq m} a_{ij}^J \langle e_J, \widehat{\mathbb{Z}}_t \rangle \mathrm{d}t$$

for some $m \in \mathbb{N}$, $a_{ij}^J \in \mathbb{R}$ and where $\widehat{Z}_t = (t, Z_t)$.

Notice that this assumption is in particular satisfied if X is a polynomial process as in Assumption 2.1.3, correlated with B. As shown in the next proposition it allows for the aforementioned tractable representation of $S(\ell)$.

Proposition 2.4.5. Let $S = (S_t)_{t\geq 0}$ satisfy (2.1.1) with $S_0 = 1$, and $\sigma^S = (\sigma_t^S)_{t\geq 0}$ satisfy (2.1.2). Suppose additionally that $Z = (X, B)$ satisfies Assumption 2.4.4. Then,

$$\log(S_t(\ell)) = -\frac{1}{2}\ell^\top Q^0(t)\ell + \sum_{|I|\leq n} \ell_I \langle \tilde{e}_I^B, \widehat{\mathbb{Z}}_t \rangle, \tag{2.4.3}$$

where

$$\tilde{e}_\emptyset^B := e_{d+1}, \qquad \tilde{e}_I^B := e_I \otimes e_{d+1} - \sum_{|J|\leq m} \frac{a_{i_{|I|}(d+1)}^J}{2} (e_{I'} \sqcup\!\sqcup e_J) \otimes e_0,$$

for each $|I| > 0$, and the components of the matrix $Q^0(t) \in \mathbb{R}^{(d+1)_n \times (d+1)_n}$ are given by

$$Q^0_{\mathscr{L}(I)\mathscr{L}(J)}(t) = \langle (e_I \sqcup\!\sqcup e_J) \otimes e_0, \widehat{\mathbb{X}}_t \rangle,$$

for a labeling function $\mathscr{L} : \{I : |I| \leq n\} \to \{1, \dots, (d+1)_n\}$.

Proof. Under our assumptions we can compute

$$\begin{aligned}
\log(S_t(\ell)) &= -\frac{1}{2}\int_0^t V_s(\ell)\mathrm{d}s + \int_0^t \sigma_s^S(\ell)\mathrm{d}B_s \\
&= -\frac{1}{2}\sum_{|I|,|J|\leq n} \ell_I \ell_J \int_0^t \langle e_I \sqcup\!\sqcup e_J, \widehat{\mathbb{X}}_s \rangle \mathrm{d}s + \sum_{|I|\leq n} \ell_I \int_0^t \langle e_I, \widehat{\mathbb{X}}_s \rangle \mathrm{d}B_s \\
&\overset{(*)}{=} -\frac{1}{2}\sum_{|I|,|J|\leq n} \ell_I \ell_J \langle (e_I \sqcup\!\sqcup e_J) \otimes e_0, \widehat{\mathbb{X}}_t \rangle + \sum_{|I|\leq n} \ell_I \langle \tilde{e}_I^B, \widehat{\mathbb{Z}}_t \rangle \\
&= -\frac{1}{2}\ell^\top Q^0(t)\ell + \sum_{|I|\leq n} \ell_I \langle \tilde{e}_I^B, \widehat{\mathbb{Z}}_t \rangle,
\end{aligned}$$

where for $(*)$ we used that $\int_0^t \langle e_I, \widehat{\mathbb{X}}_s \rangle \mathrm{d}B_s = \langle \tilde{e}_I^B, \widehat{\mathbb{Z}}_t \rangle$ by Lemma 1.2.10. □

Remark 2.4.6. Consider again the model described in Remark 2.1.1. Then the results of Proposition 2.4.5 still hold with

$$Q^0_{\mathscr{L}(I)\mathscr{L}(J)}(t) := \int_0^t \langle e_I \sqcup\!\sqcup e_J, \widehat{\mathbb{X}}_{s-\varepsilon,s}\rangle \mathrm{d}s,$$

and $\int_0^t \langle e_I, \widehat{\mathbb{X}}_{s-\varepsilon,s}\rangle \mathrm{d}B_s$ instead of $\langle \tilde{e}_I^B, \widehat{\mathbb{Z}}_t\rangle$. Since the proof follows closely the proof of the original result, we omit it.

Remark 2.4.7.

- Observe that since the matrix $(\langle e_I \sqcup\!\sqcup e_J, \widehat{\mathbb{X}}_t\rangle)_{|I|,|J|\leq n}$ is positive semidefinite, by monotonicity of the time integral on $[0,t]$ for some $t>0$, we also have
$$\ell^\top Q^0(t)\ell \geq 0,$$
for all $\ell \in \mathbb{R}^{(d+1)_n}$. This means that for any $t>0$, we can rewrite the log-price as
$$\log(S_t) = -\frac{1}{2}\|(U_t^0)^\top \ell\|^2 + \sum_{|I|\leq n} \ell_I \langle \tilde{e}_I^B, \widehat{\mathbb{Z}}_t\rangle,$$
where U_t^0 is the upper-triangular matrix of the Cholesky decomposition of $Q^0(t)$.
- Notice that the log-price model in (2.4.3), it is not exactly a signature-based model in the sense of Chapter 1, as here it is given by a linear part in the parameters ℓ and an additional quadratic part. It can also be rewritten as
$$\mathrm{d}\log(S_t) = -\frac{1}{2}\ell^\top \tilde{Q}(t)\ell \mathrm{d}t + \ell^\top \mathbf{vec}(\widehat{\mathbb{X}}_t^n)\mathrm{d}B_t,$$
where $\tilde{Q}$ is given by
$$\tilde{Q}_{\mathscr{L}(I)\mathscr{L}(J)}(t) := \langle e_I \sqcup\!\sqcup e_J, \widehat{\mathbb{X}}_t\rangle.$$
- In order to sample the log-price at maturity, consistently with the VIX, we follow the following road map. We simulate $\widehat{\mathbb{Z}}$ and compute $\langle \tilde{e}_I^B, \widehat{\mathbb{Z}}\rangle$ for each I as specified above. Next, we drop from the samples of $\widehat{\mathbb{Z}}$ the terms where B appears, i.e. the components corresponding to indices containing the letter $d+1$. The result coincides with a sampling of $\widehat{\mathbb{X}}$ and is then used to work with both Q and Q^0.

  This is equivalent to sampling $\widehat{\mathbb{X}}$ for the variance process and to compute an additional Itô integral as in (2.1.1).

In the following corollary we state the form of $\tilde{e}_I^B$ when X is d-dimensional OU-process. We omit the proof for sake of brevity.

Corollary 2.4.8. *Let X be a d-dimensional OU-process as in Example 2.2.5 driven by a d-dimensional Brownian motion with correlation matrix ρ. Then Assumption 1.2.6 is satisfied and $\tilde{e}_I^B$ is given by*

$$\tilde{e}_I^B = e_I \otimes e_{d+1} - \frac{1}{2} 1_{\{i_{|I|}\neq 0\}} (\sigma^{i_{|I|}} \rho_{i_{|I|}d+1}) e_{I'} \otimes e_0,$$

for any multi-index $I \neq \emptyset$.

Remark 2.4.9 (Variance reduction for pricing SPX options)**.** Observe that a possible control variate for reducing the variance of the Monte Carlo estimator for pricing SPX options is the value at maturity of the log-price process. This means,

$$\Phi^{cv}(\ell, T, K) := c_{T,K}\Big(\log(S_T(\ell)) + \frac{1}{2}\ell^\top Q^{0,cv}(T)\ell\Big),$$

where, using that the linear part (in ℓ) of $\log(S_T(\ell))$ vanishes under the risk-neutral expectation, we have

$$Q^{0,cv}_{\mathscr{L}(I)\mathscr{L}(J)}(T) = \mathbf{vec}((e_I \sqcup\!\sqcup e_J)\otimes e_0)^\top e^{TG^\top}\mathbf{vec}(\widehat{\mathbb{X}}_0^{2n+1}),$$

for $G \in \mathbb{R}^{(d+1)_{2n+1}\times(d+1)_{2n+1}}$ denoting the $(d+1)_{2n+1}$-dimensional matrix representative of the dual operator corresponding to $\widehat{\mathbb{X}}$. We choose the optimal $c^*_{T,K} \in \mathbb{R}$ as

$$c^*_{T,K} = \frac{\mathrm{Cov}((S_T(\ell)-K)^+, \log(S_T(\ell)))}{\mathrm{Var}(\log(S_T(\ell)))}.$$

2.4.1. Exploiting the affine nature of the signature: Fourier pricing of SPX and VIX options

In the present section we observe how the linear structure of the parametrizations for the log-price and of the variance process in $\widehat{\mathbb{Z}}$ allows to consider Fourier pricing within this family of signature-based models. Consider again the model given by (2.1.1)-(2.1.2) and let $(Z_t)_{t\geq 0}$ denote the primary process given by $Z_t = (X_t, B_t)$ introduced after (2.1.2). Suppose for simplicity that

$$\mathrm{d}Z^j_t = \kappa^j(\theta^j - Z^j_t)\mathrm{d}t + \sigma^j \mathrm{d}W^j_t, \qquad Z^j_0 = 0,$$

for each $j = 1, \dots, d+1$, where W denotes a $(d+1)$-dimensional Brownian motion with $W^{d+1} = B$. All parameters $\kappa^j, \theta^j, \sigma^j$ are in $\mathbb{R}$ with $\kappa^{d+1} = \theta^{d+1} = 0$ and $\sigma^{d+1} = 1$ so that $Z^{d+1} = W^{d+1} = B$. Note that we do not account for correlations.

We illustrate now how to apply the results of Cuchiero et al. (2023b) in the present setting. Since $(\widehat{Z}_t)_{t\geq 0}$ is a polynomial process, by Lemma 2.2.1 there are $\mathbf{b} \in (T((\mathbb{R}^{d+2})))^{d+2}$ and $\mathbf{a} \in (T((\mathbb{R}^{d+2})))^{(d+2)\times(d+2)}$ such that

$$\mathrm{d}\widehat{Z}^j_t = \langle \mathbf{b}_j, \widehat{\mathbb{Z}}_t\rangle \mathrm{d}t + \sqrt{\langle \mathbf{a}_{jj}, \widehat{\mathbb{Z}}_t\rangle}\mathrm{d}W^j_t,$$

where $\mathbf{b}_j = \kappa^j\theta^j e_\emptyset - \kappa_j e_j$ and $\mathbf{a}_{jj} = (\sigma^j)^2 e_\emptyset$, using that (with a small abuse of notation) $\kappa^0\theta^0 := 1$, $\kappa^j := 0$ and $\sigma^0 := 0$. Using the notation of (1.1.3) consider then the Riccati operator $\mathcal{R}$ given by

$$\begin{aligned}\mathcal{R}(\mathbf{u}) &= \mathbf{b}^\top \sqcup\!\sqcup \mathbf{u}^{(1)} + \frac{1}{2}\mathrm{Tr}(\mathbf{a} \sqcup\!\sqcup (\mathbf{u}^{(2)} + (\mathbf{u}^{(1)})^\top \sqcup\!\sqcup \mathbf{u}^{(1)})) \\ &= \sum_{j=0}^{d+1}\sum_{|I|\geq 0}\Big(\kappa^j\theta^j \mathbf{u}_{(Ij)}e_I + \kappa^j \mathbf{u}_{(Ij)} e_j \sqcup\!\sqcup e_I + \frac{1}{2}(\sigma^j)^2(\mathbf{u}_{(Ijj)}e_I + \mathbf{u}^2_{(Ij)}e_I \sqcup\!\sqcup e_I)\Big).\end{aligned}$$

In this equation $\mathrm{Tr} : \mathbb{R}^{(d+2)\times(d+2)} \to \mathbb{R}$ denotes the trace operator and is applied component wise to the elements of $(T((\mathbb{R}^{d+2})))^{(d+2)\times(d+2)}$. By Theorem 4.23 in Cuchiero et al. (2023b), we expect that

$$\mathbb{E}[\exp(\langle \mathbf{u}, \widehat{\mathbb{Z}}_T\rangle)] = \exp(\boldsymbol{\psi}(T)_\emptyset),$$

where $\boldsymbol{\psi}$ is a solution[1] of the extended tensor algebra valued Riccati equation

$$\partial_t \boldsymbol{\psi}(t) = \mathcal{R}(\boldsymbol{\psi}(t)), \quad \boldsymbol{\psi}(0) = \mathbf{u}.$$

Choosing $\mathbf{u}$ as

$$\mathbf{u}(\ell) := -\frac{1}{2}(\ell \sqcup\!\sqcup \ell) \otimes e_0 + \tilde{\ell}$$

where $\tilde{\ell} := \sum_{|I|\leq n} \ell_I \tilde{e}_I^B$, by Proposition 2.4.5 we get

$$\log(S_t(\ell)) = \langle \mathbf{u}(\ell), \widehat{\mathbb{Z}}_t\rangle.$$

The representation of the Fourier-Laplace transform described above can then be used for Fourier pricing. We dedicate the remaining part of this section to illustrate how this can be done.

From Fourier analysis we know that for $K > 0$ and $C < 0$ it holds

$$(K - e^y)^+ = \frac{1}{2\pi}\int_{\mathbb{R}} e^{(i\lambda + C)y} \frac{K^{-C+1-i\lambda}}{(i\lambda + C)(i\lambda + C - 1)} \mathrm{d}\lambda.$$

This in particular implies that

$$\begin{aligned}
\mathbb{E}[(K - S_T(\ell))^+] &= \frac{1}{2\pi}\int_{\mathbb{R}} \mathbb{E}[e^{(i\lambda+C)\log(S_T(\ell))}] \frac{K^{-C+1-i\lambda}}{(i\lambda + C)(i\lambda + C - 1)} \mathrm{d}\lambda \\
&= \frac{1}{2\pi}\int_{\mathbb{R}} \mathbb{E}[e^{\langle \mathbf{u}_\lambda, \widehat{\mathbb{Z}}_T\rangle}] \frac{K^{-C+1-i\lambda}}{(i\lambda + C)(i\lambda + C - 1)} \mathrm{d}\lambda \\
&= \frac{1}{2\pi}\int_{\mathbb{R}} e^{\boldsymbol{\psi}_\lambda(T)_\emptyset} \frac{K^{-C+1-i\lambda}}{(i\lambda + C)(i\lambda + C - 1)} \mathrm{d}\lambda,
\end{aligned}$$

where $\mathbf{u}_\lambda := (i\lambda + C)\mathbf{u}(\ell)$ and $\boldsymbol{\psi}_\lambda$ is a solution of the Riccati equation with initial condition $\boldsymbol{\psi}_\lambda(0) = \mathbf{u}_\lambda$.

Let us now consider the case of VIX options where Fourier pricing can be applied by computing the Fourier-Laplace transform of VIX squared, see also Sepp (2008); Papanicolaou and Sircar (2014); Cao et al. (2020); Bondi et al. (2022b) and references therein for a Fourier-based approach to pricing VIX options. Fix a labelling injective function $\mathscr{L} : \{I \colon |I| \leq n\} \to \{1, \ldots, (d+1)_{(2n+1)}\}$ as introduced before (1.1.6) and recall that by Theorem 2.3.1(ii) it holds

$$\mathrm{VIX}_T^2(\ell) = \frac{1}{\Delta}\ell^\top Q(T, \Delta)\ell$$

[1] We refer to Cuchiero et al. (2023b) for the appropriate solution concept.

for

$$Q_{\mathscr{L}(I)\mathscr{L}(J)}(T,\Delta) = \sum_{e_K=(e_I \sqcup e_J)\otimes e_0} \sum_{|H|\leq 2n+1} (e^{\Delta G^\top} - \mathrm{Id})_{\mathscr{L}(K)\mathscr{L}(H)} \langle e_H, \widehat{\mathbb{X}}_T \rangle.$$

where G denotes the $(d+1)_{(2n+1)}$-dimensional matrix representative of the dual operator corresponding to $\widehat{\mathbb{X}}$. Setting for $|I|, |J| \leq n$

$$I(\Delta)\mathbf{u} := \sum_{|K|,|H|\leq 2n+1} (e^{\Delta G^\top} - \mathrm{Id})_{\mathscr{L}(K)\mathscr{L}(H)} \mathbf{u}_K e_H$$

we can write

$$\begin{aligned}
\mathrm{VIX}_T^2(\ell) &= \frac{1}{\Delta} \sum_{|I|,|J|\leq n} \ell_I \ell_J \sum_{e_K=(e_I \sqcup e_J)\otimes e_0} \sum_{|H|\leq 2n+1} (e^{\Delta G^\top} - \mathrm{Id})_{\mathscr{L}(K)\mathscr{L}(H)} \langle e_H, \widehat{\mathbb{X}}_T \rangle \\
&= \frac{1}{\Delta} \sum_{|K|,|H|\leq 2n+1} (e^{\Delta G^\top} - \mathrm{Id})_{\mathscr{L}(K)\mathscr{L}(H)} \langle e_K, (\ell \sqcup \ell) \otimes e_0 \rangle \langle e_H, \widehat{\mathbb{X}}_t \rangle \\
&= \frac{1}{\Delta} \langle I(\Delta)((\ell \sqcup \ell) \otimes e_0), \widehat{\mathbb{X}}_T \rangle.
\end{aligned}$$

Also in this case since $\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{i\lambda y}(K - \sqrt{|y|})^+ \mathrm{d}y = \frac{\sqrt{\frac{2}{\pi}} F_S(K\sqrt{|\lambda|})}{|\lambda|^{3/2}}$ is integrable for $F_S(u) = \int_0^u \sin(z^2)\mathrm{d}z$, Fourier analysis yields

$$(K - \sqrt{y})^+ = \frac{1}{\pi} \int_{\mathbb{R}} e^{-i\lambda y} \frac{F_S(K\sqrt{|\lambda|})}{|\lambda|^{3/2}} \mathrm{d}\lambda,$$

for each $y \geq 0$. This in particular implies that

$$\mathbb{E}[(K - \mathrm{VIX}_T(\ell))^+] = \frac{1}{\pi} \int_{\mathbb{R}} \mathbb{E}[e^{-i\lambda \mathrm{VIX}_T^2(\ell)}] \frac{F_S(K\sqrt{|\lambda|})}{|\lambda|^{3/2}} \mathrm{d}\lambda = \frac{1}{\pi} \int_{\mathbb{R}} e^{\boldsymbol{\psi}_\lambda(T)\emptyset} \frac{F_S(K\sqrt{|\lambda|})}{|\lambda|^{3/2}} \mathrm{d}\lambda,$$

where $\boldsymbol{\psi}_\lambda$ is a solution of the Riccati equation with initial condition $\boldsymbol{\psi}_\lambda(0) = -i\lambda \mathbf{v}(\ell)$ for $\mathbf{v}(\ell) := \frac{1}{\Delta} I(\Delta)((\ell \sqcup \ell) \otimes e_0)$.

Analogous formulations in terms of the error function are also possible, see for instance Cao et al. (2020); Bondi et al. (2022b). In the same spirit one can also obtain a representation of future prices. We here do not provide an implementation of this Fourier pricing approach but numerical experiments can be found in Cuchiero et al. (2023b).

2.4.2. The case of time-varying parameters

Analogously to Section 2.3.5, we now further exploit Proposition 2.4.5 by allowing the parameters ℓ to depend on the maturity.

Proposition 2.4.10. Let $S = (S_t)_{t\geq 0}$ satisfy (2.1.1) with $S_0 = 1$, and $(\sigma_t^S)_{t\geq 0}$ satisfy (2.3.16) for a set of maturities $\mathcal{T}^{\mathrm{VIX}} = \{T_1, \ldots, T_N\}$. Recall that in this case $V = (V_t)_{t\geq 0}$ satisfy

$$V_t(\ell) = \sum_{i=0}^{N} \sum_{|J|,|I|\leq n} \ell_I(T_i)\ell_J(T_i) 1_{[T_i,T_{i+1})}(t) \langle e_I \sqcup e_J, \widehat{\mathbb{X}}_t \rangle.$$

Let $Z_t = (X_t, B_t)$ for all $t \geq 0$, where $X = (X_t)_{t\geq 0}$ is a d-dimensional continuous semimartingale. Then, under Assumptions 1.2.6 we get the following recursion for the log-price process

$$\log(S_t(\ell^{<m+1})) = \sum_{i=0}^{N} \Big[-\frac{1}{2}\ell(T_i)^\top \big(Q^0(t\wedge T_{i+1}) - Q^0(t\wedge T_i)\big)\ell(T_i) + \sum_{|I|\leq n} \ell_I(T_i)\langle \tilde{e}_I^B, \widehat{\mathbb{Z}}_{t\wedge T_{i+1}} - \widehat{\mathbb{Z}}_{t\wedge T_i}\rangle \Big]$$

for each $t \geq 0$, where $T_0 := 0$, $\ell^{<m+1} := \{\ell(0), \dots, \ell(T_m)\}$, $m = \max\{j : T_j < t\}$, and

$$Q^0_{\mathscr{L}(I)\mathscr{L}(J)}(t) = \langle (e_I \sqcup\!\sqcup e_J) \otimes e_0, \widehat{\mathbb{X}}_t\rangle,$$

for a labeling function $\mathscr{L} : \{I : |I| \leq n\} \to \{1, \dots, (d+1)_{2n+1}\}$.

Proof of Proposition 2.4.10. We know that

$$\log(S_t(\ell)) = -\frac{1}{2}\int_0^t V_s(\ell)\mathrm{d}s + \int_0^t \sigma_s^S(\ell)\mathrm{d}B_s$$

and we will calculate each integral separately. We start with the first one.

$$\begin{aligned}
\int_0^t V_s(\ell)\mathrm{d}s &= \int_0^t \sum_{i=0}^{N} \sum_{|I|,|J|\leq n} \ell_I(T_i)\ell_J(T_i) 1_{[T_i,T_{i+1})} \langle e_I \sqcup\!\sqcup e_J, \widehat{\mathbb{X}}_s\rangle \mathrm{d}s \\
&= \sum_{i=0}^{N} \sum_{|I|,|J|\leq n} \ell_I(T_i)\ell_J(T_i) \int_{t\wedge T_i}^{t\wedge T_{i+1}} \langle e_I \sqcup\!\sqcup e_J, \widehat{\mathbb{X}}_s\rangle \mathrm{d}s \\
&= \sum_{i=0}^{N} \sum_{|I|,|J|\leq n} \ell_I(T_i)\ell_J(T_i) \left(\int_0^{t\wedge T_{i+1}} \langle e_I \sqcup\!\sqcup e_J, \widehat{\mathbb{X}}_s\rangle \mathrm{d}s - \int_0^{t\wedge T_i} \langle e_I \sqcup\!\sqcup e_J, \widehat{\mathbb{X}}_s\rangle \mathrm{d}s \right) \\
&= \sum_{i=0}^{N} \sum_{|I|,|J|\leq n} \ell_I(T_i)\ell_J(T_i) \left(\langle (e_I \sqcup\!\sqcup e_J) \otimes e_0, \widehat{\mathbb{X}}_{t\wedge T_{i+1}}\rangle - \langle (e_I \sqcup\!\sqcup e_J) \otimes e_0, \widehat{\mathbb{X}}_{t\wedge T_i}\rangle \right) \\
&= \sum_{i=0}^{N} \left[\ell(T_i)^\top \left(Q^0(t\wedge T_{i+1}) - Q^0(t\wedge T_i) \right) \ell(T_i) \right].
\end{aligned}$$

Using similar arguments and Lemma 3.10 in Cuchiero et al. (2022b), the second integral

yields

$$\begin{aligned}\int_0^t \sigma_s^S(\ell)\mathrm{d}B_s &= \int_0^t \sum_{i=0}^{N}\sum_{|I|\le n} \ell_I(T_i)1_{[T_i,T_{i+1})}\langle e_I, \widehat{\mathbb{X}}_s\rangle \mathrm{d}B_s \\ &= \sum_{i=0}^{N}\sum_{|I|\le n} \ell_I(T_i)\int_{t\wedge T_i}^{t\wedge T_{i+1}} \langle e_I, \widehat{\mathbb{X}}_s\rangle \mathrm{d}B_s \\ &= \sum_{i=0}^{N}\sum_{|I|\le n} \ell_I(T_i)\left(\int_0^{t\wedge T_{i+1}} \langle e_I, \widehat{\mathbb{X}}_s\rangle \mathrm{d}B_s - \int_0^{t\wedge T_i} \langle e_I, \widehat{\mathbb{X}}_s\rangle \mathrm{d}B_s\right) \\ &= \sum_{i=0}^{N}\sum_{|I|\le n} \ell_I(T_i)\langle \tilde{e}_I^B, \widehat{\mathbb{Z}}_{t\wedge T_{i+1}} - \widehat{\mathbb{Z}}_{t\wedge T_i}\rangle,\end{aligned}$$

and the claim follows. □

2.5. Joint calibration of SPX and VIX options

Here we consider again the model introduced in (2.1.1)-(2.1.2) under Assumption 2.1.3. Note that we just work with call options, but the setup can easily be extended also to other liquid options on the market. Again we denote by $\mathcal{T}^{\mathrm{VIX}}$ and $\mathcal{T}^{\mathrm{SPX}}$ the maturities set for options written on VIX and SPX, respectively. Similarly we use the notation $\mathcal{K}^{\mathrm{VIX}}$ and $\mathcal{K}^{\mathrm{SPX}}$ for the corresponding strikes. The functional to be minimized in order to achieve a joint calibration of the SPX/VIX options reads as follows:

$$L_{\mathrm{joint}}(\ell,\lambda) := \lambda L_{\mathrm{SPX}}(\ell) + (1-\lambda)L_{\mathrm{VIX}}(\ell), \tag{2.5.1}$$

where $\lambda \in (0,1)$ and

- $L_{\mathrm{VIX}}(\ell)$ is as in (2.3.12), i.e.

  $$L_{\mathrm{VIX}}(\ell) := \sum_{T\in\mathcal{T}^{VIX}, K\in\mathcal{K}^{VIX}} \mathcal{L}\left(\pi_{\mathrm{VIX}}^{\mathrm{model}}(\ell,T,K), \pi_{\mathrm{VIX}}^{b,a}(T,K), \sigma_{\mathrm{VIX}}^{b,a}(T,K), F_{\mathrm{VIX}}^{\mathrm{model}}(\ell,T), F_{\mathrm{VIX}}^{mkt}(T)\right),$$

  with π_{VIX}^{model} and F_{VIX}^{model} as in (2.3.11) for $\mathrm{VIX}_T(\ell,\omega_i)$ defined as in (2.3.4);

- $L_{\mathrm{SPX}}(\ell)$ is the SPX loss function given by

  $$L_{\mathrm{SPX}}(\ell) := \sum_{T\in\mathcal{T}^{\mathrm{SPX}}, K\in\mathcal{K}^{\mathrm{SPX}}} \mathcal{L}(\pi_{\mathrm{SPX}}^{\mathrm{model}}(\ell,T,K), \pi_{\mathrm{SPX}}^{b,a}(T,K), \sigma_{\mathrm{SPX}}^{b,a}(T,K)),$$

  for a real-valued function $\mathcal{L}$. Observe that with a slight abuse of notation we denote this function as the one for L_{VIX}, but for SPX options we do not have to calibrate to futures, hence the last term of (2.3.14) vanishes.

By Proposition 2.4.5 the SPX call option payoff with maturity $T > 0$ and a strike price $K > 0$ reads in our model as follows

$$e^{-rT}(\tilde{S}_T(\ell) - K)^+ = e^{-rT}\left(\exp\left\{(r-q)T - \frac{1}{2}\ell^\top Q^0(t)\ell + \sum_{|I|\leq n} \ell_I \langle \tilde{e}_I^B, \widehat{\mathbb{Z}}_T\rangle\right\} - K\right)^+,$$

where $\tilde{S}$ denotes the undiscounted process as discussed in Remark 2.1.2 and $r, q > 0$ the interest rate and the dividends, respectively. Recall also that the call option payoff written on the VIX is given by

$$e^{-rT}(\mathrm{VIX}_T(\ell) - K)^+ = e^{-rT}\left(\sqrt{\frac{1}{\Delta}\ell^\top Q(T,\Delta)\ell} - K\right)^+ = e^{-rT}\left(\frac{1}{\sqrt{\Delta}}\|U_T^\top \ell\| - K\right)^+,$$

where U_T denotes the upper-triangular matrix of the Cholesky decomposition of the symmetric positive semidefinite matrix $Q(T,\Delta)$.

2.5.1. Numerical results

In the present section we discuss two different ways of approaching the joint calibration problem that can be found in the recent literature.

(i) The first approach consists in choosing for instance the first maturity of SPX and VIX to coincide (or differ up to two days, see Remark 0.0.1), i.e., $T_1^{\mathrm{SPX}} = T_1^{\mathrm{VIX}}$ and then for $j \geq 2$, $T_j^{\mathrm{SPX}} = T_{j-1}^{\mathrm{VIX}} + \Delta$, see for instance Guyon (2021); Guo et al. (2021); Guyon and Lekeufack (2022).

(ii) The second approach is to consider $\mathcal{T}^{\mathrm{SPX}} = \mathcal{T}^{\mathrm{VIX}}$, i.e., to choose the same (or close together, see Remark 0.0.1) maturities both for SPX and VIX options. This perspective has been adopted for instance by Gatheral et al. (2018); Rosenbaum and Zhang (2021); Grzelak (2022); Bondi et al. (2022b).

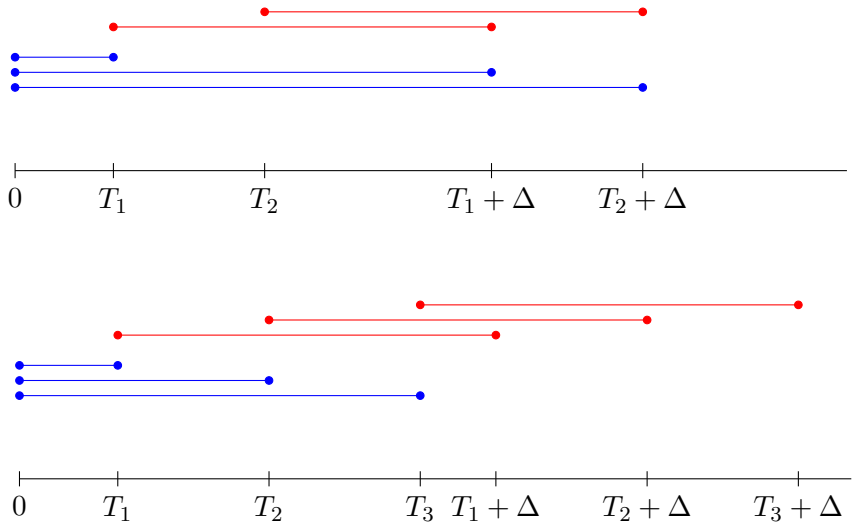


Figure 2.3.: The blue lines denote the time interval where the dynamics of the variance process influence the SPX option up to the maturity time. For instance the shortest blue line denotes the time interval where the dynamics of the variance process enter up to maturity T_1. Similarly the red lines denote the corresponding ones for the VIX, as for instance the variance process enters here in the time integral on $[T_1, T_1 + \Delta]$, see (2.3.3). On the upper graph a representation of the joint calibration approach (i) is given where we notice that the maturities of the VIX are chosen so that there is a maximal overlap with the ones of the SPX. On the lower graph a representation of approach (ii) is given where the maturities $\mathcal{T} = \{T_1, T_2, T_3\}$ are considered. We observe that there is less overlap between the maturities of the SPX and VIX than in approach (i).

Both approaches deal with the joint modeling of SPX and VIX options and in order to be consistent with both viewpoints taken in the literature, we show how our signature-based model solves the joint calibration within both settings. For this reason we split the rest of the section in two subsection and discuss them separately.

First approach

Here we consider call options for both indices on the trading day 02/06/2021, as in Guyon and Lekeufack (2022). Maturities are reported in the following tables with the corresponding range of strikes (in percentage) with respect to the spot and the market's futures prices.

$T_1^{\text{VIX}} = 0.0383$	$T_2^{\text{VIX}} = 0.0767$
(90%,220%)	(90%,220%)

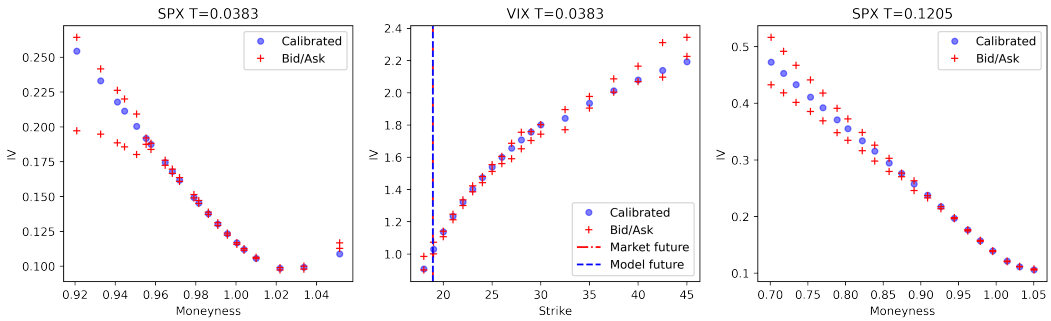
$T_1^{\text{SPX}} = 0.0383$	$T_2^{\text{SPX}} = 0.1205$	$T_3^{\text{SPX}} = 0.1588$
(92%,105%)	(70%,105%)	(80%,120%)

We stress that the shortest maturity considered is of 14 days for both SPX and VIX, then the second and third maturity of the SPX are 44 days and 58 days, respectively, and the second one for the VIX is 28 days. Moreover, we consider a high moneyness level (up to 220%) for VIX options, usually rather difficult to fit. Regarding our modeling choice we fix $d = 3$, $n = 3$ and choose the primary process X to be a three Ornstein-Uhlenbeck processes (see Example 2.2.5) with parameters

$$\kappa = (0.1, 25, 10)^\top, \qquad \theta = (0.1, 4, 0.08)^\top, \qquad \sigma = (0.7, 10, 5)^\top,$$

$$\rho = \begin{pmatrix} 1 & 0.213 & -0.576 & 0.329 \\ \cdot & 1 & -0.044 & -0.549 \\ \cdot & \cdot & 1 & -0.539 \\ \cdot & \cdot & \cdot & 1 \end{pmatrix}, \qquad X_0 = (1, 0.08, 2)^\top.$$

Note that with this configuration we need to calibrate 85 parameters, i.e., $\ell \in \mathbb{R}^{85}$. Concerning the calibration task, we solve (2.5.1) with $\lambda = 0.35$ with $N_{MC} = 80000$ Monte Carlo samples for the previous maturities and strikes. Furthermore we specify the loss function $\mathcal{L}$ as in (2.3.14) for $\beta = 1$ both for SPX and VIX.



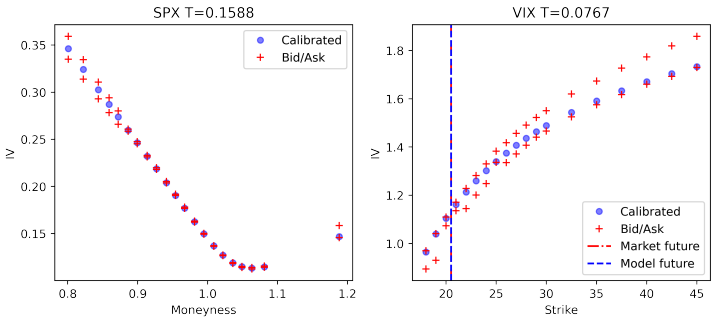


Figure 2.4.: In blue the calibrated implied volatilities smiles from top-left at maturities $T_1^{\text{SPX}}, T_1^{\text{VIX}}, T_2^{\text{SPX}}, T_3^{\text{SPX}}, T_2^{\text{VIX}}$. In red the corresponding bid-ask spreads. In the graphs of the VIX smiles the red dashed line indicates the market future price at maturity and the blue dashed line the calibrated one.

We report also the relative absolute error between the market future prices and the calibrated ones as defined in (2.3.15):

$T_1^{\text{VIX}} = 0.0383$	$T_2^{\text{VIX}} = 0.0767$
$\varepsilon_{T_1^{\text{VIX}}} = 9.8 \cdot 10^{-6}$	$\varepsilon_{T_2^{\text{VIX}}} = 6.6 \cdot 10^{-8}$

Sample time-series of SPX and VIX Let $\ell^\star \in \mathbb{R}^{85}$ be the calibrated parameters found for the fit in Figure 2.4. We then fix $T = 60$ days the longest considered maturity for the SPX and sample a trajectory for $(V_t(\ell^\star))_{t\in[0,T]}$, $(\text{VIX}_t(\ell^\star))_{t\in[0,T]}$, $(S_t(\ell^\star))_{t\in[0,T]}$. Precisely, we sample 12 grid points per day, i.e. we consider a 2 hours sampling per calendar day, for a total of $N = 720$ grid points. The results are reported in Figure 2.5.

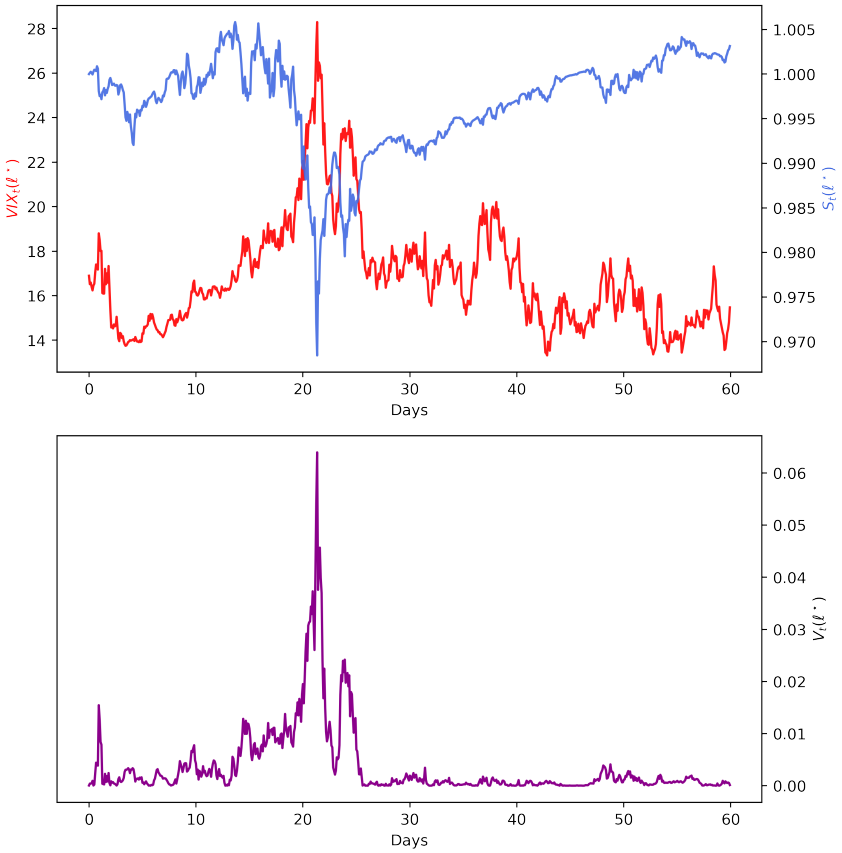


Figure 2.5.: On the top: one realization if our calibrated model $S(\ell^\star)$ for the SPX (in blue) and the corresponding calibrated VIX (in red). On the bottom: the corresponding realization of the calibrated variance process $V(\ell^\star)$.

Observe that even though ℓ^* was only calibrated to option prices, the trajectories produced by the model are economically reasonable and also in line with several stylized fact, such as negative correlation between SPX and VIX or volatility clustering. To obtain the dynamics under the physical measure, these trajectories could still be adjusted by an appropriate market prices of risk, but the path dependent quantities like the volatility of volatility or the correlation stay the same.

Next, we consider the case of time-varying parameters as introduced in Section 2.3.5 and Section 2.4.2 for VIX and SPX, respectively.

The case of time-varying parameters Although the joint calibration is mainly considered for short-dated options in the literature as VIX options are then more liquid, it is even more challenging to provide an accurate fit for both, short and long maturities. Allowing the parameters ℓ of our model to depend on time, in particular on the maturities, we are able to calibrate additionally to longer maturities than the ones considered in Figure 2.4.

We consider for the choice of the primary process the same configuration as we used for Figure 2.4. The procedure of our time-varying calibration routine is as follows:

1. Calibrate jointly $T_1^{\mathrm{SPX}}, T_1^{\mathrm{VIX}}$ and T_2^{SPX}.

2. Use the parameters from the calibration of T_j^{SPX} and T_{j-1}^{VIX} to fit jointly the maturities T_{j+1}^{SPX} and T_j^{VIX} for $j = 2, \ldots, J$.

We consider $J = 4$, where the last maturity for the SPX is 170 days, and the last maturity for the VIX is 77 days. For the first two maturities of the SPX and the first of the VIX we consider the same moneyness ranges as in Figure 2.4, hence we specify here only the ranges for the longer maturities:

$T_2^{\mathrm{VIX}} = 0.1342$	$T_3^{\mathrm{VIX}} = 0.2875$	$T_4^{\mathrm{VIX}} = 0.3833$
(90%,330%)	(78%,395%)	(80%,405%)

$T_3^{\mathrm{SPX}} = 0.2163$	$T_4^{\mathrm{SPX}} = 0.3696$	$T_5^{\mathrm{SPX}} = 0.4654$
(75%,125%)	(60%,135%)	(50%,145%)

We observe that for this choice of maturities Assumption 2.3.10 is satisfied. Hence the second expression for the time-varying VIX is used from Proposition (2.3.11). On the other hand in order to compute the price of the SPX options in the time-varying case we use the representation of the log-price provided in Proposition 2.4.10. We employ $\lambda = 0.25$ for each calibration within the rolling procedure and we consider always as loss function $\mathcal{L}^{\beta}$ as introduced in (2.3.14) for $\beta = 0$. It is worth mentioning that the initial parameter search discussed in Remark 2.3.7, has been employed for calibrating jointly $T_1^{\mathrm{SPX}}, T_1^{\mathrm{VIX}}$ and T_2^{SPX}, while for the next slices we have considered the previously calibrated parameters as starting point of the optimization.

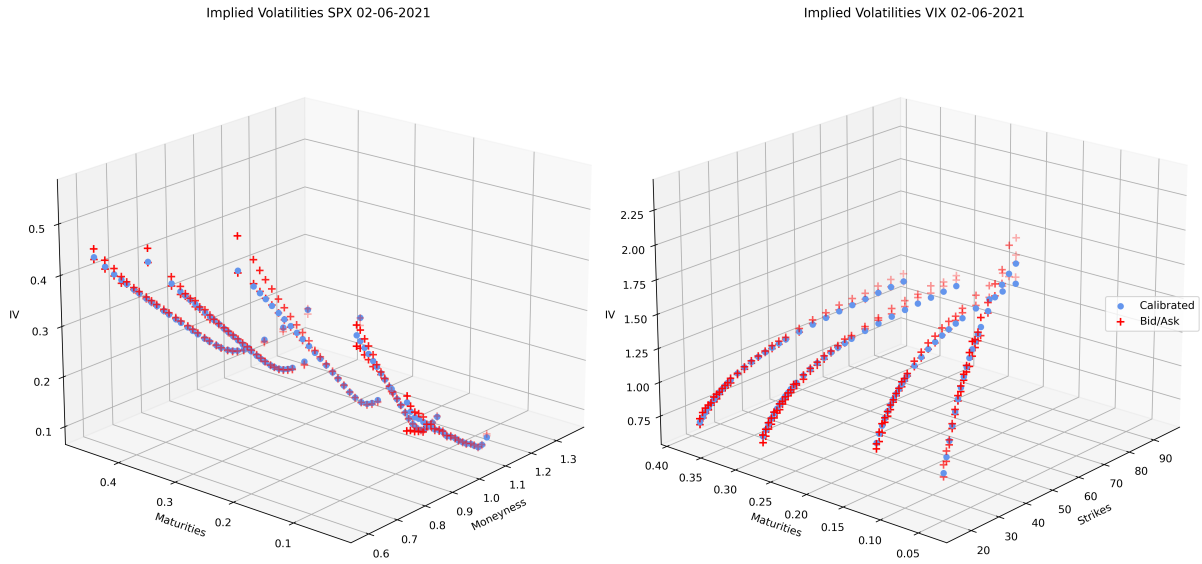


Figure 2.6.: On the left hand side: SPX smiles, in blue the calibrated implied volatilities and in red the bid-ask spreads. On the right hand side: VIX smiles, in blue the calibrated implied volatilities and in red the bid-ask spreads.

Finally we report the absolute relative error on the VIX futures:

$T_1^{\text{VIX}} = 0.0383$	$T_2^{\text{VIX}} = 0.1205$	$T_3^{\text{VIX}} = 0.1588$	$T_4^{\text{VIX}} = 0.2108$
$\varepsilon_{T_1^{\text{VIX}}} = 1.2 \cdot 10^{-11}$	$\varepsilon_{T_2^{\text{VIX}}} = 2.3 \cdot 10^{-5}$	$\varepsilon_{T_3^{\text{VIX}}} = 3.9 \cdot 10^{-5}$	$\varepsilon_{T_4^{\text{VIX}}} = 2.9 \cdot 10^{-6}$

Second approach

Let us now consider the second approach described at the beginning of Section 2.5.1. Specifically, we consider a unique set of maturities for both SPX and VIX on the trading day of 02/06/2021. For this study, we do not consider time-varying parameters. In the following table we report the moneyness ranges for SPX options in the second row and on the last row the ones for VIX options:

$T_1 = 0.0192$	$T_2 = 0.0383$	$T_3 = 0.0575$	$T_4 = 0.0767$
(97%,104%)	(92%,105%)	(90%,110%)	(85%,110%)
(90%,200%)	(90%,220%)	(90%,290%)	(90%,290%)

We consider $\lambda = 0.5$ and as loss function $\mathcal{L}$ we employ (2.3.14) with $\beta = 1$ for VIX options and the same (without futures) for SPX options.

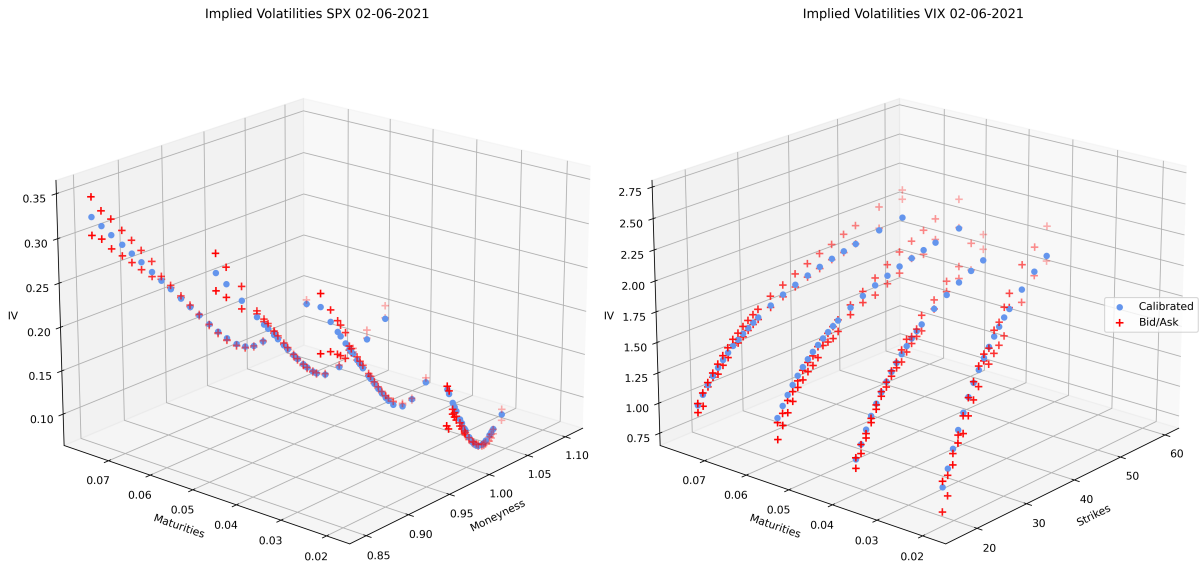


Figure 2.7.: On the left hand side: SPX smiles, in blue the calibrated implied volatilities and in red the bid-ask spreads. On the right hand side: VIX smiles, in blue the calibrated implied volatilities and in red the bid-ask spreads.

We additionally report the relative absolute error of the calibrated VIX futures:

$T_1^{\text{VIX}} = 0.0192$	$T_2^{\text{VIX}} = 0.0383$	$T_3^{\text{VIX}} = 0.0575$	$T_4^{\text{VIX}} = 0.0767$
$\varepsilon_{T_1^{\text{VIX}}} = 6.2 \cdot 10^{-3}$	$\varepsilon_{T_2^{\text{VIX}}} = 1.2 \cdot 10^{-5}$	$\varepsilon_{T_3^{\text{VIX}}} = 1.3 \cdot 10^{-2}$	$\varepsilon_{T_4^{\text{VIX}}} = 1.6 \cdot 10^{-3}$

3. Comments on the code for calibration of signature-based models

A crucial role when tackling the calibration of financial models, is played by an efficient numerical implementation of pricing functionals and of the sampling schemes for models themselves.

This self-contained chapter is to point out than one of the main contribution are open-source codes for the calibration tasks discussed in Chapter 1 and Chapter 2.

The codes for Chapter 1 have been written by the author of the present thesis and can be found in the GitHub repository https://github.com/GuidoGazzani-ai/sigsde_calibration, where the following implementations are reported.

- The tilde-transformation introduced in Section 1.2.1 for multi-dimensional processes.
- Extrapolation of the Brownian motions as discussed in Example 1.4.1 and Example 1.4.3 for a given parameters' configuration of the Heston model and of the SABR model.
- Calibration to simulated time-series data from the previous models as introduced in Section 1.4.1.
- An efficient (memory-wise) parallelized sampling procedure for the primary process at maturities and its tilde-transformation.
- An efficient (memory-wise) parallelized sampling procedure for the primary process at maturities and its tilde-transformation for calibration to market option prices.
- A fast sampling routine for the model described in Equation (1.4.9).
- The adaptation considered for the joint calibration of time-series data and option prices, as described in Section 1.4.3, for data generated by a Heston model.

The codes for Chapter 2 have been written by the author of the present thesis and by Janka Möller include the following implementations and are available upon request.

- A code that returns the conditional expected signature of a time-extended d-dimensional Ornstein-Uhlenbeck process as in Theorem 2.2.4.
- Computation of VIX_T for several $T > 0$ as in Theorem 2.3.1, in particular efficient computation of the matrix $Q(T, \Delta)$.
- Calibration task for both VIX and SPX options, jointly and separate with a flexible code and the time-varying parameter version of it.

Part II.

Robust risk measures

4. Risk measures under model uncertainty: a Bayesian viewpoint

This chapter is based on Cuchiero et al. (2022a), a joint work with Christa Cuchiero and Irene Klein.

4.1. Preliminaries on risk measures

This section is dedicated to recall some basic concepts and definitions of risk measures. After introducing some notation, we start here with risk measures being defined in a pointwise way (see Section 4.1.2).

4.1.1. Notation

Let $(\Omega, \mathcal{F})$ be a measurable space and let $\mathcal{M}_1 = \mathcal{M}_1(\Omega, \mathcal{F})$ be the set of probability measures defined on it. We equip it with the topology of weak convergence (in the probabilistic sense), which corresponds to the functional analytic notion of weak-$*$-convergence when Ω is compact. Our aim is to quantify the risk of a random variable X on $(\Omega, \mathcal{F})$ that describes a certain financial risky position. As a first step we will look at risky claims taking values in the following set:

$$\mathcal{X} = \{X : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R})), \text{ measurable and bounded}\},$$

where $\mathcal{B}(\mathbb{R})$ denotes the Borel-σ-algebra on $\mathbb{R}$. By "bounded" we mean that for each $X \in \mathcal{X}$ there exists $K > 0$ such that $|X(\omega)| \leq K$, for *all* $\omega \in \Omega$. Later we will relax the condition to an almost sure condition. If a probability measure $P \in \mathcal{M}_1$ is given, we will write $L^\infty(P)$ for the space $L^\infty(\Omega, \mathcal{F}, P)$.

4.1.2. Pointwise definition of risk measures

Let us recall in the following the definition of monetary risk measures defined on the space $\mathcal{X}$. We follow closely the approach of Föllmer and Schied (2011) in terms of axioms of a risk measure. We point out that in the literature, in particular in the context of insurance related topics, there is present another approach, see for instance Embrechts et al. (2002); Pflug and Pichler (2016). The difference is basically a change of sign which results to reversed monotonicity assumptions. The change of sign is due to the view of losses in insurance as non-negative random variables whereas in the case of our definition losses are given by non-positive random variables.

Definition 4.1.1. A map $\rho : \mathcal{X} \to \mathbb{R}$ is said to be a monetary risk measure if:

(i) $X \geq 0$ then $\rho(X) \leq 0$, for all $X \in \mathcal{X}$;

(ii) $X \geq Y$ then $\rho(X) \leq \rho(Y)$, for all $X, Y \in \mathcal{X}$;

(iii) $\rho(a + X) = \rho(X) - a$, for every $a \in \mathbb{R}$ and for all $X \in \mathcal{X}$.

Notice that we mean that each property holds for all $\omega \in \Omega$.
Given a monetary risk measure ρ the corresponding acceptance set is defined as follows:

$$C = \{X \in \mathcal{X} : \rho(X) \leq 0\}. \tag{4.1.1}$$

Note that we get ρ back from the acceptance set via the equation

$$\rho(X) = \inf\{m \in \mathbb{R} : m + X \in C\}.$$

Let us also state the definition of a convex risk measure, where the convexity assumption is economically meaningful since diversification should not increase the risk of a financial position. We also recall the main results concerning their dual representation.

Definition 4.1.2. A map $\rho : \mathcal{X} \to \mathbb{R}$ is said to be a convex risk measure if it is a monetary risk measure and if for every $\lambda \in (0, 1)$ and for all $X, Y \in \mathcal{X}$

$$\rho(\lambda X + (1 - \lambda)Y) \leq \lambda\rho(X) + (1 - \lambda)\rho(Y).$$

The acceptance set of a convex risk measure is defined exactly as for monetary risk measures in (4.1.1).
Assuming moreover, that a convex risk measure ρ on $\mathcal{X}$ is continuous from below in the sense that for any sequence $(X_n)_{n\in\mathbb{N}} \subset \mathcal{X}$,

$$X_n \uparrow X \text{pointwise} \quad \Rightarrow \quad \rho(X_n) \downarrow \rho(X), \tag{4.1.2}$$

as $n \to +\infty$, then we obtain the following dual representation

$$\rho(X) = \sup_{Q\in\mathcal{M}_1} \{\mathbb{E}_Q[-X] - \alpha(Q)\}, \qquad \forall X \in \mathcal{X}, \tag{4.1.3}$$

where $\alpha : \mathcal{M}_1 \to \mathbb{R} \cup \{+\infty\}$ is a penalty function, see Proposition 4.21 in Föllmer and Schied (2011). The penalty function itself it is not unique in general. Note however that due to the continuity from below any penalty function is concentrated on $\mathcal{M}_1$ instead of finitely additive set functions denoted by $\mathcal{M}_{1,f}$, meaning that $\alpha(Q) = \infty$ for all $Q \in \mathcal{M}_{1,f} \setminus \mathcal{M}_1$ so that it is sufficient to take the supremum in (4.1.3) only over $\mathcal{M}_1$. The penalty function can be chosen to attain the smallest values by setting

$$\alpha_{\min}(Q) := \sup_{X\in C} \mathbb{E}_Q[-X] = \sup_{X\in\mathcal{X}} \{\mathbb{E}_Q[-X] - \rho(X)\}, \qquad \forall Q \in \mathcal{M}_1, \tag{4.1.4}$$

i.e. $\alpha_{\min}$ corresponds to the Fenchel–Legendre transform or conjugate function of the convex function ρ on $\mathcal{X}$, see Theorem 4.15 as well as Remark 4.16 and Remark 4.17 in Föllmer and Schied (2011).

Finally let us recall the notion of coherent risk measures.

Definition 4.1.3. A convex risk measure $\rho : \mathcal{X} \to \mathbb{R}$ is called coherent risk measure if it satisfies positive homogeneity, i.e. for $\lambda \geq 0$ and all $X \in \mathcal{X}$, $\rho(\lambda X) = \lambda \rho(X)$.

Note that by Corollary 4.18 in Föllmer and Schied (2011) the minimal penalty function of a coherent measure that is continuous from below always takes the form

$$\alpha(Q) = \begin{cases} 0 & \text{if } Q \in \mathcal{Q}, \\ +\infty & \text{otherwise,} \end{cases} \tag{4.1.5}$$

for some convex set $\mathcal{Q} \subseteq \mathcal{M}_1$. For the minimal penalty function the set $\mathcal{Q}$ is maximal. Note that it could potentially be also represented by another α of form (4.1.5) on a smaller set $\mathcal{Q}$ (see Proposition 4.14 in Föllmer and Schied (2011)).

4.2. Risk measures with respect to reference probability measures

From now on, we additionally suppose that we are given a probability measure P on $(\Omega, \mathcal{F})$ with respect to which we shall define risk measures such that the domination properties mentioned in the introduction hold true. We shall pursue two different approaches. In both of them we start by fixing a monetary risk measure ρ on the space $\mathcal{X}$ of bounded measurable random variables. We then define two possible ways how to deduce from ρ a risk measure depending on P. The first approach works for general monetary risk measures, whereas the second assumes convexity.

In Section 4.2.3 and Section 4.2.4, we shall then compare these two approaches in the case of convex as well a convex and law-invariant risk measures.

4.2.1. First approach applicable to general monetary risk measures

The first approach consists in defining a risk measure ρ^P as follows:

Definition 4.2.1. Let $P \in \mathcal{M}_1$ and ρ a monetary risk measure. Then the corresponding risk measure with respect to P is denoted by $\rho^P : L^\infty(P) \to \mathbb{R}$ and defined by

$$\rho^P(X) = \inf_{\{\widetilde{X} \in \mathcal{X} : P(\widetilde{X} = X) = 1\}} \rho(\widetilde{X}),$$

for all $X \in L^\infty(P)$. Moreover, the corresponding acceptance set is given by

$$\begin{aligned} C^P &= \{X \in L^\infty(P) : \inf_{\{\widetilde{X} \in \mathcal{X} : P(\widetilde{X} = X) = 1\}} \rho(\widetilde{X}) \leq 0\} \\ &= \{X \in L^\infty(P) : \rho^P(X) \leq 0\}. \end{aligned}$$

Observe that the definition of C^P implies that ρ^P satisfies the properties in Definition 4.1.1 *not* for all $\omega \in \Omega$, but just for the scenarios $\omega \in \Omega$ that occur P-a.s.

Remark 4.2.2. Let $P, P' \in \mathcal{M}_1$ be such that $P' \ll P$. Then, clearly

$$\{\widetilde{X} \in \mathcal{X} : P(\widetilde{X} = X) = 1\} \subseteq \{\widetilde{X} \in \mathcal{X} : P'(\widetilde{X} = X) = 1\},$$

thus $C^P \subseteq C^{P'}$ and

$$\rho^{P'}(X) \leq \rho^P(X), \qquad \forall X \in L^\infty(P).$$

In particular, if $P \sim P'$, we have $\rho^{P'}(X) = \rho^P(X)$. When the 'worst case risk measure' takes the role of P, this leads to the desired properties (i) and (ii) mentioned in the introduction.

The following lemma states that the property of being a monetary or convex risk measure also translates to ρ^P.

Lemma 4.2.3. If ρ is a monetary (convex respectively) risk measure, then ρ^P is a monetary (convex respectively) risk measure.

Proof. We will prove the properties of a monetary risk measure P-a.s., namely (i)-(iii) and finally address the convexity here denoted by (iv).

(i) Let $X \geq 0$ P-a.s. We can choose a representative $\bar{X} = X1_{\{X \geq 0\}}$, then $P(X = \bar{X}) = 1$. $\bar{X}(\omega) \geq 0$ for all ω, hence $\rho(\bar{X}) \leq 0$. By definition it follows that $\rho^P(X) \leq \rho(\bar{X}) \leq 0$.

(ii) Let $X, Y \in L^\infty(P)$ such that $X \geq Y$ P-a.s. Since $\rho^P(Y) = \inf_{\{\widetilde{Y} \in \mathcal{X} : P(\widetilde{Y} = Y) = 1\}} \rho(\widetilde{Y})$, for each $n \geq 1$, we can choose $\widetilde{Y}^n \in \mathcal{X}$, where $P(\widetilde{Y}^n = Y) = 1$ and such that $\rho(\widetilde{Y}^n) \leq \rho^P(Y) + \frac{1}{n}$. Consider now any $\widetilde{X}$ with $P(\widetilde{X} = X) = 1$ and modify this for each n as follows:

$$\widetilde{X}^n = X1_{\{\widetilde{Y}^n = Y\} \cap \{X \geq Y\}} + \max\{\widetilde{X}, \widetilde{Y}^n\} 1_{\{\widetilde{Y}^n \neq Y\} \cup \{X < Y\}}.$$

Clearly, $\widetilde{X}^n \in \mathcal{X}$ and $P(\widetilde{X}^n = X) = P((\widetilde{Y}^n = Y) \cap (X \geq Y)) = 1$, as $P(X \geq Y) = 1$. Then we get, for all ω,

$$\widetilde{X}^n(\omega) \geq Y1_{\{\widetilde{Y}^n = Y\} \cap \{X \geq Y\}}(\omega) + \widetilde{Y}^n 1_{\{\widetilde{Y}^n \neq Y\} \cup \{X < Y\}}(\omega) = \widetilde{Y}^n(\omega).$$

Therefore as ρ satisfies (ii) we get $\rho(\widetilde{X}^n) \leq \rho(\widetilde{Y}^n) \leq \rho^P(Y) + \frac{1}{n}$. This implies

$$\rho^P(X) \leq \inf_{n \geq 1} \rho(\widetilde{X}^n) \leq \rho^P(Y),$$

and (ii) follows for ρ^P.

(iii) Observe that the property of cash-invariance or translation-invariance holds trivially as for all $X \in L^\infty(P)$ and $a \in \mathbb{R}$ we have

$$\rho^P(X + a) = \inf_{\{\widetilde{X} \in \mathcal{X} : P(\widetilde{X} = X + a) = 1\}} \rho(\widetilde{X}) = \inf_{\{\widetilde{X} \in \mathcal{X} : P(\widetilde{X} = X) = 1\}} \rho(\widetilde{X} + a) = \rho^P(X) - a.$$

(iv) Let us now assume that ρ is convex. Suppose X and Y are in $L^\infty(P)$ and $\lambda \in (0,1)$. Let $\widetilde{X}, \widetilde{Y}$ be in $\mathcal{X}$ such that $P(X = \widetilde{X}) = P(Y = \widetilde{Y}) = 1$. Define $\bar{Z} = \lambda\widetilde{X} + (1-\lambda)\widetilde{Y}$. Then, clearly, $\bar{Z} \in \{\widetilde{Z} \in \mathcal{X} : P(\widetilde{Z} = \lambda X + (1-\lambda)Y) = 1\}$. It follows from the convexity of ρ that

$$\begin{aligned}\rho^P(\lambda X + (1-\lambda)Y) &\leq \inf_{\{\bar{Z}\in\mathcal{X}:\exists\widetilde{X},\widetilde{Y}:\bar{Z}=\lambda\widetilde{X}+(1-\lambda)\widetilde{Y}\}} \rho(\bar{Z}) \\ &\leq \inf_{\{(\widetilde{X},\widetilde{Y})\in\mathcal{X}\times\mathcal{X}:P(X=\widetilde{X})=P(Y=\widetilde{Y})=1\}} \{\lambda\rho(\widetilde{X}) + (1-\lambda)\rho(\widetilde{Y})\} \\ &= \lambda\rho^P(X) + (1-\lambda)\rho^P(Y),\end{aligned}$$

showing that ρ^P is convex.

□

4.2.2. Second approach applicable to convex risk measures: fixing a penalty function

Let us now turn to the second approach how to introduce risk measures depending on some reference measure, which works in the case of convex risk measures that are continuous from below in the sense of (4.1.2). Indeed, fix some penalty function $\alpha : \mathcal{M}_1 \to \mathbb{R} \cup \{+\infty\}$ and define a convex risk measure $\rho : \mathcal{X} \to \mathbb{R}$ via

$$\rho(X) := \sup_{Q\in\mathcal{M}_1} \{\mathbb{E}_Q[-X] - \alpha(Q)\}, \qquad \forall X \in \mathcal{X}. \tag{4.2.1}$$

Here, instead of looking at the risk measure ρ^P, we modify α, which then gives rise to a new risk measure $\widehat{\rho}^P$, depending on the set of absolutely continuous measures with respect to P, denoted by

$$\mathcal{P}^P = \{Q \in \mathcal{M}_1 : Q \ll P\}.$$

To be precise, let α be the fixed penalty function as of (4.2.1) and define $\widehat{\rho}^P : L^\infty(P) \to \mathbb{R}$ as follows:

$$\widehat{\rho}^P(X) = \sup_{Q\in\mathcal{P}^P} \{\mathbb{E}_Q[-X] - \alpha(Q)\}. \tag{4.2.2}$$

This means that we actually have to use the following penalty function $\alpha^P(Q)$ in order to express the supremum again over all probability measures $\mathcal{M}_1$. Indeed, define

$$\alpha^P(Q) := \begin{cases} \alpha(Q) & \text{for } Q \in \mathcal{P}^P, \\ +\infty & \text{for } Q \in \mathcal{M}_1 \setminus \mathcal{P}^P, \end{cases} \tag{4.2.3}$$

then we exactly get

$$\widehat{\rho}^P(X) = \sup_{Q\in\mathcal{M}_1} \{\mathbb{E}_Q[-X] - \alpha^P(Q)\}, \quad \forall X \in L^\infty(P) \tag{4.2.4}$$

As before the corresponding acceptance set of $\widehat{\rho}^P$ is defined as

$$\widehat{C}^P = \{X \in L^\infty(P) : \widehat{\rho}^P(X) \leq 0\}. \tag{4.2.5}$$

Remark 4.2.4. Let $P, P' \in \mathcal{M}_1$ such that $P' \ll P$. Then, obviously, if $Q \ll P'$ then $Q \ll P$ as well. Hence $\mathcal{P}^{P'} \subseteq \mathcal{P}^P$. This implies that for all $X \in L^\infty(P)$

$$\widehat{\rho}^{P'}(X) = \sup_{Q \in \mathcal{P}^{P'}} \{\mathbb{E}_Q[-X] - \alpha(Q)\} \leq \sup_{Q \in \mathcal{P}^P} \{\mathbb{E}_Q[-X] - \alpha(Q)\} = \widehat{\rho}^P(X).$$

In particular, if $P \sim P'$, then $\widehat{\rho}^P = \widehat{\rho}^{P'}$. When the 'worst case risk measure' takes the role of P, this leads again to the desired properties (i) and (ii) mentioned in the introduction.

4.2.3. Comparing the two approaches for convex risk measures

Our goal is now to compare these two different approaches in the case of convex risk measures and establish conditions under which they coincide. To this end we let ρ be the convex risk measure given by (4.2.1) for some fixed α.

According to the first approach define now for $X \in L^\infty(P)$

$$\rho^P(X) = \inf_{\{\widetilde{X}: P(\widetilde{X}=X)=1\}} \rho(\widetilde{X}) = \inf_{\{\widetilde{X}: P(\widetilde{X}=X)=1\}} \sup_{Q \in \mathcal{M}_1} \{\mathbb{E}_Q[\widetilde{X}] - \alpha(Q)\}.$$

The obvious question at that point is when does the following minimax inequality

$$\begin{aligned} \widetilde{\rho}^P(X) :&= \sup_{Q \in \mathcal{M}_1} \inf_{\{\widetilde{X} \in \mathcal{X}: P(\widetilde{X}=X)=1\}} \{\mathbb{E}_Q[-\widetilde{X}] - \alpha(Q)\} \\ &\leq \inf_{\{\widetilde{X} \in \mathcal{X}: P(\widetilde{X}=X)=1\}} \sup_{Q \in \mathcal{M}_1} \{\mathbb{E}_Q[-\widetilde{X}] - \alpha(Q)\} = \rho^P(X). \end{aligned} \tag{4.2.6}$$

reduce to an equality. Before clarifying this, we shall first relate these risk measures with $\widehat{\rho}^P$ given by (4.2.2) or equivalently by (4.2.4) . Indeed, the following lemma shows that $\widehat{\rho}^P$ coincides with $\widetilde{\rho}^P$ but that we have a potential strict inequality with ρ^P.

Lemma 4.2.5. For $X \in L^\infty(P)$, we have $\rho^P(X) \geq \widetilde{\rho}^P(X) = \widehat{\rho}^P(X)$ and thus $C^P \subseteq \widehat{C}^P$.

Proof. Let $X \in L^\infty(P)$. Then for all $\widetilde{X} \in \mathcal{X}$ with $P(\widetilde{X} = X) = 1$ and all $Q \in \mathcal{P}^P$

$$\mathbb{E}_Q[-X] - \alpha(Q) = \mathbb{E}_Q[-\widetilde{X}] - \alpha(Q).$$

Hence, also

$$\mathbb{E}_Q[-X] - \alpha(Q) = \inf_{\{\widetilde{X}: P(\widetilde{X}=X)=1\}} \mathbb{E}_Q[-\widetilde{X}] - \alpha(Q).$$

Therefore we get the following inequalities

$$\begin{aligned} \widehat{\rho}^P(X) &= \sup_{Q \in \mathcal{P}^P} \{\mathbb{E}_Q[-X] - \alpha(Q)\} = \sup_{Q \in \mathcal{P}^P} \inf_{\{\widetilde{X}: P(\widetilde{X}=X)=1\}} \{\mathbb{E}_Q[-\widetilde{X}] - \alpha(Q)\} \\ &\leq \underbrace{\sup_{Q \in \mathcal{M}_1} \inf_{\{\widetilde{X}: P(\widetilde{X}=X)=1\}} \{\mathbb{E}_Q[-\widetilde{X}] - \alpha(Q)\}}_{\widetilde{\rho}^P(X)} \leq \underbrace{\inf_{\{\widetilde{X}: P(\widetilde{X}=X)=1\}} \sup_{Q \in \mathcal{M}_1} \{\mathbb{E}_Q[-\widetilde{X}] - \alpha(Q)\}}_{\rho^P(X)}. \end{aligned}$$

Hence if $X \in C^P$, meaning that $\rho^P(X) \leq 0$, then also $\widehat{\rho}^P(X) \leq 0$ so that $C^P \subseteq \widehat{C}^P$.

Let us now prove the missing equality $\widetilde{\rho}^P = \widehat{\rho}^P$. Let $Q \in \mathcal{M}_1$ and consider the Lebesgue decomposition into $Q = \widetilde{Q}_1 + \widetilde{Q}_2$ where $\widetilde{Q}_1$ and $\widetilde{Q}_2$ are nonnegative measures such that $\widetilde{Q}_1 \ll P$ and $\widetilde{Q}_2 \perp P$. Moreover denote by Q_i the probability measure $Q_i := \frac{\widetilde{Q}_i}{\widetilde{Q}_i(\Omega)}$ for $i = 1, 2$. Then for every fixed Q with $\widetilde{Q}_2(\Omega) > 0$, there is $B_Q \in \mathcal{F}$ with $\widetilde{Q}_2(B_Q) = \widetilde{Q}_2(\Omega)$ and $P(B_Q^c) = 1$. Define, for every $n \geq 1$, $\widetilde{X}_Q^n = n1_{B_Q} + X1_{B_Q^c} \in \mathcal{X}$. Clearly $P(\widetilde{X}_Q^n = X) = P(B_Q^c) = 1$. We have that

$$\mathbb{E}_Q[-\widetilde{X}_Q^n] = -nQ(B_Q) + \mathbb{E}_Q[-X1_{B_Q^c}] = -n\widetilde{Q}_2(\Omega) + \widetilde{Q}_1(\Omega)\mathbb{E}_{Q_1}[-X].$$

Moreover we know that $\inf_{R\in\mathcal{M}_1} \alpha(R) > -\infty$, say $-N := \inf_{R\in\mathcal{M}_1} \alpha(R)$. Note that, as $X \in L^\infty(P) \subseteq L^\infty(Q_1)$ we have that $|\mathbb{E}_{Q_1}[-X]| \leq \|X\|_{L^\infty(P)}$. It follows that

$$\begin{aligned}\inf_{\{\widetilde{X}:P(\widetilde{X}=X)=1\}} \{\mathbb{E}_Q[-\widetilde{X}] - \alpha(Q)\} &\leq \inf_{n\geq 1} \{\mathbb{E}_Q[-\widetilde{X}_Q^n] - \alpha(Q)\} \\ &\leq \inf_{n\geq 1} -n\widetilde{Q}_2(\Omega) + \widetilde{Q}_1(\Omega)\mathbb{E}_{Q_1}[-X] + N = -\infty.\end{aligned}$$

Taking the supremum over $Q \in \mathcal{M}_1$ therefore just means to restrict the set to $\mathcal{P}^P$ so that $\widetilde{Q}_2 \equiv 0$. Hence,

$$\begin{aligned}\widetilde{\rho}^P(X) &= \sup_{Q\in\mathcal{M}_1} \inf_{\{\widetilde{X}:P(\widetilde{X}=X)=1\}} \{\mathbb{E}_Q[-\widetilde{X}] - \alpha(Q)\} \\ &= \sup_{Q\in\mathcal{P}^P} \inf_{\{\widetilde{X}:P(\widetilde{X}=X)=1\}} \{\mathbb{E}_Q[-\widetilde{X}] - \alpha(Q)\} = \widehat{\rho}^P(X),\end{aligned}$$

yielding the assertion.

□

By the above lemma, equality of $\widehat{\rho}^P = \rho^P$ is thus equivalent to a minimax equality. In the following we thus apply a (standard) minimax theorem to conclude equality of $\widehat{\rho}^P = \widetilde{\rho}^P = \rho^P$.

Proposition 4.2.6. Let Ω be compact and $Q \mapsto \alpha(Q)$ lower semicontinuous with respect to the weak-$*$-topology on $\mathcal{M}_1$ and convex. Then the minimax equality

$$\begin{aligned}\widetilde{\rho}^P(X) &= \sup_{Q\in\mathcal{M}_1} \inf_{\{\widetilde{X}\in\mathcal{X}:P(\widetilde{X}=X)=1\}} \{\mathbb{E}_Q[-\widetilde{X}] - \alpha(Q)\} \\ &= \inf_{\{\widetilde{X}\in\mathcal{X}:P(\widetilde{X}=X)=1\}} \sup_{Q\in\mathcal{M}_1} \{\mathbb{E}_Q[-\widetilde{X}] - \alpha(Q)\} = \rho^P(X),\end{aligned}$$

holds true. Hence, under these conditions $\widehat{\rho}^P = \rho^P$.

Proof. Note that

$$Q \mapsto \mathbb{E}_Q[\widetilde{X}] - \alpha(Q)$$

is concave and upper semicontinous and

$$\widetilde{X} \mapsto \mathbb{E}_Q[\widetilde{X}] - \alpha(Q)$$

is linear and hence convex. Moreover, $\{\widetilde{X} \in \mathcal{X} : P(\widetilde{X} = X) = 1\}$ and $\mathcal{M}_1$ are convex. By compactness of Ω, $\mathcal{M}_1$ is also (weak-$*$-) compact. Therefore Sion's minimax theorem (see e.g. Sion (1958); Ricceri and Simons (2013)) yields the first assertion and the second one follows from Lemma 4.2.5. □

Remark 4.2.7. Note that if Ω is compact and α happens to be $\alpha_{\min}$ as introduced in Section 4.1.2, all conditions of Proposition 4.2.6 are automatically satisfied. Indeed, we can argue along the lines of Remark 4.17 in Föllmer and Schied (2011) and apply the general duality theorem for the Fenchel–Legendre transform to the convex function ρ, however here on the Banach space of continuous functions on the compact set Ω, denoted by $C(\Omega)$. This allows to conclude that

$$\alpha_{\min}(Q) = \sup_{X \in C(\Omega)} \{\mathbb{E}_Q[-X] - \rho(X)\}.$$

Then due to Theorem 2.3.1 in Zalinescu (2002), $\alpha_{\min}$ is convex and weak-$*$-lower semicontinuous.

In the following example lower semicontinuity of α fails and this shows that in general C^P and $\widehat{C}^P$ do not coincide, namely C^P is strictly contained in $\widehat{C}^P$ but not equal. Hence there is a gap between $\widehat{\rho}^P$ and ρ^P, meaning that $\widehat{\rho}^P = \widetilde{\rho}^P$ can in general take strictly smaller values than ρ^P.

Example 4.2.8. Let $(\Omega, \mathcal{F}) = ([0,1], \mathcal{B}([0,1]))$ and the reference measure P be the Lebesgue measure restricted to $\mathcal{B}([0,1])$. We fix a penalty function $\alpha : \mathcal{M}_1 \to \mathbb{R}$ such that

$$\alpha(Q) := \begin{cases} 0, & \text{if } Q \in \mathcal{P}^P, \\ -1, & \text{if } Q \in \mathcal{M}_1 \setminus \mathcal{P}^P. \end{cases}$$

Note that α fails to be lower semicontinuous. Indeed, take a sequence of measures $Q_n = \sum_{i=1}^n \alpha_i^n \delta_{x_i^n}$, where $\alpha_i^n \geq 0$, $\sum_{i=1}^n \alpha_i^n = 1$ and δ_x is the Dirac measure at x, converging to some measure Q that is absolutely continuous with respect to the Lebesgue measure. Indeed, for example, for each $n \geq 1$, we can partition the interval $[0,1]$ into intervals of length α_i^n, $1 \leq i \leq n$, and take a point x_i^n in the i-th interval. Then, for every continuous function $f : [0,1] \to \mathbb{R}$ we have that $\int_0^1 f \mathrm{d}Q_n = \sum_{i=1}^n \alpha_i^n f(x_i^n) \to \int_0^1 f(x)\mathrm{d}x = \int_0^1 f \mathrm{d}P$, hence $Q^n \to P$ in the weak-$*$-topology. Being sums of Dirac-measures it is clear that $\alpha(Q_n) = -1$ and thus

$$-1 = \liminf_{n\to\infty} \alpha(Q_n) \leq \alpha(Q) = 0,$$

whence α is not lower semicontinuous.

We now claim that there exists $X \in \widehat{C}^P \setminus C^P$. Indeed, define $X_k = k$ P-a.s. with some fixed number $k \in (0,1)$. Then $X_k \in \widehat{C}^P$. Note that

$$\widehat{C}^P = \{X \in L^\infty(P) : \mathbb{E}_Q[-X] \leq 0, \ \forall Q \in \mathcal{P}^P\}$$

and, trivially, for each $Q \ll P$:

$$\mathbb{E}_Q[-X_k] = -kQ(\Omega) = -k < 0.$$

However $X_k \notin C^P$.

Indeed, we cannot find a version $\widetilde{X} = k$ P-a.s. such that

$$\int_0^1 \widetilde{X}(\omega)\delta_x(\mathrm{d}\omega) \geq 1, \qquad \forall x \in [0,1].$$

Here, δ_x denotes the Dirac distribution centered in $x \in [0,1]$, which lies for all $x \in \mathcal{M}_1 \setminus \mathcal{P}^P$. Therefore, for any $\widetilde{X}$ with $P(X_k = \widetilde{X}) = 1$

$$\sup_{Q \in \mathcal{M}_1} \mathbb{E}_Q[-\widetilde{X}] - \alpha(Q) \geq -k + 1 > 0$$

and thus clearly also for the infimum. This proves the claim and shows together with Lemma 4.2.5 that the minimax inequality is strict.

The above example shows that lower semicontinuity of the penalty function α is crucial. By replacing it by the corresponding $\alpha_{\min}$, which is here $\alpha_{\min} \equiv -1$ (compare with the shifted worst case risk measure as defined in Section 4.3.3 for $a = 1$), the previous gap $\rho^P(X) > \widehat{\rho}^P(X) = \widetilde{\rho}^P(X)$ disappears, which is a consequence of Remark 4.2.7.

4.2.4. Convex law invariant risk measures: worst case properties and comparison of the two approaches

We now specialize the setting further and consider *law invariant risk measures*. To define law-invariance we need to start with a reference measure $\mathbb{P}$ right from the beginning, which could correspond to some 'worst case probability measure' used by the regulator.

Recall that a risk measure ρ is called law-invariant if

$$\rho(X) = \rho(Y),$$

whenever X and Y have the same distribution under $\mathbb{P}$. For a convex law invariant risk measure (that satisfies automatically the Fatou property, see Jouini et al. (2006)), any penalty function is of the following form

$$\alpha(Q) = \begin{cases} \widetilde{\alpha}(Q) & \text{for } Q \in \mathcal{Q}^{\mathbb{P}} \subseteq \mathcal{P}^{\mathbb{P}}, \\ +\infty & \text{for } Q \in \mathcal{M}_1 \setminus \mathcal{Q}^{\mathbb{P}}, \end{cases} \tag{4.2.7}$$

where $\mathcal{Q}^{\mathbb{P}}$ denotes some weak-$*$-closed subset of $\mathcal{P}^{\mathbb{P}}$ and $\widetilde{\alpha}(Q)$ satisfies

$$\widetilde{\alpha}(Q) < \infty, \quad \forall Q \in \mathcal{Q}^{\mathbb{P}}$$

and

$$\inf_{Q \in \mathcal{Q}^{\mathbb{P}}} \widetilde{\alpha}(Q) > -\infty.$$

This follows from (Föllmer and Schied, 2011, Lemma 4.30 and Theorem 4.31) and the Fatou property.

Instead of an arbitrary penalty function, we here fix $\mathbb{P}$ and consider penalty functions α of the more specific form (4.2.7). Hence, throughout this section ρ denotes the (pointwise defined) convex risk measure as of (4.2.1) now with α given by (4.2.7) with the fixed $\mathbb{P}$. Its acceptance set is given by

$$C = \{X \in \mathcal{X} : \sup_{Q \in \mathcal{Q}^{\mathbb{P}}} \{\mathbb{E}_Q[-X] - \widetilde{\alpha}(Q)\} \leq 0\}. \tag{4.2.8}$$

The goal is now to analyze ρ^P and $\widehat{\rho}^P$ defined with respect to some new measure P. Here, according to the first approach, ρ^P is defined via

$$\rho^P(X) = \inf_{\{\widetilde{X}: P(\widetilde{X}=X)=1\}} \rho(\widetilde{X}) = \inf_{\{\widetilde{X}: P(\widetilde{X}=X)=1\}} \sup_{Q \in \mathcal{Q}^{\mathbb{P}}} \{\mathbb{E}_Q[-\widetilde{X}] - \widetilde{\alpha}(Q)\}, \quad X \in L^\infty(P),$$

while $\widehat{\rho}^P$ is given by (4.2.2) or equivalently by (4.2.4) with penalty function

$$\alpha^P(Q) = \begin{cases} \widetilde{\alpha}(Q) & \text{for } Q \in \mathcal{Q}^{\mathbb{P}} \cap \mathcal{P}^P, \\ +\infty & \text{for } Q \in \mathcal{M}_1 \setminus (\mathcal{Q}^{\mathbb{P}} \cap \mathcal{P}^P). \end{cases} \tag{4.2.9}$$

The corresponding acceptance set C^P and $\widehat{C}^P$ are defined as in Section 4.2.1 and Section 4.2.2.

Remark 4.2.9. In the current setup we could introduce a third approach where we consider the sets $\mathcal{Q}^{\mathbb{P}}$ as a certain function of the measures and define an alternative penalty function $\overline{\alpha}^P$ via

$$\overline{\alpha}^P(Q) = \begin{cases} \widetilde{\alpha}(Q) & \text{for } Q \in \mathcal{Q}^P, \\ +\infty & \text{for } Q \in \mathcal{M}_1 \setminus \mathcal{Q}^P \end{cases}$$

so that $\overline{\alpha}^{\mathbb{P}}$ is equal to (4.2.7). To give an example, in the case of Expected Shortfall (see Section 4.3.2) the sets $\mathcal{Q}^P$ are given by

$$\mathcal{Q}^P = \{Q \in \mathcal{M}_1 : Q \ll P, \text{ and } \frac{\mathrm{d}Q}{\mathrm{d}P} \leq \lambda^{-1}, \mathbb{P}\text{-a.s.}\}$$

for some $\lambda \in (0,1)$. Note that in general $\mathcal{Q}^P \neq \mathcal{Q}^{\mathbb{P}} \cap \mathcal{P}^P$. However, when $P \ll \mathbb{P}$, there are certain situations where $\mathcal{Q}^P = \mathcal{Q}^{\mathbb{P}} \cap \mathcal{P}^P$ holds. Indeed, the superhedging price (see Section 4.3.1) is one example for this equality. In the case of the Expected Shortfall it however does not hold true.

As already stated in Remark 4.2.4, a crucial property of our second approach is that $\mathcal{Q}^{\mathbb{P}} \cap \mathcal{P}^{P'} \subseteq \mathcal{Q}^{\mathbb{P}} \cap \mathcal{P}^P$ whenever $P' \ll P$ so that we get the corresponding ordering $\widehat{\rho}^{P'} \leq \widehat{\rho}^P$, which would not necessarily hold true if we pursued the above outlined third approach.

Our first result of this section relates ρ^P and $\widehat{\rho}^P$ with the starting risk measure ρ and shows the following order relation

$$\widehat{\rho}^P|_{\mathcal{X}} \leq \rho^P|_{\mathcal{X}} \leq \rho = \rho^{\mathbb{P}}|_{\mathcal{X}} = \widehat{\rho}^{\mathbb{P}}|_{\mathcal{X}} \tag{4.2.10}$$

for *all* $P \in \mathcal{M}_1$ (and not only absolutely continuous ones). This justifies the motivation given in the introduction that the reference measure $\mathbb{P}$ should correspond to some 'worst case probability measure' capturing all the risk considered to be relevant since $\rho = \rho^{\mathbb{P}}|_{\mathcal{X}} = \widehat{\rho}^{\mathbb{P}}|_{\mathcal{X}}$ always yields the maximal risk. We emphasize again that also in cases where $\mathbb{P}$ is not dominating the measure P the risk computed with respect to P is *always* smaller or equal to the risk with respect to $\mathbb{P}$, which is a consequence of the definitions of ρ^P and $\widehat{\rho}^P$ (see also Remark 4.2.11 (iii) below).

Concerning the comparison of ρ^P and $\widehat{\rho}^P$, we obtain different results depending on the relation between $\mathbb{P}$ and P. If P is either dominating or singular with respect to $\mathbb{P}$, then we always have $\rho^P = \widehat{\rho}^P$, which is also asserted in the next proposition. The proposition also shows whenever $P \perp \mathbb{P}$, the risk of both ρ^P and $\widehat{\rho}^P$ is $-\infty$. Moreover, if $\mathbb{P} \ll P$, then $\rho^P|_{\mathcal{X}} = \widehat{\rho}^P|_{\mathcal{X}}$ is equal to $\rho = \rho^{\mathbb{P}}|_{\mathcal{X}} = \widehat{\rho}^{\mathbb{P}}|_{\mathcal{X}}$, which is of course in line with (4.2.10).

Proposition 4.2.10. Let $\mathbb{P} \in \mathcal{M}_1$ be a reference measure and let α be of form (4.2.7) with $\mathcal{Q}^{\mathbb{P}}$ some weak-$*$-closed subset of $\mathcal{P}^{\mathbb{P}}$. Then for all $P \in \mathcal{M}_1$ which satisfy either $\mathbb{P} \ll P$ or $\mathbb{P} \perp P$, it holds that $C^P = \widehat{C}^P$ and thus $\rho^P = \widehat{\rho}^P = \widetilde{\rho}^P$. Hence, the minimax equality holds true

$$\begin{aligned}\widetilde{\rho}^P(X) &= \sup_{Q \in \mathcal{Q}^{\mathbb{P}}} \inf_{\{\widetilde{X} \in \mathcal{X} : P(\widetilde{X} = X) = 1\}} \{\mathbb{E}_Q[-\widetilde{X}] - \widetilde{\alpha}(Q)\} \\ &= \inf_{\{\widetilde{X} \in \mathcal{X} : P(\widetilde{X} = X) = 1\}} \sup_{Q \in \mathcal{Q}^{\mathbb{P}}} \{\mathbb{E}_Q[-\widetilde{X}] - \widetilde{\alpha}(Q)\} = \rho^P(X).\end{aligned}$$

In particular, if $\mathbb{P} \ll P$, the restriction of $\rho^P = \widehat{\rho}^P = \widetilde{\rho}^P$ to $\mathcal{X}$ is equal to ρ. If $\mathbb{P} \perp P$, then $\rho^P = \widehat{\rho}^P = \widetilde{\rho}^P \equiv -\infty$.
Moreover, for all $P \in \mathcal{M}_1$ we have

$$\widehat{\rho}^P|_{\mathcal{X}} \leq \rho^P|_{\mathcal{X}} \leq \rho = \rho^{\mathbb{P}}|_{\mathcal{X}} = \widehat{\rho}^{\mathbb{P}}|_{\mathcal{X}}.$$

and also

$$\widehat{\rho}^P \leq \rho^P = \rho^{\mathbb{P}} = \widehat{\rho}^{\mathbb{P}}, \quad \text{on} \quad L^\infty(P) \cap L^\infty(\mathbb{P}).$$

Remark 4.2.11. (i) Observe that in contrast to the minimax result as of Proposition 4.2.6 and Corollary 4.2.14 below, here we do not need to assume any properties on α or Ω.

(ii) Note that Proposition 4.2.10 implies that $\rho^{\mathbb{P}} = \widehat{\rho}^{\mathbb{P}}$ *always* holds if α is of form (4.2.7).

(iii) Notice that for general P, we do not necessarily have $\rho^P = \widehat{\rho}^P$. For $\widehat{\rho}^P$ the penalty function α^P is given by (4.2.9) and $\mathcal{Q}^{\mathbb{P}} \cap \mathcal{P}^P = \mathcal{Q}^{\mathbb{P}} \cap \mathcal{P}^{P_1}$ where P_1 denotes the normalized absolutely continuous part of the Lebesgue decomposition of P with respect to $\mathbb{P}$. Then it is easy to see that $\widehat{\rho}^P = \widehat{\rho}^{P_1}$. Similarly we have $\rho^P = \rho^{P_1}$ as shown in the proof below. But we do not necessarily have $\rho^{P_1} = \widehat{\rho}^{P_1}$ (compare also Proposition 4.2.12).

Proof. We start by relating ρ^P with ρ for the cases $\mathbb{P} \ll P$ and $\mathbb{P} \perp P$. Recall that C^P is defined by

$$C^P = \{X \in L^\infty(P) : \inf_{\{\widetilde{X} \in \mathcal{X} : P(\widetilde{X}=X)=1\}} \rho(\widetilde{X}) \leq 0\},$$

which we now compare with C given in (4.2.8).

If $\mathbb{P} \ll P$, we have $C = C^P \cap \mathcal{X}$. Indeed, it always holds that $C \subseteq C^P \cap \mathcal{X}$. To prove $C^P \cap \mathcal{X} \subseteq C$, let $X \in C^P \cap \mathcal{X}$. Since all P-nullsets are $\mathbb{P}$-nullsets and in turn also Q-nullsets for all $Q \in \mathcal{Q}^{\mathbb{P}}$, we have for all $Q \in \mathcal{Q}^{\mathbb{P}}$ and for *all* $\widetilde{X}$ satisfying $P(\widetilde{X} = X) = 1$, that $\mathbb{E}_Q[-\widetilde{X}] = \mathbb{E}_Q[-X]$ and thus in turn

$$\rho(\widetilde{X}) = \sup_{Q \in \mathcal{Q}^{\mathbb{P}}} \{\mathbb{E}_Q[-X] - \widetilde{\alpha}(Q)\} = \rho(X).$$

Since $\rho(\widetilde{X})$ assumes the same value for every $\widetilde{X}$, we have $\rho^P(X) = \rho(\widetilde{X})$ which is ≤ 0 as $X \in C^P \cap \mathcal{X}$. Thus $\rho(X) \leq 0$ as well, whence $X \in C$. This proves the claim and implies that for all $\mathbb{P} \ll P$, $\rho^P|_{\mathcal{X}} = \rho^{\mathbb{P}}|_{\mathcal{X}} = \rho$ and similarly $\rho^P = \rho^{\mathbb{P}}|_{L^\infty(P)}$.

In the case $\mathbb{P} \perp P$, $C \subset C^P = L^\infty(P)$. Indeed, since $\mathbb{P} \perp P$, there exists a set B with $P(B) = 1$ and $\mathbb{P}(B^c) = 1$. Let $X \in L^\infty(P)$ and choose $\widetilde{X} \in \mathcal{X}$ such $X = \widetilde{X}$ on B and $\widetilde{X} \geq -\inf \widetilde{\alpha}(Q)$ on B^c. For example, take, for each $n \geq 1$, $\widetilde{X}^n = X1_B + n1_{B^c}$. As $-\inf \widetilde{\alpha}(Q) < \infty$ we have that, for all n large enough, $n \geq -\inf \widetilde{\alpha}(Q)$. By taking the infimum over all n, it follows that

$$\rho^P(X) = \inf_{\{\widetilde{X} \,|\, P(\widetilde{X}=X)=1\}} \sup_{Q \in \mathcal{Q}^{\mathbb{P}}} \{\mathbb{E}_Q[-\widetilde{X}] - \widetilde{\alpha}(Q)\} \leq \inf_{n \geq 1}(-n) - \inf_{Q \in \mathcal{Q}^{\mathbb{P}}} \widetilde{\alpha}(Q) \equiv -\infty.$$

Next we analyze the relation between $\widehat{\rho}^P$ and ρ in the cases $\mathbb{P} \ll P$ and $\mathbb{P} \perp P$, which in turn will allow us to relate also ρ^P with $\widehat{\rho}^P$. For $\widehat{\rho}^P$, the penalty function α^P is given by (4.2.9).

If $\mathbb{P} \ll P$, then $\mathcal{Q}^{\mathbb{P}} \cap \mathcal{P}^P = \mathcal{Q}^{\mathbb{P}}$, thus $\alpha^P = \alpha$ and $\widehat{\rho}^P|_{\mathcal{X}} = \rho$ and $\widehat{\rho}^P = \rho^{\mathbb{P}}|_{L^\infty(P)}$. Since $\rho^P = \rho^{\mathbb{P}}|_{L^\infty(P)}$ as proved above, we also get $\widehat{\rho}^P = \rho^P$.

If $\mathbb{P} \perp P$, then $\mathcal{Q}^{\mathbb{P}} \cap \mathcal{P}^P = \emptyset$, yielding $\alpha^P(Q) = +\infty$ for all $Q \in \mathcal{M}_1$ and in turn $\widehat{\rho}^P \equiv -\infty$ such that $\widehat{C}^P = L^\infty(P)$. This already proves the equality $C^P = \widehat{C}^P$ and $\rho^P = \widehat{\rho}^P$ for $\mathbb{P} \perp P$. The assertion concerning the minimax equality follows from Lemma 4.2.5.

Finally, consider an arbitrary $P \in \mathcal{M}_1$. Then by Lebesgue's decomposition theorem we can decompose P into $P = \widetilde{P}_1 + \widetilde{P}_2$ with $\widetilde{P}_1 \ll P$ and $\widetilde{P}_2 \perp P$. We denote by P_1 and by P_2 the corresponding normalized probability measures. As above there exists some set B such that $P_2(B) = 1$ and $\mathbb{P}(B) = P_1(B) = 0$. Hence $\mathbb{P}(B^c) = 1$ and $P_1(B^c) = 1$. Thus, for all $Q \in \mathcal{Q}^{\mathbb{P}}$ and for all $\widetilde{X}$ with $P(\widetilde{X} = X) = 1$, we have $\mathbb{E}_Q[-\widetilde{X}] = \mathbb{E}_Q[-\widetilde{X}1_{B^c}]$ since $\mathbb{E}_Q[-\widetilde{X}1_B] = 0$ due to the singularity between P_2 and Q. We therefore get

$$\rho(\widetilde{X}) = \sup_{Q \in \mathcal{Q}^{\mathbb{P}}} \{\mathbb{E}_Q[-\widetilde{X}] - \widetilde{\alpha}(Q)\} = \rho(\widetilde{X}1_{B^c}),$$

whence $\rho^P = \rho^{P_1}$ on $L^\infty(P)$ since we only need that $\widetilde{X} = X$, P_1-almost surely. As $P_1 \ll \mathbb{P}$, we thus have $\rho^P \leq \rho^{\mathbb{P}}$ on $L^\infty(P) \cap L^\infty(\mathbb{P})$. This proves the last assertion since $\widehat{\rho}^P \leq \rho^P$ holds by Lemma 4.2.5. □

As indicated in Remark 4.2.11 (iii) the relation between ρ^P and $\widehat{\rho}^P$ is more subtle for general $P \in \mathcal{M}_1$. We consider in the next proposition the case when P is absolutely continuous with respect to the reference measure $\mathbb{P}$ and derive sufficient conditions when equality holds true. By Remark 4.2.11 (iii) this then translates also to the general case.

Proposition 4.2.12. Let $\mathbb{P} \in \mathcal{M}_1$ be a reference measure and let α be of form (4.2.7) with $\mathcal{Q}^{\mathbb{P}}$ some weak-$*$-closed subset of $\mathcal{P}^{\mathbb{P}}$. Let $P \ll \mathbb{P}$ and consider for $Q \in \mathcal{Q}^{\mathbb{P}} \setminus (\mathcal{Q}^{\mathbb{P}} \cap \mathcal{P}^P)$ the Lebesgue decomposition into $Q = \widetilde{Q}_1 + \widetilde{Q}_2$ where $\widetilde{Q}_1, \widetilde{Q}_2$ are nonnegative measures such that $\widetilde{Q}_1 \ll P$ and $\widetilde{Q}_2 \perp P$. If

(i) there exists a set $B \in \mathcal{F}$ such that $P(B) = 1$ and $\widetilde{Q}_2(B) = 0$ for the singular parts $\widetilde{Q}_2$ of all $Q \in \mathcal{Q}^{\mathbb{P}} \setminus (\mathcal{Q}^{\mathbb{P}} \cap \mathcal{P}^P)$,

(ii) $\inf_{\mathcal{Q}^{\mathbb{P}}} \widetilde{\alpha}(Q) = \inf_{\mathcal{Q}^{\mathbb{P}} \cap \mathcal{P}^P} \widetilde{\alpha}(Q)$,

then $C^P = \widehat{C}^P$ and thus $\rho^P = \widehat{\rho}^P = \widetilde{\rho}^P$. Hence, the minimax equality

$$\begin{aligned}\widetilde{\rho}^P(X) &= \sup_{Q \in \mathcal{Q}^{\mathbb{P}}} \inf_{\{\widetilde{X} \in \mathcal{X} : P(\widetilde{X} = X) = 1\}} \{\mathbb{E}_Q[-\widetilde{X}] - \widetilde{\alpha}(Q)\} \\ &= \inf_{\{\widetilde{X} \in \mathcal{X} : P(\widetilde{X} = X) = 1\}} \sup_{Q \in \mathcal{Q}^{\mathbb{P}}} \{\mathbb{E}_Q[-\widetilde{X}] - \widetilde{\alpha}(Q)\} = \rho^P(X),\end{aligned}$$

holds true.

Remark 4.2.13. Concerning Condition (i) of Proposition 4.2.12, recall that $\widetilde{Q}_2 \perp P$ means that there are disjoint sets $A(\widetilde{Q}_2)$ and $B(\widetilde{Q}_2)$ such that $A(\widetilde{Q}_2) \cup B(\widetilde{Q}_2) = \Omega$ and $P(A(\widetilde{Q}_2)) = 0$ and $\widetilde{Q}_2(B(\widetilde{Q}_2)) = 0$. If $\bigcup_{\widetilde{Q}_2} A(\widetilde{Q}_2)$ can be described by a countable union, i.e.

$$\bigcup_{\widetilde{Q}_2} A(\widetilde{Q}_2) = \bigcup_{i=1}^{\infty} A(\widetilde{Q}_2^i),$$

then $P(\bigcup_{\widetilde{Q}_2} A(\widetilde{Q}_2)) = 0$ and the set B with $P(B) = 1$ can be chosen as $B = \bigcap_{i=1}^{\infty} B(\widetilde{Q}_2^i)$. This clearly always works when Ω is countable and thus the Condition (i) is in this case always satisfied.

Observe additionally that Condition (ii) is satisfied for any coherent risk measure where $\widetilde{\alpha}(Q)$ is chosen to be 0 for all $Q \in \mathcal{Q}^{\mathbb{P}}$.

Proof. As shown in Lemma 4.2.5, $\rho^P \geq \widehat{\rho}^P$, hence we only have to prove the converse inequality. As $\alpha^P(Q)$ is given by (4.2.9), we have

$$\widehat{\rho}^P(X) = \sup_{Q \in \mathcal{Q}^{\mathbb{P}} \cap \mathcal{P}^P} \{\mathbb{E}_Q[-X] - \widetilde{\alpha}(Q)\},$$

while ρ^P is given by

$$\rho^P(X) = \inf_{\{\widetilde{X} \in \mathcal{X} : P(\widetilde{X} = X) = 1\}} \sup_{Q \in \mathcal{Q}^{\mathbb{P}}} \{\mathbb{E}_Q[-\widetilde{X}] - \widetilde{\alpha}(Q)\}.$$

Consider a measure $Q \in \mathcal{Q}^{\mathbb{P}} \setminus (\mathcal{Q}^{\mathbb{P}} \cap \mathcal{P}^P)$ and denote by Q_i its normalized Lebesgue decomposition, i.e. $Q_i = \frac{\widetilde{Q}_i}{\widetilde{Q}_i(\Omega)}$ for $i = 1, 2$. Then due to the assumption that there exists a set B with $P(B) = 1$ and $\widetilde{Q}_2(B) = 0$ for the singular parts of all $Q \in \mathcal{Q}^{\mathbb{P}} \setminus (\mathcal{Q}^{\mathbb{P}} \cap \mathcal{P}^P)$, we can consider $\widetilde{X} = X$ on B and $\widetilde{X} = n \in \mathbb{R}$ on B^c. This yields

$$\begin{aligned}\sup_{Q\in\mathcal{Q}^{\mathbb{P}}} \{\mathbb{E}_Q[-\widetilde{X}] - \widetilde{\alpha}(Q)\} &= \sup_{Q\in\mathcal{Q}^{\mathbb{P}}} \{\widetilde{Q}_1(\Omega)\mathbb{E}_{Q_1}[-\widetilde{X}1_B] + \widetilde{Q}_2(\Omega)\mathbb{E}_{Q_2}[-\widetilde{X}1_{B^c}] - \widetilde{\alpha}(Q)\} \\ &\leq \sup_{\widetilde{Q}_2(\Omega)\in[0,1]} \big\{(1-\widetilde{Q}_2(\Omega)) \sup_{Q\in\mathcal{Q}^{\mathbb{P}}\cap\mathcal{P}^P} \mathbb{E}_Q[-\widetilde{X}1_{\{B\}}] \\ &\quad - \widetilde{Q}_2(\Omega)n - \inf_{Q\in\mathcal{Q}^{\mathbb{P}}} \widetilde{\alpha}(Q)\big\}.\end{aligned}$$

We now optimize the last expression over $\widetilde{Q}_2(\Omega)$, which leads to

$$\widetilde{Q}_2(\Omega) = \begin{cases} 0 & \text{if } n \geq -\sup_{Q\in\mathcal{Q}^{\mathbb{P}}\cap\mathcal{P}^P} \mathbb{E}_Q[-X], \\ 1 & \text{else.} \end{cases}$$

and

$$\begin{aligned}&\sup_{Q\in\mathcal{Q}^{\mathbb{P}}} \{\mathbb{E}_Q[-\widetilde{X}] - \widetilde{\alpha}(Q)\} \\ &\quad\leq \begin{cases} \sup_{Q\in\mathcal{Q}^{\mathbb{P}}\cap\mathcal{P}^P} \mathbb{E}_Q[-X] - \inf_{Q\in\mathcal{Q}^{\mathbb{P}}} \widetilde{\alpha}(Q) & \text{if } n \geq -\sup_{Q\in\mathcal{Q}^{\mathbb{P}}\cap\mathcal{P}^P} \mathbb{E}_Q[-X], \\ -n - \inf_{Q\in\mathcal{Q}^{\mathbb{P}}} \widetilde{\alpha}(Q) & \text{else.} \end{cases}\end{aligned}$$

Taking the infimum over n on both sides (recall that $\widetilde{X} = n$ on B^c) then yields,

$$\begin{aligned}\rho^P(X) &= \inf_{\{\widetilde{X}\in\mathcal{X}:P(\widetilde{X}=X)=1\}} \sup_{Q\in\mathcal{Q}^{\mathbb{P}}} \{\mathbb{E}_Q[-\widetilde{X}] - \widetilde{\alpha}(Q)\} \leq \inf_{\{\widetilde{X}=X1_B+n1_{B^c}\}} \sup_{Q\in\mathcal{Q}^{\mathbb{P}}} \{\mathbb{E}_Q[-\widetilde{X}] - \widetilde{\alpha}(Q)\} \\ &\leq \sup_{Q\in\mathcal{Q}^{\mathbb{P}}\cap\mathcal{P}^P} \mathbb{E}_Q[-X] - \inf_{Q\in\mathcal{Q}^{\mathbb{P}}} \widetilde{\alpha}(Q) = \sup_{Q\in\mathcal{Q}^{\mathbb{P}}\cap\mathcal{P}^P} \{\mathbb{E}_Q[-X] - \widetilde{\alpha}(Q)\} \\ &= \widehat{\rho}^P(X)\end{aligned}$$

where the second last equality follows from Assumption (ii) and shows $\rho^P \leq \widehat{\rho}^P$. The assertion concerning the minimax equality also follows from this lemma. □

Alternatively to the above propositions one can apply again Sion's minimax result to determine conditions for the equality of ρ^P and $\widehat{\rho}^P$. In this case fewer assumptions are needed on the set $\mathcal{Q}^{\mathbb{P}}$ but more on the penalty function $\widetilde{\alpha}$ and Ω. Note that we do not need to distinguish between the different relations to the initial measure $\mathbb{P}$.

Corollary 4.2.14. *Suppose that Ω is compact. Let $\mathbb{P} \in \mathcal{M}_1$ be a reference measure and let α be of form (4.2.7) with $\mathcal{Q}^{\mathbb{P}}$ some weak-$*$-closed convex subset of $\mathcal{P}^{\mathbb{P}}$. If $Q \mapsto \widetilde{\alpha}(Q)$ is lower semicontinuous (with respect to the weak-$*$-topology) and convex, then for any*

$P \in \mathcal{M}_1$ *the minimax equality*

$$\widehat{\rho}^P(X) = \widetilde{\rho}^P(X) = \sup_{Q \in \mathcal{Q}^{\mathbb{P}}} \inf_{\{\widetilde{X} \in \mathcal{X} : P(\widetilde{X} = X) = 1\}} \{\mathbb{E}_Q[-\widetilde{X}] - \widetilde{\alpha}(Q)\}$$
$$= \inf_{\{\widetilde{X} \in \mathcal{X} : P(\widetilde{X} = X) = 1\}} \sup_{Q \in \mathcal{Q}^{\mathbb{P}}} \{\mathbb{E}_Q[-\widetilde{X}] - \widetilde{\alpha}(Q)\} = \rho^P(X),$$

holds true.

Proof. The proof is analogous to Proposition 4.2.6 by replacing α by $\widetilde{\alpha}$ and noting that the weak-$*$-closed set $\mathcal{Q}^{\mathbb{P}} \subseteq \mathcal{M}_1$ is compact, since Ω is compact. □

Remark 4.2.15. Similarly as in Remark 4.2.7 all conditions of Corollary 4.2.14 are automatically satisfied, if Ω is compact and α happens to be $\alpha_{\min}$. This is because $\alpha_{\min}$ is convex and weak-$*$-lower semicontinuous (see Theorem 2.3.1 in Zalinescu (2002)), which translates to $\widetilde{\alpha}$ then.

4.3. Examples

In the following we shall give well-known examples of risk measures to illustrate the above comparison results.

We here additionally discuss possible differences between starting with a fixed acceptance set or rather with a fixed penalty function to define the initial risk measure.

4.3.1. Superhedging price as risk measure

Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0})$ be fixed and $\mathbb{P}$ a reference measure. Let $S = (S_t)_{t \geq 0}$ denote the price process of an asset and let T be a finite time horizon. We denote by $\mathcal{H}$ the class of $(\mathcal{F}_t)_{t \geq 0}$-predictable processes representing the set of admissible trading strategies, (see e.g. Föllmer and Schied (2011)) and by $(H \bullet S)$ the stochastic integral of $H \in \mathcal{H}$ with respect to S, representing the gains and losses up to time T, obtained from trading in S according to the strategy H. We recall here the standard definition of the superhedging price.

Definition 4.3.1 (Superhedging price)**.** Let $\mathbb{P} \in \mathcal{M}_1$ and $X \in L^\infty(\mathbb{P})$ be a European contingent claim. Then the superhedging price $\overline{\pi}(-X) \in \mathbb{R}$ of $-X$ is defined as

$$\overline{\pi}(-X) = \inf\{x \in \mathbb{R} : \exists H \in \mathcal{H} \text{ with } -X \leq x + (H \bullet S)_T, \mathbb{P}\text{-a.s.}\}. \tag{4.3.1}$$

Notice that by the superhedging duality (see e.g., Delbaen and Schachermayer (1999)), $\overline{\pi}(-X)$ can be equivalently written as

$$\overline{\pi}(-X) = \sup_{Q \in \mathcal{M}_a(\mathbb{P})} \mathbb{E}_Q[-X],$$

where $\mathcal{M}_a(\mathbb{P})$ denotes the set of absolutely continuous separating measures for S with respect to the reference measure $\mathbb{P}$, see for instance Kabanov (1997); Delbaen and Schachermayer (1999). This already gives the dual representation (with the corresponding α_{min}).

By the previous definition, for a financial agent "superhedging" means to find a self-financing strategy with minimal initial capital which covers any possible future obligation resulting from selling a European contingent claim.

We now show how the superhedging price can be embedded in our frameworks. There are two possibilities: one which does not exploit convexity and starts just with a monetary risk measure coming from an acceptance set to introduce $\rho_1^{\mathbb{P}}$, fully in spirit of Section 4.2.1; and the second which uses convexity and the minimal penalty function, in spirit of Section 4.2.2 to define in turn $\rho_2^{\mathbb{P}}$ and $\widehat{\rho}_2^{\mathbb{P}}$. We shall show that they all coincide.

We start by defining an acceptance set via

$$C := \{X \in \mathcal{X} : \exists H \in \mathcal{H} \text{ with } X(\omega) \geq -(H \bullet S)_T(\omega),\ \forall \omega \in \Omega\},$$

where the stochastic integral is here understood in discrete time, so that it is also well-defined in a pointwise sense. This then induces the corresponding risk measure $\rho_1 : \mathcal{X} \to \mathbb{R}$ via

$$\rho_1(X) = \inf\{m \in \mathbb{R} : m + X \in C\}. \tag{4.3.2}$$

For the reference measure $\mathbb{P}$, we then introduce according to Definition 4.2.1

$$\rho_1^{\mathbb{P}}(X) = \inf_{\{\widetilde{X} \in \mathcal{X} : \mathbb{P}(\widetilde{X} = X) = 1\}} \rho_1(\widetilde{X}),$$

which is equal to (4.3.1) since there the inequality has to hold only $\mathbb{P}$-a.s.

Since the superhedging price is a coherent risk measure, and thus convex we can introduce it via its dual representation. Fix α in (4.2.1) to be the following (minimal) penalty function

$$\alpha(Q) = \begin{cases} 0 & \text{for } Q \in \mathcal{M}_a(\mathbb{P}), \\ +\infty & \text{for } Q \in \mathcal{M}_1 \setminus \mathcal{M}_a(\mathbb{P}), \end{cases} \tag{4.3.3}$$

with $\mathcal{M}_a(\mathbb{P})$ denoting again the set of absolutely continuous separating measures with respect to $\mathbb{P}$, which correspond in the setup of discrete time to absolutely continuous martingale measures with respect to $\mathbb{P}$. Pursuing the second approach we can define a risk measure ρ_2 (pointwise) via (4.2.1) using the penalty function (4.3.3). Then $\alpha^{\mathbb{P}}$ as of (4.2.3) is equal to α and thus

$$\begin{aligned} \widehat{\rho}_2^{\mathbb{P}}(X) &= \sup_{Q \in \mathcal{M}_1} \{\mathbb{E}_Q[-X] - \alpha^{\mathbb{P}}(Q)\} \\ &= \sup_{Q \in \mathcal{M}_a(\mathbb{P})} \mathbb{E}_Q[-X]. \end{aligned}$$

Note that we could have started with any function α such that $\alpha^{\mathbb{P}}$ as defined in (4.2.3) is equal to the right hand side of (4.3.3) to get the same expression for $\widehat{\rho}_2^{\mathbb{P}}$. In particular the values outside $\mathcal{P}^{\mathbb{P}}$ do not matter. Using ρ_2, we can of course also consider $\rho_2^{\mathbb{P}}$ defined via

$$\rho_2^{\mathbb{P}}(X) = \inf_{\{\widetilde{X}\in\mathcal{X}:\mathbb{P}(\widetilde{X}=X)=1\}} \rho_2(\widetilde{X}) = \inf_{\{\widetilde{X}\in\mathcal{X}:\mathbb{P}(\widetilde{X}=X)=1\}} \sup_{Q\in\mathcal{M}_1} \{\mathbb{E}[-\widetilde{X}] - \alpha(Q)\}.$$

Due to the current choice of α, we can then verify that $\rho_1^{\mathbb{P}} = \rho_2^{\mathbb{P}} = \widehat{\rho}_2^{\mathbb{P}}$. Indeed, since α is of form (4.2.7), $\rho_2^{\mathbb{P}} = \widehat{\rho}_2^{\mathbb{P}}$ follows from Proposition 4.2.10 by setting $P = \mathbb{P}$, and $\rho_1^{\mathbb{P}} = \rho_2^{\mathbb{P}}$ by checking that the respective acceptance sets are equal.

In the following we introduce an additional measure $P \ll \mathbb{P}$ and consider the setup of Proposition 4.2.12 as well as Corollary 4.2.14 for the superhedging price in a simple trinomial model.

Example 4.3.2. Let us consider an arbitrage-free one-period trinomial model $S = (S_t)_{t=0,1}$, with zero interest rate, under some reference measure $\mathbb{P}$ which assigns positive probability to each scenario ω_i for $i = 1, 2, 3$. Here, ω_1 corresponds to the up-scenario, ω_2 to the middle one and ω_3 to the down-scenario. We suppose that $\mathcal{F}_0$ is trivial. Let $P \ll \mathbb{P}$ be such that $P(\{\omega_1, \omega_3\}) = 1$ and $P(\{\omega_2\}) = 0$. We consider as risk measure ρ again the superhedging price and introduce it via (4.2.1) with α of form (4.3.3).

Note that in this case the conditions of Proposition 4.2.12 hold true. Indeed, the first condition is clearly satisfied by Remark 4.2.13 as $|\Omega| = 3$. The second condition is also fulfilled since $\widetilde{\alpha} \equiv 0$ on $\mathcal{Q}^{\mathbb{P}} = \mathcal{M}_a(\mathbb{P})$. The assumption of Corollary 4.2.14 on $\widetilde{\alpha}$ is thus clearly also satisfied.

Consider now the concrete example $S_0 = 2$, $S_1(\omega_1) = 4$, $S_1(\omega_2) = 3$ and $S_1(\omega_3) = 1$. Then for all $Q \in \mathcal{M}_a(\mathbb{P})$, we have $Q(\{\omega_1\}) \in [0, \frac{1}{3}]$, $Q(\{\omega_2\}) = \frac{1-3Q(\{\omega_1\})}{2}$ and $Q(\{\omega_3\}) = \frac{1+Q(\{\omega_1\})}{2}$. For $Q(\{\omega_1\}) = \frac{1}{3}$ it holds that $Q \in \mathcal{M}_a(\mathbb{P}) \cap \mathcal{P}^P$ and this is actually the only element therein. Take now some claim $X \in L^\infty(P)$. Then since α^P is given by (4.2.9)

$$\widehat{\rho}^P(X) = -\frac{1}{3}X(\omega_1) - \frac{2}{3}X(\omega_3),$$

while

$$\begin{aligned}
\rho^P(X) &= \inf_{X(\omega_2)\in\mathbb{R}} \sup_{Q(\{\omega_1\})\in[0,\frac{1}{3}]} \left(-Q(\{\omega_1\})X(\omega_1) - \frac{1-3Q(\{\omega_1\})}{2}X(\omega_2) - \frac{1+Q(\{\omega_1\})}{2}X(\omega_3)\right) \\
&= \inf_{X(\omega_2)\in\mathbb{R}} \sup_{Q(\{\omega_1\})\in[0,\frac{1}{3}]} \left(-\frac{1}{2}(X(\omega_2)+X(\omega_3)) + Q(\{\omega_1\})(-X(\omega_1) + \frac{3}{2}X(\omega_2) - \frac{1}{2}X(\omega_3))\right) \\
&= \inf_{X(\omega_2)\in\mathbb{R}} \begin{cases} -\frac{1}{2}(X(\omega_2)+X(\omega_3)) & \text{if } X(\omega_2) \le \frac{2}{3}X(\omega_1) + \frac{1}{3}X(\omega_3), \\ -\frac{1}{3}X(\omega_1) - \frac{2}{3}X(\omega_3) & \text{else.} \end{cases} \\
&= -\frac{1}{3}X(\omega_1) - \frac{2}{3}X(\omega_3),
\end{aligned}$$

showing that we have indeed equality.

In the following we provide an example showing that Assumption (ii) of Lemma 4.2.12 or the lower-semicontinuity assumption of Theorem 4.2.14 on $\widetilde{\alpha}$ is crucial to have equality.

Example 4.3.3. Consider exactly the same trinomial model as in Example 4.3.2 above. Define however the penalty function $\alpha : \mathcal{M}_1 \to \mathbb{R} \cup \{+\infty\}$ as follows

$$\alpha(Q) = \begin{cases} \frac{Q(\{\omega_2\})^2}{2} - 1 & \text{for } Q \in \mathcal{M}_a(\mathbb{P}) \text{ s.t. } Q(\{\omega_2\}) > 0, \\ 0 & \text{for } Q \in \mathcal{M}_a(\mathbb{P}) \text{ s.t. } Q(\{\omega_2\}) = 0, \\ +\infty & \text{for } Q \in \mathcal{M}_1 \setminus \mathcal{M}_a(\mathbb{P}). \end{cases}$$

Note that with this choice $\widehat{\rho}^P$ still corresponds to the superhedging price, which is in this case the price of the perfect replication. To show that ρ^P is different, consider the trivial claim $X \in L^\infty(P)$, i.e. $X(\omega_1) = 0$ and $X(\omega_3) = 0$. Then $\widehat{\rho}^P = 0$ while

$$\rho^P(X) = \inf_{X(\omega_2)\in\mathbb{R}} \sup_{Q(\{\omega_2\})\in[0,\frac{1}{2}]} \left(-Q(\{\omega_2\})X(\omega_2) - \frac{Q(\{\omega_2\})^2}{2} + 1 \right) = 1,$$

showing that Condition (ii) of Proposition 4.2.12 is in general needed for equality. Note that $Q(\{\omega_2\}) \mapsto \alpha(Q(\{\omega_2\}))$ fails to be lower semi-continuous at 0, whence the Condition of Corollary 4.2.14 are also not satisfied.

4.3.2. Expected Shortfall

As next example we consider Expected Shortfall (ES) of level $\lambda \in (0, 1)$ for some reference measure $\mathbb{P}$, another example of a coherent risk measure. Fix its (minimal) penalty function

$$\alpha(Q) = \begin{cases} 0 & \text{for } Q \in \mathcal{Q}^{\mathbb{P}}, \\ +\infty & \text{for } Q \in \mathcal{M}_1 \setminus \mathcal{Q}^{\mathbb{P}}, \end{cases} \tag{4.3.4}$$

where $\mathcal{Q}^{\mathbb{P}} := \{Q \in \mathcal{M}_1 : Q \ll \mathbb{P}, \text{ and } \frac{\mathrm{d}Q}{\mathrm{d}\mathbb{P}} \leq \lambda^{-1}, \mathbb{P}\text{-a.s.}\}$. Then we define ρ via (4.2.1), i.e. $\rho(X) = \sup_{Q\in\mathcal{M}_1}\{\mathbb{E}_Q[-X] - \alpha(Q)\}$ for $X \in \mathcal{X}$ and since $\alpha = \alpha^{\mathbb{P}}$, $\widehat{\rho}^{\mathbb{P}}$ is given by

$$\widehat{\rho}^{\mathbb{P}}(X) = \sup_{Q\in\mathcal{Q}^{\mathbb{P}}} \mathbb{E}_Q[-X], \quad X \in L^\infty(\mathbb{P}).$$

Additionally, we introduce $\rho^{\mathbb{P}}$ according to the first approach as

$$\rho^{\mathbb{P}}(X) = \inf_{\{\widetilde{X}\in\mathcal{X}:\mathbb{P}(\widetilde{X}=X)=1\}} \rho(X) = \inf_{\{\widetilde{X}\in\mathcal{X}:\mathbb{P}(\widetilde{X}=X)=1\}} \sup_{Q\in\mathcal{M}_1} \{\mathbb{E}_Q[-\widetilde{X}] - \alpha(Q)\}.$$

As α is of form (4.2.7), we can conclude by applying Proposition 4.2.10 with $P = \mathbb{P}$ that $\rho^{\mathbb{P}} = \widehat{\rho}^{\mathbb{P}}$.

Also in the context of ES we now introduce a new measure $P \ll \mathbb{P}$ and consider the setup of Lemma 4.2.12 and Corollary 4.2.14 in a simple trinomial model, where by the same arguments as in Example 4.3.2 all conditions are satisfied. We here also aim to show the equality of ρ^P and $\widehat{\rho}^P$ by an explicit calculation.

Example 4.3.4. For some measure $P \ll \mathbb{P}$, let α^P be given by (4.2.9) with α of form (4.3.4) and consider

$$\widehat{\rho}^P(X) := \sup_{Q \in \mathcal{M}_1} \{\mathbb{E}_Q[-X] - \alpha^P(Q)\}.$$

Our goal is to show explicitly equality with

$$\rho^P(X) = \inf_{\{\widetilde{X} \in \mathcal{X} : P(\widetilde{X} = X) = 1\}} \sup_{Q \in \mathcal{M}_1} \{\mathbb{E}_Q[-\widetilde{X}] - \alpha(Q)\}$$

in the concrete trinomial setup of Example 4.3.2. Let $\mathbb{P}$ be such that $\mathbb{P}(\{\omega_i\}) = \frac{1}{3}$ for all $i = 1, 2, 3$ and P as above, i.e. $P(\{\omega_1, \omega_3\}) = 1$ and $P(\{\omega_2\}) = 0$. Then the elements $Q \in \mathcal{Q}^{\mathbb{P}} \cap \mathcal{P}^P$ are simply such that

$$Q(\{\omega_2\}) = 0, \qquad \frac{Q(\{\omega_i\})}{\mathbb{P}(\{\omega_i\})} \leq \frac{1}{\lambda}, \text{ for } i = 1, 3,$$

with the constraints $Q(\{\omega_i\}) \geq 0$ for $i = 1, 3$ and $Q(\{\omega_1\}) + Q(\{\omega_3\}) = 1$. Notice that if we fix $\lambda = 2/3$, then $\mathcal{Q}^{\mathbb{P}} \cap \mathcal{P}^P = \{Q^*\}$ with

$$Q^*(\{\omega_1\}) = Q^*(\{\omega_3\}) = \frac{1}{2}, \qquad Q^*(\{\omega_2\}) = 0.$$

For a claim $X \in L^\infty(P)$, we thus have

$$\widehat{\rho}^P(X) = -\frac{1}{2}(X(\omega_1) + X(\omega_3))$$

while

$$\rho^P(X) = \inf_{X(\omega_2) \in \mathbb{P}} \sup_{Q \in \mathcal{Q}^{\mathbb{P}}} \{-Q(\{\omega_1\})X(\omega_1) - Q(\{\omega_2\})X(\omega_2) - Q(\{\omega_3\})X(\omega_3)\}.$$

To compute $\rho^P(X)$, let us consider first the inner part which can be seen as a linear programming problem. Observe that the objective functional is a linear function in $Q(\omega_i)$ for all $i = 1, 2, 3$ and that $X(\omega_1)$ and $X(\omega_3)$ are fixed. Additionally recall that the set of probability measures with Ω as above, can be identified with the two-dimensional simplex and that $\mathcal{Q}^{\mathbb{P}} = \{Q \in \mathcal{M}_1 : Q(\{\omega_i\}) \leq \frac{1}{2},\ \forall i = 1, 2, 3\}$. Therefore the supremum will be attained on one of the three extremal points of $\mathcal{Q}^{\mathbb{P}}$, namely in those probability measures under which two scenarios occur with probability $\frac{1}{2}$ and the third with probability zero. For these reasons

$$\sup_{Q \in \mathcal{Q}^{\mathbb{P}}} -\sum_{j=1}^{3} Q(\{\omega_j\})X(\omega_j) = -\frac{1}{2}\max\{X(\omega_1) + X(\omega_3), X(\omega_1) + X(\omega_2), X(\omega_3) + X(\omega_2)\}.$$

Let us now consider the outer problem with the infimum, hence $\rho^P(X)$ can be equivalently rewritten as

$$\rho^P(X) = -\frac{1}{2} \sup_{X(\omega_2) \in \mathbb{R}} \min\{X(\omega_1) + X(\omega_3), X(\omega_1) + X(\omega_2), X(\omega_3) + X(\omega_2)\}.$$

Let us consider without loss of generality two cases, namely $X(\omega_1) < X(\omega_3)$ and $X(\omega_1) \geq X(\omega_3)$. In the first one we have

$$\min\{X(\omega_1)+X(\omega_3), X(\omega_1)+X(\omega_2), X(\omega_3)+X(\omega_2)\} = \begin{cases} X(\omega_1) + X(\omega_3) & \text{if } X(\omega_3) \leq X(\omega_2), \\ X(\omega_1) + X(\omega_2) & \text{if } X(\omega_3) > X(\omega_2). \end{cases}$$

Similarly in the second case, we get

$$\min\{X(\omega_1)+X(\omega_3), X(\omega_1)+X(\omega_2), X(\omega_3)+X(\omega_2)\} = \begin{cases} X(\omega_1) + X(\omega_3) & \text{if } X(\omega_1) \leq X(\omega_2), \\ X(\omega_2) + X(\omega_3) & \text{if } X(\omega_1) > X(\omega_2). \end{cases}$$

Therefore in both cases by taking the supremum over $X(\omega_2) \in \mathbb{R}$ we have that

$$\rho^P(X) = -\frac{1}{2}(X(\omega_1) + X(\omega_3)),$$

which coincides with $\widehat{\rho}^P(X)$ as claimed.

4.3.3. The worst case risk measure

The worst case risk measure is an additional example of a risk measure which is part of our framework. Similarly as for the superhedging price it can be introduced via an acceptance set, in spirit of Section 4.2.1 or via a particular penalty function in spirit of Section 4.2.2.

In the following we fix $a \geq 0$ to cover a shifted version of the original worst case risk measure, but we stress that for $a = 0$ the definition coincides with the one considered in Föllmer and Schied (2011). Let C_a be the convex acceptance set defined as

$$C_a := \{X \in \mathcal{X} : X(\omega) \geq a, \ \forall \omega \in \Omega\}.$$

Then the corresponding risk measure is given by

$$\rho_{\max,1}(X) = \inf\{m \in \mathbb{R} : m + X \in C_a\} \tag{4.3.5}$$

or equivalently by

$$\rho_{\max,1}(X) = -\inf_{\omega\in\Omega} X(\omega) + a.$$

If we fix the penalty function $\alpha(Q) = -a$ for all $Q \in \mathcal{M}_1$, we can define

$$\rho_{max,2}(X) := \sup_{Q\in\mathcal{M}_1} \{\mathbb{E}_Q[-X] - \alpha(Q)\} = \sup_{Q\in\mathcal{M}_1} \mathbb{E}_Q[-X] + a$$

and see that this coincides with (4.3.5).

For a reference measure $\mathbb{P}$, we can then introduce according to Definition 4.2.1 the acceptance set to be

$$C^{\mathbb{P}} = \{X \in L^\infty(\mathbb{P}) : X \geq a, \ \mathbb{P}\text{-a.s.}\}$$

and

$$\rho^{\mathbb{P}}_{\max,1}(X) := \inf_{\{\widetilde{X}\in\mathcal{X}:\mathbb{P}(\widetilde{X}=X)=1\}} \rho_{\max,1}(\widetilde{X}) = -\mathbb{P}\text{- ess inf}_{\omega\in\Omega} X(\omega) + a. \tag{4.3.6}$$

Since $\rho_{\max,1} = \rho_{\max,2}$, it clearly holds that

$$\rho^{\mathbb{P}}_{\max,1}(X) = \rho^{\mathbb{P}}_{\max,2}(X) = \inf_{\{\widetilde{X}\in\mathcal{X}:\mathbb{P}(\widetilde{X}=X)=1\}} \sup_{Q\in\mathcal{M}_1} \mathbb{E}[-\widetilde{X}] + a.$$

Starting from the second approach we can additionally introduce for all $X \in L^\infty(\mathbb{P})$,

$$\widehat{\rho}^{\mathbb{P}}_{\max,2}(X) = \sup_{Q\in\mathcal{P}^{\mathbb{P}}} \mathbb{E}[-X] + a.$$

This also coincides with $\rho^{\mathbb{P}}_{\max,1}(X) = \rho^{\mathbb{P}}_{\max,2}(X)$ which follows from Proposition 4.2.10 by setting $\widetilde{\alpha} = -a$ on $\mathcal{P}^{\mathbb{P}}$.

4.4. Robust risk measures and their classical representation

We shall now use the above approaches of the defining risk measures with respect to certain sets of reference measures to include model uncertainty. To this end assume that we are given a family of probability measures $\mathcal{M} \subset \mathcal{M}_1$ defined on the same measurable space $(\Omega, \mathcal{F})$. In order to be able to consider all risk measures for all $P \in \mathcal{M}$, we choose X to be in the space $\mathcal{X}^{\mathcal{M}}$, with

$$\mathcal{X}^{\mathcal{M}} := \bigcap_{P\in\mathcal{M}} L^\infty(P), \tag{4.4.1}$$

where the elements of the previous have to be understood as intersections of equivalence classes. Note that if $X \in \mathcal{X}^{\mathcal{M}}$ then, for all $P \in \mathcal{M}$, we have that $P(X \in \mathcal{X}) = 1$. Let us now define the robust risk measure $\rho^{\mathcal{M}}$ as the supremum over all risk measures ρ^P, with $P \in \mathcal{M}$ and similarly $\widehat{\rho}^{\mathcal{M}}$ as the supremum over all risk measures $\widehat{\rho}^P$.

Definition 4.4.1. Let $X \in \mathcal{X}^{\mathcal{M}}$. Define the robust risk measure of X according to the class of models $\mathcal{M}$ as

$$\rho^{\mathcal{M}}(X) := \sup_{P\in\mathcal{M}} \rho^P(X).$$

Similarly, let $X \in \mathcal{X}^{\mathcal{M}}$ and fix $\alpha : \mathcal{M}_1 \to \mathbb{R} \cup \{+\infty\}$. Define the robust risk measure of X according to the class of model $\mathcal{M}$ as

$$\begin{aligned}\widehat{\rho}^{\mathcal{M}}(X) :&= \sup_{P\in\mathcal{M}} \widehat{\rho}^P(X) \\ &= \sup_{Q\in\mathcal{M}_1} \{\mathbb{E}_Q[-X] - \alpha^{\mathcal{M}}(Q)\},\end{aligned}$$

where

$$\alpha^{\mathcal{M}}(Q) = \begin{cases} \alpha(Q) & \text{for } Q \in \mathcal{P}^{\mathcal{M}}, \\ +\infty & \text{for } Q \in \mathcal{M}_1 \setminus \mathcal{P}^{\mathcal{M}}, \end{cases}$$

with $\mathcal{P}^{\mathcal{M}} := \bigcup_{P \in \mathcal{M}} \mathcal{P}^P$.

Remark 4.4.2. Observe that the previous robust risk measures fall into the setup of *generalized risk measures* introduced recently by Fadina et al. (2021). Let $\mathcal{M}_1$ be the set of probability measure here assumed to be atomless, and $2^{\mathcal{M}_1}$ the collection of subsets of $\mathcal{M}_1$. A generalized risk measure is a map $\Psi : \mathcal{X} \times 2^{\mathcal{M}_1} \to \mathbb{R}$. Consider the following axioms

(A1) $\Psi(X|\mathcal{Q}) \leq \Psi(X|\mathcal{R})$, for all $X \in \mathcal{X}$, $\mathcal{Q} \subset \mathcal{R} \subset \mathcal{M}_1$;

(A2) $\Psi(X|\{P\}) \leq \Psi(Y|\{P\})\ \forall P \in \mathcal{Q} \implies \Psi(X|\mathcal{Q}) \leq \Psi(Y|\mathcal{Q})$, for any $X, Y \in \mathcal{X}$;

(A3) $\Psi(X|\mathcal{Q}) \leq \sup_{P \in \mathcal{Q}} \Psi(X|\{P\})$, for all $X \in \mathcal{X}$ and $\mathcal{Q} \subset \mathcal{M}_1$.

In Fadina et al. (2021) is shown that if Ψ satisfies A1-A3, then $\Psi(X|\mathcal{Q}) = \sup_{P \in \mathcal{Q}} \Psi(X|\{P\})$. By setting $\mathcal{Q} = \mathcal{M}$ and $\Psi(X|\{P\}) := \rho^P(X)$ in our setup we notice immediately that

$$\Psi(X|\mathcal{M}) = \sup_{P \in \mathcal{M}} \Psi(X|\{P\}) = \sup_{P \in \mathcal{M}} \rho^P(X),$$

can be seen as generalized risk measure on $\mathcal{X} \times 2^{\mathcal{M}_1}$. Similar considerations hold for $\widehat{\rho}^P$. In Section 4.4.1, under certain assumptions on $\mathcal{M}$, we will address the problem of finding a measure $\mathbb{P}$ such that

$$\Psi(X|\mathcal{M}) = \Psi(X|\{\mathbb{P}\}), \qquad \forall X \in L^\infty(\mathbb{P}),$$

i.e. we ask when the above generalized risk measure has a representation as a classical risk measure.

Let us now introduce an extended set of probability measures induced by $\mathcal{M}$, namely

$$c\mathcal{M} := \overline{conv}(\mathcal{M}), \tag{4.4.2}$$

where $\overline{conv}$ denotes the set of all probability measures which are countable convex combinations of the elements in $\mathcal{M}$.

Sometimes we shall replace $\mathcal{M}$ via the set $c\mathcal{M}$ in the definition of the robust risk measure $\rho^{\mathcal{M}}$, i.e.

$$\rho^{c\mathcal{M}}(X) = \sup_{P \in c\mathcal{M}} \rho^P(X), \qquad \forall X \in \mathcal{X}^{c\mathcal{M}}$$

and similarly for $\widehat{\rho}^{c\mathcal{M}}$. Observe that $\rho^{\mathcal{M}}(X) \leq \rho^{c\mathcal{M}}(X)$ for $X \in \mathcal{X}^{c\mathcal{M}}$ with a possible strict inequality for certain $X \in \mathcal{X}^{c\mathcal{M}}$ and the analogous statement holds for $\widehat{\rho}^{c\mathcal{M}}$.

In the following we give an example where the inequality is strict even with finite convex combinations. We consider in particular the case of Ω being compact and we choose the risk measure to be coherent. We work with $\widehat{\rho}^P$ but remark that the next example also holds true for ρ^P, which follows from Corollary 4.2.14 or Proposition 4.2.10 with $R = P$.

Example 4.4.3. Let $(\Omega, \mathcal{F}) = ([0,1], \mathcal{B}([0,1]))$ and suppose $\mathcal{M}$ is induced by a parameter set Θ with a σ-algebra $\mathcal{D}$, i.e., $(\Theta, \mathcal{D}) = ([0,1], \mathcal{B}([0,1]))$ as well. Let $N > 2$ and $\gamma = \frac{1}{N}$. We define $\mathcal{M} := \{P^\theta : \theta \in \Theta\}$ where P^θ is given as follows. For all $A \in \mathcal{B}([0,1])$

$$P^\theta(A) := \begin{cases} \frac{2}{\gamma}\Lambda(A \cap [\theta, \theta + \frac{\gamma}{2}]) & \text{for all } \theta \in [0, 1 - \frac{\gamma}{2}], \\ \frac{2}{\gamma}\Lambda(A \cap [\theta, 1]) & \text{for all } \theta \in (1 - \frac{\gamma}{2}, 1], \end{cases}$$

where Λ denotes the Lebesgue measure on $[0,1]$. We set in the following $\widehat{\rho}^\theta := \widehat{\rho}^{P^\theta}$ for all $\theta \in \Theta$ and $\widehat{\rho}^{\mathcal{M}}$ as above. Additionally with VaR_γ^Λ we refer to the Value at Risk of level $\gamma \in (0,1)$ under Λ.

Let $X := -1_{[0,2\gamma]}$, then

$$VaR_\gamma^\Lambda(X) = \inf\{m \in \mathbb{R} : \Lambda(X + m < 0) \leq \gamma\} = 1.$$

Indeed for all $\varepsilon > 0$,

$$\Lambda(-1_{[0,2\gamma]} + 1 - \varepsilon < 0) = \Lambda(-\varepsilon 1_{[0,2\gamma]} + (1-\varepsilon)1_{(2\gamma,1]} < 0) = \Lambda([0, 2\gamma]) = 2\gamma > \gamma,$$

and

$$\Lambda(-1_{[0,2\gamma]} + 1 < 0) = 0 < \gamma.$$

Moreover, for all $\beta \in (0, \gamma]$ we have that $VaR_\beta^\Lambda(X) = 1$, thus $\rho(X) = 1$ where ρ is the Expected Shortfall of level γ as defined in 4.3.2 via the set $\mathcal{Q}^\Lambda$. Define similarly as in Section 4.3.2

$$\widehat{\rho}^\theta(X) = \sup_{Q \in \mathcal{M}_1} \{\mathbb{E}_Q[-X] - \alpha^\theta(Q)\},$$

where

$$\alpha^\theta(Q) = \begin{cases} 0 & \text{for } Q \in \mathcal{Q}^\Lambda \cap \mathcal{P}^\theta, \\ +\infty & \text{for } Q \in \mathcal{M}_1 \setminus (Q^\Lambda \cap \mathcal{P}^\theta), \end{cases}$$

with

$$\mathcal{Q}^\Lambda = \{Q \in \mathcal{M}_1 : Q \ll \Lambda, \text{ and } \frac{\mathrm{d}Q}{\mathrm{d}\Lambda} \leq \gamma^{-1}, \Lambda - a.s.\}.$$

Let us consider now $Q \in \mathcal{Q}^\Lambda \cap \mathcal{P}^\theta$. First of all since $Q \in \mathcal{Q}^\Lambda$ we notice that

$$\begin{aligned} Q\left([\theta, \theta + \frac{\gamma}{2}]\right) &= \mathbb{E}_\Lambda\left[\frac{\mathrm{d}Q}{\mathrm{d}\Lambda} 1_{[\theta, \theta + \gamma/2]}\right] \\ &\leq \frac{1}{\gamma} \cdot \frac{\gamma}{2} = \frac{1}{2}. \end{aligned}$$

On the other hand since $Q \in \mathcal{P}^\theta$, we have that $\mathrm{supp}(Q) \subseteq [\theta, \theta + \frac{\gamma}{2}]$ yielding a contradiction as

$$Q\left([\theta, \theta + \frac{\gamma}{2}]\right) = 1.$$

Hence $\mathcal{Q}^\Lambda \cap \mathcal{P}^\theta = \emptyset$ for all $\theta \in [0,1]$, meaning that the penalty function always attains infinite values, namely $\alpha^\theta(Q) = +\infty$ for all $Q \in \mathcal{M}_1$ and all $\theta \in [0,1]$. Therefore $\widehat{\rho}^\theta(X) = -\infty$ and consequently $\widehat{\rho}^{\mathcal{M}}(X) = -\infty$.
Let us now consider $c\mathcal{M} = \overline{conv}(\mathcal{M})$ and show that $\Lambda \in c\mathcal{M}$. To this extent fix the sequence $(\tilde{\theta}_k)_{k=0}^{2N} \subseteq [0,1]$ such that for all $k = 0, \dots, 2N$ we define $\tilde{\theta}_k := k\frac{\gamma}{2} = \frac{k}{2N}$. Then for every $A \in \mathcal{B}([0,1])$

$$\Lambda(A) = \frac{\gamma}{2} \sum_{k=1}^{2N} \frac{2}{\gamma} \Lambda\left(A \cap \left[\tilde{\theta}_k, \tilde{\theta}_k + \frac{\gamma}{2}\right]\right) = \sum_{k=1}^{2N} \frac{\gamma}{2} P^{\tilde{\theta}_k}(A),$$

hence $\Lambda \in c\mathcal{M}$. In particular $\widehat{\rho}^{c\mathcal{M}}(X) = \widehat{\rho}^{\Lambda}(X) = \rho(X) = 1 > \widehat{\rho}^{\mathcal{M}}(X)$.

4.4.1. Classical representation

Let us now come to one of our main goals namely to show equality of the robust risk measure (with respect to a set $\mathcal{M}$) with a classical risk measure with respect to one single probability measure $\mathbb{P}$. In this section we analyze under which conditions this is the case.

Let us start by recalling a version of the well known result of Halmos and Savage (1949) (see Lemma 7).

Theorem 4.4.4 (Halmos and Savage (1949))**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let $\mathcal{M}$ be a set of $\mathbb{P}$-absolutely continuous probability measures on $(\Omega, \mathcal{F})$ that is closed under countable convex combinations. Suppose that, for each $A \in \mathcal{F}$ with $\mathbb{P}(A) > 0$, there exists $P \in \mathcal{M}$ (depending on A) with $P(A) > 0$. Then there exists $P^0 \in \mathcal{M}$ such that, for all $A \in \mathcal{F}$ with $\mathbb{P}(A) > 0$, we have that $P^0(A) > 0$, that is P^0 and $\mathbb{P}$ are equivalent probability measures.*

Next we introduce the notion of *generalized equivalence*, a version of absolute continuity between the elements of $\mathcal{M}$ and a given probability measure which does not necessarily belong to $\mathcal{M}$. It means in particular that there exists a dominating measure for the set $\mathcal{M}$.

Definition 4.4.5. Let $\mathcal{M}$ be a family of probability measures defined on $(\Omega, \mathcal{F})$ and let $\mathbb{P}$ be a some probability measure defined on $(\Omega, \mathcal{F})$. The model $\mathcal{M}$ and the measure $\mathbb{P}$ are said to be *generalized equivalent*

(i) if $P \ll \mathbb{P}$ for all $P \in \mathcal{M}$, denoted by $\mathcal{M} \ll \mathbb{P}$, and

(ii) if, for $A \in \mathcal{F}$, the condition $P(A) = 0$ for all $P \in \mathcal{M}$ implies $\mathbb{P}(A) = 0$, denoted by $\mathbb{P} \ll \mathcal{M}$.

Using Theorem 4.4.4 and the notion of generalized equivalence now allows to formulate the precise statement when $\rho^{\mathbb{P}}(X) = \rho^{\mathcal{M}}(X)$ holds.

Theorem 4.4.6. *Let $\mathcal{M} \subseteq \mathcal{M}_1$ and $\mathbb{P}$ a probability measure on $(\Omega, \mathcal{F})$. Suppose $\mathbb{P}$ is generalized equivalent to $\mathcal{M}$ and that $\mathcal{M}$ is closed under countable convex combinations. Then for all $X \in L^\infty(\mathbb{P})$,*

$$\rho^{\mathbb{P}}(X) = \rho^{\mathcal{M}}(X).$$

Proof. By Theorem 4.4.4 there exists $P^0 \in \mathcal{M}$ such that $\mathbb{P}$ is equivalent to P^0. Since in particular $\mathbb{P}$ is dominated by P^0, by Remark 4.2.2 we have that for all $X \in L^\infty(\mathbb{P})$

$$\rho^{\mathbb{P}}(X) = \rho^{P^0}(X) \leq \rho^{\mathcal{M}}(X).$$

On the other hand since for all $P \in \mathcal{M}$, $P \ll \mathbb{P}$ then, for all $X \in L^\infty(\mathbb{P})$

$$\rho^{\mathcal{M}}(X) = \sup_{P \in \mathcal{M}} \rho^P(X) \leq \rho^{\mathbb{P}}(X),$$

which concludes the proof. □

Remark 4.4.7. Note that the above result also holds true when $\rho^{\mathbb{P}}$ and $\rho^{\mathcal{M}}$ are replaced by $\widehat{\rho}^{\mathbb{P}}$ and $\widehat{\rho}^{\mathcal{M}}$, simply by applying the same proof. As it also follows from the next more general result, we only formulated it for $\rho^{\mathbb{P}}$ and $\rho^{\mathcal{M}}$.

In the following theorem we characterize when equality of $\rho^{\mathbb{P}}(X)$ with $\rho^{\mathcal{M}}(X)$ holds.

Theorem 4.4.8. *Let $\mathcal{M} \subseteq \mathcal{M}_1$ and $\mathbb{P}$ a probability measure on $(\Omega, \mathcal{F})$. Fix a penalty function $\alpha : \mathcal{M}_1 \to \mathbb{R} \cup \{+\infty\}$.*

- *Then, the following assertions are equivalent.*
  - *(i) For all $X \in \mathcal{X}^{\mathcal{M}} \cap L^\infty(\mathbb{P})$, it holds $\widehat{\rho}^{\mathcal{M}}(X) = \widehat{\rho}^{\mathbb{P}}(X)$.*
  - *(ii)*
$$\bigcup_{P \in \mathcal{M}} \mathcal{P}^P \cap \{Q : \alpha(Q) < \infty\} = \mathcal{P}^{\mathbb{P}} \cap \{Q : \alpha(Q) < \infty\}.$$
- *The above assertions are implied by the following condition:*
  - *(iii) All $P \in \mathcal{M}$ are dominated by $\mathbb{P}$ and there is a measure $P_0 \in \mathcal{M}$ such that $\mathbb{P} \ll P_0$.*

  Moreover, if α is finitely valued on $\bigcup_{P \in \mathcal{M}} \mathcal{P}^P \cup \mathcal{P}^{\mathbb{P}}$, (iii) is equivalent to (i) and (ii).
- *Condition (iii) is implied if $\mathcal{M}$ is closed under countable convex combinations and $\mathbb{P}$ generalized equivalent to $\mathcal{M}$. In this case $\mathcal{X}^{\mathcal{M}} \cap L^\infty(\mathbb{P}) = L^\infty(\mathbb{P})$.*

Proof. $(i) \implies (ii)$. Let $X \in \mathcal{X}^{\mathcal{M}} \cap L^\infty(\mathbb{P})$. Consider

$$\widehat{\rho}^{\mathcal{M}}(X) = \sup_{P \in \mathcal{M}} \sup_{Q \in \mathcal{M}_1} \{\mathbb{E}_Q[-X] - \alpha^P(Q)\} = \sup_{Q \in \mathcal{M}_1} \{\mathbb{E}_Q[-X] - \alpha^{\mathcal{M}}(Q)\},$$

where

$$\alpha^{\mathcal{M}}(Q) = \begin{cases} \inf_{P \in \mathcal{M}} \alpha^P(Q) & \text{for } Q \in \bigcup_{P \in \mathcal{M}} \mathcal{P}^P, \\ +\infty & \text{for } Q \in \mathcal{M}_1 \setminus \bigcup_{P \in \mathcal{M}} \mathcal{P}^P, \end{cases} \tag{4.4.3}$$

and

$$\widehat{\rho}^{\mathbb{P}}(X) = \sup_{Q \in \mathcal{M}_1} \{\mathbb{E}_Q[-X] - \alpha^{\mathbb{P}}(Q)\},$$

with

$$\alpha^{\mathbb{P}}(Q) = \begin{cases} \alpha(Q) & \text{for } Q \in \mathcal{P}^{\mathbb{P}}, \\ +\infty & \text{for } Q \in \mathcal{M}_1 \setminus \mathcal{P}^{\mathbb{P}}. \end{cases} \tag{4.4.4}$$

If (i) holds true then,

$$\{X \in \mathcal{X}^{\mathcal{M}} \cap L^{\infty}(\mathbb{P}) : \widehat{\rho}^{\mathbb{P}}(X) \leq 0\} = \{X \in \mathcal{X}^{\mathcal{M}} \cap L^{\infty}(\mathbb{P}) : \widehat{\rho}^{\mathcal{M}}(X) \leq 0\}.$$

By the definition of the risk measures the previous yields

$$\begin{aligned} &\{X \in \mathcal{X}^{\mathcal{M}} \cap L^{\infty}(\mathbb{P}) : \mathbb{E}_Q[-X] \leq \alpha^{\mathbb{P}}(Q), \forall Q \in \mathcal{M}_1\} \\ &\quad = \{X \in \mathcal{X}^{\mathcal{M}} \cap L^{\infty}(\mathbb{P}) : \mathbb{E}_Q[-X] \leq \alpha^{\mathcal{M}}(Q), \forall Q \in \mathcal{M}_1\}. \end{aligned}$$

Hence, by the definition of the penalty functions (4.4.3) and (4.4.4) we obtain

$$\begin{aligned} &\{X \in \mathcal{X}^{\mathcal{M}} \cap L^{\infty}(\mathbb{P}) : \mathbb{E}_Q[-X] \leq \alpha(Q),\ \forall Q \in \mathcal{P}_f^{\mathbb{P}}\} \\ &\quad = \{X \in \mathcal{X}^{\mathcal{M}} \cap L^{\infty}(\mathbb{P}) : \mathbb{E}_Q[-X] \leq \inf_{P \in \mathcal{M}} \alpha^P(Q),\ \forall Q \in \bigcup_{P \in \mathcal{M}} \mathcal{P}_f^P\} \\ &\quad = \{X \in \mathcal{X}^{\mathcal{M}} \cap L^{\infty}(\mathbb{P}) : \mathbb{E}_Q[-X] \leq \alpha(Q),\ \forall Q \in \bigcup_{P \in \mathcal{M}} \mathcal{P}_f^P\}, \end{aligned}$$

where $\mathcal{P}_f^P := \mathcal{P}^P \cap \{Q \in \mathcal{M}_1 : \alpha(Q) < +\infty\}$ for any $P \in \mathcal{M}_1$. Hence we can conclude that $\mathcal{P}_f^{\mathbb{P}} = \bigcup_{P \in \mathcal{M}} \mathcal{P}_f^P$.
$(ii) \implies (i)$. Let $X \in \mathcal{X}^{\mathcal{M}} \cap L^{\infty}(\mathbb{P})$. Then,

$$\begin{aligned} \widehat{\rho}^{\mathcal{M}}(X) = \sup_{P \in \mathcal{M}} \widehat{\rho}^P(X) &= \sup_{Q \in \mathcal{M}_1} \{\mathbb{E}_Q[-X] - \alpha^{\mathcal{M}}(Q)\} \\ &= \sup_{Q \in \bigcup_{P \in \mathcal{M}} \mathcal{P}^P} \{\mathbb{E}_Q[-X] - \inf_{P \in \mathcal{M}} \alpha^P(Q)\} \\ &= \sup_{Q \in \bigcup_{P \in \mathcal{M}} \mathcal{P}_f^P} \{\mathbb{E}_Q[-X] - \alpha(Q)\} \\ &\overset{(*)}{=} \sup_{Q \in \mathcal{P}_f^{\mathbb{P}}} \{\mathbb{E}_Q[-X] - \alpha(Q)\} \\ &= \widehat{\rho}^{\mathbb{P}}(X), \end{aligned}$$

where in $(*)$ the hypothesis (ii) is employed, yielding the identity between the risk measures. Observe that if (iii) holds, this implies

$$\bigcup_{P \in \mathcal{M}} \mathcal{P}^P = \mathcal{P}^{\mathbb{P}}$$

and thus (ii) as the previous trivially gives $\bigcup_{P \in \mathcal{M}} \mathcal{P}_f^P = \mathcal{P}_f^{\mathbb{P}}$. Notice additionally that if $\mathcal{M}$ is closed under countable convex combination and $\mathbb{P}$ is generalized equivalent to $\mathcal{M}$, then (iii) is satisfied as a consequence of Halmos-Savage's Theorem. □

In the following we apply the above result to the situation considered in Section 4.2.4, which shows again that the reference measure with respect to which we consider law-invariance qualifies as 'worst case probability measure'. Note that for this result the generalized equivalence property is not necessary and it thus works in a non-dominated situation.

Corollary 4.4.9. *Let $\mathcal{M} \subseteq \mathcal{M}_1$ and let $\mathbb{P} \in \mathcal{M}$ a probability measure on $(\Omega, \mathcal{F})$ with respect to which we consider law-invariance as in Section 4.2.4. Fix a penalty function $\alpha : \mathcal{M}_1 \to \mathbb{R} \cup \{+\infty\}$ as in (4.2.7) with respect to $\mathbb{P}$. Then for all $X \in \mathcal{X}^{\mathcal{M}} \cap L^\infty(\mathbb{P})$ we have*

$$\widehat{\rho}^{\mathcal{M}}(X) = \widehat{\rho}^{\mathbb{P}}(X) = \rho^{\mathbb{P}}(X).$$

Proof. We apply Theorem 4.4.8 and verify (ii). Note that

$$\bigcup_{P \in \mathcal{M}} \mathcal{P}^P \cap \{Q : \alpha(Q) < \infty\} = \Big(\bigcup_{P \in \mathcal{M}} \mathcal{P}^P \Big) \cap \mathcal{Q}^{\mathbb{P}} = \mathcal{Q}^{\mathbb{P}},$$

where the latter follows from the fact that $\mathbb{P} \in \mathcal{M}$. Moreover, we clearly also have

$$\mathcal{P}^{\mathbb{P}} \cap \{Q : \alpha(Q) < \infty\} = \mathcal{Q}^{\mathbb{P}},$$

whence property (ii) is satisfied and the claim follows. The equality $\widehat{\rho}^{\mathbb{P}}(X) = \rho^{\mathbb{P}}(X)$ for all $X \in \mathcal{X}^{\mathcal{M}} \cap L^\infty(\mathbb{P})$ follows from Proposition 4.2.10. □

Note that Corollary 4.4.9 could of course also have been deduced from Proposition 4.2.10. This result yields additionally the same assertion for $\rho^{\mathcal{M}}$ which we state here for completeness. Note again that the generalized equivalence property is not required.

Corollary 4.4.10. *Let $\mathcal{M} \subseteq \mathcal{M}_1$ and let $\mathbb{P} \in \mathcal{M}$ a probability measure on $(\Omega, \mathcal{F})$ with respect to which we consider law-invariance as in Section 4.2.4. Fix a penalty function $\alpha : \mathcal{M}_1 \to \mathbb{R} \cup \{+\infty\}$ as in (4.2.7) with respect to $\mathbb{P}$. Then for all $X \in \mathcal{X}^{\mathcal{M}} \cap L^\infty(\mathbb{P})$ we have*

$$\widehat{\rho}^{\mathcal{M}}(X) = \widehat{\rho}^{\mathbb{P}}(X) = \rho^{\mathbb{P}}(X) = \rho^{\mathcal{M}}(X).$$

Proof. This is a consequence of the last statement of Proposition 4.2.10. □

Remark 4.4.11. (i) Let us stress the significance of the above results. Whenever $\mathbb{P}$ and $\mathcal{M}$ are generalized equivalent and $\mathcal{M}$ is closed under countable convex combinations, then the robust risk measure for a given $X \in L^\infty(\mathbb{P})$ is nothing else than the classical risk measure for the probability $\mathbb{P}$. So the current robust theory, i.e., the theory under model uncertainty, just reduces to the classical theory. A particular case of this situation is if $\mathcal{M} \subset \mathcal{M}_1$ is countable, see Example 4.4.12 below.

(ii) In the setup of Section 4.2.4 the probability measure with respect to which we consider law-invariance is the natural candidate to achieve a classical representation *without domination property and convexity assumptions* provided that it is contained in the set $\mathcal{M}$.

(iii) Recall the space of random variables defined in (4.4.1). Concerning the assumption $X \in L^\infty(\mathbb{P})$, this cannot be replaced without additional considerations by $X \in \mathcal{X}^{\mathcal{M}}$, as we see in Example 4.4.13 below. This example shows that $X \in \mathcal{X}^{\mathcal{M}}$ does not imply $X \in L^\infty(\mathbb{P})$ not even in the case where $\mathcal{M}$ is countable. It only holds for $\mathcal{M}$ finitely generated.

Example 4.4.12. Let Θ be a countable parameter space i.e., $\Theta = \{\theta_i, i \in \mathbb{N}\}$ and $\mathcal{M}^\Theta := \{P^\theta : \theta \in \Theta\}$ a collection of probability measures. Consider as the dominating measure $\mathbb{P}$ any countable convex combination that gives strictly positive weight to all θ_i, e.g.

$$\mathbb{P}(A) := \sum_{i=1}^{+\infty} 2^{-i} P^{\theta_i}(A), \qquad \forall A \in \mathcal{F}.$$

Then for all $i \in \mathbb{N}$ we have that $P^{\theta_i} \ll \mathbb{P}$. Furthermore if $P^{\theta_i}(A) = 0$, for all $i \in \mathbb{N}$, then $\mathbb{P}(A) = 0$, which implies $\mathbb{P} \ll \mathcal{M}^\Theta$.

Example 4.4.13. Let $(\Omega, \mathcal{F}) = ([0,1], \mathcal{B}([0,1]))$ and let $\mathcal{M} = \{P^n : n \in \mathbb{N} \cup \{0\}\}$ where

$$\frac{\mathrm{d}P^n}{\mathrm{d}\Lambda}(\omega) = 2^{n+1} 1_{(2^{-(n+1)}, 2^{-n}]}(\omega),$$

for all $\omega \in [0,1]$ where Λ here denotes the Lebesgue measure. Define

$$\mathbb{P}(A) := \sum_{n=0}^{+\infty} 2^{-(n+1)} P^n(A), \qquad \forall A \in \mathcal{B}[0,1].$$

Then $\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\Lambda} = 1_{[0,1]}$ Λ–a.s., hence $\mathbb{P}$ can be identified with the Lebesgue-measure. Let $X(\omega) = \frac{1}{\omega}$. We have that, for all n, $P^n(X \leq 2^{n+1}) = 1$, hence $X \in \mathcal{X}^{\mathcal{M}} = \bigcap_{n \geq 0} L^{+\infty}(P^n)$. But on the other hand $X \notin L^\infty(\mathbb{P})$.

4.5. Mixture probability measure

We now include model uncertainty assuming that we are given a family of probability measures $\mathcal{M}^\Theta = \{P^\theta : \theta \in \Theta\}$ defined on the same measurable space $(\Omega, \mathcal{F})$ and where Θ is a parameter space. We assume that, for all $A \in \mathcal{F}$, $\theta \mapsto P^\theta(A)$ is a measurable map from $(\Theta, \mathcal{D}) \to ([0,1], \mathcal{B}([0,1]))$, where $\mathcal{B}([0,1])$ are the Borel sets on $[0,1]$ and $\mathcal{D}$ is a σ-algebra on Θ. We assume that we have a priori subjective beliefs in the degree of validity of the possible measures P^θ, $\theta \in \Theta$. This means we are given a probability measure ν on Θ. We understand uncertainty here from a Bayesian viewpoint, i.e., we average different measures with respect to a given prior distribution ν.

4.5.1. Definition and properties

Definition 4.5.1. Let $\mathbb{P}$ on $(\Omega, \mathcal{F})$ be the mixture probability measure defined as

$$\mathbb{P}(A) := \int_\Theta P^\theta(A) \nu(\mathrm{d}\theta), \qquad \forall A \in \mathcal{F}.$$

Remark 4.5.2. We have the following property for all measures P^θ, $\theta \in \Theta$, which can be seen as a weak form of absolute continuity with respect to $\mathbb{P}$: for each $A \in \mathcal{F}$ with $\mathbb{P}(A) = 0$ there exists a subset $\Theta^A \subset \Theta$ with $\nu(\Theta^A) = 1$ such that $P^\theta(A) = 0$ for all $\theta \in \Theta^A$.

Suppose we are now given two initial believes on $(\Theta, \mathcal{D})$, i.e. two different prior probability measures ν, μ. We introduce the respective mixture probability measures according to Definition 4.5.1 for both the priors as follows

$$\mathbb{P}_\nu(A) := \int_\Theta P^\theta(A)\nu(\mathrm{d}\theta), \qquad \mathbb{P}_\mu(A) := \int_\Theta P^\theta(A)\mu(\mathrm{d}\theta), \tag{4.5.1}$$

for all $A \in \mathcal{F}$.

Lemma 4.5.3. Let μ and ν be two probability measures defined on the parameter space $(\Theta, \mathcal{D})$. Then,

$$\mu \ll \nu \implies \mathbb{P}_\mu \ll \mathbb{P}_\nu,$$

where $\mathbb{P}_\mu$, $\mathbb{P}_\nu$ are defined as in (4.5.1) and the last inequality holds for every $X \in \mathcal{X}$.

Proof. If $\mu \ll \nu$, i.e. $\nu(D) = 1$ implies $\mu(D) = 1$ for all $D \in \mathcal{D}$, then for any fixed $A \in \mathcal{F}$, if $\mathbb{P}_\nu(A) = 0$ there exists $\Theta^A \subset \Theta$ with $\nu(\Theta^A) = 1$ and $P^\theta(A) = 0$ for all $\theta \in \Theta^A$. Hence

$$\mathbb{P}_\mu(A) = \int_\Theta P^\theta(A)\mu(\mathrm{d}\theta) = \int_{\Theta^A} P^\theta(A)\mu(\mathrm{d}\theta) = 0,$$

which yields $\mathbb{P}_\mu \ll \mathbb{P}_\nu$. □

Clearly under the assumption $\mu \ll \nu$, we also have that $C^{\mathbb{P}_\nu} \subseteq C^{\mathbb{P}_\mu}$. Moreover if $\mu \sim \nu$ then by the previous result $\mathbb{P}_\mu \sim \mathbb{P}_\nu$.

Remark 4.5.4. Notice that by Lemma 4.5.3 and Remarks 4.2.2, 4.2.4 we have that

$$\rho^{\mathbb{P}_\mu}(X) \leq \rho^{\mathbb{P}_\nu}(X), \qquad \widehat{\rho}^{\mathbb{P}_\mu}(X) \leq \widehat{\rho}^{\mathbb{P}_\nu}(X),$$

for all $X \in L^\infty(\mathbb{P}_\nu)$.

We introduce as follows the notion of “contiguity" between sequences of probability measures, in our case priors on the parameter space, in order to extend the concept of absolute continuity to an asymptotic framework.

Definition 4.5.5 (Contiguity)**.** Let $(\Theta_n, \mathcal{D}_n)_{n\in\mathbb{N}}$ be a sequence of measurable space and let $(\nu_n)_{n\in\mathbb{N}}$, $(\mu_n)_{n\in\mathbb{N}}$ be two sequences of distributions such that ν_n and μ_n are defined on $(\Omega_n, \mathcal{D}_n)$ for each $n \in \mathbb{N}$. We say that $(\mu_n)_{n\in\mathbb{N}}$ is contiguous with respect to $(\nu_n)_{n\in\mathbb{N}}$ if

$$\lim_{n\to+\infty} \nu_n(D_n) = 0 \implies \lim_{n\to+\infty} \mu_n(D_n) = 0,$$

for every sequence of measurable sets D_n. This is denoted with $\mu_n \triangleleft \nu_n$. The two sequences are said to be mutually contiguous or bi-contiguous if $\mu_n \triangleleft \nu_n$ and $\nu_n \triangleleft \mu_n$. This is denoted by $\mu_n \triangleleft\triangleright \nu_n$.

Remark 4.5.6. "Contiguity" does not mean that the two sequences of probability measures are close to each other. Indeed, denoting with $\|\cdot\|_{\mathrm{TV}}$ the total variation distance between probability measures, if on one side $\|\nu_n - \mu_n\|_{\mathrm{TV}} \to 0$ as $n \to +\infty$ implies $(\mu_n)_{n\in\mathbb{N}} \triangleleft\triangleright (\nu_n)_{n\in\mathbb{N}}$, the converse is not true in general. We address the reader to Van der Vaart (2000) and Le Cam (2012) for results concerning contiguity.

The underlying idea is to extend the result in Lemma 4.5.3 to an asymptotic setup, since absolute continuity can be seen as a particular case of contiguity with constant sequences of probability measures and trivial sequences of measurable sets.

Lemma 4.5.7. Let $(\Theta_n, \mathcal{D}_n)_{n\in\mathbb{N}}$ be a sequence of measurable space and let $(\nu_n)_{n\in\mathbb{N}}$, $(\mu_n)_{n\in\mathbb{N}}$ be two sequences of distributions. Then,

$$\mu_n \triangleleft \nu_n, \implies \mathbb{P}_{\mu_n} \triangleleft \mathbb{P}_{\nu_n}.$$

Proof. We aim to show that

$$\lim_{n\to+\infty} \mathbb{P}_{\nu_n}(A_n) = 0 \implies \lim_{n\to+\infty} \mathbb{P}_{\mu_n}(A_n),$$

for all $(A_n)_{n\in\mathbb{N}}$ sequence of measurable sets. For every $\varepsilon > 0$ we know that

$$0 \leq \varepsilon \nu_n(\{\theta \in \Theta_n : P^\theta(A_n) > \varepsilon\}) \leq \int_{\Theta_n} P^\theta(A_n)\nu_n(\mathrm{d}\theta) = \mathbb{P}_{\nu_n}(A_n).$$

Observe in particular that the set $D = \{\theta \in \Theta_n : P^\theta(A_n) > \varepsilon\}$ is $\mathcal{D}_n$-measurable as the map $\theta \mapsto P^\theta$ is measurable. Notice that if $\mathbb{P}_{\nu_n}(A_n) \to 0$ as $n \to +\infty$, then

$$\lim_{n\to+\infty} \nu_n(\{\theta \in \Theta_n : P^\theta(A_n) > \varepsilon\}) = 0 \overset{(*)}{\implies} \lim_{n\to+\infty} \mu_n(\{\theta \in \Theta_n : P^\theta(A_n) > \varepsilon\}) = 0,$$

where $(*)$ follows by the assumption $\mu_n \triangleleft \nu_n$. Moreover we can decompose the mixture measure with respect to μ_n as follows for each $n \in \mathbb{N}$,

$$\begin{aligned}\mathbb{P}_{\mu_n}(A_n) &= \int_{\Theta_n} P^\theta(A_n) 1_{\{\theta\in\Theta_n : P^\theta(A_n)\leq\varepsilon\}} \mu_n(\mathrm{d}\theta) + \int_{\Theta_n} P^\theta(A_n) 1_{\{\theta\in\Theta_n : P^\theta(A_n)>\varepsilon\}} \mu_n(\mathrm{d}\theta) \\ &\leq \varepsilon + \int_{\Theta_n} P^\theta(A_n) 1_{\{\theta\in\Theta_n : P^\theta(A_n)>\varepsilon\}} \mu_n(\mathrm{d}\theta).\end{aligned}$$

By $(*)$ the second term of the previous inequality goes to zero as $n \to +\infty$ since the set where the integral is defined has asymptotically zero measure, which leads to the fact that for every $\varepsilon > 0$ we know that

$$0 \leq \limsup_{n\to+\infty} \mathbb{P}_{\mu_n}(A_n) \leq \varepsilon,$$

yielding the thesis as claimed:

$$\lim_{n\to+\infty} \mathbb{P}_{\mu_n}(A_n) = 0, \qquad \forall (A_n)_{n\in\mathbb{N}} \subset \mathcal{F}.$$

□

4.5.2. Classical representation for the mixture probability measure

As in Section 4.4.1 we discuss now the classical representation of robust risk measures for $\mathcal{M}^\Theta = \{P^\theta : \theta \in \Theta\}$, with respect to the mixture probability measure.

In order to prove similar results in the case where $\mathcal{M}^\Theta = \{P^\theta : \theta \in \Theta\}$ is not supposed to be dominated a priori, we need some additional assumptions. Let the parameter set be a subset of a Polish space with $\mathcal{D} = \mathcal{B}(\Theta)$ the sigma-algebra of the Borel sets, and let $d(\cdot, \cdot)$ denote a distance on the Polish space. Our starting point builds on the (rather strong) assumption of continuity in total variation of the map $\theta \mapsto P^\theta$, which will in turn imply that $\mathcal{M}^\Theta$ can be dominated, however not necessarily by the mixture probability measure. Indeed, the mixture probability measure will only serve as dominating measure for a subset of $\mathcal{M}^\Theta$ induced from a subset of Θ that has ν-measure 1. We then show by means of a specific example (see Example 4.5.15) that exactly this property can also be obtained *without* the continuity assumption and *without* the existence of a dominating measure for the whole set, implying that these are not necessary conditions to obtain classical representations with respect to the mixture probability measure as of Theorem 4.5.13.

Moreover, when specializing the setup to the case of law-invariance and considering the mixture probability measure $\mathbb{P}$ as the measure with respect to which the penalty function in (4.2.7) is defined, we can prove a classical representation with respect to $\mathbb{P}$ under mild conditions, see Corollary 4.5.17.

Let us now start with the announced continuity assumption:

Assumption 4.5.8. *We assume that the map $\theta \mapsto P^\theta$ is continuous from $\Theta \to \mathcal{M}_1$ with respect to the total variation norm, this means that for each $\varepsilon > 0$ there exists $\delta > 0$ such that for $d(\theta, \theta') < \delta$ we have that*

$$\sup_{A \in \mathcal{F}} |P^\theta(A) - P^{\theta'}(A)| < \varepsilon.$$

Recall Definition 4.5.1 for the mixture probability measure: for each $A \in \mathcal{F}$

$$\mathbb{P}(A) = \int_\Theta P^\theta(A) \mathrm{d}\nu(\theta).$$

By Remark 4.5.2 we know that for each $A \in \mathcal{F}$ with $\mathbb{P}(A) = 0$ there exists Θ^A with $\nu(\Theta^A) = 1$ such that $P^\theta(A) = 0$ for all $\theta \in \Theta^A$. We would like to have the property that $P^\theta \ll \mathbb{P}$ for ν-almost all θ. The following example shows a phenomenon that we avoid by the previous continuity assumption.

Example 4.5.9. Let $(\Omega, \mathcal{F}) = ([0,1], \mathcal{B}([0,1]))$. Let $\Theta = [0,1]$ and define for each $\theta \in \Theta$ the measure $P^\theta = \delta_{\{\theta\}}$ (i.e. the Dirac measure in the point $\theta \in \Theta$). Define the set $c\mathcal{M}$ as the convex hull (with respect to even countable convex combinations) of these Dirac measures, i.e.

$$c\mathcal{M} = \left\{ \sum_{i=1}^\infty \alpha_i \delta_{\{\theta_i\}} : \ 0 \leq \alpha_i \leq 1, \ \text{ with } \ \sum_{i=1}^\infty \alpha_i = 1 \text{ and } \ \theta_i \in [0,1] \text{ for all } i \geq 1 \right\}.$$

That means $c\mathcal{M} = \overline{conv}(\mathcal{M}^\Theta)$. Let ν be the Lebesgue measure on $[0,1]$. Then it holds that

$$\mathbb{P}(A) = \int_{\theta\in[0,1]} P^\theta(A)\mathrm{d}\nu(\theta) = \int_{\theta\in[0,1]} 1_{\{\theta\in A\}}\mathrm{d}\nu(\theta) = \nu(A).$$

By Remark 4.5.2 it holds that for each $A \in \mathcal{F}$ there exists Θ^A such that $\nu(\Theta^A) = 1$ and $P^\theta(A) = 0$ for all $\theta \in \Theta^A$. But each Dirac measure P^θ is singular to the Lebesgue measure and the same holds for each countable convex combination of Dirac measures. Obviously, $\theta \mapsto P^\theta = \delta_{\{\theta\}}$ is not continuous in θ. Indeed, for each $\theta \in [0,1]$, we can find $\theta' \in [0,1]$ arbitrarily close to θ and $A \in \mathcal{F}$ such that $\theta \in A$, $\theta' \notin A$ and hence $P^\theta(A) = 1$ and $P^{\theta'}(A) = 0$.

Note that in the previous example all measures P^θ are mutually singular. A similar phenomenon occurs for instance in the context of volatility uncertainty.

Example 4.5.10. Consider stock price models with $\mathrm{d}S_t = S_t\theta\mathrm{d}W_t$, where W is a standard Brownian motion and θ in $[0,1]$. Its laws denoted by P^θ are all mutually singular and thus Assumption 4.5.8 cannot be satisfied.

Let us now show that there is a simple choice of a dominating measure for the set $\mathcal{M}^\Theta$.

Lemma 4.5.11. Let Θ be a subset of a Polish space. Under Assumption 4.5.8 there exists a dominating measure $\mathbb{P}$ which is a countable convex combination of measures P^θ, $\theta \in \Theta$.

Proof. By assumption there exists a countable dense subset of Θ which we denote by $\{\theta_k, k \in \mathbb{N}\}$. Define $\mathbb{P} = \sum_{k=1}^\infty 2^{-k}P^{\theta_k}$. This measure is in $c\mathcal{M}^\Theta$. We will show that this measure dominates $\mathcal{M}^\Theta$. Indeed, let $\mathbb{P}(A) = 0$. Then $P^{\theta_k}(A) = 0$ for all $k \in \mathbb{N}$. Fix an arbitrary $\theta \in \Theta$, then for each $\varepsilon > 0$ there exists $k \in \mathbb{N}$ such that θ_k is close enough to θ such that $|P^{\theta_k}(A) - P^\theta(A)| < \varepsilon$. This follows by the density of the subset $\{\theta_k, k \in \mathbb{N}\}$ and by Assumption 4.5.8. Hence, for each $\varepsilon > 0$ we get that $P^\theta(A) < \varepsilon$ and so $P^\theta(A) = 0$. Clearly $\mathbb{P}$ is even generalized equivalent with respect to $\mathcal{M}^\Theta$.

□

The above lemma gives a certain prior ν on Θ. Indeed it is the measure that gives the weight 2^{-k} to the parameter θ_k. The measure $\mathbb{P}$ in Lemma 4.5.11 above is the mixture probability measure for this prior.

Our goal is now to show a similar result for any mixture probability measure induced by an arbitrary prior ν. Indeed, we can generalize the above result and show that under Assumption 4.5.8 the set $\mathcal{M}^\Theta$ is ν-a.s. dominated by the mixture probability measure.

Let us now state a relevant consequence of Assumption 4.5.8. Indeed under this hypothesis, we can find set of measures that has ν-measure 1 and that can be dominated by the mixture probability measure $\mathbb{P}$.

Lemma 4.5.12. Let Θ be a subset of a Polish space and define $\Theta^0 = \{\theta \in \Theta : P^\theta \not\ll \mathbb{P}\}$. Under Assumption 4.5.8 the subset $\Theta^0 \subseteq \Theta$ is Borel-measurable and satisfies $\nu(\Theta^0) = 0$.

Proof. We assume without loss of generality $\Theta^0 \neq \emptyset$. Indeed if $\Theta^0 = \emptyset$, then it is trivially Borel-measurable and all $P^\theta \ll \mathbb{P}$.

(i) First we will show the measurability of Θ^0. Define, for each $m \geq 1$,

$$\Theta^m = \{\theta \in \Theta^0 : \exists A \in \mathcal{F} \text{ with } \mathbb{P}(A) = 0 \text{ and } P^\theta(A) > 2^{-m}\}.$$

Notice that since $\Theta^0 \neq \emptyset$, then $\Theta^m \neq \emptyset$ for some $m \geq 1$. It is clear that $\Theta^m \uparrow \Theta^0$. Moreover, each Θ^m is Borel-measurable. Indeed, fix $m \geq 1$ and choose $\theta \in \Theta^m$. There exists A with $\mathbb{P}(A) = 0$ and $P^\theta(A) > 2^{-m}$. Because of the strict inequality there exists $\varepsilon > 0$ such that $P^\theta(A) \geq 2^{-m} + \varepsilon$. Let us denote in the following with $B_\delta(\theta)$ the open ball with center $\theta \in \Theta$ and radius $\delta > 0$. By Assumption 4.5.8 there exists $\delta > 0$ small enough such that for all $\theta' \in B_\delta(\theta)$ we have that $|P^\theta(A) - P^{\theta'}(A)| < \frac{\varepsilon}{2}$. Hence $P^{\theta'}(A) > 2^{-m} + \varepsilon - \frac{\varepsilon}{2} > 2^{-m}$ and each $P^{\theta'} \in \Theta^m$, i.e. Θ^m is open and therefore Borel-measurable. In particular we can conclude that Θ^0 is open and thus Borel-measureable.

(ii) We show now that $\nu(\Theta^0) = 0$. Suppose by contradiction that $\nu(\Theta^0) = \gamma > 0$. By the above we have that $\nu(\Theta^0) = \lim_{m\to\infty} \nu(\Theta^m)$ and hence there exists $m_0 \geq 1$ such that $\nu(\Theta^{m_0}) \geq \frac{\gamma}{2}$. Let D be a dense countable subset of Θ and define $\tilde{\Theta}^{m_0} = \Theta^{m_0} \cap D$. Obviously this set is dense in Θ^{m_0} and countable. Let $(q_k)_{k\geq 0}$ be any enumeration of this countable set, i.e., $\tilde{\Theta}^{m_0} = \{q_k, k \geq 1\}$. As $q_k \in \Theta^{m_0}$, $k \geq 1$, we have that for each $k \geq 1$ there exists $A_k \in \mathcal{F}$ such that $\mathbb{P}(A_k) = 0$ and $P^{q_k}(A_k) > 2^{-m_0} =: \epsilon$. Define

$$A_0 = \bigcup_{k=1}^{\infty} A_k \in \mathcal{F}.$$

As A_0 is a countable union of $\mathbb{P}$-nullsets we still have that

$$\mathbb{P}(A_0) = 0. \tag{4.5.2}$$

Choose an arbitrary $\theta \in \Theta^{m_0}$. By Assumption 4.5.8, there exists $\delta > 0$ such that for all $\theta' \in B_\delta(\theta)$ we have that

$$\sup_{A\in\mathcal{F}} |P^\theta(A) - P^{\theta'}(A)| < \frac{\varepsilon}{2}. \tag{4.5.3}$$

As the sequence $(q_k)_{k\geq 1}$ is dense there exists $k \geq 1$ such that $q_k \in B_\delta(\theta)$. Since (4.5.3) holds for all $A \in \mathcal{F}$ we have that $P^\theta(A_k) \geq P^{q_k}(A_k) - \frac{\varepsilon}{2} = \frac{\varepsilon}{2}$. Hence $P^\theta(A_0) \geq P^\theta(A_k) \geq \frac{\epsilon}{2}$. As θ was arbitrary, we have that, for all $\theta \in \Theta^{m_0}$,

$$P^\theta(A_0) \geq \frac{\epsilon}{2}.$$

Let us now use the definition of the mixture probability measure to calculate $\mathbb{P}(A_0)$:

$$\mathbb{P}(A_0) = \int_\Theta P^\theta(A_0)\mathrm{d}\nu(\theta) \geq \int_{\Theta^{m_0}} P^\theta(A_0)\mathrm{d}\nu(\theta) \geq \frac{\varepsilon}{2}\nu(\Theta^{m_0}) \geq \frac{\varepsilon}{2}\cdot\frac{\gamma}{2} > 0,$$

which is a contradiction to (4.5.2).

Therefore we showed that $\nu(\Theta^0) = \nu(\{\theta \in \Theta : P^\theta \not\ll \mathbb{P}\}) = 0$.

□

Theorem 4.5.13. *Let Θ be a subset of a Polish space and let Assumption 4.5.8 hold. Then there exists a Borel-measurable subset $\Theta^1 \subset \Theta$ with $\nu(\Theta^1) = 1$ such that, for all $\theta \in \Theta^1$, $P^\theta \ll \mathbb{P}$, where $\mathbb{P}$ denotes the mixture probability measure introduced in Definition 4.5.1. Let $c\mathcal{M}^1 = \overline{conv}(\mathcal{M}^{\Theta^1})$ where $\mathcal{M}^{\Theta^1} = \{P^\theta : \theta \in \Theta^1\}$. Then, for all $X \in L^\infty(\mathbb{P})$ it holds that*

$$\widehat{\rho}^{c\mathcal{M}^1}(X) := \sup_{P \in c\mathcal{M}^1} \widehat{\rho}^P(X) = \widehat{\rho}^{\mathbb{P}}(X).$$

and

$$\rho^{c\mathcal{M}^1}(X) := \sup_{P \in c\mathcal{M}^1} \rho^P(X) = \rho^{\mathbb{P}}(X).$$

Proof. This is a simple consequence of Lemma 4.5.12, Theorem 4.4.6 and Theorem 4.4.8, because $\mathbb{P}$ is generalized equivalent to $c\mathcal{M}^1$. Indeed, as we restrict ourselves to Θ^1, we know that $P^\theta \ll \mathbb{P}$ for all $\theta \in \Theta^1$ (which satisfies $\nu(\Theta^1) = 1$ by Lemma 4.5.12). Clearly, if A is such that $P^\theta(A) = 0$ for all $\theta \in \Theta^1$ then $\mathbb{P}(A) = 0$, implying that also the second property of generalized equivalence is satisfied. Theorem 4.4.8 and Theorem 4.4.6 now give the existence of a probability measure $P^0 \in c\mathcal{M}^1$ with $\mathbb{P} \ll P^0$. Hence the statements follow. □

Remark 4.5.14. (i) Note that for the measure $P^0 \in c\mathcal{M}^1$ as introduced in the proof of Theorem 4.4.6 and Theorem 4.4.8, it holds that $P^0 \sim \mathbb{P}$ and thus by Remarks 4.2.2 and 4.2.4

$$\widehat{\rho}^{P^0}(X) = \widehat{\rho}^{\mathbb{P}}(X), \quad \text{and} \quad \rho^{P^0}(X) = \rho^{\mathbb{P}}(X).$$

In other words, under Assumption 4.5.8 there is a measure in $c\mathcal{M}^1$ equivalent to $\mathbb{P}$ yielding the same risk. This measure is a countable convex combination of measures with parameters in Θ^1 with $\nu(\Theta^1) = 1$. Comparing this to the first countable convex combination of Lemma 4.5.11 the difference here is that P^0 can be interpreted as a countable new prior that fits to our prior ν. So, if we think of a dynamical procedure, where we improve our knowledge using data this goes into the prior ν. With the corresponding P^0 we then find a tailor-made countable prior. As the set Θ^1 might be smaller than Θ the risk measure with respect to P^0 might therefore yield a smaller risk than the first countable convex combination of Lemma 4.5.11.

(ii) Note that the mixture probability measure $\mathbb{P}$ does not necessarily lie in $c\mathcal{M}^1$, see e.g. Example 4.5.16 below. Hence the measure P^0 we find can be understood as a measure with respect to a countable prior that gives the same risk. By a countable prior $\tilde{\nu}$ for P^0 we mean a probability measure of the form $\tilde{\nu} = \sum_{k=1}^\infty \gamma^k \delta_{\theta^k}$ such that $\gamma^k = \tilde{\nu}(\theta^k)$ and $P^0 = \sum_{k=1}^\infty \gamma^k P^{\theta^k}$.

The following example shows that Assumption 4.5.8 is only sufficient to the get the assertion of Theorem 4.5.13. Indeed, in this example we still get that the mixture probability measure is a dominating measure for ν-almost all θ even though Assumption 4.5.8 is not satisfied.

Example 4.5.15. The following example satisfies

(i) There does not exist a dominating measure for $\{P^\theta : \theta \in \Theta\}$.

(ii) There exists $\widetilde{\Theta} \subseteq \Theta$ such that $\nu(\widetilde{\Theta}) = 1$ and such that there exists a dominating measure $\mathbb{P}$ for $\mathcal{M}^{\widetilde{\Theta}} := \{P^\theta : \theta \in \widetilde{\Theta}\}$. This dominating measure is also generalized equivalent with respect to $\mathcal{M}^{\widetilde{\Theta}}$. In particular, $\mathbb{P}$ can be chosen to be the mixture probability measure.

(iii) Assumption 4.5.8 does not hold on.

(iv) There are uncountably many P^θ that are mutually singular.

Hence we see that we can have a ν-a.s. dominating measure also in the case that Assumption 4.5.8 is not satisfied. Moreover a closer look at the example shows that we can have uncountably many singular measures as long as the corresponding parameters are in a ν-nullset $\Theta \setminus \widetilde{\Theta}$. Moreover, the example shows that if there are uncountably many P^θ that are mutually singular we cannot find a dominating measure *for all* P^θ, $\theta \in \Theta$ (only on a set of ν-measure 1).

The construction is as follows. Let $(\Theta, \mathcal{D}, \nu)$ be $([0,1], \mathcal{B}([0,1]), \Lambda)$ where Λ is again the Lebesgue-measure on $[0,1]$. $F \subset [0,1]$ be the Cantor set. It is well-known that F is uncountable and $\Lambda(F) = 0$. Define the measures P^θ, $\theta \in [0,1]$ as follows:

$$P^\theta = \begin{cases} \delta_{\{\theta\}} & \text{for } \theta \in F \\ R^\theta & \text{for } \theta \in [0,1] \setminus F, \end{cases}$$

where $\frac{\mathrm{d}R^\theta}{\mathrm{d}\Lambda} = \frac{1}{\theta} 1_{[0,\theta]}$. Note that $0 \in F$, hence the R^θ are well-defined. We define $\widetilde{\Theta} = [0,1] \setminus F$. Obviously $\nu(\widetilde{\Theta}) = \Lambda(\widetilde{\Theta}) = 1$. Let us check the properties.

Concerning (ii) an obvious dominating measure is the Lebesgue measure Λ as all $R^\theta \ll \Lambda$. Observe that the mixture probability measure $\mathbb{P}$ is equivalent to Λ. Indeed since all $R^\theta \ll \Lambda$ and $\Lambda(F) = 0$, then clearly $\mathbb{P} \ll \Lambda$. Notice that for every $A \in \mathcal{F}$,

$$\begin{aligned} \mathbb{P}(A) = \int_\Theta P^\theta(A)\nu(\mathrm{d}\theta) = \int_0^1 R^\theta(A)\mathrm{d}\theta &= \int_0^1 \int_A \frac{1}{\theta} 1_{[0,\theta]}(s)\mathrm{d}s\mathrm{d}\theta \\ &= \int_A \int_s^1 \frac{1}{\theta}\mathrm{d}\theta\mathrm{d}s \\ &= \int_A -\ln(s)\mathrm{d}s, \end{aligned}$$

therefore if $\mathbb{P}(A) = 0$, since the integrand is positive we obtain that $\Lambda(A) = 0$. Hence we can also use $\mathbb{P}$ as dominating measure. It is also straightforward to see that $\mathcal{M}^{\widetilde{\Theta}} \ll \Lambda$. Indeed, if $P^\theta(A) = 0$ for all $\theta \in \widetilde{\Theta}$, then we have that $\Lambda(A \cap [0,\theta]) = 0$ for all $\theta \in \widetilde{\Theta}$. Because $\nu(\widetilde{\Theta}) = 1$ we get that $\Lambda(A) = 0$ as we can choose $\theta \in \widetilde{\Theta}$ arbitrarily close to 1.

For (iii) let $\theta' \in F$ and $\theta \in [0,1] \setminus F$ such that $|\theta - \theta'| < \delta$ with δ small. If $\theta < \theta'$ we can choose $A = [0, \theta]$, then $P^\theta(A) = R^\theta(A) = 1$ and $P^{\theta'}(A) = \delta_{\{\theta'\}}(A) = 0$. If $\theta' < \theta$ we can choose $A = \{\theta'\}$ to see that $P^\theta(A) = 0$ whereas $P^{\theta'}(A) = 1$. This works for arbitrarily small δ. If we choose θ and θ' both in F then the mutual singularity of the measures gives a similar result. Therefore the function $\theta \mapsto P^\theta$ cannot be continuous in total variation.

(iv) is obvious as F is uncountable.

Let us now give a proof of (i). Suppose there would exists a dominating measure R for all P^θ, $\theta \in \Theta$. In particular we would have $P^\theta \ll R$ for all $\theta \in F$. Suppose A is a nullset of R, i.e. $R(A) = 0$. Suppose $\theta \in A \cap F$. Clearly, $P^\theta(A) = \delta_\theta(A) = 1$, which is a contradiction to $P^\theta(A) = 0$. It follows that if $R(A) = 0$ then $A \cap F = \emptyset$. This implies that for all $x \in F$ we have that $R(\{x\}) > 0$. Such a probability measure cannot exist. Indeed, for all $x \in F$ there exists $\varepsilon > 0$ such that $R(\{x\}) \geq \varepsilon$. Fix $k \geq 1$ and define $F_k := \{x \in F : R(\{x\}) \geq 2^{-k}\}$. As R is a probability measure there cannot be more than 2^k points in F_k, i.e. $|F_k| \leq 2^k$. Define $\widetilde{F} = \bigcup_{k \geq 1} F_k$. This set is at most countable as it is a countable union of finite sets. Hence $F \setminus \widetilde{F} \neq \emptyset$. Now, take any $x \in F \setminus \widetilde{F}$. We have that $R(\{x\}) < 2^{-k}$ for all $k \geq 1$ and it follows that $R(\{x\}) = 0$, a contradiction. This proof can be reproduced whenever we have uncountably many singular measures.

As already indicated in Remark 4.5.14, the following example shows that the mixture probability measure does not necessarily have to lie in $c\mathcal{M}$.

Example 4.5.16. We consider the previous example without the Cantor set, i.e., $(\Theta, \mathcal{D}, \nu) = ([0,1], \mathcal{B}([0,1]), \Lambda)$ where Λ is again the Lebesgue-measure on $[0,1]$. Define the measures P^θ, $\theta \in [0,1]$ as the measures R^θ from before, i.e., $\frac{\mathrm{d}P^\theta}{\mathrm{d}\Lambda} = \frac{1}{\theta} 1_{[0,\theta]}$, but now for all $\theta \in (0,1]$. Let $\mathcal{M} = \{P^\theta : \theta \in (0,1]\}$. As before the Lebesgue measure Λ is a dominating measures for $\mathcal{M}$. The set $c\mathcal{M}$ consists again of all countable convex combination of measures in $\mathcal{M}$. As, trivially, for $\theta = 1$, $P^1 = \Lambda \in \mathcal{M}$ and as it is equivalent to itself there exists a measure in $c\mathcal{M}$ that is equivalent to Λ. The mixture probability measure $\mathbb{P}$ satisfies, as before, that

$$\mathbb{P}(A) = \int_A -\ln(s) \mathrm{d}s,$$

and so it is equivalent to Λ. But it is not an element of $c\mathcal{M}$ as we will show now.

Suppose we could find a countable convex combination $R = \sum_{n=1}^\infty \alpha_n P^{\theta^n}$ such that $R = \mathbb{P}$. Then we would have that R and $\mathbb{P}$ agree on all sets $[0, c)$ with $c \in (0,1)$. Observe that

$$\mathbb{P}([0,c)) = \int_0^c -\ln(s)) \mathrm{d}s = c(1 - \ln(c)) =: f(c),$$

where $f(c)$ is a strictly concave increasing differentiable function on $(0,1)$. On the other hand $R((0,c])$ can be expressed as follows

$$R([0,c)) = \sum_{n=1}^\infty \alpha_n P^{\theta^n}([0,c)) = \sum_{n=1}^\infty \alpha_n g^n(c) =: g(c),$$

where $g^n(c) = \frac{c \wedge \theta^n}{\theta^n}$. In order to show that the two probability measures are equal we would have to show that $f(c) = g(c)$ for all $c \in (0,1)$.

We will get a contradiction. Indeed, we will find c_0 with $0 < c_0 < 1$ such that g is not differentiable in c_0. Then f and g cannot agree for all $c \in (0,1)$ as f is differentiable.

If R should equal $\mathbb{P}$ then there has to exist n_0 in the indices of the convex combination such that $\alpha_{n_0} > 0$ for a $0 < \theta_{n_0} < 1$. Indeed, otherwise $R = \Lambda$ and this is not equal to $\mathbb{P}$. Let $c_0 = \theta^{n_0}$. Moreover, not all the weights for the indices n with $\theta^n > c_0$ can be zero. Otherwise the support of R would be $[0, c_0]$, and then R could not be equal to $\mathbb{P}$ which has full support $[0,1]$. Hence it has to hold that $\sum_{n \geq 1, \theta^n > c_0} \alpha^n = \delta$, for some $\delta > 0$.

Clearly, by definition of c_0, the function g^{n_0} is not differentiable in c_0. Note that the right hand derivative of g^{n_0} in c_0 is equal to 0, whereas the left hand derivative is equal to $\frac{1}{\theta^{n_0}} = \frac{1}{c_0}$. All other functions g^{θ_n}, $n \geq 1$, $n \neq n_0$, are differentiable in c_0, with

$$g^{n\prime}(c_0) = \begin{cases} 0 & \text{for those } n \text{ with } \theta^n < c_0, \\ \frac{1}{\theta^n} & \text{for those } n \text{ with } \theta^n > c_0. \end{cases}$$

Recall that $g(c) = \sum_{n=1}^{\infty} \alpha_n g^n(c)$. Then, we get:

$$\lim_{h \downarrow 0} \frac{g(c_0 + h) - g(c_0)}{h} = \sum_{n, \theta^n > c_0} \alpha_n \frac{1}{\theta_n} =: A,$$

where A is a constant with $0 < A < \frac{\delta}{c_0}$.

On the other hand

$$\lim_{h \uparrow 0} \frac{g(c_0 + h) - g(c_0)}{h} = \frac{\alpha_{n_0}}{\theta^{n_0}} + \sum_{n, \theta^n > c_0} \alpha_n \frac{1}{\theta_n} = \frac{\alpha_{n_0}}{c_0} + A,$$

which is different from A as $\alpha_{n_0} > 0$. Hence g is not differentiable in c_0.

Note that the interchange of limits in the above calculations is justified for $h \downarrow 0$ as well as $h \uparrow 0$. To show this, introduce an artificial probability space $\tilde{\Omega} = \mathbb{N}$ with the probability measure that gives, to each $n \geq 1$, the weight $\tilde{P}(n) = \alpha_n$. Define, for h, a random variable X_h on $\tilde{\Omega}$ as $X_h(n) = \frac{g^n(c_0+h) - g^n(c_0)}{h}$. Hence, it holds, for all $n \neq n_0$ that $\lim_{h \to 0} X_h(n) = g^{n\prime}(c_0)$. For n_0 we have that $\lim_{h \downarrow 0} X_h(n_0) = 0$ and $\lim_{h \uparrow 0} X_h(n_0) = \frac{1}{c_0}$. Define $X(n) = g^{n\prime}(c_0)$ for $n \neq n_0$ and $X(n_0) = 0$. Then, pointwise, $X_h \to X$ for $h \downarrow 0$ and $X_h \to X + \frac{1}{c_0} 1_{\{n_0\}}$ for $h \uparrow 0$. Moreover, it is easy to see that $|X_h| \leq \frac{1}{c_0}$ for all $h > 0$ and $|X_h| \leq \frac{2}{c_0}$, for all $h < 0$ small enough such that $c_0 + h > \frac{c_0}{2}$. Then, we can interpret the above interchange of limits as applications of the dominated convergence theorem for the measure $\tilde{P}$. Indeed,

$$\lim_{h \downarrow 0} \frac{g(c_0 + h) - g(c_0)}{h} = \lim_{h \downarrow 0} \mathbb{E}_{\tilde{P}}[X_h] = \mathbb{E}_{\tilde{P}}[X] = A,$$

with A as above and

$$\lim_{h\uparrow 0}\frac{g(c_0+h)-g(c_0)}{h}=\lim_{h\uparrow 0}\mathbb{E}_{\tilde{P}}[X_h]=\mathbb{E}_{\tilde{P}}[X+\frac{1}{c_0}1_{\{n_0\}}]=A+\frac{\alpha_{n_0}}{c_0}.$$

Finally, we specialize our setup to the one of convex law-invariant risk measures as of Section 4.2.4 and consider the mixture probability measure $\mathbb{P}$ as the measure with respect to which the penalty function in (4.2.7) is defined. Just by requiring that $\mathcal{M}^\Theta$ is closed under countable convex combinations and that the mass of sets of positive measure under $\mathbb{P}$ do not only stem from singular parts (see Condition 4.5.4 below), we obtain the following result.

Corollary 4.5.17. *Let $\mathcal{M}^\Theta \subseteq \mathcal{M}_1$ be a set of probability measures induced by a parameter space Θ and closed under countable convex combinations. Let $\mathbb{P}$ denote the mixture probability measure as of Definition 4.5.1 and consider law-invariance as in Section 4.2.4 with respect to it, i.e. fix a penalty function $\alpha : \mathcal{M}_1 \to \mathbb{R}\cup\{+\infty\}$ as in (4.2.7) with respect to $\mathbb{P}$. Suppose additionally that for all $A \in \mathcal{F}$ with $\mathbb{P}(A) > 0$*

$$\mathbb{P}(A) > \int_\Theta \widetilde{P}_2^\theta(A)\nu(\mathrm{d}\theta), \tag{4.5.4}$$

where $\widetilde{P}_2^\theta$ denotes the (unnormalized) singular part of the Lebesgue decomposition of P^θ with respect to $\mathbb{P}$. Then for all $X \in \mathcal{X}^{\mathcal{M}^\Theta} \cap L^\infty(\mathbb{P})$ we have

$$\widehat{\rho}^{\mathcal{M}}(X)=\widehat{\rho}^{\mathbb{P}}(X)=\rho^{\mathbb{P}}(X)=\rho^{\mathcal{M}}(X).$$

Proof. We start by proving $\rho^{\mathbb{P}}=\rho^{\mathcal{M}}$. Note that the only difference with respect to Corollary 4.4.10 is that $\mathbb{P}$ does not need to lie in $\mathcal{M}^\Theta$. Instead of this, we assumed that $\mathbb{P}$ is the mixture probability measure and that $\mathcal{M}^\Theta$ is closed under countable convex combinations, the latter with the goal to apply Theorem 4.4.4. To this end we need additionally a family of probability measures which is absolutely continuous with respect to $\mathbb{P}$. Consider therefore instead of P^θ always P_1^θ denoting the normalized absolutely continuous part of the Lebesgue decomposition with respect to $\mathbb{P}$. Then, as visible from the proof of Proposition 4.2.10 (see also Remark 4.2.11 (iii))

$$\rho^{P_1^\theta}=\rho^{P^\theta}$$

and thus $\rho^{\mathcal{M}^\Theta}=\sup_{P_1^\theta}\rho^{P_1^\theta}$ on $\mathcal{X}^{\mathcal{M}^\Theta}$. Hence is suffices to consider the family $\mathcal{M}_1^\Theta=\{P_1^\theta : \theta\in\Theta\}$, which inherits the property of being closed under countable convex combinations. In order to apply Theorem 4.4.4 we need additionally that the condition $P_1^\theta(A)=0$ for all $P_1^\theta\in\mathcal{M}_1^\Theta$ implies $\mathbb{P}(A)=0$. Denoting by $\widetilde{P}_i^\theta$ the unnormalized parts of the Lebesgue decomposition of P^θ with respect to $\mathbb{P}$, we have

$$\mathbb{P}(A)=\int_\Theta(\widetilde{P}_1^\theta(A)+\widetilde{P}_2^\theta(A))\nu(\mathrm{d}\theta)=\int_\Theta\widetilde{P}_2^\theta(A)\nu(\mathrm{d}\theta),$$

where the last equality holds if $P_1^\theta(A)=0$ for all $P_1^\theta\in\mathcal{M}_1^\Theta$. Therefore, Condition 4.5.4 yields $\mathbb{P}(A)=0$. Theorem 4.4.4 thus implies that there exists some $P_0\in\mathcal{M}_1^\Theta$ which is

equivalent to $\mathbb{P}$ and hence $\rho^{\mathbb{P}} = \rho^{P_0}$. Moreover, using $P_0 \in \mathcal{M}_1^\Theta$ and Proposition 4.2.10, we get

$$\rho^{P_0} \leq \rho^{\mathcal{M}_1^\Theta} = \rho^{\mathcal{M}^\Theta} \leq \rho^{\mathbb{P}}.$$

As $\rho^{\mathbb{P}} = \rho^{P_0}$, all inequalities reduce to equalities. The same reasoning holds for $\widehat{\rho}$ and we can thus conclude since $\rho^{\mathbb{P}} = \widehat{\rho}^{\mathbb{P}}$ again by Proposition 4.2.10. □

5. The conditional case

This chapter is based on joint on-going work with Irene Klein and concerns a possible extension of the results in Cuchiero et al. (2022a).

We will fix first some useful notation and recall basic concepts about conditional risk measures. We will start here with conditional risk measures being defined in a pointwise way (see Section 5.1). Observe that this approach is different to the one adopted for instance in Föllmer and Schied (2011); Acciaio and Penner (2011); Bion-Nadal and Di Nunno (2020); Di Nunno and Rosazza Gianin (2023).

Notation Let us fix a time horizon $T \in \mathbb{N}$ and let us consider the time set defined as $\mathbb{T} := \{0, \dots, T\}$. Let us fix a filtered probability space with a reference probability measure P, namely $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in\mathbb{T}}, P)$ where $\mathcal{F}_0 = \{0, \Omega\}$ and $\mathcal{F} := \mathcal{F}_T$. For all $t \in \mathbb{T}$, we additionally introduce the following notation for the Lebesgue space:

$$L_t^\infty(P) := L^\infty(\Omega, \mathcal{F}_t, P),$$
$$L^\infty(P) := L^\infty(\Omega, \mathcal{F}, P).$$

As in the previous chapter, let $\mathcal{M}_1 = \mathcal{M}_1(\Omega, \mathcal{F})$ be the set of probability measures defined on $(\Omega, \mathcal{F})$. We equip it with the topology of weak convergence (in the probabilistic sense), which corresponds to the functional analytic notion of weak-$*$-convergence when Ω is compact. Recall from the previous chapter $\mathcal{X}$, i.e. the space of pointwise bounded and $\mathcal{F}$-measurable risky claims, and introduce $\mathcal{X}_t$ the space of point-wise bounded and $\mathcal{F}_t$-measurable risky claims, for some $t \in \mathbb{T}$.

5.1. Conditional risk measures

Let us recall in the following the definition of monetary risk measures defined on the space $\mathcal{X}$. Throughout the section we will work with a fixed $t \in \mathbb{T}$ unless specified otherwise.

Definition 5.1.1. Let $\rho_t : \mathcal{X} \to \mathcal{X}_t$. We say that ρ_t is a conditional monetary risk measure, if for all $X, Y \in \mathcal{X}$ it satisfies for all $\omega \in \Omega$ the following conditions:

(i) for all $m_t \in \mathcal{X}_t$; $\rho_t(X + m_t) = \rho_t(X) - m_t$.

(ii) $X \leq Y \implies \rho_t(X) \geq \rho_t(Y)$;

(iii) $\rho_t(0) = 0$.

If in addition to *(i)-(iii)* ρ_t is such that

(iv) for all $\lambda_t \in [0,1]$ with $\lambda_t \in \mathcal{X}_t$, $\rho_t(\lambda_t X + (1-\lambda_t)Y) \leq \lambda_t \rho_t(X) + (1-\lambda_t)\rho_t(Y)$,

then it is said to be a conditional convex risk measure. If in addition to (i)-(iii) ρ_t fulfills

(v) for all $\lambda_t \in \mathcal{X}_t$ with $\lambda_t \geq 0$, $\rho_t(\lambda_t X) = \lambda_t \rho_t(X)$,

then it is said to be a conditional coherent risk measure.

Notice that we mean that each property holds for all $\omega \in \Omega$. In the present work we will mainly focus on the class of conditional convex risk measures. Let us fix a conditional monetary risk measure $\rho_t : \mathcal{X} \to \mathcal{X}_t$ and define the corresponding acceptance set

$$C_t := \{X \in \mathcal{X} : \rho_t(X)(\omega) \leq 0,\ \forall \omega \in \Omega\}. \tag{5.1.1}$$

By the previous, a conditional monetary risk measure ρ_t can be understood as the conditional capital requirement needed at time $t > 0$ to make a financial position X acceptable at that time. Let us now consider the case where we are given a probability measure $P \in \mathcal{M}_1(\Omega, \mathcal{F})$.

Definition 5.1.2. Let P be any probability measure on $(\Omega, \mathcal{F})$. Let the risk measure with respect to P given by $\rho_t^P : L^\infty(P) \to L_t^\infty(P)$ defined as follows

$$\rho_t^P(X) := P - \operatorname{ess\,inf}_{\{\tilde{X} \in \mathcal{X} : P(\tilde{X} = X) = 1\}} \rho_t(\tilde{X}),$$

and let its corresponding acceptance set be

$$C_t^P := \{X \in L^\infty(P) : \rho_t^P(X) \leq 0,\ P - \text{a.s.}\}.$$

We introduce additionally a more compact notation to denote the set of P-versions of $X \in L^\infty(P)$, namely $\mathcal{X}^{X,P} := \{\tilde{X} \in \mathcal{X} : P(\tilde{X} = X) = 1\}$.

Observe that trivially for $X \in L^\infty(P)$ and any $\tilde{X} \in \mathcal{X}^{X,P}$ we have that

$$\rho_t^P(X) \leq \rho_t(\tilde{X}), \qquad P - a.s.$$

Proposition 5.1.3. Let ρ_t be a conditional monetary risk measure, and fix $X \in L^\infty(P)$ for some $P \in \mathcal{M}_1$. Then, there exists a decreasing sequence $(\rho_t(\tilde{X}_n))_{n \in \mathbb{N}}$ such that

$$\rho_t^P(X) = \lim_{n \to +\infty} \rho_t(\tilde{X}_n), \qquad P - a.s, \tag{5.1.2}$$

where $(\tilde{X}_n)_{n \in \mathbb{N}} \subset \mathcal{X}^{X,P}$. In addition, the sequence $(\tilde{X}_n)_{n \in \mathbb{N}}$ can be chosen without loss of generality to be increasing pointwise.

Proof. To show the claim we first have to prove that the subset of $\mathcal{X}_t$

$$\mathscr{R}_t^{X,P} := \{\rho_t(\tilde{X}) : \tilde{X} \in \mathcal{X}^{X,P}\},$$

for $X \in L^\infty(P)$, is *directed downwards*. This means that for $\rho_t(\tilde{X}_1), \rho_t(\tilde{X}_2) \in \mathscr{R}_t^{X,P}$ there exists $\tilde{X} \in \mathcal{X}^{X,P}$ such that, pointwise,

$$\rho_t(\tilde{X}) \leq \rho_t(\tilde{X}_1) \wedge \rho_t(\tilde{X}_2).$$

Let $\tilde{X}_1, \tilde{X}_2 \in \mathcal{X}^{P,X}$, then $\tilde{X}_1 \vee \tilde{X}_2 \in \mathcal{X}^{P,X}$, as $P(X = \tilde{X}_1)$ and $P(X = \tilde{X}_2)$ implies $P(X = \tilde{X}_1 \vee \tilde{X}_2) = 1$. Set $\tilde{X} := \tilde{X}_1 \vee \tilde{X}_2$. Notice that for $i = 1, 2$ pointwise we have

$$\rho_t(\tilde{X}) \leq \rho_t(\tilde{X}_i),$$

and thus $\rho_t(\tilde{X}) \leq \rho_t(\tilde{X}_1) \wedge \rho_t(\tilde{X}_2)$. Observe that by Theorem A.32 in Föllmer and Schied (2011) it follows (5.1.2) as well as the second part of the statement. □

We aim now to show that whenever we fix a conditional convex risk measure in the sense of Definition 5.1.1 and we are given a probability measure P, the resulting ρ_t^P is still a conditional convex risk measure where the properties in Definition 5.1.1 hold P-a.s. This is proved to hold in the next Lemma.

Lemma 5.1.4. Let $\rho_t : \mathcal{X} \to \mathcal{X}_t$ be a conditional convex risk measure and let P be a reference measure on $(\Omega, \mathcal{F})$. Then, $\rho_t^P : L^\infty(P) \to L_t^\infty(P)$ is a conditional convex risk measure.

Proof. Let us consider first the monotonicity, i.e., (ii) of Definition 5.1.1. Let $X, Y \in L^\infty(P)$ such that $X \geq Y$ P-a.s. and we aim to show that $\rho_t^P(X) \leq \rho_t^P(Y)$ P-a.s. Recall that

$$\rho_t^P(Y) = P - \operatorname{ess\,inf}_{\tilde{Y} \in \mathcal{X}^{P,Y}} \rho_t(\tilde{Y}),$$

then by Proposition 5.1.3, there exists a decreasing sequence $(\rho_t(\tilde{Y}_n))_{n\in\mathbb{N}}$ such that

$$\rho_t^P(Y) = \lim_{n\to+\infty} \rho_t(\tilde{Y}_n), \qquad P - a.s.$$

Choose without loss of generality $(\tilde{Y}_n)_{n\in\mathbb{N}} \subset \mathcal{X}^{P,Y}$ to be pointwise increasing. Similarly for X we consider $(\tilde{X}_n)_{n\in\mathbb{N}} \subset \mathcal{X}^{P,Y}$ increasing such that

$$\rho_t^P(X) = \lim_{n\to+\infty} \rho_t(\tilde{X}_n), \qquad P - a.s.$$

Define now

$$\tilde{Z}_n := X 1_{A_n} + \max(\tilde{X}_n, \tilde{Y}_n) 1_{A_n^c},$$

where $A_n = \{X = \tilde{X}_n\} \cap \{Y = \tilde{Y}_n\} \cap \{X \geq Y\}$. Notice that by definition $P(A_n) = 1$, and thus $P(\tilde{Z}_n = X) = 1$, i.e., $\tilde{Z}_n \in \mathcal{X}^{P,X}$. Moreover it holds true that $\tilde{Z}_n \geq \tilde{X}_n$ pointwise and that $(\tilde{Z}_n)_{n\in\mathbb{N}}$ is increasing. Let us show in the following the latter statement. Let $n \in \mathbb{N}$, then

$$\tilde{Z}_{n+1} = X 1_{A_{n+1}} + \max(\tilde{X}_{n+1}, \tilde{Y}_{n+1}) 1_{A_{n+1}^c} \geq X 1_{A_{n+1}} + \max(\tilde{X}_n, \tilde{Y}_n) 1_{A_{n+1}^c}.$$

Furthermore for all $\omega \in A_{n+1}$ observe that $X(\omega) = \tilde{X}_{n+1}(\omega) \geq \tilde{X}_n(\omega)$ and $X(\omega) \geq Y(\omega) = \tilde{Y}_{n+1}(\omega) \geq \tilde{Y}_n(\omega)$, hence

$$X(\omega) \geq \max(\tilde{X}_n, \tilde{Y}_n)(\omega), \qquad \forall \omega \in A_{n+1},$$

yielding $\tilde{Z}_{n+1}(\omega) \geq \max(\tilde{X}_n, \tilde{Y}_n)(\omega)$ for all $\omega \in \Omega$. Therefore

$$\tilde{Z}_{n+1}(\omega) \geq \begin{cases} \max(\tilde{X}_n, \tilde{Y}_n)(\omega) = X(\omega), & \text{if } \omega \in A_n, \\ \max(\tilde{X}_n, \tilde{Y}_n)(\omega), & \text{if } \omega \in A_n^c, \end{cases}$$

i.e., $\tilde{Z}_{n+1} \geq \tilde{Z}_n$ pointwise. The previous implies trivially that $(\rho_t(\tilde{Z}_n))_{n\in\mathbb{N}}$ is pointwise decreasing and additionally $\rho_t(\tilde{Z}_n) \leq \rho_t(\tilde{X}_n)$ pointwise. Notice moreover that

$$\lim_{n\to+\infty} \rho_t(\tilde{Z}_n) = \rho_t^P(X), \qquad P-a.s.$$

Indeed, $\rho_t(\tilde{Z}_n) \in \mathscr{R}_t^{P,X}$ for all $n \in \mathbb{N}$

$$\rho_t(\tilde{Z}_n) \geq \rho_t^P(X), \qquad P-a.s.,$$

and since the following holds true P-a.s.

$$\rho_t^P(X) \leq \limsup_{n\to+\infty} \rho_t(\tilde{Z}_n) \leq \limsup_{n\to+\infty} \rho_t(\tilde{X}_n) = \rho_t^P(X).$$

From the previous we obtain the desired inequality to show monotonicity:

$$\rho_t^P(Y) = \lim_{n\to+\infty} \rho_t(\tilde{Y}_n) \geq \lim_{n\to+\infty} \rho_t(\tilde{Z}_n) = \rho_t^P(X),$$

where the previous inequality follows by the fact that pointwise

$$\tilde{Y}_n = \tilde{Y}_n 1_{A_n} + \tilde{Y}_n 1_{A_n^c} = Y 1_{A_n} + \tilde{Y}_n 1_{A_n^c} \leq X 1_{A_n} + \max(\tilde{X}_n, \tilde{Y}_n) 1_{A_n^c} = \tilde{Z}_n.$$

Let us now focus on property (i), i.e, we have to show that for $m_t \in L_t^\infty(P)$ and $X \in L^\infty(P)$ it holds that $\rho_t^P(X + m_t) = \rho_t^P(X) - m_t$ P-a.s. Let $\tilde{X} \in \mathcal{X}^{P,X}$ and $\tilde{m} \in \mathcal{X}_t^{P,m_t}$, then clearly $\tilde{X} + \tilde{m} \in \mathcal{X}^{P,X+m_t}$, hence $\mathcal{X}^{P,X} + \mathcal{X}^{P,m_t} \subseteq \mathcal{X}^{P,X+m_t}$. This yields

$$\begin{aligned} \rho_t^P(X + m_t) = P - \operatorname{ess\,inf}_{\tilde{Z}\in\mathcal{X}^{P,X+m_t}} \rho_t(\tilde{Z}) &\leq P - \operatorname{ess\,inf}_{\tilde{X}\in\mathcal{X}^{P,X},\tilde{m}\in\mathcal{X}_t^{P,m_t}} \rho_t(\tilde{X} + \tilde{m}) \\ &= P - \operatorname{ess\,inf}_{\tilde{X}\in\mathcal{X}^{P,X},\tilde{m}\in\mathcal{X}_t^{P,m_t}} \rho_t(\tilde{X}) - \tilde{m} \\ &= \rho_t^P(X) - m_t, \end{aligned}$$

where the last identity holds P-a.s. Let us turn to the other direction. By Proposition 5.1.3 there exists $(\tilde{Z}_n)_{n\in\mathbb{N}} \subset \mathcal{X}^{P,X+m_t}$ such that

$$\lim_{n\to+\infty} \rho_t(\tilde{Z}_n) = \rho_t^P(X + m_t), \qquad P-a.s.$$

Introduce the sets $B_n := \{\tilde{Z}_n = X + m_t\}$, and choose an increasing sequence $(\tilde{X}_n)_{n\in\mathbb{N}} \subset \mathcal{X}^{P,X}$ such that $\rho_t(\tilde{X}_n) \to \rho_t^P(X)$ P-a.s. Fix $K := \|m_t\|_{L^\infty(P)}$ and define

$$Z_n^{'} := X_n^{'} + m_n^{'},$$

where

$$\begin{aligned} X_n^{'} &:= \max(\tilde{X}_n, \tilde{Z}_n - m_t) 1_{B_n \cap \{m_t \leq K\}} + \max(\tilde{X}_n, \tilde{Z}_n) 1_{B_n^c \cup \{m_t \geq K\}} \\ m_n^{'} &= m_t 1_{B_n \cap \{m_t \leq K\}}, \end{aligned}$$

then the following hold true:

- $X_n^{'} \in \mathcal{X}^{X,P}$, $m_n^{'} \in \mathcal{X}_t^{P,m_t}$, for all $n \in \mathbb{N}$;
- $Z_n^{'} \geq \tilde{Z}_n$, pointwise, for all $n \in \mathbb{N}$;
- $\rho_t(X_n^{'}) \to \rho_t^P(X)$ as $n \to +\infty$, P-a.s.

Observe that the statements in the first point hold true as $P(X_n^{'} = X) = P(\{\tilde{X}_n = X\} \cap B_n \cap \{m_t \leq K\}) = 1$ and $P(m_n' = m_t) = P(B_n \cap \{m_t \leq K\}) = 1$ as $P(B_n) = P(m_t \leq K) = P(\tilde{X}_n = X) = 1$. Moreover, clearly, m_n' is $\mathcal{F}_t$-measurable.

The second point can be shown to hold directly as for all $n \in \mathbb{N}$

$$Z_n^{'} \geq (\tilde{Z}_n - m_t)1_{B_n \cap \{m_t \leq K\}} + m_t 1_{B_n \cap \{m_t \leq K\}} + \tilde{Z}_n 1_{B_n^c \cup \{m_t > K\}} = \tilde{Z}_n.$$

The last point of the previous follows from the fact that by definition

$$\rho_t(X_n^{'}) \geq \rho_t^P(X) = \lim_{n \to +\infty} \rho_t(\tilde{X}_n),$$

and $X_n^{'} \geq \tilde{X}_n$ for all $n \in \mathbb{N}$, thus

$$\rho_t^P(X) \leq \limsup_{n \to +\infty} \rho_t(X_n^{'}) \leq \lim_{n \to +\infty} \rho_t(\tilde{X}_n) = \rho_t^P(X), \qquad P - a.s.$$

By the previous properties we obtain P-a.s. that $\lim_{n\to\infty} m_n' = \lim_{n\to\infty} m_t = m_t$ and further

$$\rho_t^P(X) - m_t = \lim_{n \to +\infty} (\rho_t(X_n^{'}) - m_n^{'}) = \lim_{n \to +\infty} \rho_t(Z_n^{'}) \leq \lim_{n \to +\infty} \rho_t(\tilde{Z}_n) = \rho_t^P(X + m_t),$$

which shows the remaining direction for cash invariance.

We are left to show the convexity property, i.e., (iv). To this extent fix $X, Y \in L^\infty(P)$, $\lambda_t \in L_t^\infty(P)$ such that $P(0 \leq \lambda_t \leq 1) = 1$ and choose two increasing sequences $(\tilde{X}_n)_{n\in\mathbb{N}}, (\tilde{Y}_n)_{n\in\mathbb{N}}$ such that $\rho_t(\tilde{X}_n) \to \rho_t^P(X)$ and $\rho_t(\tilde{Y}_n) \to \rho_t^P(Y)$ P-a.s., respectively. Define

$$\tilde{\lambda} := \lambda_t 1_{\{\lambda_t \in [0,1]\}} \in \mathcal{X}_t^{\lambda_t, P},$$

then for all $n \in \mathbb{N}$ it holds that $\tilde{\lambda}\tilde{X}_n + (1 - \tilde{\lambda})\tilde{Y}_n \in \mathcal{X}^{P, \lambda_t X + (1-\lambda_t)Y}$. Moreover,

$$\rho_t(\tilde{\lambda}\tilde{X}_n + (1 - \tilde{\lambda})\tilde{Y}_n) \leq \tilde{\lambda}\rho_t(\tilde{X}_n) + (1 - \tilde{\lambda})\rho_t(\tilde{Y}_n) \to \lambda_t \rho_t^P(X) + (1 - \lambda)\rho_t^P(Y), \quad (5.1.3)$$

as $n \to +\infty$, P-a.s. Hence convexity holds. □

Remark 5.1.5. Concerning property (iii), observe that we get the following. Let $P(X = 0) = 1$. Then, obviously the random variable that is 0 for all ω is in $\mathcal{X}^{P,X}$. Therefore, we get

$$\rho_t^P(X) = P - \operatorname{ess\,inf}_{\tilde{X} \in \mathcal{X}^{P,X}} \rho_t(\tilde{X}) \leq \rho_t(0) = 0, \quad P\text{-a.s.},$$

hence $\rho_t^P(0) \leq 0$ P-a.s. We cannot get the other inequality but we can always normalize the dynamic risk measure ρ_t^P in the following way: let $\alpha_t := \rho_t^P(0)$ which is $\mathcal{F}_t$-measurable. Define

$$\hat{\rho}_t^P(Y) := \rho_t^P(Y) - \alpha_t.$$

Then $\hat{\rho}_t^P$ is a conditional risk measure with $\hat{\rho}_t^P(0) = 0$, by the property of cash invariance.

The next lemma shows that ρ_t^P is uniquely determined by its acceptance set.

Lemma 5.1.6. For any $X \in L^\infty(P)$,

$$\rho_t^P(X) = \operatorname{ess\,inf}\{Y \in L_t^\infty(P) : X + Y \in C_t^P,\ P - \text{a.s.}\}.$$

Proof. First note that

$$\rho_t^P(X + \rho_t^P(X)) = \rho_t^P(X) - \rho_t^P(X) = 0 \quad P - \text{a.s},$$

hence $X + \rho_t^P(X) \in C_t^P$, P-a.s., and $\rho_t^P(X)$ is greater equal than the essential infimum of the Lemma. For the other direction take any $Y \in L_t^\infty(P)$ such that $X + Y \in C_t^P$, P-a.s. Then $\rho_t^P(X + Y) \leq 0$ and, as $\rho_t^P(X + Y) = \rho_t^P(X) - Y$ we get $\rho_t^P(X) \leq Y$, everything P-a.s. Hence

$$\rho_t^P(X) \leq \operatorname{ess\,inf}\{Y \in L_t^\infty(P) : X + Y \in C_t^P,\ P - \text{a.s.}\}$$

□

As follows we provide a possible example of pointwise conditional risk measure as introduced above.

Example 5.1.7. Let Ω be countably or finitely generated, i.e., there exists $(\tilde{\Omega}_k)_{k=1}^K$, such that

$$\Omega = \bigcup_{k=1}^{K} \tilde{\Omega}_k,$$

with $K < +\infty$ or $K = +\infty$. Define the filtration as generated by $(\tilde{\Omega}_k)_{k=1}^K$, i.e., $\mathcal{F}_K := \sigma(\tilde{\Omega}_k, k = 1, \ldots, K)$. Let $\mathcal{F}_0 = \{\emptyset, \Omega\}$ and introduce without loss of generality for $1 \leq n < K$ $\mathcal{F}_n = \sigma(\Omega_1^n, \ldots, \Omega_{k_n}^n)$ where

$$\Omega = \bigcup_{i=1}^{k_n} \Omega_i^n,$$

meaning that all Ω_i^n are unions of some $\tilde{\Omega}_k$. Moreover let for all $n < K$ be $\Omega_1^n, \ldots, \Omega_{k_n}^n \in \mathcal{F}_{n+1}$ this ensures the filtration $\mathcal{F} = (\mathcal{F}_n)_n$ to be directed upwards. This setup includes for instance the generated filtration of the binomial or trinomial model.
Let a_n be $\mathcal{F}_n$-measurable, i.e., $a_n \in \mathcal{X}_n$ and $a_n(\omega) \geq 0$ for all $\omega \in \Omega$. As a_n is $\mathcal{F}_n$-measurable,

$$a_n(\omega) = \sum_{i=1}^{k_n} a_i^n 1_{\Omega_i^n},$$

where $a_i^n \in \mathbb{R}_+$. Define now,

$$\rho_0^{\max}(X) := -\inf_{\omega \in \Omega} X(\omega) + a,$$

with $a \geq 0$ and for all $n \geq 1$,

$$\rho_n^{\max}(X)(\omega) := \sum_{i=1}^{k_n} 1_{\Omega_i^n}(\omega) \left(- \inf_{\omega \in \Omega} X(\omega) + a_i^n \right). \tag{5.1.4}$$

As usual, define

$$\begin{aligned} C_n :&= \{X \in \mathcal{X}_n : \rho_n^{\max}(X)(\omega) \leq 0, \ \forall \omega \in \Omega\} \\ &= \{X \in \mathcal{X}_n : X(\omega) \geq -a_i^n, \text{ for } \omega \in \Omega_i^n, \ i = 1, \ldots, k_n\}. \end{aligned}$$

In the following we prove the next two facts:

- $\rho_n^{\max}$ as defined in (5.1.4) is a conditional monetary risk measure, see Lemma 5.1.8.
- For all $\omega \in \Omega$ and $X \in \mathcal{X}$ it holds,

$$\rho_n^{\max}(X)(\omega) = \sup_{Q \in \mathcal{M}_1} \mathbb{E}_Q[-X|\mathcal{F}_n](\omega) + a_n(\omega),$$

  see Lemma 5.1.9.

Lemma 5.1.8. Let $\rho_n^{\max}$ be defined as in (5.1.4). Then, $\rho_n^{\max}$ is a conditional monetary risk measure.

Proof. We follow the properties as outlined in Definition 5.1.1. For (i), fix $m_n \in \mathcal{X}_n$ such that for $m_i^n \in \mathbb{R}$,

$$m_n = \sum_{i=1}^{k_n} m_i^n 1_{\Omega_i}.$$

Then for $\omega \in \Omega_i^n$,

$$\begin{aligned} \rho_n^{\max}(X + m_n)(\omega) &= - \inf_{\omega \in \Omega_i} (X(\omega) + m_n^i) + a_i^n \\ &= - \inf_{\omega \in \Omega_i} X(\omega) + a_i^n - m_i^n \\ &= \rho_n^{\max}(X)(\omega) - m_n(\omega), \end{aligned}$$

which holds true for $i = 1, \ldots, k_n$. Concerning monotonicity, let $X(\omega) \leq Y(\omega)$ for all $\omega \in \Omega$, hence

$$\inf_{\omega \in \Omega_i} X(\omega) \leq \inf_{\omega \in \Omega_i} Y(\omega) \implies - \inf_{\omega \in \Omega_i} X(\omega) + a_i^n \geq - \inf_{\omega \in \Omega_i^n} Y(\omega) + a_i^n,$$

yielding $\rho_n^{\max}(X) \geq \rho_n^{\max}(Y)$ as it holds for all $\omega \in \Omega_i^n$ and i is arbitrary. The normalization property hold if $a_i^n = 0$ for each i. □

Lemma 5.1.9. Let $\rho_n^{\max}$ be defined as in (5.1.4). Then, for all $\omega \in \Omega$ and $X \in \mathcal{X}$ it holds,

$$\rho_n^{\max}(X)(\omega) = \sup_{Q \in \mathcal{M}_1} \mathbb{E}_Q[-X|\mathcal{F}_n](\omega) + a_n(\omega).$$

Proof. Let $Q \in \mathcal{M}_1$ be arbitrary, then,

$$\begin{aligned}\mathbb{E}_Q[-X|\mathcal{F}_n](\omega) &\leq \sum_{i=1}^{k_n} \frac{\mathbb{E}_Q[-\inf_{\omega\in\Omega} X(\omega)1_{\Omega_i^n}]}{Q(\Omega_i^n)}1_{\Omega_i^n} \\ &= \sum_{i=1}^{k_n} -\inf_{\omega\in\Omega_i} X(\omega)\frac{Q(\Omega_i^n)}{Q(\Omega_i^n)}1_{\Omega_i^n} \\ &= \sum_{i=1}^{k_n} -\inf_{\omega\in\Omega_i} X(\omega)1_{\Omega_i^n},\end{aligned}$$

yielding

$$\begin{aligned}\sup_{Q\in\mathcal{M}_1} \mathbb{E}_Q[-X|\mathcal{F}_n](\omega) + a_n(\omega) &\leq \sum_{i=1}^{k_n} \left(-\inf_{\omega\in\Omega_i^n} X(\omega)\right)1_{\Omega_i^n} + a_n(\omega) \\ &= \rho_n^{\max}(X)(\omega).\end{aligned}$$

For the other direction let us fix Ω_i^n and let $x_i := \inf_{\omega\in\Omega_i^n} X(\omega)$. By definition, for all $\varepsilon > 0$, there exists $\omega_i^\varepsilon \in \Omega_i$ such that $X(\omega_i^\varepsilon) \leq x_i + \varepsilon$.

Define $Q^\varepsilon := \frac{1}{k_n}\sum_{i=1}^{k_n} \delta_{\{\omega_i^\varepsilon\}} \in \mathcal{M}_1$. Then,

$$\begin{aligned}\mathbb{E}_{Q^\varepsilon}[-X|\mathcal{F}_n] &= \sum_{i=1}^{k_n} \frac{\mathbb{E}_{Q^\varepsilon}[-X1_{\Omega_i^n}]}{Q^\varepsilon(\Omega_i^n)}1_{\Omega_i^n} \\ &= \sum_{i=1}^{n} \frac{1}{k_n}\frac{-X(\omega_i^\varepsilon)}{1/k_n}1_{\Omega_i^n} \\ &\geq \sum_{i=1}^{n}(-x_i - \varepsilon)1_{\Omega_i^n} \\ &= \sum_{i=1}^{n}\left(-\inf_{\omega\in\Omega_i} X(\omega)\right)1_{\Omega_i^n} - \varepsilon,\end{aligned}$$

which yields

$$\mathbb{E}_{Q^\varepsilon}[-X|\mathcal{F}_n] + a_n \geq \rho_n^{\max}(X) - \varepsilon,$$

i.e., we can always find for all $\varepsilon > 0$ a probability measure $Q_\varepsilon \in \mathcal{M}_1$ such that the previous holds, so pointwise

$$\sup_{Q\in\mathcal{M}_1} \mathbb{E}_Q[-X|\mathcal{F}_n] + a_n \geq \rho_n^{\max}(X).$$

□

Remark 5.1.10. We underline in the present the fact that the supremum in Lemma 5.1.9 is well-defined. Indeed consider $\Omega = \bigcup_{k=1}^{+\infty} \tilde{\Omega}_k$, then each $Q \in \mathcal{M}_1$ is a sequence

of real numbers $q = (q_k)_{k=1}^{+\infty}$, with $q_k \in [0,1]$ and $\sum_{k=1}^{+\infty} q_k = 1$. In particular $q \in B_1(0)$, where $B_1(0)$ here denotes a closed ball in ℓ^1. We show that,

$$\sup_{Q\in\mathcal{M}_1} \mathbb{E}_Q[-X|\mathcal{F}_n](\omega) = \sup_{q\in S} \mathbb{E}_q[-X|\mathcal{F}_n](\omega), \qquad \forall \omega \in \Omega$$

where $S = B_1(0) \cap \ell^1_+ \cap \{q : q_k \in \mathbb{Q}, \forall k \geq 1\}$, which is a countable set. This would then imply that the supremum is measurable and well-defined. Observe that clearly

$$\sup_{Q\in\mathcal{M}_1} \mathbb{E}_Q[-X|\mathcal{F}_n](\omega) \geq \sup_{q\in S} \mathbb{E}_q[-X|\mathcal{F}_n](\omega), \qquad \forall \omega \in \Omega.$$

For the other direction let us fix $Q \in \mathcal{M}_1$ with $Q = (q_k)_{k\geq 1}$. Then, for all $\varepsilon > 0$ we choose $\tilde{q}_k$ s.t. $\tilde{q}_k \in \mathbb{Q}$, $\sum_{k\geq 1} \tilde{q}_k = 1$ and $\sum_{k\geq 1} |q_k - \tilde{q}_k| < \varepsilon$. For $\omega \in \Omega_i^n$ with $i \in \{1, \dots, k_n\}$ then

$$\mathbb{E}_{\tilde{Q}^\varepsilon}[-X|\mathcal{F}_n](\omega) \geq \mathbb{E}_Q[-X|\mathcal{F}_n](\omega) + \varepsilon \inf_{\omega\in\Omega} X(\omega) > -\infty,$$

where $\tilde{Q}^\varepsilon = (\tilde{q}_k)_{k\geq 1}$, such that $|Q(\Omega_i^n) - \tilde{Q}^\varepsilon(\Omega_i^n)| < \varepsilon$, showing the other direction as $\varepsilon \to 0$ since X is bounded.

Let $\mathcal{M} \subset \mathcal{M}_1$ be a collection of probability measure. We state some assumptions on the set $\mathcal{M}$ to which we will refer whenever they are in force.

Assumption 5.1.11. *The set $\mathcal{M}$ is convex.*

Assumption 5.1.12. *The set $\mathcal{M}$ is dominated, i.e., there exists $\mathbb{P} \in \mathcal{M}_1$ such that $P \ll \mathbb{P}$ for all $P \in \mathcal{M}$.*

Definition 5.1.13. Under Assumption 5.1.12, for any $P \in \mathcal{M}$ we introduce a version of the dynamic risk measure ρ_t^P with respect to $\mathbb{P}$ where $\mathbb{P}$ is the dominating measure of $\mathcal{M}$. To be precise:

$$\rho_t^{\mathbb{P},P}(X) := \mathbb{P} - \operatorname{ess\,inf}_{\tilde{X}\in\mathcal{X}^{P,X}} \rho_t(\tilde{X}), \qquad \mathbb{P} - a.s.$$

Observe that whenever it is not clear with respect to which probability measure $P \in \mathcal{M}_1$ we are considering the essential infimum or supremum we shall write $P - \operatorname{ess\,inf}$ or $P - \operatorname{ess\,sup}$, respectively.

Remark 5.1.14. Notice that, since for any $P \in \mathcal{M}$ and $P \ll \mathbb{P}$ we have that $\mathcal{X}^{\mathbb{P},X} \subseteq \mathcal{X}^{P,X}$ for all $X \in L^\infty(\mathbb{P})$, then it holds true that

$$\rho_t^{\mathbb{P}}(X) \geq \rho_t^{\mathbb{P},P}(X), \qquad \mathbb{P} - a.s.$$

Lemma 5.1.15. Suppose Assumption 5.1.12 holds true. Let $P \in \mathcal{M}$ and let $\mathbb{P}$ be the dominating measure of $\mathcal{M}$. Then, for all $X \in L^\infty(\mathbb{P})$

$$\rho_t^{\mathbb{P},P}(X) = \rho_t^P(X), \qquad P - a.s.$$

Proof. By definition, we have that, for all $\tilde{X} \in \mathcal{X}^{P,X}$, $\rho_t^{\mathbb{P},P}(X) \leq \rho_t(\tilde{X})$ $\mathbb{P}$-a.s. As $P \ll \mathbb{P}$ this holds P-a.s. as well. Therefore we get

$$\rho_t^{\mathbb{P},P}(X) \leq P - \operatorname{ess\,inf}_{\tilde{X} \in \mathcal{X}^{P,X}} \rho_t(\tilde{X}) = \rho_t^P(X), \quad P - a.s.$$

For the other direction we use Proposition 5.1.3: there exist sequences $(\tilde{X}_n)_{n\in\mathbb{N}}), (\tilde{Y}_n)_{n\in\mathbb{N}})$ in $\mathcal{X}^{P,X}$ such that

$$\rho_t^P(X) = \lim_{n\to+\infty} \rho_t(\tilde{X}_n), \quad P - a.s., \qquad \rho_t^{\mathbb{P},P}(X) = \lim_{n\to+\infty} \rho_t(\tilde{Y}_n), \quad \mathbb{P} - a.s.,$$

respectively. As $P \ll \mathbb{P}$ the second limit in the above is also P-a.s. For all $n \in \mathbb{N}$, set $\tilde{Z}_n := \max(\tilde{X}_n, \tilde{Y}_n)$, then still $P(\tilde{Z}_n = X) = 1$. Therefore

$$\rho_t^P(X) \leq \rho_t(\tilde{Z}_n), \quad P - \text{a.s.},$$

yielding

$$\rho_t^P(X) \leq P - \limsup_{n\to+\infty} \rho_t(\tilde{Z}_n) \leq P - \limsup_{n\to+\infty} \rho_t(\tilde{Y}_n) = \rho_t^{\mathbb{P},P}(X), \quad P - a.s.,$$

and thus the claim follows and $\rho_t^{\mathbb{P},P} = \rho_t^P(X)$ P-a.s. □

Clearly by Remark 5.1.14 we have that under Assumption 5.1.12, that $C_t^{\mathbb{P}} \subset C_t^P$ for $P \ll \mathbb{P}$. Indeed suppose $X \in C_t^{\mathbb{P}}$, then since $P \ll \mathbb{P}$ by Lemma 5.1.15

$$\rho_t^P(X) = \rho_t^{\mathbb{P},P}(X), \qquad P - a.s.,$$

and by Remark 5.1.14

$$\rho_t^{\mathbb{P},P}(X) \leq \rho_t^{\mathbb{P}}(X) \leq 0, \qquad \mathbb{P} - a.s.,$$

and thus P-a.s. Hence one can conclude that $X \in C_t^P$.

Corollary 5.1.16. *Under Assumptions 5.1.11, 5.1.12 if ρ_t is a conditional (convex) risk measure, then for all $X \in L^\infty(\mathbb{P})$*

$$\rho_t^{\mathcal{M}}(X) := \mathbb{P} - \operatorname{ess\,sup}_{P\in\mathcal{M}} \rho_t^{\mathbb{P},P}(X),$$

is a conditional (convex) risk measure. We call the previous a robust conditional risk measure.

Proof. The proof follows from two main steps. Let $X \in L^\infty(\mathbb{P})$ be fixed. First of all notice that the set $\{\rho_t^{\mathbb{P},P}(X) : P \in \mathcal{M}\}$ is directed upwards. Indeed if $P_1, P_2 \in \mathcal{M}$, then

$$\mathcal{X}^{\tilde{P},X} \subseteq \mathcal{X}^{P_1,X} \cap \mathcal{X}^{P_2,X},$$

with $\tilde{P} := \frac{P_1+P_2}{2} \in \mathcal{M}$ by assumption. Thus

$$\rho_t^{\mathbb{P},\tilde{P}}(X) \geq \rho_t^{\mathbb{P},P_1}(X) \vee \rho_t^{\mathbb{P},P_2}(X), \qquad \mathbb{P} - a.s.$$

Then, there exists a sequence $(P_n)_{n\in\mathbb{N}} \subset \mathcal{M}$ such that

$$\rho_t^{\mathcal{M}}(X) = \lim_{n\to+\infty} \rho_t^{\mathbb{P},P_n}(X), \qquad \mathbb{P}-a.s.$$

The second step translates into showing that, given ρ_t a conditional convex risk measure, then also $\rho_t^{\mathbb{P},P}$ is a conditional convex risk measure where now all the properties hold $\mathbb{P}$-almost surely. Observe that this follows as in Lemma 5.1.4 where all the properties hold $\mathbb{P}$-a.s. □

Theorem 5.1.17. *Let $\mathcal{M} \subseteq \mathcal{M}_1$ and $\mathbb{P}$ a probability measure on $(\Omega, \mathcal{F})$. Suppose $\mathbb{P}$ is generalized equivalent to $\mathcal{M}$ and that $\mathcal{M}$ is closed under countable convex combinations. Then for all $X \in L^\infty(\mathbb{P})$,*

$$\rho_t^{\mathbb{P}}(X) = \rho_t^{\mathcal{M}}(X), \qquad \mathbb{P}-a.s.$$

Proof. By Halmos-Savage's theorem there exists $P^0 \in \mathcal{M}$ such that $P^0 \sim \mathbb{P}$. Moreover since the fact that $\mathbb{P}$ is generalized equivalent to $\mathcal{M}$ and $\mathcal{M}$ is closed under countable convex combinations, implies that Assumptions 5.1.11, 5.1.12 are satisfied. Hence

$$\rho_t^{\mathcal{M}}(X) \geq \rho_t^{\mathbb{P},P_0}(X) = \rho_t^{\mathbb{P}}(X),$$

holds $P^0 - a.s.$ and thus $\mathbb{P} - a.s.$ as $P^0 \sim \mathbb{P}$. On the other hand as $P \ll \mathbb{P}$ for all $P \in \mathcal{M}$, we have that for all $P \in \mathcal{M}$

$$\rho_t^{\mathbb{P},P}(X) \leq \rho_t^{\mathbb{P}}(X), \qquad \mathbb{P}-a.s.,$$

hence,

$$\rho_t^{\mathcal{M}}(X) = \mathbb{P}-\operatorname{ess\,sup}_{P\in\mathcal{M}} \rho_t^P(X) \leq \rho_t^{\mathbb{P}}(X), \qquad \mathbb{P}-a.s.,$$

for any $X \in L^\infty(\mathbb{P})$, yielding the thesis. □

A. Appendix

A.1. Conditional and unconditional expected signature of correlated Brownian motions

We explain now how to compute the expected signature of the time-extended polynomial processes $\widehat{X}$, i.e. how to compute $\mathbb{E}[\langle e_I, \widehat{\mathbb{X}}_t\rangle]$ for each multi-index I.

A.1.1. The expected signature

A possible choice for the primary process X is given by a d-dimensional correlated Brownian motion W with correlation matrix ρ. Since Assumption 1.2.6 is satisfied, we follow Definition 1.2.1(ii) and set $\widehat{W}_t := (t, W_t)$. The signature of $\widehat{W}$ up to time t is denoted by $\widehat{\mathbb{W}}_t$.

Theorem A.1.1. *Consider a multi-index I admitting the representation*

$$e_I = e_0^{\otimes k_0} \otimes e_{J_1} \otimes e_0^{\otimes k_1} \otimes e_{J_2} \otimes \cdots \otimes e_0^{\otimes k_m}, \tag{A.1.1}$$

for some $J_i \in \{1, \dots, d\}^{2h_i}$ and $h_i, k_i \in \mathbb{N}_0$. Then

$$\mathbb{E}[\langle e_I, \widehat{\mathbb{W}}_t\rangle] = \frac{t^{\sum_{i=0}^m k_i + \sum_{i=1}^m h_i}}{(\sum_{i=0}^m k_i + \sum_{i=1}^m h_i)!} \left(\frac{1}{2}\right)^{\sum_{i=1}^m h_i} \prod_{i=1}^m \rho(J_i),$$

where $\rho(J) := \prod_{k=1}^{|J|/2} \rho_{j_{2k-1}, j_{2k}}$. Moreover, if I does not admit representation (A.1.1) then $\mathbb{E}[\langle e_I, \widehat{\mathbb{W}}_t\rangle] = 0$.

Proof. We first prove that the truncated signature $\widehat{\mathbb{W}}_t^n$ is a polynomial process in the sense of Definition 2.1 in Filipović and Larsson (2016). By Lemma 2.2 in Filipović and Larsson (2016), it suffices to show that for each $|I| \leq n$ it holds

$$\mathrm{d}\langle e_I, \widehat{\mathbb{W}}_t^n\rangle = b(\widehat{\mathbb{W}}_t^n)\mathrm{d}t + \sigma(\widehat{\mathbb{W}}_t^n)\mathrm{d}W_t \tag{A.1.2}$$

for some linear maps b, σ. Observe that

$$\begin{aligned}\langle e_I, \widehat{\mathbb{W}}^n_t\rangle &= \int_0^t \langle e_{I'}, \widehat{\mathbb{W}}^n_s\rangle \circ \mathrm{d}\langle e_{i_{|I|}}, \widehat{W}_s\rangle \\ &= \int_0^t \langle e_{I'}, \widehat{\mathbb{W}}^n_s\rangle \mathrm{d}\langle e_{i_{|I|}}, \widehat{W}_s\rangle + \frac{\rho_{i_{|I|},i_{|I|-1}}}{2} 1_{\{i_{|I|},i_{|I|-1}\neq 0\}} \int_0^t \langle e_{I''}, \widehat{\mathbb{W}}^n_s\rangle \mathrm{d}s \\ &= \int_0^t \langle 1_{\{i_{|I|}=0\}} e_{I'} + \frac{\rho_{i_{|I|},i_{|I|-1}}}{2} 1_{\{i_{|I|},i_{|I|-1}\neq 0\}} e_{I''}, \widehat{\mathbb{W}}^n_s\rangle \mathrm{d}s \\ &\quad + \int_0^t \langle e_{I'}, \widehat{\mathbb{W}}^n_s\rangle 1_{\{i_{|I|}\neq 0\}} \mathrm{d}W^{i_{|I|}}_s.\end{aligned}$$

Since $|I'|, |I''| \leq n$ we can conclude that $\widehat{\mathbb{W}}^n$ satisfies (A.1.2) and is thus a polynomial process.

Our next step consists in applying Theorem 3.1 in Filipović and Larsson (2016). Observe that by Itô's formula we have that the generator $\mathcal{A}$ of $\widehat{\mathbb{W}}^n$ satisfies

$$\mathcal{A}\langle e_I, \cdot\rangle := \langle 1_{\{i_{|I|}=0\}} e_{I'} + \frac{\rho_{i_{|I|},i_{|I|-1}}}{2} 1_{\{i_{|I|},i_{|I|-1}\neq 0\}} e_{I''}, \cdot\rangle = \sum_{e_{I_1}\otimes e_{I_2} = e_I} \langle e_{I_1}, Q\rangle \langle e_{I_2}, \cdot\rangle,$$

for $Q := e_0 + \frac{1}{2}\sum_{i,j=1}^d \rho_{i,j} e_i \otimes e_j$. Shortly, this can be written as $\mathcal{A}\langle e_I, \cdot\rangle = \langle e_I, Q \otimes (\cdot)\rangle$. Theorem 3.1 in Filipović and Larsson (2016) yields then that

$$\mathbb{E}[\langle e_I, \widehat{\mathbb{W}}_t\rangle] = \sum_{k=0}^\infty \frac{1}{k!} (t\mathcal{A})^k \langle e_I, \cdot\rangle(\widehat{\mathbb{W}}_0) = \sum_{k=0}^\infty \frac{t^k}{k!} \langle e_I, Q^{\otimes k} \otimes \widehat{\mathbb{W}}_0\rangle = \sum_{k=0}^{|I|} \frac{t^k}{k!} \langle e_I, Q^{\otimes k}\rangle.$$

Observe that $\langle e_I, Q^{\otimes k}\rangle = \sum_{e_{I_1}\otimes\cdots\otimes e_{I_k} = e_I} \prod_{i=1}^k \langle e_{I_i}, Q\rangle$ is nonzero if and only if I can be decomposed in multi-indices of length 1 whose elements are all 0, and multi-indices of length 2, whose elements are all strictly larger than 0, i.e. if and only if I satisfies (A.1.1). If this is the case, by definition of Q it holds

$$\mathbb{E}[\langle e_I, \widehat{\mathbb{W}}_t\rangle] = \frac{t^{\sum_{i=0}^m k_i + \sum_{i=1}^m h_i}}{(\sum_{i=0}^m k_i + \sum_{i=1}^m h_i)!} \langle e_0, Q\rangle^{\sum_{i=0}^m k_i} \prod_{i=1}^m \langle e_{J_i}, Q^{\otimes h_i}\rangle,$$

and since $\langle e_0, Q\rangle = 1$ and $\langle e_{J_i}, Q^{\otimes h_i}\rangle = (\frac{1}{2})^{h_1} \prod_{k=1}^{h_i} \rho_{j_{2k-1},j_{2k}}$ the claim follows. □

A.1.2. The conditional expected signature

A combination of the result of Theorem A.1.1 with Chen's identity (Lemma 1.1.10) yields the following expression for the conditional expected signature.

Lemma A.1.2. For each $s \leq t$ it holds

$$\mathbb{E}[\langle e_I, \widehat{\mathbb{W}}_{s+t}\rangle | \mathcal{F}_s] = \sum_{e_{I_1}\otimes e_{I_2} = e_I} \langle e_{I_1}, \widehat{\mathbb{W}}_s\rangle \mathbb{E}[\langle e_{I_2}, \widehat{\mathbb{W}}_t\rangle].$$

Proof. Observe that applying Chen's identity and the definition of the tensor product we can compute

$$\begin{aligned}\mathbb{E}[\langle e_I, \widehat{\mathbb{W}}_{s+t}\rangle|\mathcal{F}_s] &= \mathbb{E}[\langle e_I, \widehat{\mathbb{W}}_s \otimes \widehat{\mathbb{W}}_{s,s+t}\rangle|\mathcal{F}_s] \\ &= \sum_{e_{I_1}\otimes e_{I_2}=e_I} \mathbb{E}[\langle e_{I_1}, \widehat{\mathbb{W}}_s\rangle\langle e_{I_2}, \widehat{\mathbb{W}}_{s,s+t}\rangle|\mathcal{F}_s] \\ &= \sum_{e_{I_1}\otimes e_{I_2}=e_I} \langle e_{I_1}, \widehat{\mathbb{W}}_s\rangle\mathbb{E}[\langle e_{I_2}, \widehat{\mathbb{W}}_{s,s+t}\rangle|\mathcal{F}_s].\end{aligned}$$

Note that for each suitable integrand H and each $u \in [0,t]$ it holds

$$\int_s^{s+u} H_r \mathrm{d}\widehat{W}_r^i = \int_0^u H_{s+r}\mathrm{d}\widehat{B}_r^i,$$

where $\widehat{B}$ is the time extended Brownian motion given by $\widehat{B}_r := \widehat{W}_{s+r} - \widehat{W}_s$. Since B is independent of $\mathcal{F}_s$ we can conclude that $\mathbb{E}[\langle e_{I_2}, \widehat{\mathbb{W}}_{s,s+t}\rangle|\mathcal{F}_s] = \mathbb{E}[\langle e_{I_2}, \widehat{\mathbb{W}}_t\rangle]$, and the claim follows. □

A.2. Calibration using sig-payoffs

We now discuss the procedure proposed in Perez Arribas et al. (2020) which consists in calibrating the model to sig-payoffs that are intended to approximate the true payoffs. For general payoff functions F,

$$F : (\widehat{S}_t)_{t\in[0,T]} \mapsto F\Big((\widehat{S}_t)_{t\in[0,T]}\Big),$$

the first step consists in approximating F with a sig-payoff such that

$$F\Big((\widehat{S}_t)_{t\in[0,T]}\Big) \approx f_\emptyset + \sum_{0<|J|\leq m} f_J \langle e_J, \widehat{\mathbb{S}}_T\rangle, \tag{A.2.1}$$

for some $m \geq 0$ and $f_\emptyset, f_J \in \mathbb{R}$. Here, $\widehat{\mathbb{S}}_T$ denotes the signature of $\widehat{S} = (t, S_t)$ at time T. For our purposes, the approximation of the call payoffs has to be accurate at least with high probability on the set of models

$$\{\widehat{S}_n(\ell) \colon \ell \text{ enters in the optimisation}\}.$$

Under the hypothesis that this can be achieved and proceeding as in Section 1.3, the model call prices $C^{\text{model}}(T, K, \ell)$ (appearing in (1.4.6)) can then be (approximately) computed via

$$C^{\text{model}}(T, K, \ell) = \mathbb{E}_{\mathbb{Q}}[(S_n(\ell)_T - K)^+] \approx f_\emptyset + \sum_{0<|J|\leq m} f_J \widetilde{P}_J(\ell, \mathbb{E}_{\mathbb{Q}}[\widehat{\mathbb{X}}_T]).$$

The benefits from a computational perspective are immediate.

To achieve the approximation of the payoff functions, in Lyons et al. (2020) and Perez Arribas (2020) it is suggested to use an auxiliary stochastic process $(Y_t)_{t\geq 0}$ playing the role of a market generator under $\mathbb{Q}$. A standard example for Y is given by the Black-Scholes model

$$\mathrm{d}Y_t = \sigma Y_t \mathrm{d}W_t, \qquad Y_0 = S_0,$$

where W denotes a Brownian motion and σ is sampled uniformly in the range of the implied volatility surface of the market, or as in Perez Arribas (2020) between 5% and 40% of the spot price. We set again $\widehat{Y}_t := (t, Y_t)$ and denote by $\widehat{\mathbb{Y}}$ the signature of $\widehat{Y}$. The procedure to find the coefficients f of (A.2.1) consists in a linear regression, fitting $N_{MC} > 0$ realizations of the simulated sig-payoffs to the corresponding realizations of the simulated payoff. The corresponding loss function is given by

$$L_{\text{payoff}}(f) := \sum_{j=1}^{N_{MC}} \left((Y_T(\omega_j) - K)^+ - \left(f_\emptyset + \sum_{0<|J|\leq m} f_J \langle e_J, \widehat{\mathbb{Y}}_T(\omega_j)\rangle \right) \right)^2.$$

In Perez Arribas (2020) the author suggests to choose $N_{MC} = 10^4$, $m = 5$ and to sample $\sigma \sim \mathcal{U}([0.05, 0.4])$.

We applied this procedure and then compared the call prices obtained under the Black Scholes model with $\sigma = 0.25$ with the the approximation via sig-payoffs denoted by $F(T, K)$. For $T = 0.5$, $Y_0 = 100$, and $K = 120$, we obtain the following results:

$\mathbb{E}_{\mathbb{Q}}[(S_T - K)^+]$	$\frac{1}{N_{MC}} \sum_{i=1}^{N_{MC}} (S_T(\omega_i) - K)^+$	$\mathbb{E}_{\mathbb{Q}}[F(T,K)]$	$\frac{1}{N_{MC}} \sum_{i=1}^{N_{MC}} F(T,K)(\omega_i)$
1.5155	1.5145	1.6567	1.6550

Here, $\mathbb{E}_{\mathbb{Q}}[(S_T - K)^+]$ was computed via the Black-Scholes formula and $\mathbb{E}_{\mathbb{Q}}[F(T,K)]$ using the analytic expression of the expected signature of the time extended Black-Scholes model. We also compare the corresponding Monte Carlo prices using $N_{MC} = 10^6$. We observed that the absolute percentage error between the Black-Scholes price and the sig-payoff option price is greater than 8%. The difference can become even more significant if the Black-Scholes model is replaced by $S_n(\ell)$ for ℓ in a certain parameter range over which we optimize. So even if the model can provide an accurate fit to the sig-payoffs, this does not mean that the true implied volatility surface is well approximated, as the gap between call option prices and the approximate sig option prices is simply too large. Therefore we opted for the procedure outlined in Section 1.4.2.

A.3. Numerical results for VIX options in the Brownian motion case

In the present we report the numerical results corresponding to the calibration to only VIX options as in Section 2.3.4, where the components of X are simply correlated Brownian motions as done in Chapter 1. To be precise, we here model the volatility process σ^S as a linear function of the time extended signature of X, where X is 2-dimensional Brownian motion with correlation matrix given, as in Section 2.3.4, by

$$\rho = \begin{pmatrix} 1 & -0.577 & 0.3 \\ \cdot & 1 & -0.6 \\ \cdot & \cdot & 1 \end{pmatrix},$$

i.e., where we consider additionally the correlation with the Brownian motion $B = (B_t)_{t\geq 0}$ driving the price process. We truncate the signature at level $n = 3$, we sample $N_{MC} = 80000$ trajectories for Monte Carlo pricing and we minimize the loss function (2.3.14) with $\beta = 1$ to fit the same data-set employed in Section 2.3.4.

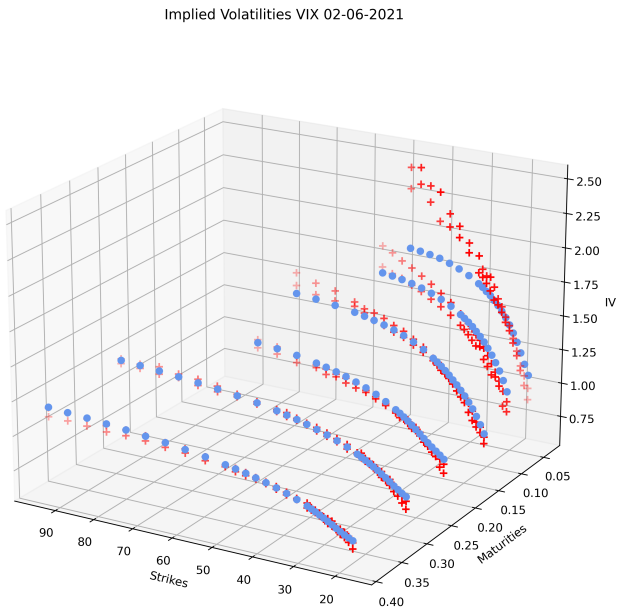


Figure A.1.: The red crosses denote the bid-ask spread (of the implied volatilities) for each maturity, while the azure dots denote the calibrated implied volatilities of the model. On the x-axis we find the strikes and on the y-axis we find the maturities.

We observe how with only correlated Brownian motion the model is not able to calibrate both to all futures' market prices (see Figure A.2 below) and the model implied volatilities

do not fall in general in the bid-ask spread, in particular for high strikes and short maturities.

$T_1 = 0.0383$	$T_2 = 0.0767$	$T_3 = 0.1342$
$\varepsilon_{T_1} = 2.1 \times 10^{-5}$	$\varepsilon_{T_2} = 2.7 \times 10^{-2}$	$\varepsilon_{T_3} = 2.1 \times 10^{-7}$

$T_4 = 0.2108$	$T_5 = 0.2875$	$T_6 = 0.3833$
$\varepsilon_{T_4} = 1.6 \times 10^{-6}$	$\varepsilon_{T_5} = 6.8 \times 10^{-3}$	$\varepsilon_{T_6} = 2.1 \times 10^{-3}$

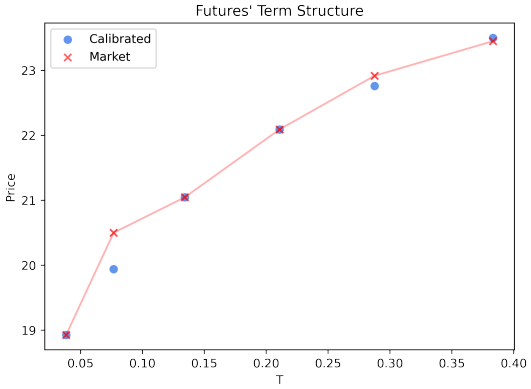


Figure A.2.: The blue circles denote the calibrated futures prices and the red crosses the futures prices on the market, in between a linear interpolation is reported.

Bibliography

E. Abi Jaber, M. Larsson, and S. Pulido. Affine Volterra processes. *Annals of Applied Probability*, 29(5):3155–3200, 2019.

E. Abi Jaber, C. Illand, and S. Li. Joint SPX-VIX calibration with Gaussian polynomial volatility models: deep pricing with quantization hints. *Preprint arXiv:2212.08297*, 2022a.

E. Abi Jaber, C. Illand, and S. Li. The quintic Ornstein-Uhlenbeck volatility model that jointly calibrates SPX & VIX smiles. *Preprint arXiv:2212.10917*, 2022b.

B. Acciaio and I. Penner. Dynamic risk measures. In *Advanced mathematical methods for finance*, pages 1–34. Springer, 2011.

E. Akyildirim, M. Gambara, J. Teichmann, and S. Zhou. Applications of Signature Methods to Market Anomaly Detection. *Preprint arXiv:2201.02441*, 2022.

M. A. Arias-Serna, J.-M. Loubes, and F. J. Caro-Lopera. Risk measures estimation under Wasserstein Barycenter. *arXiv preprint arXiv:2008.05824*, 2020.

P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath. Coherent measures of risk. *Mathematical Finance*, 9(3):203–228, 1999.

J. Baldeaux and A. Badran. Consistent modelling of VIX and equity derivatives using a 3/2 plus jumps model. *Applied Mathematical Finance*, 21(4):299–312, 2014.

C. Bayer, B. Horvath, A. Muguruza, B. Stemper, and M. Tomas. On deep calibration of (rough) stochastic volatility models. *Preprint arXiv:1908.08806*, 2019.

C. Bayer, P. Hager, S. Riedel, and J. Schoenmakers. Optimal stopping with signatures. *Preprint arXiv:2105.00778*, 2021.

J. Belles-Sampera, M. Guillén, and M. Santolino. Beyond value-at-risk: GlueVaR distortion risk measures. *Risk Analysis*, 34(1):121–134, 2014.

J. Bion-Nadal and G. Di Nunno. Fully-dynamic risk-indifference pricing and no-good-deal bounds. *SIAM Journal on Financial Mathematics*, 11(2):620–658, 2020.

C. Birghila and G. Pflug. Optimal XL-insurance under Wasserstein-type ambiguity. *Insurance: Mathematics and Economics*, 88:30–43, 2019.

H. Boedihardjo, X. Geng, T. Lyons, and D. Yang. The signature of a rough path: uniqueness. *Advances in Mathematics*, 293:720–737, 2016.

H. Boedihardjo, J. Diehl, M. Mezzarobba, and H. Ni. The expected signature of Brownian motion stopped on the boundary of a circle has finite radius of convergence. *Bulletin of the London Mathematical Society*, 53(1):285–299, 2021.

A. Bondi, G. Livieri, and S. Pulido. Affine volterra processes with jumps. *Preprint arXiv:2203.06400*, 2022a.

A. Bondi, S. Pulido, and S. Scotti. The rough Hawkes Heston stochastic volatility model. *Preprint arXiv:2210.12393*, 2022b.

O. Bonesini, G. Callegaro, and A. Jacquier. Functional quantization of rough volatility and applications to the VIX. *Available at SSRN 3822933*, 2021.

F. Bourgey and S. De Marco. Multilevel Monte Carlo simulation for VIX options in the rough Bergomi model. *Preprint arXiv:2105.05356*, 2021.

H. Buehler, B. Horvath, T. Lyons, I. Perez Arribas, and B. Wood. A data-driven market simulator for small data environments. *Preprint arXiv:2006.14498*, 2020.

H. Cao, A. Badescu, Z. Cui, and S. K. Jayaraman. Valuation of VIX and target volatility options with affine GARCH models. *Journal of Futures Markets*, 40(12):1880–1917, 2020.

K. T. Chen. Integration of paths, geometric invariants and a generalized Baker-Hausdorff formula. *Annals of Mathematics*, page 163–178, 1957.

K. T. Chen. Iterated path integrals. *Bulletin of the American Mathematical Society*, page 831–879, 1977.

I. Chevyrev and P. K. Friz. Canonical RDEs and general semimartingales as rough paths. *The Annals of Probability*, 47(1):420–463, 2019.

S. N. Cohen, C. Reisinger, and S. Wang. Arbitrage-free neural-sde market models. *Preprint arXiv:2105.11053*, 2021.

R. Cont and S. Ben Hamida. Recovering volatility from option prices by evolutionary optimization. *Journal of Computational Finance*, 8(4):43–76, 2005.

R. Cont and P. Das. Quadratic variation and quadratic roughness. *Bernoulli*, 29(1): 496–522, 2023.

R. Cont and P. Tankov. *Financial modelling with jump processes*, volume 2. Chapman-Hall, 2004.

C. Cuchiero and J. Möller. Signature methods for stochastic portfolio theory. *Working paper*, 2023.

C. Cuchiero and S. Svaluto-Ferro. Infinite-dimensional polynomial processes. *Finance and Stochastics*, 25(2):383–426, 2021.

C. Cuchiero and J. Teichmann. Markovian lifts of positive semidefinite affine Volterra type processes. *Decisions in Economics and Finance*, 42(2):407–448, 2019.

C. Cuchiero and J. Teichmann. Generalized Feller processes and Markovian lifts of stochastic Volterra processes: the affine case. *Journal of Evolution Equations*, pages 1–48, 2020.

C. Cuchiero, M. Keller-Ressel, and J. Teichmann. Polynomial processes and their applications to mathematical finance. *Finance and Stochastics*, 16:711–740, 2012.

C. Cuchiero, W. Khosrawi, and J. Teichmann. A generative adversarial network approach to calibration of local stochastic volatility models. *Risks*, 8(4):101, 2020.

C. Cuchiero, L. Gonon, L. Grigoryeva, J.-P. Ortega, and J. Teichmann. Discrete-Time Signatures and Randomness in Reservoir Computing. *IEEE Transactions on Neural Networks and Learning Systems*, pages 1–10, 2021a.

C. Cuchiero, L. Gonon, L. Grigoryeva, J.-P. Ortega, and J. Teichmann. Expressive power of randomized signature. In *The Symbiosis of Deep Learning and Differential Equations*, 2021b.

C. Cuchiero, F. Guida, L. Di Persio, and S. Svaluto-Ferro. Measure-valued affine and polynomial diffusions. *Preprint arXiv:2112.15129*, 2021c.

C. Cuchiero, G. Gazzani, and I. Klein. Risk measures under model uncertainty: a Bayesian viewpoint. *Preprint arXiv:2204.07115*, 2022a.

C. Cuchiero, G. Gazzani, and S. Svaluto-Ferro. Signature-based models: theory and practice. *Preprint arXiv:2207.13136*, 2022b.

C. Cuchiero, F. Primavera, and S. Svaluto-Ferro. Universal approximation theorems for continuous functions of càdlàg paths and Lévy-type signature models. *Preprint arXiv:2208.02293*, 2022c.

C. Cuchiero, G. Gazzani, J. Möller, and S. Svaluto-Ferro. Joint calibration of SPX and VIX options with signature-based models. *Working paper*, 2023a.

C. Cuchiero, S. Svaluto-Ferro, and J. Teichmann. Signature SDEs from an affine and polynomial perspective. *Working Paper*, 2023b.

F. Delbaen and W. Schachermayer. A general version of the fundamental theorem of asset pricing. *Mathematische Annalen*, 300(1):463–520, 1994.

F. Delbaen and W. Schachermayer. The fundamental theorem of asset pricing for unbounded stochastic processes. *Mathematische Annalen*, 312:215–250, 1999.

G. Di Nunno and E. Rosazza Gianin. Fully-dynamic risk measures: horizon risk, time-consistency, and relations with BSDEs and BSVIEs. *Preprint arXiv:2301.04971*, 2023.

J. Diehl, T. Lyons, R. Preiß, and J. Reizenstein. Areas of areas generate the shuffle algebra. *Preprint arXiv:2002.02338*, 2020.

B. Dupire and V. Tissot-Daguette. Functional Expansions. *Preprint arXiv:2212.13628*, 2022.

K. Ebrahimi-Fard and F. Patras. Cumulants, free cumulants and half-shuffles. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 471(2176): 20140843, 2015.

P. Embrechts, A. McNeil, and D. Straumann. Correlation and dependence in risk management: properties and pitfalls. *Risk management: value at risk and beyond*, 1: 176–223, 2002.

P. Embrechts, G. Puccetti, and L. Rüschendorf. Model uncertainty and VaR aggregation. *Journal of Banking & Finance*, 37(8):2750–2764, 2013.

P. Embrechts, H. Liu, and R. Wang. Quantile-based risk sharing. *Operations Research*, 66(4):936–949, 2018.

D. Escobar and G. Pflug. The distortion principle for insurance pricing: properties, identification and robustness. *Annals of Operations Research*, pages 1–24, 2018.

T. Fadina, Y. Liu, and R. Wang. A framework for measures of risk under uncertainty. *Available at SSRN 3943660*, 2021.

T. Fawcett. Problems in stochastic analysis. Connections between rough paths and non-commutative harmonic analysis. *PhD Thesis, Univ. Oxford*, 2003.

A. Fertis, M. Baes, and H.-J. Lüthi. Robust risk management. *European Journal of Operational Research*, 222(3):663–672, 2012.

D. Filipović and M. Larsson. Polynomial diffusions and applications in finance. *Finance and Stochastics*, 20(4):931–972, 2016.

H. Föllmer and A. Schied. Convex measures of risk and trading constraints. *Finance and Stochastics*, 6(4):429–447, 2002.

H. Föllmer and A. Schied. *Stochastic finance: an introduction in discrete time*. Walter de Gruyter, 2011.

J.-P. Fouque and Y. Saporito. Heston stochastic vol-of-vol model for joint calibration of VIX and S&P 500 options. *Quantitative Finance*, 18(6):1003–1016, 2018.

M. Frittelli and E. R. Gianin. Putting order in risk measures. *Journal of Banking & Finance*, 26(7):1473–1486, 2002.

P. K. Friz and N. B. Victoir. *Multidimensional Stochastic Processes as Rough Paths: Theory and Applications*. Cambridge University Press, 2010.

P. K. Friz, P. Hager, and N. Tapia. Unified signature cumulants and generalized magnus expansions. *Forum of Mathematics, Sigma*, 10, 2022.

J. Gatheral. Consistent modeling of SPX and VIX options. *Bachelier Congress*, 2008.

J. Gatheral. *The volatility surface: a practitioner's guide*. John Wiley & Sons, 2011.

J. Gatheral, T. Jaisson, and M. Rosenbaum. Volatility is rough. *Quantitative Finance*, 18 (6):933–949, 2018.

J. Gatheral, P. Jusselin, and M. Rosenbaum. The quadratic rough Heston model and the joint S&P 500/VIX smile calibration problem. *Preprint arXiv:2001.01789*, 2020.

L. Ghaoui, M. Oks, and F. Oustry. Worst-case value-at-risk and robust portfolio optimization: A conic programming approach. *Operations Research*, 51(4):543–556, 2003.

P. Gierjatowicz, M. Sabate-Vidales, D. Siska, L. Szpruch, and Z. Zuric. Robust pricing and hedging via neural SDEs. *Available at SSRN 3646241*, 2020.

P. Glasserman. *Monte Carlo methods in financial engineering*, volume 53. Springer, 2004.

L. A. Grzelak. On Randomization of Affine Diffusion Processes with Application to Pricing of Options on VIX and S&P 500. *Preprint arXiv:2208.12518*, 2022.

H. Guerreiro and J. Guerra. VIX pricing in the rBergomi model under a regime switching change of measure. *Preprint arXiv:2201.10391*, 2022.

I. Guo, G. Loeper, J. Obloj, and S. Wang. Optimal transport for model calibration. *Preprint arXiv:2107.01978*, 2021.

I. Guo, G. Loeper, J. Obloj, and S. Wang. Joint Modeling and Calibration of SPX and VIX by Optimal Transport. *SIAM Journal on Financial Mathematics*, 13(1):1–31, 2022.

J. Guyon. Inversion of convex ordering in the VIX market. *Quantitative Finance*, 20(10): 1597–1623, 2020a.

J. Guyon. The joint S&P 500/VIX smile calibration puzzle solved. *Risk, April*, 2020b.

J. Guyon. Dispersion-Constrained Martingale Schrödinger Bridges: Joint Entropic Calibration of Stochastic Volatility Models to S&P 500 and VIX Smiles. *Available at SSRN 4165057*, 2021.

J. Guyon and F. Bourgey. Fast Exact Joint S&P 500/VIX Smile Calibration in Discrete and Continuous Time. *Available at SSRN 4315084*, 2022.

J. Guyon and P. Henry-Labordere. *Nonlinear option pricing*. CRC Press, 2013.

J. Guyon and J. Lekeufack. Volatility Is (Mostly) Path-Dependent. *Available at SSRN 4174589*, 2022.

J. Guyon and S. Mustapha. Neural Joint S&P 500/VIX Smile Calibration. *Available at SSRN 4309576*, 2022.

P. S. Hagan, D. Kumar, A. S. Lesniewski, and D. Woodward. Managing smile risk. *The Best of Wilmott*, 1:249–296, 2002.

P. Halmos and L. Savage. Application of the Radon-Nikodym theorem to the theory of sufficient statistics. *Annals of Mathematical Statistics*, 20(2):225–241, 1949.

P. Henry-Labordere. From (Martingale) Schrödinger bridges to a new class of Stochastic Volatility Models. *Preprint arXiv:1904.04554*, 2019.

S. L. Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The Review of Financial studies*, 6(2):327–343, 1993.

J. Jacod and P. Protter. *Discretization of processes*, volume 67. Springer Science & Business Media, 2011.

A. Jacquier, A. Muguruza, and A. Pannier. Rough multifactor volatility for SPX and VIX options. *Preprint arXiv:2112.14310*, 2021.

S. Jaimungal, S. Pesenti, Y. S. Wang, and H. Tatsat. Robust risk-aware reinforcement learning. *SIAM Journal on Financial Mathematics*, 13(1):213–226, 2022.

E. Jouini, W. Schachermayer, and N. Touzi. Law invariant risk measures have the Fatou property. In *Advances in Mathematical Economics*, pages 49–71. Springer, 2006.

Y. M. Kabanov. On the FTAP of Kreps–Delbaen–Schachermayer. *Proceedings of Steklov Mathematical Institute Seminar*, pages 191—-203, 1997.

J. Kalsi, T. Lyons, and P.-A. I. Optimal execution with rough path signatures. *SIAM Journal on Financial Mathematics*, 11(2):470–493, 2020.

C. Kardaras, D. Kreher, and A. Nikeghbali. Strict local martingales and bubbles. *The Annals of Applied Probability*, 25(4):1827–1867, 2015.

P. Kidger and T. Lyons. Signatory: differentiable computations of the signature and logsignature transforms, on both CPU and GPU. In *International Conference on Learning Representations*, 2020.

F. J. Király and H. Oberhauser. Kernels for sequentially ordered data. *Journal of Machine Learning Research*, 20(31):1–45, 2019.

P. E. Kloeden and E. Platen. Stochastic differential equations. In *Numerical solution of stochastic differential equations*, pages 103–160. Springer, 1992.

T. Kokholm and M. Stisen. Joint pricing of VIX and SPX options with stochastic volatility and jump models. *The Journal of Risk Finance*, 2015.

L. Le Cam. *Asymptotic methods in statistical decision theory*. Springer Science & Business Media, 2012.

D. Levin, T. Lyons, and H. Ni. Learning from the past, predicting the statistics for the future, learning an evolving system. *Preprint arXiv:1309.0260*, 2013.

A. Lipton. The vol smile problem. *Risk Magazine*, 15(2):61–66, 2002.

C. Litterer and H. Oberhauser. On a Chen–Fliess approximation for diffusion functionals. *Monatshefte für Mathematik*, 175(4):577–593, 2014.

F. Liu, T. Mao, R. Wang, and L. Wei. Inf-convolution, Optimal Allocations, and Model Uncertainty for Tail Risk Measures. *Mathematics of Operations Research, forthcoming*, 2022.

P. Liu, A. Schied, and R. Wang. Distributional transforms, probability distortions, and their applications. *Mathematics of Operations Research*, 46(4):1490–1512, 2021.

H. Luschgy and G. Pagès. Functional quantization of Gaussian processes. *Journal of Functional Analysis*, 196(2):486–531, 2002.

T. Lyons. Differential equations driven by rough signals. *Revista Matemática Iberoamericana*, 14(2):215–310, 1998.

T. Lyons and H. Ni. Expected signature of Brownian motion up to the first exit time from a bounded domain. *The Annals of Probability*, 43(5):2729–2762, 2015.

T. Lyons and N. Victoir. Cubature on Wiener space. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 460(2041):169–198, 2004.

T. Lyons, M. Caruana, and T. Lévy. *Differential equations driven by rough paths*. Springer, 2007.

T. Lyons, S. Nejad, and I. Perez Arribas. Non-parametric Pricing and Hedging of Exotic Derivatives. *Applied Mathematical Finance*, 27(6):457–494, 2020.

M. Min and R. Hu. Signatured deep fictitious play for mean field games with common noise. *Preprint arXiv:2106.03272*, 2021.

H. Ni. The expected signature of a stochastic process. *Ph.D. thesis, University of Oxford*, 2012.

H. Ni, L. Szpruch, M. Sabate-Vidales, B. Xiao, M. Wiese, and S. Liao. Sig-Wasserstein GANs for Time Series Generation. *Preprint arXiv:2111.01207*, 2021.

J. Obłój and J. Wiesel. Robust estimation of superhedging prices. *The Annals of Statistics*, 49(1):508–530, 2021.

C. Pacati, G. Pompa, and R. Renò. Smiling twice: the Heston++ model. *Journal of Banking & Finance*, 96:185–206, 2018.

G. Pagès and J. Printems. Functional quantization for numerics with an application to option pricing. *Monte Carlo Methods and Applications*, 4(11):407–446, 2005.

A. Papanicolaou and R. Sircar. A regime-switching Heston model for VIX and S&P 500 implied volatilities. *Quantitative Finance*, 14(10):1811–1827, 2014.

Z. Pei, Y. Wang, X.and Xu, and X. Yue. A worst-case risk measure by G-VaR. *Acta Mathematicae Applicatae Sinica, English Series*, 37(2):421–440, 2021.

S. Peng, S. Yang, and J. Yao. Improving Value-at-Risk Prediction Under Model Uncertainty. *Journal of Financial Econometrics*, 2020.

I. Perez Arribas. *Signatures in machine learning and finance*. PhD thesis, University of Oxford, 2020.

I. Perez Arribas, C. Salvi, and L. Szpruch. Sig-SDEs model for quantitative finance. In *Proceedings of the First ACM International Conference on AI in Finance*, pages 1–8, 2020.

G. Pflug and A. Pichler. Time-consistent decisions and temporal decomposition of coherent risk functionals. *Mathematics of Operations Research*, 41(2):682–699, 2016.

A. Quarteroni, R. Sacco, and F. Saleri. *Numerical mathematics*, volume 37. Springer Science & Business Media, 2010.

R. Ree. Lie elements and an algebra associated with shuffles. *Annals of Mathematics*, pages 210–220, 1958.

J. Reizenstein and B. Graham. Algorithm 1004: The iisignature library: Efficient calculation of iterated-integral signatures and log signatures. *ACM Transactions on Mathematical Software (TOMS)*, 46(1):1–21, 2020.

Y. Ren, D. Madan, and M. Q. Qian. Calibrating and pricing with embedded local volatility models. *Risk Magazine*, 20(9):138, 2007.

R. Rhoads. *Trading VIX derivatives: trading and hedging strategies using VIX futures, options, and exchange-traded notes*, volume 503. John Wiley & Sons, 2011.

B. Ricceri and S. Simons. *Minimax theory and applications*, volume 26. Springer Science & Business Media, 2013.

L. Rogers. Things we think we know. In *Options — 45 Years since the Publication of the Black–Scholes–Merton Model*, chapter 9, pages 173–184.

S. Rømer. Empirical analysis of rough and classical stochastic volatility models to the SPX and VIX markets. *Quantitative Finance*, pages 1–34, 2022.

M. Rosenbaum and J. Zhang. Deep calibration of the quadratic rough Heston model. *Preprint arXiv:2107.01611*, 2021.

E. Rossi Ferrucci and T. Cass. On the Wiener Chaos Expansion of the Signature of a Gaussian Process. *Preprint arXiv:2207.08422*, 2022.

A. Sepp. VIX option pricing in a jump-diffusion model. *Risk*, pages 84–89, 2008.

A. Sepp. Achieving consistent modeling of VIX and equities derivatives. *Global Derivatives Conf. Barcelona*, 2012.

M. Sion. On general minimax theorems. *Pacific Journal of mathematics*, 8(1):171–176, 1958.

E. M. Stein and J. C. Stein. Stock price distributions with stochastic volatility: an analytic approach. *The review of financial studies*, 4(4):727–752, 1991.

V. Tissot-Daguette. Projection of Functionals and Fast Pricing of Exotic Options. *SIAM Journal on Financial Mathematics*, 13(2):SC74–SC86, 2022.

A. Van der Vaart. *Asymptotic statistics*, volume 3. Cambridge University Press, 2000.

R. Wang and J. Ziegel. Scenario-based risk evaluation. *Available at SSRN 3235450*, 2018.

R. Wang, Y. Wei, and G. E. Willmot. Characterization, robustness, and aggregation of signed Choquet integrals. *Mathematics of Operations Research*, 45(3):993–1015, 2020.

W. Wang and H. Xu. Robust spectral risk optimization when information on risk spectrum is incomplete. *SIAM Journal on Optimization*, 30(4):3198–3229, 2020.

C. Zalinescu. *Convex analysis in general vector spaces*. World Scientific, 2002.

S. Zhu and M. Fukushima. Worst-case conditional value-at-risk with application to robust portfolio management. *Operations Research*, 57(5):1155–1168, 2009.

S. Zymler, K. D., and R. B. Worst-case value at risk of nonlinear portfolios. *Management Science*, 59(1):172–188, 2013.