Zeliade Systems

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

Zeliade Systems

Title:	A closed form model-free approximation for the Initial Margin of option portfolios
Author:	Zeliade Systems
Number of pages:	50
Date:	2023-06-22
Version:	1.0

Contents

Abstract

Central clearing counterparty houses (CCPs) play a fundamental role in mitigating the counterparty risk for exchange traded options. CCPs cover for possible losses during the liquidation of a defaulting member's portfolio by collecting initial margins from their members. In this article we analyze the current state of the art in the industry for computing initial margins for options, whose core component is generally based on a VaR or Expected Shortfall risk measure. We derive an approximation formula for the VaR at short horizons in a model-free setting. This innovating formula has promising features and behaves in a much more satisfactory way than the classical Filtered Historical Simulation-based VaR in our numerical experiments. In addition, we consider the neural-SDE model for normalized call prices proposed by [Cohen et al., arXiv:2202.07148, 2022] and obtain a quasi-explicit formula for the VaR and a closed formula for the short term VaR in this model, due to its conditional affine structure.

1. Introduction

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

We concentrate on the initial margin, which is supposed to cover for the losses incurred in case of the liquidation of a given portfolio in normal market conditions. Article 41 of [9] requires CCPs to collect margins from the parties entering a transaction in a measure to be sufficient to cover the CCP potential exposures while liquidating the position. The margin must also be sufficient to cover at least 99.5% of these exposures in the case of OTC derivatives, and 99% for other financial products over the Margin Period Of Risk (MPOR), as recommended in article 24 of [10].

Since the drafting of EMIR regulation, CCPs have put in practice different ways to compute margins for option portfolios. A first notorious methodology for complex portfolios is the SPAN algorithm of CME Group, which simulates joint risk scenarios for the underlier and the implied volatility and infers a conservative margin from these scenarios. However, this methodology has been overcome by more refined ones, which in most cases apply a Filtered Historical Simulation (FHS)-type algorithm [2] to selected risk factors in order to generate scenarios consistent with historical moves (examples are the SPAN2 by CME and the IRM2 by ICE). FHS is widely used among CCPs but its use on option markets is tricky and questionable. In particular, a straightforward use of FHS breaks the structural relationships between risk factors, possibly generating highly implausible scenarios.

Different techniques rather than FHS for options margining have been studied in theory and eventually implemented, as the procyclicality control model by Wong and Zhang from OCC [17], which relies on a dynamic scaling factor adjusting the dynamics of the ATM IV log returns to be higher during low-volatility periods and lower during high-volatility periods. More academic papers such as [7, 11] also look at the issue of computing option initial margins, additionally ensuring the absence of arbitrage for the generated scenarios. Indeed, in [11] the authors describe a generic algorithm which penalizes arbitrageable scenarios (in a static sense) which can be simply upgraded to any scenarios generation algorithm already in production. In particular, the authors apply it to Generative Adversarial Networks to simulate arbitrage-free implied volatility surfaces. In [7], an affine factor model for normalized call option prices is firstly defined and then calibrated minimizing dynamic and static arbitrages. Scenarios are subsequently generated by neural networks which constrain the paths to live in the polytope defined by the no static arbitrage conditions. In this panorama, it is worth including the works by Bergeron at al. [4, 16] on the Variational Autoencoders used to reconstruct missing data on implied volatility surface (eventually with no arbitrage), and which can be tweaked to simulate scenarios based on historical movements.

In the present work we have two main objectives: the first one is to provide a practical and concrete panorama in options margining; the second, more ambitious, is to design a closed formula for the VaR of option portfolios, which is easy to understand and to implement. Specifically, we compute a short-term VaR formula which is completely model-free and coincide with the exact one in the Stochastic Volatility model and the particular affine factor model for normalized call prices proposed in [7] as the neural-SDE model. For the latter model, we show that it is actually possible to directly infer the VaR formula without any need of simulating scenarios, so that once the parameters of the model are calibrated, these can be plugged into a quasi-explicit formula to obtain the required margin. Also, considering the limit for small time steps, the formula becomes closed and it has the same form of our short-term model-free formula. Testing the short-term model-free formula, we obtain well-behaved margins which actually beat the classic FHS ones in terms of regularity and adaptation to the market current behavior. For these reasons, we believe that the suggested short-term model-free formula could lay the foundations to a practical model-free formula for options margining.

In the first part of this paper, we look at the mechanism of options' initial margin adopted by CCPs in section 2. In section 3 we go into detail in the practical implementations used by CCPs to calculate initial margins, followed by an assessment of their pros and cons. In the second part of the paper, we firstly describe the short-term model-free formula in section 4 and secondly derive the closed margin formula in the neural-SDE model for normalized option prices in section 5. We conclude by performing numerical experiments in section 6.

We thank Zeliade Systems and in particular Ismail Laachir for his detailed explanations on how the margining methodology works in CCPs, and Pierre Cohort and Niels Escarfail for being always available for helping with coding. A particular thank goes to Nicolò Filippas who, as an intern at Zeliade, has firstly studied and explained Cohen et al. papers to the team, hinting us to deepen our investigation of the model.

We also thank Stefano De Marco who especially helped in the theoretical adjustment of proofs and in the overall structuring of the article. All remaining errors are ours.

2. The mechanism of initial margin for options

CCPs charge clearing members, on a daily or intra-daily basis, with total risk requirements that are computed from initial margins. The initial margin aims at covering possible losses in the portfolio value when liquidating it after a default, under normal market conditions, and it is estimated considering a tail risk.

Consider a portfolio, at time t, with possibly both long $(L_i)_i$ and short $(S_j)_j$ option positions (both L_i and S_j are positive) with different strikes and expiries. In the case of default at time t, the CCP has to liquidate the portfolio in a Margin Period of Risk (MPOR) of say h days (h is usually 2 days for exchange-traded options). At date $t+h$, the portfolio could have undergone market movements, so that the CCP has to estimate its payoff after liquidation.

The initial margin (IM) is then the Value-at-Risk (VaR) or Expected-Shortfall (ES) at a confidence level of generally 0.99 of the portfolio predicted P&Ls:

$$\mathrm{IM}(t) = -\mathrm{VaR}_{0.99}\big(\mathrm{P\&L}(t+h)\big)$$

where the minus sign ensures a positive margin value.

At this point, the total risk requirement charged by the CCP does not solely include the initial margin. Indeed, the CCP eventually adds to the latter quantity some add-ons to take into account risks that are not directly related to market moves. Among these, we typically find the Wrong Way Risk add-on, the liquidity and concentration risk add-on and possibly other specific add-ons:

$$\text{Total margin}(t) = \mathrm{IM}(t) + \text{Add-ons}(t).$$

Now, the total margin is floored by the Short Option Minimum (SOM). Deep short OTM positions have very little risk since they will probably stay OTM along the MPOR. However, their extreme risk is still not 0 and the methodology should be able to capture it. This is generally not the case for strikes very far from the ATM, because of the lack of historical liquid data on these strikes. Then, to assure an enough conservative margin, a secure floor should be applied to the risk requirement. The SOM is generally the sum along all short positions of the calibrated extreme costs for these options:

$$\text{Refined total margin}(t) = \max\big(\text{Total margin}(t);\ \mathrm{SOM}(t)\big).$$

At this point, the final total risk requirement is the refined total margin adjusted by two other terms: minus the Net Option Value (NOV) on equity-style options (options for which the premium is paid in full at the settlement date, i.e. one or two days after the option trade) and the Unpaid Premium (UP), described below:

$$\text{Total risk requirement}(t) = \max\big(\text{Refined total margin}(t) - \mathrm{NOV}(t) + \mathrm{UP}(t);\ 0\big)$$

where the floor by zero is to avoid that the CCP pays and

$$\begin{aligned}\mathrm{NOV}(t) &= \sum_i L_i O_i(t) - \sum_j S_j O_j(t)\\ \mathrm{UP}(t) &= \sum_{i \text{ unpaid}} L_i O_i(t) - \sum_{j \text{ undelivered}} S_j O_j(t)\end{aligned}$$

with O denoting the option prices.

For the NOV, let us consider for instance the case of a short option position, where the option is of equity-style. In the case of a default, the liquidation of this position would require to buy the option in the market, which amounts to the CCP paying the option price at the time of liquidation. This means that the initial margin for a short option position should aim to cover the largest option price, up to a fixed confidence level. On the other hand, the liquidation of a long option position will always result in a positive inflow for the CCP, because the CCP will sell the long option position and receive the option price.

The reason why the CCP applies the NOV can alternatively be explained observing that the liquidation at the end of the MPOR, at time $t+h$, will result for the CCP in the monetary flow:

$$\text{Liquidation P\&L}(t+h) = \sum_i L_i O_i(t+h) - \sum_j S_j O_j(t+h).$$

The Liquidation P&L can be expressed as the sum of the NOV and the portfolio's value increment:

$$\begin{aligned}\text{Liquidation P\&L}(t+h) &= \Big(\sum_i L_i O_i(t) - \sum_j S_j O_j(t)\Big) +\\ &\quad + \Big(\sum_i L_i\big(O_i(t+h) - O_i(t)\big) - \sum_j S_j\big(O_j(t+h) - O_j(t)\big)\Big)\\ &= \mathrm{NOV}(t) + \text{P\&L}(t+h).\end{aligned}$$

Then, the CCP has to charge to the clearing member minus the liquidation profit, i.e. the predicted losses (the initial margin appropriately adjusted by the add-ons and the SOM) minus the NOV on equity-style options.

The UP is charged by the CCP to cover from the risk of default of the counterparts before the settlement date of the option premium, and it corresponds to the net position of accrued option premiums which are still unpaid (because the settlement date has not passed yet). In this way, the difference between the UP and the NOV can be seen as a (Contingent) Variation Margin for Options (VMO) not yet settled.

Consider a defaulting clearing member which is long an option before the settlement date. The CCP will need to pay the premium to its counterpart in the trade, and this will be done re-selling the option and collecting its new premium, and using the VMO previously required to the buyer. This latter component is needed to account for the difference between the initially established option premium and today's one. Similarly, the CCP has to liquidate defaulting short positions on not yet settled options buying the option and delivering it to the buyer counterpart. To do so, it will use the money from the buyer plus the VMO from the defaulting seller.

All in all, the final formula for the total risk requirement is

$$\text{Total risk requirement}(t) = \max\big(\max\big(\mathrm{IM}(t) + \text{Add-ons}(t);\ \mathrm{SOM}(t)\big) - \mathrm{NOV}(t) + \mathrm{UP}(t);\ 0\big).$$

In this article we will focus on the IM component of the total risk requirement.

3. Initial margin for options in the industry: a short survey

The total risk requirement mechanism and its different layers is essentially the same across all CCPs, with possible differences in wording. What really makes the difference among CCPs' requirements is the way the IMs (and the add-ons) are computed. A notorious parametric model for margining has been proposed by CME Group under the name of SPAN.[1] It consists in computing the P&Ls of the portfolio under different risk scenarios depending on the combination of underlying price changes and implied volatility changes. A similar model has been implemented by ICE with the name of IRM. These models are particularly tricky and overconservative, and for these reasons nowadays CCPs are passing to new models. In particular, both SPAN and IRM models have been upgraded to the corresponding SPAN2[2] and IRM2[3] models, which both use the Filtered Historical Simulation (FHS) techniques to create risk scenarios. Indeed, the majority of CCPs is now adopting the FHS to compute IMs for option portfolios.

3.1 Filtered Historical Simulation

The FHS has recently become the standard approach for VaR computations among CCPs, especially on cash equity markets. The FHS technique is indeed particularly efficient in cash equity and fixed income markets for spot instruments, but it becomes more subtle in derivatives clearing.

The FHS model is particularly appreciated since it is essentially data-driven and model-free, and it relies on few requirements to be satisfied. For a given instrument to be cleared, firstly the CCP must choose the risk factors which drive its price; let r_s denote their returns, either logarithmic, absolute, or relative depending on the risk factor. A key property that scaled returns must satisfy is stationarity (see [1]). Indeed, the FHS model relies on the hypothesis that risk factors' returns at tomorrow's date $t+1$ behave as

$$r_{t+1} = \eta \sigma_{t+1}$$

where σ_{t+1} is the returns' simulated conditional volatility at day $t+1$ and η is drawn from the historical observations

$$\eta_s = \frac{r_s}{\sigma_s}.$$

In other words, the past historical return is *re-contextualized* to the current volatility context by the FHS devoling/revoling steps.

Generally, the industry standard is to use an Exponentially Weighted Moving Average (EWMA) variance estimator for the volatility. A EWMA volatility with decay factor λ is computed as

$$\mathrm{EWMA}_s = \sqrt{(1-\lambda)(r_s)^2 + \lambda \mathrm{EWMA}_{s-1}^2},$$

[1] https://www.cmegroup.com/clearing/risk-management/span-overview.html
[2] https://www.cmegroup.com/clearing/risk-management/span-overview/span-2-methodology.html
[3] https://www.theice.com/clearing/margin-models/irm-2/methodology

with an eventual flooring in case of too low values. Then, the historical volatility σ_s used to scale historical returns can be calculated with two possible formulations: $\sigma_s = \text{EWMA}_s$ and $\sigma_s = \text{EWMA}_{s-1}$ respectively. The two alternatives are discussed in [14], section 7.1, where it is acknowledged that they will lead to significantly different outcomes.

When computing the IM for portfolios of options using the FHS methodology, the CCP has to choose a set of risk factors, assessing the stationarity property in particular. Together with the underlying value (and possibly the interest/repo rate), also the Implied Volatility (IV) has to be taken as a risk factor. Since the IV is actually a surface which behaves differently depending on the strike and the maturity of the option, two alternatives can be considered in order to generate IV scenarios:

1. Identify a fixed two-dimensional grid for the IV surface and define each point as a risk factor.
2. Choose a model for option prices and take its parameters as risk factors.

For deeper insights on the VaR computation for options in a FHS approach see [13].

3.1.1 Implied Volatility anchor points

In the first alternative, the anchor points on the grid can be chosen with fixed time-to-maturity or fixed rolling index as first coordinate, and fixed log-forward moneyness, or fixed delta, or (equivalently) fixed ratio between log-forward moneyness and square-root of time-to-maturity as second coordinate. Since market data is more dense around the ATM point for shortest maturities and spreads out for increasing maturities, the fixed delta grid is generally preferred. Indeed, it implies a grid in log-forward moneyness with a triangular shape, with a range that starts from the ATM point and spreads out as the time-to-maturity increases.

If choosing a dense grid guarantees a more precise fit of IVs in between the anchor points, it highly decreases computation performances and makes it difficult to identify general historical patterns in the IV surface dynamics. For this reason, Principal Component Analysis (PCA) can be performed in order to model the shifts of the surface. To cite an example, in [18], the authors test the FHS method on the (PCA) principal components in a Karhunen-Loève decomposition and find that scenarios satisfy the conditions of no butterfly arbitrage (i.e. the requirement that for a fixed maturity, call prices must be non-increasing and convex with respect to the strike).

Once scenarios are generated on the anchor points, the model still needs an interpolation/extrapolation criterion to predict future prices on points outside the grid. The criterion could be either a model for the implied volatility (such as SVI) or for prices (such as SABR), which has to be calibrated from the scenarios grid, or classic interpolations via b-splines. The choice can be driven by arguments of non-arbitrability of prices, or of best fit and computation efficiency of the algorithm.

3.1.2 Implied Volatility models

In the second alternative, the CCP chooses a pricing model for options and, once the stationarity property on the model parameters' returns is verified, performs an FHS on the model parameters.

As an example, the SABR model is an industry standard and it is driven by three parameters α, β and ρ. Generally, the β parameter is fixed a priori based on historical observations, so that only the α and ρ

parameters need to be estimated. After showing the stationarity of their returns in the target market, the CCP can apply the FHS technique on the historical observations of α and ρ, and use their drawn values to simulate future prices.

Similarly, the Stochastic Volatility Inspired (SVI) model by Gatheral is largely used among CCPs, and also among crypto funds, to model the implied total variance. Its sub-model Slices SVI (SSVI) is sometimes preferred since it has more tractable arbitrage-free requirements and since it still fits data pretty well. SSVI has three parameters θ, φ and ρ per each maturity, so that if the stationarity of their returns is verified, the FHS technique can be applied to obtain simulated prices. An example of this application can be found in [13].

Lastly, a model which is sometimes considered is the Gaussian lognormal mixture model as described by Glasserman and Pirjol in [12]. It consists in a convex combination of Black-Scholes prices and the number of parameters depends on chosen number of basis prices. Even though it is easy to implement, it guarantees no arbitrage for slices and it has very good fitting ability, the model is not easy to extend to full surfaces and it hides potential issues when extrapolating in extreme events. Indeed, one it has the theoretical property to have the same constant asymptotic level in the two wings of the smile (Proposition 5.1 of [12]), so that while the calibration of market smiles could suggest a decreasing shape, the calibrated smile with a lognormal Gaussian mixtures would necessarily increase at large strikes, a pure model artefact. As a consequence, while calibration fit could be good for liquid market data (concentrated around the ATM point), in contexts such as the computation of tail risks as in margins, the extrapolation at extreme strikes would be misleading in those circumstances.

3.1.3 Limitations of the FHS

The FHS technique works well when a large history of risk factors is stored, which is in fact a first immediate practical limitation. Indeed, the possible number of scenarios for the FHS cannot be larger than the available history, since the normalized returns are drawn from past observations.

A second important drawback of the FHS methodology for complex products is the capture of the joint dynamics of risk factors. Indeed, in the FHS model, risk factors are re-scaled according to their own intrinsic volatility, without any reference to other risk factors and this may cause a discrepancy in the relationships between the risk factors, and in particular in their correlations, as explained in section 7.2 of [14]. Furthermore, while the property of stationarity of the normalized returns of single assets is generally historically satisfied, this is more hardly the case for the returns of IV points, resulting in more unstable and unnatural results for the FHS methodology.

Thirdly, FHS is relatively straightforward to implement, as far as the risk factors under study do not have structural relationships which could be destroyed by the core FHS algorithm. Unfortunately, this is exactly the case for futures' curves and implied volatility surfaces.

Indeed, in the case of futures' curves, the FHS simulation considers a set of fixed pillars (i.e. futures' time-to-maturities) of the curves today and apply the re-scaled corresponding past returns. For each simulation, the simulated vector of futures values on the fixed pillars should be consistent between the spot returns and the future returns. However, this consistency is not guaranteed by the FHS simulations.

Similarly, when using the IV anchor points as risk factors, even if the calibrated volatility surfaces are perfectly calibrated and arbitrage-free, the volatility surfaces obtained by an FHS procedure have no reason

to be arbitrage-free in turn (and in general will not be, because arbitrage-free surfaces do not have nice additive or multiplicative properties). Furthermore, IV anchor points returns are generally considered in absolute terms, which could cause negative simulated implied volatilities. Flooring the latter quantities to 0 is not a good choice, since: 1) prices for zero volatility are always strictly lower than the market prices for European options; 2) since call option prices are decreasing functions of the strike, a zero volatility for an OTM call implies that all the calls with the same maturity and larger strikes should also have a zero volatility, so that also simulated implied volatility smiles should satisfy this property.

The possibility of generating scenarios such that each matrix of prices indexed by the moneyness and time-to-maturity grid is arbitrage-free is essentially an open question. A recent article [11] describes a weighted Monte-Carlo algorithm which penalizes arbitrageable scenarios to obtain arbitrage-free simulations with higher probability. We explain the model in section 3.3. Another alternative is to use parametric models of IV surfaces, for which no-arbitrage conditions are available, and work at the level of the parameters of such models. Yet, randomizing the model parameters may produce a lot of instability. A noteworthy attempt is the neural-SDE model of [7], that we investigate further below, which provides a consistent framework for this purpose.

Finally, the FHS methodology is known to be procyclical as shown for example in section 6 of [14]. Procyclicality has to be avoided because it implies margins which react too abruptly to market changes, and this may cause liquidity issues to the clearing members who have to post the corresponding collateral.

3.2 The procyclicality control by Wong and Zhang (Options Clearing Corporation)

Even though the FHS model is the most popular among CCPS, in the recent years some new models for the options clearing are born and CCPs are starting to look at these alternatives. An important feature in margin requirements that CCPs should always try to mitigate is procyclicality and we have seen that FHS does not properly satisfy this requirement. The EMIR Regulatory Technical Standards of 2013 dedicates Article 28 to the procyclicality control, detailing specific actions that CCPs have to adopt for its limitation.

With this in mind, Wong and Zhang from the Options Clearing Corporation (OCC) choose a model for options initial margin (see [17]) that guarantees to control procyclicality thanks to a dynamic scaling factor that behaves as an inverted S-curve and is higher during low-volatility periods and lower during high-volatility ones.

The model specifies the log-returns of the ATM IV at expiry T_j by

$$\log \frac{\sigma_{t+h}(T_j, F_t(T_j))}{\sigma_t(T_j, F_t(T_j))} := \eta_t \Big(\frac{T_j}{T_1}\Big)^{-\alpha} \sqrt{h} z_t \tag{1}$$

where z_t is a normalized innovation, centered with unit variance, $F_t(T_j)$ is today's forward for maturity T_j, and T_1 is the first quoted expiry. The factor η_t in turn is a dynamic rescaling of the CBOE VVIXSM (VVIX), in particular

$$\eta_t = D(\sigma_t)\mathrm{VVIX}_t$$

where σ_t is the S&P500 ATM IV of the short-term expiry (or any reference expiry, like the one-month), and

the scaling factor $D(\sigma_t)$ is a sigmoid function, which models a state transition from a risk point of view:

$$D(\sigma_t) = L + \frac{H}{1 + \exp\left(\kappa(\sigma_t - \sigma^*)\right)}.$$

Here L is the minimum of the ratio between the long-term mean of the historical vol-of-vol and the VVIX, H is the difference between the maximum and the minimum of the latter ratio, κ is the growth rate of the curve, and σ^* is the sigmoid's midpoint.

The IV surface is recovered from the ATM IV dynamics considering the second order approximation in $\log \frac{K}{F_t(T)}$:

$$\sigma_t(T,K) \approx \sigma_t(T, F_t(T)) + \Sigma_t(T) \log \frac{K}{F_t(T)} + C_t(T)\Big(\log \frac{K}{F_t(T)}\Big)^2,$$

where $\Sigma_t(T)$ and $C_t(T)$ are respectively the ATM skew and the ATM curvature.

In this way, knowing the distribution of z_t allows to perform simulations of implied volatility surfaces and to compute an empirical VaR.

Observe that the dynamics of the implied volatility in eq. (1) are modeled for fixed strike and expiry, i.e. for a fixed contract. This differs with the majority of other models, whose dynamics are defined for fixed time-to-maturity and log-forward moneyness.

3.3 Arbitrage-free simulations for options

When computing an IM, the priority of the CCP is to be conservative enough to cover for members' defaults, while not requiring too high margins to keep its competitiveness in the market and avoiding procyclicality. For this reason, arbitrage-free requirements are not necessarily taken into account as seen for the FHS methodology. However, simulating reliable scenarios (and so scenarios with no arbitrage) allows to estimate more plausible margins, and avoids the pitfall of paying for implausible scenarios.

The article [11] describes a cunning way to compute an empirical VaR tweaking the simulations from any model in favor of arbitrage-free simulations. The arbitrage considered in the article is the static arbitrage, that, in case of options, can arise in both the direction of time-to-maturity and the direction of log-forward moneyness. Arbitrage-free call prices should:

1. lie between the discounted intrinsic value (computed with respect to the forward) and the discounted forward;
2. increase in time-to-maturity at a fixed moneyness;
3. decrease in log-forward moneyness at a fixed time-to-maturity;
4. be a convex function of the log-forward moneyness.

Note that in the article, the authors only address the last three points, but the methodology can be easily extended to include the first one.

Per each arbitrage situation, a penalization function is defined, depending only on the normalized call prices surface on a fixed time-to-maturity and log-forward moneyness discrete grid. Penalization functions are null in case of no arbitrage and increase their value with increasing arbitrageable grid points. The target

arbitrage penalty function is the sum of the three penalization functions, and it is null if and only if the discrete call prices are free of arbitrage.

At this point, the VaR calculation algorithm is straightforward:

1. Simulate scenarios using the chosen initial model.
2. Per each simulated scenario:
   - evaluate the arbitrage penalty function;
   - compute its weight inversely proportional to the arbitrage penalty function.
3. Compute empirical VaR under the probability measure resulting from weights.

Since weights prioritize arbitrage-free scenarios, the VaR calculation hangs towards more reliable and possible values.

The methodology holds for any model that simulates scenarios. It can then be applied to both FHS and Monte-Carlo simulation models. In particular, the authors apply it to a non-parametric generative model for implied volatility surfaces called VolGAN (section 6 of [11]).

3.4 The neural-SDE model

In [8], Cohen at al. show very good empirical results on options' VaR estimation. The results are based on a specific model that the authors introduce in [7], which consists in a representation of normalized call prices via non-random linear functions of some risk factors ξ_t. The articles focus on how to calibrate and consequently generate arbitrage-free call prices surfaces via neural networks for the dynamics under consideration.

In the neural-SDE model, the normalized call prices (i.e. call prices divided by the forward and discount factor) are affinely decomposed into time-independent non-random surfaces G_i and time-dependent stochastic combining factors $\xi_{t,i} \in \mathbb{R}^d$:

$$\begin{aligned} c_t(\tau, k) &= G_0(\tau, k) + G(\tau, k) \cdot \xi_t \\ &= G_0(\tau, k) + \sum_{i=1}^{d} G_i(\tau, k)\xi_{t,i} \end{aligned} \tag{2}$$

where τ is the time-to-maturity and k is the log-forward moneyness. The underlying asset S_t and the time-dependent vector ξ_t evolve as

$$\begin{aligned} dS_t &= \alpha(\xi_t)S_t dt + \beta(\xi_t)S_t dW_{0,t} \qquad S_0 = s_0 \in \mathbb{R}, \\ d\xi_t &= \mu(\xi_t)dt + \sigma(\xi_t) \cdot dW_t \qquad\quad \xi_0 = \zeta_0 \in \mathbb{R}^d, \end{aligned} \tag{3}$$

where $W_0 \in \mathbb{R}$, $W = (W_1, \dots, W_d)^T \in \mathbb{R}^d$ are independent standard Brownian motions under real-world measure $\mathbb{P}$, and the hypothesis for the existence and uniqueness of the processes hold, i.e. $\alpha(\xi_t) \in L^1_{\text{loc}}(\mathbb{R})$, $\mu(\xi_t) \in L^1_{\text{loc}}(\mathbb{R}^d)$, $\beta(\xi_t) \in L^2_{\text{loc}}(\mathbb{R})$, $\sigma(\xi_t) \in L^2_{\text{loc}}(\mathbb{R}^{d\times d})$.

Starting from these assumptions, the factors are decoded using different PCA-based techniques to also ensure that the reconstructed prices are more likely to be arbitrage-free both in a static and in a dynamic sense. Absence of dynamical arbitrage is ensured through Heath-Jarrow-Morton-type conditions while absence of static arbitrage is ensured by imposing that each discretized normalized call prices' surface respects a set of

linear conditions $Ac \geq b$ for some matrix A and vector b (see [6]). Notice that since the decomposition of the normalized call prices is affine and the no static arbitrage conditions are linear, it is possible to rewrite the latter conditions for ξ_t as $A \cdot G \cdot \xi_t \geq b - A \cdot G_0$.

Given the history of market call prices, the factors G_i can be calibrated for every grid point (τ_j, k_j) and factors $\xi_{s,i}$ for every past day s under the no arbitrage constraints.

After the factors decoding, Cohen at al. set up a supervised learning process to estimate $\alpha(\xi_t), \beta(\xi_t), \mu(\xi_t), \sigma(\xi_t)$ via a maximal likelihood function which ensures that the time series for ξ_t evolves inside the convex polytope generated by the no static arbitrage conditions.

3.4.1 Empirical VaR in the neural-SDE model

Suppose we want to compute the VaR of a portfolio constituted of call options at MPOR date $t + h$, where $h = n\delta t$ and δt is the one day unit.

Having the drift and diffusion functions for the time series of ξ_t and the underlier from the model calibration, predictions can be made with an Euler scheme in a Monte-Carlo fashion. In particular (with an abuse of notation), processes values for the first step at $t + \delta t$ are

$$\begin{aligned} S_{t+\delta t} &= S_t \exp\left(\left(\alpha_t - \frac{\beta_t^2}{2}\right)\delta t + \beta_t(W_{0,t+\delta t} - W_{0,t})\right), \\ \xi_{t+\delta t} &= \xi_t + \mu_t \delta t + \sigma_t(W_{t+\delta t} - W_t), \end{aligned}$$

where

$$\begin{aligned} W_{0,t+\delta t} - W_{0,t} &= \sqrt{\delta t} X_0, \\ W_{t+\delta t} - W_t &= \sqrt{\delta t} X, \end{aligned}$$

with independent standard Gaussian random variables $X_0 \in \mathbb{R}$, $X \in \mathbb{R}^d$.

In order to guarantee more stability of simulations, a tamed Euler scheme can also be implemented.

The following steps are performed as above, using the latest values of S and ξ. At each step, new parameters α, β, μ and σ can be estimated using the neural network algorithm implemented in [7].

Alternatively, assuming α, β, μ and σ to be constant between t and $t + h = t + n\delta t$, simulations for S_{t+h} and ξ_{t+h} can be faster computed as

$$\begin{aligned} S_{t+h} &= S_t \exp\left(\left(\alpha_t - \frac{\beta_t^2}{2}\right)h + \beta_t(W_{0,t+h} - W_{0,t})\right), \\ \xi_{t+h} &= \xi_t + \mu_t h + \sigma_t(W_{t+h} - W_t). \end{aligned} \tag{4}$$

The predicted values of ξ_{t+h} can then be used to compute predicted values of normalized call prices using eq. (2), which can be de-normalized using the predicted values of S_{t+h}.

The number of simulations that can be performed is arbitrary, so that a stable value of the VaR can be computed as the empirical quantile of simulated call prices.

3.4.2 Limitations of the neural-SDE model

In this calibration routine of the neural-SDE model, there is an important point which, according to us, should be taken into consideration in applications: the G parameters are calibrated on the history of market prices, but given their linear role in the normalized call prices, there is little hope that a long history of call prices will be well explained by the very same G factors. Indeed, normalized call prices in this model are random linear combinations of *fixed* surfaces, so that one should probably expect the call prices to maintain these fixed parameters for no more than a typical period of one month or so, after which they should be re-calibrated. In [7], the G parameters are calibrated on a 17-years history, which might be far from being realistic in practice. As a consequence, calibration fit is not as good as in other more dynamic models. As an example, the mean absolute percentage error (MAPE) computed by the authors in Table 2 of [8] using two components of ξ_t is around 4.61% and 5.40%, while in our tests limiting the calibration window of G to one month reduces the MAPE to about 1.5 percentage points. On the other hand, it is not possible to simply calibrate the G parameters on shorter periods of past history, since then the neural-SDE on the ξ_t cannot be properly trained to estimate the model parameters, given the too low amount of historical data.

To some extent, there is therefore a trade-off between the stationarity of the model and its relevance - note though that one could argue that this is a general situation for any model.

Furthermore, this stationarity of parameters is likely to be related to the low procyclicality of the obtained VaR estimations that the authors claim: because the G parameters are the same since several years, the initial margin is indeed automatically less reactive to market changes.

This being said, the neural-SDE model provides a consistent and tractable framework which seems to us very promising.

3.5 The market data in input of the margin computation, and Market/Model add-ons

The models described above for margin computations (FHS, arbitrage-free GANs, neural-SDE) have all in common the generation of scenarios for the risk factors. In the case of options, these scenarios can only be generated after an initial calibration of market prices using any internal model, calibration which will then be reversed to get simulated prices. Indeed, the CCP needs a model and/or an interpolation scheme to get prices at any time-to-maturity and log-forward moneyness, and this scheme is used since the beginning of the IM computation. As a result, margins are based on model prices (i.e. prices calibrated/interpolated with the selected scheme), rather than market prices, and should then be adjusted by a term taking into consideration how the initial discrepancy between market and model prices propagates when computing the IM.

There can be two approaches to this issue in the context of a VaR-type model:

1. Apply the scenarios to the calibrated model prices, thus obtaining shocked model prices, and assume that the model P&Ls are a good representative of the market P&Ls, along each scenario. This means that the calibration error is assumed to be the same at the current date and at the future date along the shocked scenario.

2. Compute a *Market/Model add-on*, which incorporates risk coming from the fact that the model which has been used to estimate the IM does not perfectly match market prices. Since the final risk requirement is computed on model prices and captures future movements of model prices, so that it could differ from the actual requirement needed for market prices, the Market/Model add-on estimates how large the difference between the market IM and the model IM is and adds it to the final requirement.

In the second approach, market P&Ls can be decomposed in 3 terms:

- the difference between the portfolio price under the calibrated model and its market price: $P_t^{\text{mod}} - P_t^{\text{mkt}}$;
- the difference between the portfolio model prices along the scenario s: $\tilde{P}_{t+h,s}^{\text{mod}} - P_t^{\text{mod}}$;
- the difference between the portfolio price under the calibrated model and its market price at the simulated date along the scenario s: $\tilde{P}_{t+h,s}^{\text{mkt}} - \tilde{P}_{t+h,s}^{\text{mod}}$.

The first of the 3 terms above is known and can be readily computed; the second term is computed in the IM; the third term depends on each scenario and upon the assumption on the distance between the market and model prices at the future simulated date along each scenario. The Market/Model add-on aims at covering this third source of risk.

4. A simple short-term model-free formula

In this section we describe a new short-term model-free formula for options VaR, which only depends on market data and does not need any model-specific calibration. The short-term attribute depends on the fact that approximations are performed in the MPOR component, so that the shorter the MPOR, the more precise is the formula.

In the following we will denote with $\mathrm{DF}_t(\tau)$ and $F_t(\tau)$ the discount factor and the forward value for time-to-maturity τ evaluated at time t. We work under the hypothesis of known constant rates between today date t and the MPOR date $t+h$, so that for a given time-to-maturity, discount factors are constant and forward values are proportional to the underlier S_t. In particular we write $F_t(\tau) = f(\tau)S_t$. We call δt the one day unit and consider an MPOR $h = n\delta t$ of n days. Finally, we denote respectively by p_Y and F_Y the probability density function and the cumulative density function of a generic random variable Y. The cumulative density function and the probability density function of a standard Gaussian random variable are denoted with Φ and φ respectively. Also, when considering the distribution of the underlier S_{t+h} at time $t+h$, we actually mean the distribution conditional to quantities at time t (i.e. S_t and other risk factors ξ_t).

In the following sections we will always consider a portfolio of Vanilla calls with price at time t given by

$$\Pi_t(S_t, \xi_t) = \sum_i \pi_i C\Big(T_i - t, \log \frac{K_i}{f(T_i - t)S_t}; S_t, \xi_t\Big)$$

where $C(\tau, k; S_t, \xi_t)$ is a generic model price of a call with time-to-maturity τ and log-forward moneyness k, depending on the current value of the underlier S_t and of the other possible risk factors ξ_t (as for example the implied volatility in the short-term model-free case). The P&Ls are defined as the finite differences of the portfolio over the MPOR:

$$\mathrm{P\&L} := \Pi_{t+h}(S_{t+h}, \xi_{t+h}) - \Pi_t(S_t, \xi_t).$$

The h-days VaR at confidence level θ (close to 1) of the portfolio is the quantity $v(\theta, h)$ such that

$$P\big(\mathrm{P\&L} \le v(\theta, h)\big) = 1 - \theta.$$

Sometimes we will need to develop the above expression using conditional probabilities. In particular, it holds

$$\begin{aligned} P\big(\mathrm{P\&L} \le v(\theta, h)\big) &= E[\mathbb{1}_{\mathrm{P\&L} \le v(\theta,h)}] \\ &= E\big[E[\mathbb{1}_{\mathrm{P\&L} \le v(\theta,h)} | S_{t+h}]\big] \\ &= \int_0^\infty P\big(\mathrm{P\&L} \le v(\theta, h) | S_{t+h} = s\big)\, dF_{S_{t+h}}(s). \end{aligned} \tag{5}$$

In the case of existence of a probability function for S_{t+h}, the latter expression can be written as

$$P\big(\mathrm{P\&L} \le v(\theta, h)\big) = \int_0^\infty p_{S_{t+h}}(s) P\big(\mathrm{P\&L} \le v(\theta, h) | S_{t+h} = s\big)\, ds.$$

4.1 The Black-Scholes case and the Stochastic Volatility case

Before introducing the short-term model-free VaR formula, we firstly look at some prototypical examples such as the Black-Scholes and the Stochastic Volatility cases.

In the classic Black-Scholes case, the underlier is a geometric Brownian motion

$$dS_t = \alpha_t S_t dt + \beta_t S_t dW_t$$

under the real-world probability measure. Applying Ito's lemma, portfolio prices are processes such that

$$d\Pi_t(S_t) = \left(S_t \alpha_t \frac{d}{dS_t}\Pi_t(S_t) + \frac{1}{2} S_t^2 \beta_t^2 \frac{d^2}{dS_t^2}\Pi_t(S_t) \right) dt + S_t \beta_t \frac{d}{dS_t}\Pi_t(S_t) dW_t.$$

Writing dW_t as $\sqrt{h}X$ where X is a standard Gaussian random variable and approximating the above expression at the first order we have that the P&Ls have the form

$$\Pi_{t+h}(S_{t+h}) - \Pi_t(S_t) \approx S_t \beta_t \frac{d}{dS_t}\Pi_t(S_t)\sqrt{h}X.$$

Then, it is easy to compute the VaR with a first order approximation:

$$P\big(\text{P\&L} \leq v(\theta, h)\big) = P\left(\frac{d}{dS_t}\Pi_t(S_t) X \leq \frac{v(\theta, h)}{S_t \beta_t \sqrt{h}} \right)$$

so that

$$v(\theta, h) = \Phi^{-1}(1-\theta) S_t \beta_t \left| \frac{d}{dS_t}\Pi_t(S_t) \right| \sqrt{h}.$$

The above reasoning can actually be generalized to Stochastic Volatility models where the volatility of the underlier is a stochastic process with volatility σ_t:

$$\begin{aligned} dS_t &= \alpha_t S_t dt + \xi_t S_t dW_{0,t} \\ d\xi_t &= \mu_t dt + \sigma_t dW_t \\ dW_{0,t} dW_t &= \rho_t dt. \end{aligned}$$

In the above formulation we have dropped the dependency of volatility parameters in the volatility itself, i.e. $\mu_t = \mu_t(\xi_t)$ and $\sigma_t = \sigma_t(\xi_t)$. Indeed, in order to guarantee the positivity of the volatility there must be such a dependency. Applying Ito's lemma to the portfolio $\Pi_t(S_t, \xi_t)$ of option prices generated by the pricing version of the Stochastic Volatility model, one finds

$$d\Pi_t(S_t, \xi_t) = a_t dt + S_t \xi_t \frac{d}{dS_t}\Pi_t(S_t, \xi_t) dW_{0,t} + \sigma_t \frac{d}{d\xi_t}\Pi_t(S_t, \xi_t) dW_t$$

where

$$a_t = \alpha_t S_t \frac{d}{dS_t}\Pi_t(S_t,\xi_t) + \mu_t \frac{d}{d\xi_t}\Pi_t(S_t,\xi_t) + \frac{\xi_t^2 S_t^2}{2}\frac{d^2}{dS_t^2}\Pi_t(S_t,\xi_t) + \\ + \frac{\sigma_t^2}{2}\frac{d^2}{d\xi_t^2}\Pi_t(S_t,\xi_t) + \xi_t S_t \sigma_t \rho_t \frac{d^2}{dS_t d\xi_t}\Pi_t(S_t,\xi_t).$$

Considering the finite increments of the portfolio and neglecting linear terms for h going to 0, the form of the P&Ls becomes

$$\Pi_{t+h}(S_{t+h},\xi_{t+h}) - \Pi_t(S_t,\xi_t) \approx S_t\xi_t \frac{d}{dS_t}\Pi_t(S_t,\xi_t)\sqrt{h}X_0 + \sigma_t \frac{d}{d\xi_t}\Pi_t(S_t,\xi_t)\sqrt{h}X,$$

where X_0 and X are standard jointly normal random variables with correlation ρ_t equal to the correlation of the two Brownian motions. Then, any combination of X_0 and X is still normal and the VaR of the portfolio is

$$v(\theta,h) = \Phi^{-1}(1-\theta)\sqrt{\Big(S_t\xi_t\frac{d}{dS_t}\Pi_t(S_t,\xi_t)\Big)^2 + \Big(\sigma_t\frac{d}{d\xi_t}\Pi_t(S_t,\xi_t)\Big)^2 + 2\rho_t S_t\xi_t\sigma_t\frac{d}{dS_t}\Pi_t(S_t,\xi_t)\frac{d}{d\xi_t}\Pi_t(S_t,\xi_t)}\sqrt{h}. \tag{6}$$

4.2 A short-term model-free formula

Driven by the results in the Black-Scholes and the Stochastic Voaltility case, we generalize the VaR formulas to a short-term model-free formula which can be applied to any historical series of spot and option prices.

With this aim, we rather work using the implied volatility, which can always be computed from market prices using a root-finding algorithm applied to the classic Black-Scholes pricing formula

$$\mathrm{BS}_t\big(k,\tau,\omega,F_t(\tau),\mathrm{DF}_t(\tau),\sigma_t^{\mathrm{imp}}(k,\tau)\big) = \omega \mathrm{DF}_t(\tau)F_t(\tau)\big(\Phi(\omega d_1) - e^k\Phi(\omega d_2)\big)$$

where $k = \log\frac{K}{F_t(\tau)}$ is the log-forward moneyness, $\tau = T-t$ is the time-to-maturity,

$$d_{1,2} = -\frac{k}{\sigma_t^{\mathrm{imp}}(k,\tau)\sqrt{\tau}} \pm \frac{\sigma_t^{\mathrm{imp}}(k,\tau)\sqrt{\tau}}{2}$$

and $\omega = +1$ if the option is a call, -1 if it is a put.

When computing risks, the implied volatility $\sigma_t^{\mathrm{imp}}(k,\tau)$ is generally considered as a risk factor together with the underlier S_t. For this reason, we write it as a function of a driving factor ξ_t: $\sigma_t^{\mathrm{imp}}(k,\tau) = \sigma(k,\tau,\xi_t)$, so that the dynamics of the two risk factors are

$$\begin{aligned} dS_t &= \alpha_t S_t dt + \beta_t S_t dW_{0,t} \\ d\xi_t &= \mu_t dt + \eta_t dW_t \\ dW_{0,t}dW_t &= \rho_t dt. \end{aligned} \tag{H1}$$

Observe that the implied volatility risk factor depends on the log-forward moneyness and the time-to-maturity rather than the contract strike and its maturity. Indeed, the time series of a fixed contract is

available since its issue date and is then limited in time. Furthermore, we do not expect its implied volatility to have any nice statistical property of stationarity that could legitimate drawing meaningful forecasts for its historical returns between time t and $t+h$. On the contrary, we expect that the market encode the implied volatility risk rather in a log-forward moneyness, time-to-maturity map, meaning that the time series of the implied volatility at a fixed point in this relative grid will have much nicer features.

Let us consider a portfolio of calls and puts written as Black-Scholes functions:

$$\begin{aligned}\Pi_t(S_t,\xi_t) &= \sum_i \pi_i \mathrm{BS}_t\Big(\log\frac{K_i}{f(T_i-t)S_t}, T_i-t, \omega_i, F_t(T_i-t), \mathrm{DF}_t(T_i-t), \sigma\Big(\log\frac{K_i}{f(T_i-t)S_t}, T_i-t, \xi_t\Big)\Big)\\ &=: \sum_i \pi_i \mathrm{BS}_t^i.\end{aligned} \tag{7}$$

Repeating the steps in section 4.1, the approximated formula for the VaR becomes

$$v(\theta,h) = \Phi^{-1}(1-\theta)\sqrt{\Big(S_t\beta_t\frac{d}{dS_t}\Pi_t(S_t,\xi_t)\Big)^2 + \Big(\eta_t\frac{d}{d\xi_t}\Pi_t(S_t,\xi_t)\Big)^2 + 2\rho_t S_t\beta_t\eta_t\frac{d}{dS_t}\Pi_t(S_t,\xi_t)\frac{d}{d\xi_t}\Pi_t(S_t,\xi_t)}\sqrt{h}.$$

This formula is far from being model-free for two reasons:

- The term $\frac{d}{dS_t}\Pi_t(S_t,\xi_t)$ is the full derivative of the portfolio Π_t with respect to S_t, which also involves the derivative of prices with respect to the implied volatility, since it depends on $k=\log\frac{K}{f(\tau)S_t}$. As a consequence, it does not correspond to the Black-Scholes delta and its expression must be made explicit.
- The term $\frac{d}{d\xi_t}\Pi_t(S_t,\xi_t)$ is the derivative of the portfolio with respect to ξ_t, and it does not coincide with what the market indicates with vega, i.e. the sensibility of the portfolio to the option volatility.

Given the above, we shall rather develop the dynamics of the portfolio as a function of S_t and $\sigma(k,\tau,\xi_t)$ where k also depends on S_t.

Observe that for fixed k and τ, we have

$$\begin{aligned}d\sigma_t &= \partial_\xi\sigma_t\, d\xi_t\\ &= \mu_t\partial_\xi\sigma_t\, dt + \eta_t\partial_\xi\sigma_t\, dW_t\end{aligned} \tag{8}$$

where $\sigma_t=\sigma(k,\tau,\xi_t)$. We define $\zeta_t(k,\tau,\xi_t) := \eta_t\partial_\xi\sigma(k,\tau,\xi_t)$.

On the other hand, writing $k=\log\frac{K}{f(\tau)S_t}$ and $\tau=T-t$, we rather find

$$\begin{aligned}d\sigma_t &= a_t\, dt - \frac{\partial_k\sigma_t}{S_t}\, dS_t + \partial_\xi\sigma_t\, d\xi_t\\ &= \big(a_t - \alpha_t\partial_k\sigma_t + \mu_t\partial_\xi\sigma_t\big)\, dt - \beta_t\partial_k\sigma_t\, dW_{0,t} + \zeta_t\, dW_t\end{aligned} \tag{9}$$

where $a_t = -\partial_\tau\sigma_t + \partial_k\sigma_t\frac{\partial_\tau f}{f} + \frac{\beta_t^2}{2}(\partial_k^2\sigma_t + \partial_k\sigma_t) + \frac{\eta_t^2}{2}\partial_\xi^2\sigma_t - \rho_t\beta_t\eta_t\partial_\xi\partial_k\sigma_t$ and $\sigma_t = \sigma\big(\log\frac{K}{f(T-t)S_t}, T-t, \xi_t\big)$.

Using eq. (9) and ignoring the terms in h in the finite scheme of the portfolio increments, the P&Ls assume the form

$$\Pi_{t+h}(S_{t+h},\xi_{t+h}) - \Pi_t(S_t,\xi_t) \approx \beta_t\big(S_t\sum_i \pi_i\partial_S\mathrm{BS}_t^i - \sum_i \pi_i\partial_k\sigma_t^i\partial_\sigma\mathrm{BS}_t^i\big)\sqrt{h}X_0 + \sum_i \pi_i\zeta_t^i\partial_\sigma\mathrm{BS}_t^i\sqrt{h}X$$

where $\sigma_t^i = \sigma\big(\log \frac{K_i}{f(T_i-t)S_t}, T_i - t, \xi_t\big)$, $\zeta_t^i = \zeta_t\big(\log \frac{K_i}{f(T_i-t)S_t}, T_i - t, \xi_t\big)$, and X_0 and X are standard jointly normal random variables with correlation ρ_t.

Let us denote $\Delta_t^i := \partial_S \mathrm{BS}_t^i$ and $\mathcal{V}_t^i := \partial_\sigma \mathrm{BS}_t^i$. Using the same proof as in the Stochastic Volatility case of section 4.1, we finally find a VaR formula on an MPOR horizon of h days of the form:

$$\begin{aligned}
\mathrm{VaR}_{\theta,t}(h) &= \Phi^{-1}(1-\theta)\sqrt{c_t^2 + q_t^2 + 2\rho_t c_t q_t}\sqrt{h} \\
c_t &= \beta_t\big(S_t \sum_i \pi_i \Delta_t^i - \sum_i \pi_i \mathcal{V}_t^i \partial_k \sigma_t^i\big) \\
q_t &= \sum_i \pi_i \zeta_t^i \mathcal{V}_t^i.
\end{aligned} \qquad \text{(Short-term formula)}$$

This expression is actually model-free. Indeed, the terms Δ_t^i and $\mathcal{V}_t^i$ are respectively the Black-Scholes delta and vega of the i-th option in the portfolio. In particular they correspond to

$$\begin{aligned}
\Delta_t\big(k, \tau, \omega, \sigma_t^{\mathrm{imp}}(k,\tau)\big) &= \omega\Phi(\omega d_1), \\
\mathcal{V}_t\big(k, \tau, F_t(\tau), \mathrm{DF}_t(\tau), \sigma_t^{\mathrm{imp}}(k,\tau)\big) &= \mathrm{DF}_t(\tau) F_t(\tau) \varphi(d_1)\sqrt{\tau}.
\end{aligned}$$

The volatility β_t of the underlying spot S_t can be computed looking at historical moves. For example, it could be a EWMA volatility on log-returns appropriately rescaled by the square-root of the returns' distance h_r (of for example one trading day): $\beta_t = \frac{\mathrm{EWMA}(r_{S,t})}{\sqrt{h_r}}$ where $r_{S,t} = \log \frac{S_t}{S_{t-h_r}}$.

Given eq. (8), the quantity ζ_t^i is simply the vol-of-vol evaluated in $\big(\log \frac{K_i}{f(T_i-t)S_t}, T_i - t, \xi_t\big)$ and it could be also computed as a EWMA volatility on historical absolute returns of the implied volatility surface at the fixed log-forward moneyness and time-to-maturity grid point, rescaled by the square-root of h_r. For liquidity reasons, it is also possible to approximate the latter quantity as the vol-of-vol at the 1M ATM point, times an appropriate factor (see section 6.3 for the description of a possible way to calibrate such a factor). In [17], the authors suggest to consider the historical series of the 1M ATM implied volatility point. A less procyclical alternative identified by the authors consists in rescaling the VVIX historical value with a sigmoid function which ensures a smooth vol-of-vol transition between high and low volatility regimes.

The correlation parameter ρ_t can be computed using a EWMA correlation between spot log-returns and absolute IV returns, where the IV point considered can be again the 1M ATM point.

Lastly, the derivative $\partial_k \sigma_t^i$ is the derivative of the smile with respect to the log-forward moneyness, evaluated in $\big(\log \frac{K_i}{f(T_i-t)S_t}, T_i - t, \xi_t\big)$. Since options are quoted in strike and maturity rather than log-forward moneyness and time-to-maturity, observe that

$$\partial_k \sigma_t^i = \partial_k \sigma_t\left(\log \frac{K_i}{f(T_i-t)S_t}, T_i - t, \xi_t\right) = K \partial_K \tilde{\sigma}_t^{\mathrm{imp}}(K_i, T_i)$$

where $\tilde{\sigma}_t^{\mathrm{imp}}(K,T) = \sigma_t^{\mathrm{imp}}\big(\log \frac{K}{f(T-t)S_t}, T-t\big)$. The derivative of the strike smile can be recovered by simple interpolation of market data (for example, using cubic B-splines or arbitrage-free smile models), or by finite differences of market data.

4.2.1 t-Student short-term model-free VaR formulation

When calculating risk, a large majority of financial players consider the distribution of the returns of an underlier S_t to be t-Student. The reason is linked to the shape of the probability functions of such distribution, which are fatter, compared to a classic normal distribution. In this way the importance of extreme events is higher and this guarantees a larger conservativeness of the risk model.

In this section we consider then a t-Student distribution for the relative returns $\frac{S_{t+h}-S_t}{S_t}$, with ν_t degrees of freedom, location parameter $\alpha_t h$ and scale parameter $\beta_t\sqrt{h}$. In particular, we consider the model

$$S_{t+h} = S_t(1 + \alpha_t h + \beta_t T_{t+h}) \tag{H1}$$

where $T_{t+h} \in \mathbb{R}$ is a t-Student with ν_t degrees of freedom, null mean and variance equal to h. Then, the probability density function of S_{t+h} is

$$p_{S_{t+h}}(s) = \frac{\Gamma\left(\frac{\nu_t+1}{2}\right)}{\Gamma\left(\frac{\nu_t}{2}\right)S_t\beta_t\sqrt{\pi h}}\left(1 + \frac{1}{\nu_t}\left(\frac{s - S_t(1+\alpha_t h)}{S_t\beta_t\sqrt{h}}\right)^2\right)^{-\frac{\nu_t+1}{2}}. \tag{10}$$

Also, for every strike K and maturity T, denoting $\tau = T - t$, $k(s) = \log\frac{K}{f(\tau)s}$, we consider the increment $\Delta\sigma_t(k(S_t),\tau) := \sigma_{t+h}(k(S_{t+h}),\tau - h) - \sigma_t(k(S_t),\tau)$ conditional to S_{t+h} to be a Gaussian random variable with mean $m_t(S_{t+h}, k(S_t),\tau)$ and variance $\zeta_t(k(S_t),\tau)^2(1-\rho_t^2)h$, with

$$m_t(s, k(S_t),\tau) = \mu_t(k(S_t),\tau)h + \frac{s - S_t(1+\alpha_t h)}{\beta_t S_t}\zeta_t(k(S_t),\tau)\rho_t.$$

In other words, we write the conditional implied volatility increments as

$$\Delta\sigma_t(k(S_t),\tau)|(S_{t+h} = s) = m_t(s, k(S_t),\tau) + \zeta_t(k(S_t),\tau)\sqrt{1-\rho_t^2}\sqrt{h}X \tag{H2}$$

where X is a standard Gaussian random variable.

Remark 4.1. *Since with the above hypothesis we only know the distribution of the implied volatility increments conditional to S_{t+h}, it could seem difficult to calibrate parameters μ_t, ζ_t and ρ_t on market data. However, given a conditional distribution, it is easy to recover the moments of the marginal distribution using the tower property in expectations. Indeed, moments up to the second order of $\Delta\sigma_t(k(S_t),\tau)$ (without the conditioning to S_{t+h}) are*

$$\begin{aligned} E[\Delta\sigma_t(k(S_t),\tau)] &= \mu_t(k(S_t),\tau)h \\ Var[\Delta\sigma_t(k(S_t),\tau)] &= \zeta_t(k(S_t),\tau)^2 h \\ Corr[S_{t+h}, \Delta\sigma_t(k(S_t),\tau)] &= \rho_t. \end{aligned}$$

This allows to easily calibrate parameters based on the historical mean and variance of $\Delta\sigma_t(k(S_t),\tau)$.

In the next paragraphs, we justify the following formula for the h-days VaR with confidence level θ under

eqs. (H1) and (H2):

$$\begin{aligned}\mathrm{VaR}_{\theta,t}(h) &= F_Z^{-1}(1-\theta)\sqrt{c_t^2+q_t^2+2\rho_t c_t q_t}\sqrt{h}\\ c_t &= \beta_t\big(S_t\sum_i \pi_i\Delta_t^i - \sum_i \pi_i \mathcal{V}_t^i \partial_k \sigma_t^i\big)\\ q_t &= \sum_i \pi_i \zeta_t^i \mathcal{V}_t^i\end{aligned} \qquad \text{(Short-term t-Student)}$$

where

$$Z = \frac{q_t\sqrt{1-\rho_t^2}X + (c_t+q_t\rho_t)Y}{\sqrt{c_t^2+q_t^2+2\rho_t c_t q_t}} \tag{11}$$

and X is a standard Gaussian random variable and Y is a standard t-Student with ν_t degrees of freedom independent of X.

Quantities that enter eq. (Short-term t-Student) are the same as in the Gaussian case: the Black-Scholes Greeks delta Δ_t^i and vega $\mathcal{V}_t^i$, the volatility β_t of the underlying spot S_t, the vol-of-vol ζ_t^i, the correlation ρ_t, and the derivative $\partial_k\sigma_t^i$ of the smile with respect to the log-forward moneyness. See section 4.2 for a description of how to compute these quantities in practice from market data.

Remark 4.2. *[3] shows in Theorem 1 that the probability density function of Z is*

$$p_Z(z) = \sum_{k=0}^{\infty} \phi_k^{(\nu_t,\gamma)} g_{k,a_1}(z)$$

where $a_1 = \frac{q\sqrt{1-\rho_t^2}}{\sqrt{c_t^2+q_t^2+2\rho_t c_t q_t}}$, $\gamma = \frac{c_t+q_t\rho_t}{q_t\sqrt{2(1-\rho_t^2)}}$ *and*

$$\begin{aligned}\phi_k^{(\nu_t,\gamma)} &= \frac{\Gamma\left(k+\frac{1}{2}\right)}{k!\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{\nu_t}{2}\right)}\int_0^\infty \exp(-f) f^{\frac{\nu_t-1}{2}}\left(f+\gamma^2\right)^{-k-\frac{1}{2}} df\\ g_{k,a_1}(z) &= \frac{\Gamma\left(\frac{1}{2}\right)}{\Gamma\left(k+\frac{1}{2}\right)a_1\sqrt{2\pi}}\left(\frac{z^2}{2a_1^2}\right)^k \exp\left(-\frac{z^2}{2a_1^2}\right).\end{aligned}$$

Quantiles of Z can also be computed empirically, simulating the distribution of the linear combination of two independent random variables distributed as a standard Gaussian and a standard t-Student with ν_t degrees of freedom. In particular, simulations of Z can be found following the steps:

1. *Simulate two independent normal random variables X and N_Y, with mean 0 and variance 1;*
2. *Turn N_Y into a uniform distribution $U_Y = \Phi(N_Y)$;*
3. *Recover the t-Student random variable via $Y = F_t^{-1}(U_Y;\nu_t)$ where $F_t(\cdot;\nu_t)$ is the cumulative density function of a t-Student with ν_t degrees of freedom;*
4. *Put $Z = \frac{q_t\sqrt{1-\rho_t^2}X+(c_t+q_t\rho_t)Y}{\sqrt{c_t^2+q_t^2+2\rho_t c_t q_t}}$.*

We now explain the rationale of eq. (Short-term t-Student). Doing a first order approximation of increments of the option portfolio $\Pi_t(S_t,\sigma_t)$, taking into consideration the dependence of every options' implied volatility

to the log-forward moneyness and so to the underlier, we find that the form of the P&Ls is

$$\begin{aligned}\Pi_{t+h}(S_{t+h},\sigma_{t+h})-\Pi_t(S_t,\sigma_t) &\approx \beta_t\big(S_t\sum_i \pi_i\partial_S \mathrm{BS}_t^i-\sum_i \pi_i\partial_k\sigma_t^i\partial_\sigma \mathrm{BS}_t^i\big)\,T_{t+h}+\sum_i \pi_i\partial_\sigma \mathrm{BS}_t^i\Delta\sigma_t^i \\ &= c_tT_{t+h}+\sum_i \pi_i\mathcal{V}_t^i\Delta\sigma_t^i,\end{aligned} \tag{12}$$

where we used the same notations as in section 4.2. Here, we do not know the distribution of the increments of the implied volatilities, so that we cannot automatically infer the distribution of the P&Ls. However, we can still compute VaRs using the relation in eq. (5).

Firstly, we can write $T_{t+h}=\sqrt{h}\tilde{Y}$ where $\tilde{Y}$ is a standard t-Student with ν_t degrees of freedom, and

$$\Delta\sigma_t(k(S_t),\tau)=\mu_t(k(S_t),\tau)h+\zeta_t(k(S_t),\tau)\sqrt{h}\tilde{X}$$

for a certain random variable $\tilde{X}$ with mean 0 and variance 1. Then, given eq. (12), the distribution of $\frac{\mathrm{P\&L}}{\sqrt{h}}$ tends to the distribution of $c_t\tilde{Y}+\sum_i \pi_i\mathcal{V}_t^i\zeta_t^i\tilde{X}$, which does not depend on h. In particular since

$$\begin{aligned}1-\theta &= P\big(\mathrm{P\&L}\le v(\theta,h)\big)\\ &= P\bigg(\frac{\mathrm{P\&L}}{\sqrt{h}}\le\frac{v(\theta,h)}{\sqrt{h}}\bigg),\end{aligned}$$

and the limiting random variable has a strictly positive density, then the function $v(\theta,h)$ is asymptotic with $\sqrt{h}$, i.e. $v(\theta,h)=u(\theta)\sqrt{h}+o(\sqrt{h})$. This is consistent with eq. (Short-term formula), where the h-days VaR is proportional to the square-root of h.

Secondly, the distribution of the P&Ls conditional to S_{t+h} is Gaussian and in particular

$$P\big(\mathrm{P\&L}\le v(\theta,h)|S_{t+h}=s\big)=\Phi\bigg(\frac{v(\theta,h)-c_tt(s)-\sum_i \pi_i\mathcal{V}_t^im_t^i(s)}{q_t\sqrt{1-\rho_t^2}\sqrt{h}}\bigg)$$

where $t(s)=\frac{s-S_t(1+\alpha_th)}{\beta_tS_t}$. Then, removing the conditionality to the probability of the P&Ls, it holds

$$\begin{aligned}P\big(\mathrm{P\&L}\le v(\theta,h)\big) &= \int_{-\infty}^{\infty} p_{S_{t+h}}(s)\Phi\bigg(\frac{v(\theta,h)-c_tt(s)-\sum_i \pi_i\mathcal{V}_t^im_t^i(s)}{q_t\sqrt{1-\rho_t^2}\sqrt{h}}\bigg)\,ds\\ &= \int_{-\infty}^{\infty} p_T(y)\Phi\bigg(\frac{v(\theta,h)-c_ty\sqrt{h}-\sum_i \pi_i\mathcal{V}_t^im_t^i\big(S_t(1+y\beta_t\sqrt{h}+\alpha_th)\big)}{q_t\sqrt{1-\rho_t^2}\sqrt{h}}\bigg)\,dy\end{aligned}$$

where $p_{S_{t+h}}$ is as in eq. (10), p_T is the probability density function of a standard t-Student with ν degrees of freedom, and we used the transformation $y=\frac{t(s)}{\sqrt{h}}$. In this way, for the Lebesgue's dominated convergence theorem and using the fact that $v(\theta,h)=u(\theta)\sqrt{h}+o(\sqrt{h})$, the right hand side of the previous relation goes to

$$\int_{-\infty}^{\infty} p_T(y)\Phi\bigg(\frac{u(\theta)-(c_t+q_t\rho_t)y}{q_t\sqrt{1-\rho_t^2}}\bigg)\,dy$$

for h going to 0. Consider two independent random variables X and Y with X a standard Gaussian and Y a standard t-Student with ν_t degrees of freedom. We can write the latter expression as

$$E\bigg[P\bigg(X\le\frac{u(\theta)-(c_t+q_t\rho_t)Y}{q_t\sqrt{1-\rho_t^2}}\bigg)\bigg]=P\bigg(X\le\frac{u(\theta)-(c_t+q_t\rho_t)Y}{q_t\sqrt{1-\rho_t^2}}\bigg).$$

Defining the random variable Z as in eq. (11), we shall look at the value of $u(\theta)$ such that

$$1-\theta = P\left(Z \leq \frac{u(\theta)}{\sqrt{c_t^2 + q_t^2 + 2\rho_t c_t q_t}}\right).$$

All in all, the short-term model-free VaR formula in the t-Student case becomes eq. (Short-term t-Student) ignoring terms in $o(\sqrt{h})$.

Remark 4.3. *In both eq.* (Short-term formula) *and eq.* (Short-term t-Student) *the vol-of-vol parameter depends on the strike and maturity of the option, while the correlation does not. This is due to the underlying hypothesis that the whole implied volatility surface is driven by one single Brownian motion, even though the magnitude of movements for each surface point depends on the point itself. The short-term model-free formulas can be generalized to the case where there is more than one Brownian motion driving the implied volatility surface (typically the target dimension is of* 2 *or* 3*).*

4.3 Properties and limitations

4.3.1 Local quantities and extreme risk: concrete practical implementation

Observe that all the above VaR estimations (the Black-Scholes formula, the Stochastic volatility formula, and the short-term model-free formula) are defined with local quantities: deltas, vegas, instantaneous volatility and correlation coefficients. The initial margin however incorporates a tail risk which looks at future moves in prices that typically correspond to large moves. Even if there is an apparent paradox here, the explanation is clear: those formulas are asymptotic formulas when the time step h goes to zero, and for sufficient small h even the tail risk will be driven by the local quantities, in so far as we deal we diffusion models.

The whole question therefore is how those asymptotic formulas will behave in practice. Obviously, the smaller the MPOR, or the less volatile the market, the better. A careful backtesting will be the clue here: it will allow to diagnose whether the coverage and procyclicality behavior of the formula are satisfactory.

In this regard, and especially from a regulatory perspective, one should keep in mind that the final IM formula will contain other components besides this core one, like a weighted Stress Historical VaR and the Short Option Minimum quantity described in section 2. In general the former component will be obtained by computing price returns along stress historical scenarios with full re-evaluation (meaning, using the Black-Scholes formula for options with the shocked underlier and implied volatility) instead of the local first order Greeks. Therefore the risk of missing a convexity behavior should be largely mitigated, if not fully eliminated. Regarding the SOM, consider a portfolio of short deep OTM options. Today, this portfolio has negligible delta and vega quantities, so that the VaR estimation is around 0, even though there actually is a tail risk. This hidden risk is far to be local, but it still should be taken into consideration in the initial margin calculation. This is the rationale of the SOM, which is already implemented by CCPs and should cover the risk of those short-term portfolios, as discussed in section 2.

4.3.2 Symmetry with respect to the portfolio

It is easy to see that all the new VaR formulas in this article are symmetrical with respect to the portfolio, i.e. being short or long on the same portfolio would produce the same VaR exposure. This could seem weird, especially when we suppose a log-normal distribution of the spot, which is not symmetric. The symmetricity appears when we take the limit for h going to 0. Indeed, the terms multiplying $\sqrt{h}$ are symmetrical in the portfolio position while the ones that should break the symmetricity multiply higher orders of h, so that they are canceled out in the limit.

However, this symmetry is not an issue when computing margins: as seen in section 2, the final total risk requirement charged by the CCP is composed of the margin computed on P&Ls (refined by the add-ons and the SOM) minus the NOV component. In this way, neglecting the add-ons and the SOM, a long portfolio $\Pi > 0$ with initial margin IM has a total risk requirement equal to $\text{IM} - \Pi$; while the same portfolio but on a short position $-\Pi < 0$ implies a total risk requirement of $\text{IM} + \Pi$.

4.3.3 Comparison with FHS

In section 3.1.3, we have seen that among its drawbacks, the FHS model is limited by the number of scenarios that it can generate, depending on the available historical data. On the other hand, the short-term model-free formula in section 4.2 does not need to compute simulated scenarios and eventually requires historical data only for the calibration of volatility parameters.

Secondly, while the FHS does not capture the joint dynamics of risk factors in complex products, the short-term model-free formula in section 4.2 considers both the singular margin impact of each risk factor and the joint margin impact affected by the correlation of risk factors. Furthermore, the short-term model-free formula for options is more natural than the FHS methodology, whose application to IV surface points is more subtle.

A third limitation is the difficulty of FHS to generate arbitrage-free scenarios, which is not an issue for the short-term model-free formula in section 4.2 since it does not require the generation of scenarios and does not face the arbitrage issue.

Finally, regarding the procyclicality of the VaR estimation, we show in numerical experiments in section 6.3 that the short-term model-free VaR is less procyclical then the FHS VaR for the tested portfolios.

We turn now to the exact computation of the VaR in the neural-SDE model.

5. Quasi-explicit formula for the VaR in the neural-SDE model

In this section we investigate the neural-SDE model described in section 3.4 and the special specification of its parameters with the aim of applying it to an IM computation. We are not interested in the calibration of arbitrage-free call prices surfaces via neural networks but to the affine factor model for normalized option prices itself, so that parameters can be calibrated with any algorithm of choice, which is not necessarily a neural network.

The model is particularly simple and it turns out to have a quasi-explicit formula for the VaR of option portfolios, as we show in section 5.1 below. In practice, this could enable rapid computations for the IM in such models, which may prove to be highly relevant when properly calibrated.

Moreover, the VaR can be approximated by a closed formula which is proportional to the square-root of the MPOR (see section 5.2). This approximated formula coincides with the VaR formula in the Stochastic Volatility model of section 4.1.

We reemphasize the fact that while the model can be calibrated also in different ways as the ones described in [7], the results in this chapter are still valid and independent from the calibration setup.

We use the same notations as in section 4. Furthermore, in the whole section, the notation $\|\cdot\|_2$ indicates the Euclidean 2-norm, i.e. $\|(a_1,\dots,a_d)^T\|_2^2 = \sum_{i=1}^d a_i^2$.

5.1 Quasi-explicit formula for the VaR

Consider a portfolio of Vanilla calls with price at time t given by

$$\begin{aligned}\Pi_t(S_t,\xi_t) &= \sum_i \pi_i C\Big(T_i - t, \log\frac{K_i}{f(T_i-t)S_t}; S_t, \xi_t\Big)\\ &= S_t \sum_i \pi_i \mathrm{DF}_t(T_i-t) f(T_i-t) c\Big(T_i - t, \log\frac{K_i}{f(T_i-t)S_t}; \xi_t\Big)\end{aligned} \tag{13}$$

where

$$c\Big(T_i - t, \log\frac{K_i}{f(T_i-t)S_t}; \xi_t\Big) = G_0\Big(T_i - t, \log\frac{K_i}{f(T_i-t)S_t}\Big) + G\Big(T_i - t, \log\frac{K_i}{f(T_i-t)S_t}\Big)\cdot\xi_t.$$

From now on, we work at time t, so that quantities S_t and ξ_t are known. The P&L $:= \Pi_{t+h}(S_{t+h},\xi_{t+h}) - \Pi_t(S_t,\xi_t)$ of the portfolio reads then

$$\mathrm{P\&L} = A(h,S_{t+h}) + B(h,S_{t+h})\cdot\big(\xi_{t+h} - \xi_t\big) \tag{14}$$

where

$$A(h,s) = s\sum_i \pi_i \mathrm{DF}_t(T_i-(t+h))f(T_i-(t+h))\Big(G_0\Big(T_i-(t+h),\log\frac{K_i}{f(T_i-(t+h))s}\Big)+$$
$$+G\Big(T_i-(t+h),\log\frac{K_i}{f(T_i-(t+h))s}\Big)\cdot\xi_t\Big)-\Pi_t(S_t,\xi_t), \tag{15}$$
$$B(h,s) = s\sum_i \pi_i \mathrm{DF}_t(T_i-(t+h))f(T_i-(t+h))G\Big(T_i-(t+h),\log\frac{K_i}{f(T_i-(t+h))s}\Big).$$

An important consequence to the representation in eq. (14) is the linearity of the P&Ls in $\xi_{t+h}-\xi_t$. In terms of VaR calculations, this means that the VaR for the P&Ls' distribution conditional to S_{t+h} is linear with respect to the VaR for the ξ_{t+h} distribution.

5.1.1 Hypothesis on the joint increments

Since from a practical perspective market data is always related to a discrete time grid, from now on, for risk calculations we consider processes defined via their Euler scheme as in section 3.4.1, i.e.

$$\begin{aligned} S_{t+h} &= S_t\exp\Big(\Big(\alpha_t-\frac{\beta_t^2}{2}\Big)h+\beta_t(W_{0,t+h}-W_{0,t})\Big),\\ \xi_{t+h} &= \xi_t+\mu_t h+\sigma_t\cdot(W_{t+h}-W_t)\end{aligned} \tag{16}$$

where $W_{0,t+h}-W_{0,t}\in\mathbb{R}$ and $W_{t+h}-W_t\in\mathbb{R}^d$ are Gaussian random variables with combined law $N(\mathbf{0},hP_t)$ and P_t is the correlation matrix

$$P_t=\begin{pmatrix}1 & P_{S,\xi,t}^T\\ P_{S,\xi,t} & P_{\xi,t}\end{pmatrix}.$$

We work at time t, so that quantities S_t, ξ_t, α_t, β_t, μ_t and σ_t are known. We will not need to observe $W_{0,t}$ and W_t.

Remark 5.1. *The Euler schemes with time step h for the processes S_t and ξ_t defined via the SDE eq. (3) are a particular case of eq. (16), therefore the results of this section hold also in this case.*

To develop eq. (5) we need a partial result regarding the distribution of the increments of ξ_t conditional to S_{t+h}.

Lemma 5.2. *For processes in eq. (16), the distribution of $\xi_{t+h}-\xi_t$ conditional to $S_{t+h}=s$ is a Gaussian $N(m_t(s),V_t)$ where*

$$\begin{aligned} m_t(s) &= \mu_t h+\frac{1}{\beta_t}\Big(\log\frac{s}{S_t}-\Big(\alpha_t-\frac{\beta_t^2}{2}\Big)h\Big)\sigma_t\cdot P_{S,\xi,t},\\ V_t &= h(\sigma_t\cdot b_t)\cdot(\sigma_t\cdot b_t)^T,\end{aligned}$$

and $b_t\in\mathbb{R}^{d\times d}$ is a matrix such that $b_t\cdot b_t^T=P_{\xi,t}-P_{S,\xi,t}\cdot P_{S,\xi,t}^T$.

In the case of independent processes, $m_t(s)=\mu_t h$ and $V_t=h\sigma_t\cdot\sigma_t^T$.

Proof. Let us consider the Brownian increments $\Delta W_{0,t} = W_{0,t+h} - W_{0,t}$ and $\Delta W_t = W_{t+h} - W_t$, which have Gaussian joint distribution $N(\mathbf{0}, hP_t)$. The Gaussian random variable $Z = \Delta W_t - \Delta W_{0,t} P_{S,\xi}$ has null mean and variance equal to $h(P_{\xi,t} - P_{S,\xi,t} \cdot P_{S,\xi,t}^T)$. Since covariance matrices are symmetric and positive semi-definite, there exists a matrix $b_t \in \mathbb{R}^{d\times d}$ such that $b_t \cdot b_t^T = P_{\xi,t} - P_{S,\xi,t} \cdot P_{S,\xi,t}^T$. Also, Z and $\Delta W_{0,t}$ are independent since uncorrelated and jointly Gaussian, and it follows that the distribution of ΔW_t conditional to $\Delta W_{0,t} = w_0$ is a Gaussian with mean $w_0 P_{S,\xi,t}$ and covariance matrix $hb_t \cdot b_t^T$.

From eq. (16), it is immediate to recover the distribution of $\xi_{t+h} - \xi_t$ conditional to $\Delta W_{0,t} = w_0$. For the conditionality with respect to $S_{t+h} = s$, it is enough to substitute w_0 with $\frac{1}{\beta_t}\big(\log \frac{s}{S_t} - \big(\alpha_t - \frac{\beta_t^2}{2}\big)h\big)$.

If processes are independent, $P_{S,\xi,t} = \mathbf{0}$, $P_{\xi,t} = I_d$, and the conclusion follows.

□

As an immediate consequence to theorem 5.2, we can write the increments of ξ_t conditional to S_{t+h} as

$$\xi_{t+h} - \xi_t | (S_{t+h} = s) = m_t(s) + \sqrt{h}\sigma_t \cdot b_t \cdot X \tag{17}$$

where $X \sim N(0, I_d)$ is a Gaussian random variable independent to S_{t+h}.

Then

$$\begin{aligned} \text{P\&L}|(S_{t+h} = s) &= A(h,s) + B(h,s) \cdot \big(m_t(s) + \sqrt{h}\sigma_t \cdot b_t \cdot X\big) \\ &= \hat{A}(h,s) + \hat{B}(h,s) \cdot X \end{aligned}$$

where

$$\begin{aligned} \hat{A}(h,s) &:= A(h,s) + B(h,s) \cdot m_t(s) \in \mathbb{R}, \\ \hat{B}(h,s) &:= B(h,s) \cdot \sqrt{h}\sigma_t \cdot b_t \in \mathbb{R}^{1\times d}. \end{aligned} \tag{18}$$

In particular, conditional to $S_{t+h} = s$, the P&L is a sum of jointly Gaussian variables, so it is also a Gaussian variable with law $N(\hat{A}(h,s), \|\hat{B}(h,s)\|_2^2)$. Then the quantity $P(\text{P\&L} \le v(\theta,h) | S_{t+h} = s)$ is the cumulative function of a Gaussian variable, and in particular it is equal to

$$P(\text{P\&L} \le v(\theta,h) | S_{t+h} = s) = \Phi\left(\frac{v(\theta,h) - \hat{A}(h,s)}{\|\hat{B}(h,s)\|_2}\right).$$

Reconsidering eq. (5), the VaR at risk level θ for the P&Ls can be computed as specified in the following lemma.

Proposition 5.3. *Under the model of eqs. (2) and (16), the h-days VaR at confidence level θ of the portfolio eq. (13) is the value of $v(\theta,h)$ which solves*

$$1 - \theta = \int_0^\infty \Phi\left(\frac{v(\theta,h) - \hat{A}(h,s)}{\|\hat{B}(h,s)\|_2}\right) dF_{S_{t+h}}(s)s. \tag{19}$$

where eqs. (15) and (18) define $\hat{A}(h,s)$ and $\hat{B}(h,s)$.

Note that theorem 5.3 gives a semi-closed formula for the VaR in the neural-SDE model, with no need of further hypothesis. As a consequence, we can compute efficiently the VaR in this model without using any approximation.

We shall notice that since losses cannot be larger than today's position, the result $v(\theta, h)$ should always be higher than minus the current value of the portfolio, i.e. $v(\theta, h) \geq -\Pi_t(S_t, \xi_t)$. This condition holds true if and only if the P&Ls' distribution is null below $-\Pi_t(S_t, \xi_t)$, or equivalently if and only if the distribution of future prices is null below 0. In particular, it must hold

$$G \cdot \xi_{t+h} > -G_0$$

for any ξ_{t+h}. This condition does not seem to be guaranteed a priori. Indeed, it depends on how the parameters of the distribution of the ξ_t are calibrated. However, if the ξ_t are calibrated such that call prices always satisfy no arbitrage conditions, then in particular prices will always be positive.

5.1.2 Calls and Puts portfolio

In the case of portfolios with both call and put options, it is sufficient to re-write put options using the put-call-parity

$$P\Big(T-t, \log \frac{K}{F_t(T-t)}\Big) = C\Big(T-t, \log \frac{K}{F_t(T-t)}\Big) - \mathrm{DF}_t(T-t)\big(F_t(T-t) - K\big).$$

In this way, the relation in eq. (14) still holds redefining quantities A and B with elementary steps. In particular for every put option position $\pi P\big(T-t, \log \frac{K}{F_t(T-t)}\big)$ in the portfolio, $A(h,s)$ adds the term

$$\begin{aligned} s\pi \mathrm{DF}_t(T-(t+h)) f(T-(t+h)) \Big(G_0\Big(T-(t+h), \log \frac{K}{f(T-(t+h))s}\Big) + \\ + G\Big(T-(t+h), \log \frac{K}{f(T-(t+h))s}\Big) \cdot \xi_t - 1 \Big) + \pi \mathrm{DF}_t(T-(t+h))K, \end{aligned}$$

with $\Pi_t(S_t, \xi_t)$ updating its value with the added puts, while $B(h,s)$ adds

$$s\pi \mathrm{DF}_t(T-(t+h)) f(T-(t+h)) G\Big(T-(t+h), \log \frac{K}{f(T-(t+h))s}\Big).$$

5.2 Closed formula for the short term VaR

In this section we start by proving that the VaR in the neural-SDE model for option prices is of the form $u(\theta)\sqrt{h}$ asymptotically with h. This formulation reflects empirical results and standard models adopted in industry and it is consistent with eq. (Short-term formula). Then, we state the main result of this section computing the explicit form of the function $u(\theta)$.

Firstly, we look at the form of the function $v(\theta, h)$ when h is small.

Lemma 5.4. *Under the model of eqs. (2) and (16), the h-days VaR at confidence level θ of the portfolio eq. (13) is asymptotic to $\sqrt{h}$ for h going to 0:*

$$VaR_{\theta,t}(h) = u(\theta)\sqrt{h} + o\big(\sqrt{h}\big)$$

for a certain function $u(\theta)$ not depending on h.

We give the proof in appendix A.

Our next result uses the quantities

$$\begin{aligned} c_t &:= S_t\beta_t \sum_i \pi_i \mathrm{DF}_t(T_i - t) f(T_i - t) \times \\ &\times \Big[G_0\Big(T_i - t, \log \frac{K_i}{f(T_i - t)S_t}\Big) + G\Big(T_i - t, \log \frac{K_i}{f(T_i - t)S_t}\Big) \cdot \xi_t - \mathbb{1}_{\mathrm{Puts}}(i) + \\ &- \partial_k \Big(G_0\Big(T_i - t, \log \frac{K_i}{f(T_i - t)S_t}\Big) + G\Big(T_i - t, \log \frac{K_i}{f(T_i - t)S_t}\Big) \cdot \xi_t \Big) \Big] + \\ &+ B(0, S_t) \cdot \sigma_t \cdot P_{S,\xi,t}, \\ q_t &:= \big\| B(0, S_t) \cdot \sigma_t \cdot b_t \big\|_2, \end{aligned} \tag{20}$$

where $\mathbb{1}_{\mathrm{Puts}}(i)$ is 1 if the index i refers to a put, otherwise it is null, and

$$B(0, S_t) = S_t \sum_i \pi_i \mathrm{DF}_t(T_i - t) f(T_i - t) G\Big(T_i - t, \log \frac{K_i}{f(T_i - t)S_t}\Big).$$

Remark 5.5. *It is easy to prove that c_t and q_t can be written with the alternative expressions:*

$$\begin{aligned} c_t &= S_t\beta_t \frac{d}{dS_t}\Pi_t(S_t, \xi_t) + \nabla_{\xi_t}\Pi_t(S_t, \xi_t)^T \cdot \sigma_t \cdot P_{S,\xi,t}, \\ q_t &= \|\nabla_{\xi_t}\Pi_t(S_t, \xi_t)^T \cdot \sigma_t \cdot b_t\|_2. \end{aligned}$$

We can now state the main result of this section.

Proposition 5.6. *Under the model of eqs. (2) and (16), the h-days VaR at confidence level θ of the portfolio eq. (13) is*

$$VaR_{\theta,t}(h) = \Phi^{-1}\big(1 - \theta\big)\sqrt{c_t^2 + q_t^2}\sqrt{h} + o\big(\sqrt{h}\big) \tag{21}$$

where c_t and q_t are defined in eq. (20).

The proof is given in appendix B.

Corollary 5.7. *Under the model of eqs. (2) and (16) with $d = 1$, the h-days VaR at confidence level θ of the portfolio eq. (13) is*

$$\begin{aligned} VaR_{\theta,t}(h) &= \Phi^{-1}\big(1 - \theta\big) \times \\ &\times \sqrt{\Big(S_t\beta_t \frac{d}{dS_t}\Pi_t(S_t, \xi_t)\Big)^2 + \Big(\sigma_t \frac{d}{d\xi_t}\Pi_t(S_t, \xi_t)\Big)^2 + 2P_{S,\xi,t}S_t\beta_t\sigma_t \frac{d}{dS_t}\Pi_t(S_t, \xi_t)\frac{d}{d\xi_t}\Pi_t(S_t, \xi_t)}\sqrt{h} + \\ &+ o\big(\sqrt{h}\big). \end{aligned}$$

Observe that this is compatible with the Stochastic Volatility model's VaR in eq. (6).

5.2.1 Normal distribution for S_{t+h}

Theorem 5.2 still holds considering normal increments for S_{t+h}:

$$S_{t+h} = S_t\Big(1 + \alpha_t h + \beta_t(W_{0,t+h} - W_{0,t})\Big),$$

appropriately redefining the quantity $m_t(s)$. Indeed, in this case $m_t(s)$ becomes $\mu_t h + \frac{s - S_t(1+\alpha_t h)}{S_t\beta_t}\sigma_t \cdot P_{S,\xi,t}$.

The results in theorems 5.4 and 5.6 are exactly the same as for the log-normal case. Indeed, the probability density function of S_{t+h} is

$$p_{S_{t+h}}(s) = \frac{1}{S_t\beta_t\sqrt{2\pi h}} \exp\left(-\left(\frac{s - S_t(1+\alpha_t h)}{S_t\beta_t\sqrt{2h}}\right)^2\right)$$

and the conclusion can be easily attained as in the previous case, using the transformation $y = \frac{s - S_t(1+\alpha_t h)}{S_t\beta_t\sqrt{h}}$ as done in appendix B.

5.2.2 t-Student distribution for S_{t+h}

As in section 4.2.1, we consider the case where the increments of the underlier S_t follow a t-Student distribution, while increments of the implied volatility σ_t are Gaussian conditional to S_{t+h}.

Theorem 5.2 holds true in the Gaussian case because the random variable $Z = \Delta W_t - \Delta W_{0,t} P_{S,\xi}$ is still Gaussian and its decorrelation with $\Delta W_{0,t}$ implies its independence. However, in this case $\Delta W_{0,t}$ is substituted with the t-Student T_t and we cannot derive the same result. We need then some additional hypothesis to derive the equivalent of theorem 5.3 when S_{t+h} is a t-Student, and in particular we shall take eq. (17) as granted a priori.

Lemma 5.8. *Consider the model of eq. (2) and the hypothesis that*

$$S_{t+h} = S_t(1 + \alpha_t h + \beta_t T_{t+h})$$

where $T_{t+h} \in \mathbb{R}$ is a t-Student with ν_t degrees of freedom, null mean and variance equal to h, and $\xi_{t+h} - \xi_t$ conditional to $S_{t+h} = s$ is a Gaussian random variable with mean $m_t(s)$ and covariance matrix V_t with

$$m_t(s) = \mu_t h + \frac{s - S_t(1+\alpha_t h)}{S_t\beta_t}\sigma_t \cdot P_{S,\xi,t},$$
$$V_t = h(\sigma_t \cdot b_t) \cdot (\sigma_t \cdot b_t)^T,$$

and $b_t \in \mathbb{R}^{d\times d}$ is a matrix such that $b_t \cdot b_t^T = P_{\xi,t} - P_{S,\xi,t} \cdot P_{S,\xi,t}^T$, for certain parameters μ_t, $P_{S,\xi,t} \in \mathbb{R}^d$, σ_t, $P_{\xi,t} \in \mathbb{R}^{d\times d}$. Then the h-days VaR at confidence level θ of the portfolio eq. (13) is the value of $v(\theta, h)$ which solves

$$1 - \theta = \int_0^\infty \Phi\left(\frac{v(\theta,h) - \hat{A}(h,s)}{\|\hat{B}(h,s)\|_2}\right) dF_{S_{t+h}}(s)s.$$

where eqs. (15) and (18) define $\hat{A}(h,s)$ and $\hat{B}(h,s)$.

The calibration of the parameters μ_t, $P_{S,\xi,t}$, σ_t and $P_{\xi,t}$ is practically difficult if performed from the distribution of $\xi_{t+h} - \xi_t$ conditional to S_{t+h}. However, as in section 4.2.1, we can recover the moments of the marginal distribution, allowing an easy calibration of the parameters based on the historical mean and covariances of $\xi_{t+h} - \xi_t$.

Corollary 5.9. *Under the model*

$$S_{t+h} = S_t(1 + \alpha_t h + \beta_t T_{t+h})$$

where $T_{t+h} \in \mathbb{R}$ is a t-Student with ν_t degrees of freedom, null mean and variance equal to h, if $\xi_{t+h} - \xi_t$ conditional to $S_{t+h} = s$ is a Gaussian random variable with mean $m_t(s)$ and covariance matrix V_t as in theorem 5.8, then

$$\begin{aligned} E[\xi_{t+h} - \xi_t] &= \mu_t h \\ Cov[\xi_{t+h} - \xi_t] &= h\sigma_t \cdot P_{\xi,t} \cdot \sigma_t^T \\ Cov\left[\begin{pmatrix} S_{t+h} \\ \xi_{t+h} - \xi_t \end{pmatrix}\right] &= \begin{pmatrix} \beta_t^2 S_t^2 h & \beta_t S_t h(\sigma_t \cdot P_{S,\xi,t})^T \\ \beta_t S_t h \sigma_t \cdot P_{S,\xi,t} & h\sigma_t \cdot P_{\xi,t} \cdot \sigma_t^T \end{pmatrix}. \end{aligned}$$

The result in theorem 5.4 still holds true in the t-Student case. The proof is equivalent to the one given in appendix A, with the adaptations regarding the distribution of S_{t+h} and $\xi_{t+h} - \xi_t$. These adaptations can be found in the proof of eq. (Short-term t-Student) in section 4.2.1. As a consequence, the function $v(\theta, h)$ is of the form $u(\theta)\sqrt{h} + o(\sqrt{h})$ for h small.

The probability density function of S_{t+h} is as in eq. (10). In this way, using the transformation $y = \frac{s - S_t(1+\alpha_t h)}{S_t \beta_t \sqrt{h}}$ and repeating the calculations in appendix B.1, we find eq. (23) with exactly the same values for c_t and q_t. We then look for the value of $u(\theta)$ such that

$$1 - \theta = E\left[P\left(\frac{q_t X + c_t Y}{\sqrt{c_t^2 + q_t^2}} \leq \frac{u(\theta)}{\sqrt{c_t^2 + q_t^2}}\right)\right].$$

The random variable $Z = \frac{q_t X + c_t Y}{\sqrt{c_t^2 + q_t^2}}$ is not Gaussian in this case since it is the sum of a standard Gaussian random variable and a standard t-Student random variable, which are independent. However, it still holds that the initial margin in the case of t-Student returns is

$$\mathrm{VaR}_{\theta,t}(h) = F_Z^{-1}(1-\theta)\sqrt{c_t^2 + q_t^2}\sqrt{h} + o(\sqrt{h}).$$

The empirical quantile of Z can be recovered as detailed in section 4.2.1.

The above observations lead us to the following.

Corollary 5.10. *Under the framework of theorem 5.8, the h-days VaR at confidence level θ of the portfolio eq. (13) is*

$$VaR_{\theta,t}(h) = F_Z^{-1}(1-\theta)\sqrt{c_t^2 + q_t^2}\sqrt{h} + o(\sqrt{h})$$

where c_t, q_t are defined in eq. (20) and $Z = \frac{qX + cY}{\sqrt{c_t^2 + q_t^2}}$ with X and Y independent and X standard Gaussian, Y standard t-Student with ν_t degrees of freedom.

6. Numerical experiments

6.1 Backtesting option portfolios

Before going into the numerical experiments, it is worth focusing on the specific issues arising while backtesting option portfolios. Indeed, options are contracts with a fixed strike and a fixed expiry, so that considering a backtest on a real fixed contract is awkward for two reasons: at maturity the option will expire and the backtest could not continue anymore; the option could become far OTM or ITM in time, completing losing interest in the market and becoming illiquid or even not traded anymore.

Moreover, in order to focus on a given risk (like the calendar spread one), and its adequate coverage by the margin model, it is better to consider a portfolio with a constant risk profile across time, and so constant specifications in terms of moneyness and time-to-maturity.

For this reason, options are generally backtested for fixed log-forward moneyness (or delta) and fixed time-to-maturity, rather than fixed contract. Of course, these desired options are not always available among the market quoted ones, so that two possibilities arise:

1. Considering the nearest in log-forward moneyness (or delta) and time-to-maturity real quoted option.
2. Considering synthetic option prices on the chosen fixed log-forward moneyness (or delta) and fixed time-to-maturity obtained via the model pricing criteria.

In the first case, the backtested portfolios will possibly change every day, depending on how much the ATM level has moved and on the rolling maturity. In the second case, the VaR estimations are compared to model P&Ls rather than real ones, so that if the calibration model is not good enough, backtesting results could be misleading.

In general, there is no preferred way to backtest option portfolios and CCPs may adopt both methodologies. We recommend to perform the two approaches for production, because the synthetic option prices, due to the complexity of the data treatments performed, may eventually not reflect fully faithfully the effective market returns when they are available, as used directly in the first approach above.

6.2 VaR formula in the Heston model

The VaR formula in eq. (6) can be applied to any Stochastic Volatility model. In this section we test it on the classic Heston model

$$\begin{aligned} dS_t &= \alpha_t S_t dt + \sqrt{\nu_t} S_t dW_{0,t} \\ d\nu_t &= \kappa(\theta - \nu_t)dt + \xi\sqrt{\nu_t} dW_t \\ dW_{0,t} dW_t &= \rho dt. \end{aligned}$$

In this case, eq. (6) has the form

$$\mathrm{VaR}_{\theta,t}(h) = \Phi^{-1}(1-\theta)\sqrt{S_t^2\nu_t\Big(\frac{d}{dS_t}\Pi_t(S_t,\nu_t)\Big)^2 + \xi^2\nu_t\Big(\frac{d}{d\nu_t}\Pi_t(S_t,\nu_t)\Big)^2 + 2\rho\xi S_t\nu_t\frac{d}{dS_t}\Pi_t(S_t,\beta_t)\frac{d}{d\nu_t}\Pi_t(S_t,\nu_t)}\sqrt{h}. \tag{22}$$

Firstly, we simulate one year (365 days) history of the process (S_t, ν_t) with a simple Euler scheme of the form

$$\begin{aligned} S_{t+\delta t} &= S_t\big(1 + \alpha_t\delta t + \sqrt{\nu_t}\sqrt{\delta t}X_0\big) \\ \nu_{t+\delta t} &= \nu_t + \kappa(\theta - \nu_t)\delta t + \xi\sqrt{\nu_t}\sqrt{\delta t}X \end{aligned}$$

where X_0, X are standard Gaussian random variables with correlation ρ. Differences are negligible using the log-formulation for S_t, i.e.

$$S_{t+\delta t} = S_t\exp\Big(\Big(\alpha_t - \frac{\nu_t}{2}\Big)\delta t + \sqrt{\nu_t}\sqrt{\delta t}X_0\Big),$$

or using a Milstein scheme instead of the Euler's one.

At this point, for different outright, calendar and butterfly portfolios, we compute daily prices using the semi-analytical formula for a call option in the Heston model described in [15]. We take null rates, so that the forward price coincides with the spot value and the discount factor is 1. Finally we compare real PnLs with 0.99-VaR estimations as in eq. (22) on different MPOR horizons. In order to compute portfolio derivatives with respect to the spot and the volatility of the spot, we use the average between the corresponding backward and forward finite differences:

$$\begin{aligned} \frac{d}{dS_t}\Pi_t(S_t,\nu_t) &\approx \frac{\Pi_t(S_t+\varepsilon,\nu_t) - \Pi_t(S_t-\varepsilon,\nu_t)}{2\varepsilon}, \\ \frac{d}{d\nu_t}\Pi_t(S_t,\nu_t) &\approx \frac{\Pi_t(S_t,\nu_t+\varepsilon) - \Pi_t(S_t,\nu_t-\varepsilon)}{2\varepsilon}. \end{aligned}$$

Note in particular that these quantities are *not* the Black-Scholes ones defined through the sensitivities of the Black-Scholes formula evaluated at the implied volatility corresponding to the Heston model price.

We use calibrated Heston parameters on S&P500 on December 2015 (see [5]), in particular $(\kappa, \sqrt{\theta}, \xi, \rho) = (6.169, 0.16168, 0.477, -0.781)$, and we start with initial values $S_0 = 2054$ and $\sqrt{\nu_0} = 0.15562$. For the Euler scheme, we choose a step δt of 10^{-1} days. The portfolios considered are outright calls with delta in $\{0.2, 0.35, \delta_{\mathrm{ATM}}, 0.65, 0.8\}$ and time-to-maturity in $\{30, 90, 180, 365\}$ days, and the resulting combinations of calendar spread and butterfly spread portfolios. In particular, a calendar spread is a portfolio with one short call at a fixed strike K and maturity T_1 and one long call with same strike K and maturity $T_2 > T_1$. For a fixed delta, the common strike of the two options is chosen to be the one related to the shortest maturity. A butterfly spread is composed of two long calls with deltas δ and $1-\delta$ respectively, and two short ATM calls. For butterfly spreads, we also test the deltas $0.1, 0.3, 0.4, 0.45$.

The number of tested portfolios is then: $5 \times 4 = 20$ outrights, $\binom{4}{2} \times 5 = 30$ calendar spreads, and $6 \times 4 = 24$ butterfly spreads; in total 74 portfolios.

For each tested portfolio, we compute the coverage ratio as the number of days where the model VaR covers the realized loss over the total number of tested days, and the size of losses as the ratio between the margin loss (difference between realized loss and model VaR) and the portfolio price. The average and the median

MPOR (days)	Coverage		Size of loss	
	Mean	Median	Mean	Median
1	0.9927	0.9945	0.0485	0.0204
2	0.9902	0.9945	0.0645	0.0353
3	0.9896	0.9945	0.0682	0.0422

Table 1: Average coverage and size of loss of eq. (22) on 74 option portfolios on simulated Heston data.

MPOR (days)	Coverage		Size of loss	
	Mean	Median	Mean	Median
1	0.9853	0.9863	0.0404	0.0194
2	0.9826	0.9822	0.0565	0.0368
3	0.9813	0.9808	0.0496	0.0255

Table 2: Average coverage and size of loss of the short-term model-free VaR eq. (Short-term formula) on 74 option portfolios on simulated Heston data, under the hypothesis of log-normal distribution of spot returns.

over all portfolios is displayed in table 1. The results are very satisfactory for all MPORs. Indeed, the average coverage meets the 0.99-VaR requirement and breaches are of a very small size below 7% of the portfolio value.

Similarly, we compute the initial margin using the short-term model-free formula described in section 4.2. In particular, we simulate 5 years history of an Heston process with same parameters as in the previous test and compute the initial margin for the same portfolios on the last year's observations (the previous history is used to calibrate parameters). The formula used is then eq. (Short-term formula), where the delta and the vega Greeks are the Black-Scholes ones, the spot volatility, the vol-of-vol and the correlation between risk factors are computed with the EWMA specification, and the derivative of the implied volatility with respect to the log-forward moneyness in computed via finite differences. The implied volatility point used to compute the EWMA correlation is the 1M ATM point.

As in the previous test, we compute the average coverage and size of loss for each tested MPOR. Results are shown in table 2. Results are less conservative than in the previous test since the coverage is around 0.983. However, the size of loss is still very small compared to the portfolio value, and actually smaller than in the previous test.

Results can actually be improved using the hypothesis of a t-Student distribution for spot returns and redefining the VaR for MPOR h as in eq. (Short-term t-Student) where the distribution of Z is obtained empirically as explained in section 4.2.1. Table 3 shows results when considering 5 degrees of freedom in the t-Student distribution of spot returns. As expected, results are more conservative than the normal case and satisfy the 0.99 coverage requirement.

MPOR (days)	Coverage		Size of loss	
	Mean	Median	Mean	Median
1	0.9907	0.9918	0.0362	0.0164
2	0.9886	0.9918	0.0596	0.0364
3	0.9889	0.9945	0.0485	0.0291

Table 3: Average coverage and size of loss of the short-term model-free VaR eq. (Short-term t-Student) on 74 option portfolios on simulated Heston data, under the hypothesis of t-Student with 5 degrees of freedom distribution of spot returns.

6.3 Coverage performances of the short-term model-free VaR

In this section we show the results of coverage of the short-term model-free 0.99 VaR formula in section 4.2 compared with the classical FHS model described in section 3.1.

We use a database of S&P500 data provided to Zeliade by the Clearify project[4] on end of the day option prices. Firstly, we clean the rough data removing options with 0 volume and use the put-call-parity on mid prices to extrapolate forward and discount factors for each quoted maturity having at least two put-call couples. Then, we remove all calls not satisfying the arbitrage bounds

$$\mathrm{DF}_t(T)(F_t(T)-K)^+ \leq C_t(T,K) \leq \mathrm{DF}_t(T)F_t(T),$$

and all puts not satisfying

$$\mathrm{DF}_t(T)(K-F_t(T))^+ \leq P_t(T,K) \leq \mathrm{DF}_t(T)K.$$

At this points, puts are transformed into calls and all the following computations are performed for call prices.

In order to get normalized historical prices, we compute synthetic historical prices on a fixed grid of time-to-maturity and log-forward moneyness. At this aim, we observe that market data is generally dense around the ATM point for short maturities and it spreads out with increasing expiry. For this reason, the most cunning fixed grid should be in delta, so we identify 17 delta points from 0.015 to 0.985 and compute the corresponding log-forward moneyness for a symbolic volatility of 0.1, over the grid of time-to-maturities of 2, 5, 10, 21, 42, 63, 126 and 252 days. Observe that in this way the grid is not constant in log-forward moneyness for different time-to-maturities. At this point, we firstly compute implied volatilities based on real prices. Then, we interpolate linearly on the log-forward moneyness direction since data is dense enough. The interpolation on the time-to-maturity direction is done on the implied total variances (squared implied volatilities times the time-to-maturity) adding synthetic points for the zero maturity equal to 0 and then interpolating linearly. The interpolated prices are then corrected to avoid static arbitrage as described in [6].

For comparison, we implemented a second interpolation scheme firstly adding synthetic points for the zero maturity (setting prices equal to their intrinsic values) and for extreme moneyness (with prices equal to the discounted forward on the left and null prices on the right); then normalizing all prices by their discounted forward; finally using monotonic cubic splines on the log-forward moneyness direction and linear interpolation on the time-to-maturity direction. The final results reported in this section do not significantly change.

We consider two portfolios: the first one is an ATM calendar spread between maturities of 1M and 6M; while the second one is a butterfly spread on maturity 3M and moneyness 0.9, 1, 1.1. We compute the VaR for an MPOR horizon of 1 day and a confidence level $\theta = 0.99$.

The short-term model-free VaR is obtained computing:

1. The spot volatility β_t via a EWMA volatility algorithm with decay factor 0.97 on spot log-returns, divided by the square-root of the daily step;

[4] Clearify is a collaboration between Zeliade and the Imperial College Mathematical Finance Department funded by an Imperial Faculty of Natural Sciences Strategic Research Funding Award.

2. The correlation ρ_t via a EWMA correlation on spot log-returns and ATM implied volatility absolute returns;

3. The vol-of-vol $\zeta_t(k,\tau)$ as F times the 1M ATM vol-of-vol obtained as a EWMA volatility on implied volatility absolute returns, divided by the square root of the daily step. In order to be conservative enough, the F factor is 1.1 times the quantile 0.9 of 5 years history of ratios between the (k,τ) vol-of-vol and the 1M ATM vol-of-vol;

4. Delta and vega quantities as the Black-Scholes deltas and vegas evaluated at the option implied volatility;

5. The derivative of the smile with respect to the log-forward moneyness as the derivative of the interpolated smile via B-splines.

Since we consider log-normal spot returns, we use the VaR formulation with normal quantiles of eq. (Short-term formula).

The FHS risk factors are the spot prices and the 17×8 implied volatility grid points for fixed log-forward moneyness and time-to-maturity. We consider log-returns for the former risk factors and absolute returns for the latter ones. The volatility of risk factors' returns is computed via EWMA with decay factor 0.97. The future discount factors and forward values are obtained under the assumption of constant risk-free rates in the MPOR horizon.

We backtest the short-term model-free VaR and the FHS VaR against the synthetic P&Ls as explained in the second approach of section 6.1.

Figure 1 shows the VaR patterns for the tested portfolios. We can see that the short-term model-free VaR has more breaches then the FHS VaR, however these are of small size and can be entirely removed setting a vol-of-vol factor $F = 1$. Alternatively, one could consider the t-Student framework described in section 4.2.1. The most noticeable feature of the short-term model-free VaR is its regularity compared to the FHS VaR. In particular, the short-term model-free VaR behaves as we expect after large moves in realized P&Ls, and it also softens its behavior, without big jumps. On the contrary, the FHS VaR is not as consistent (and sometimes seems to move without following market patterns).

Furthermore, the short-term model-free VaR looks more smooth, i.e. less procyclical. To prove this sentence, we compute the peak-to-trough ratio on the whole 2019 dates and the average n-day procyclicality measure (in percentage) for n equal to 1, 5, 10 and 20 days. In particular, the two quantities are computed as

$$\text{Peak-to-trough} = \frac{\max_t\big(-\text{VaR}_{0.99,t}(h)\big)}{\min_t\big(-\text{VaR}_{0.99,t}(h)\big)}$$
$$n\text{-day } \% = \max_t\left(\frac{-\text{VaR}_{0.99,t}(h)}{-\text{VaR}_{0.99,t-n}(h)} - 1\right) \times 100$$

where t ranges in the whole 2019 and we choose an MPOR $h = 1$. Results are displayed in table 4. We see that except for the 1-day procyclicality measure in the calendar spread portfolio, all other procyclicality measures for both portfolios are largely smaller for the short-term model-free VaR, i.e. the latter model is less procyclical than the FHS VaR.

	Calendar ATM 1M-6M		Butterfly 3M m(0.9, 1, 1.1)	
	FHS	Short-term model-free	FHS	Short-term model-free
Peak-to-trough	5.32	3.27	2.26	2.77
1-day %	42.17	45.38	31.34	17.95
5-day %	90.11	89.04	70.47	44.23
10-day %	143.88	109.28	106.69	52.17
20-day %	139.75	118.53	90.35	44.56

Table 4: Comparison between FHS VaR and short-term model-free VaR peak-to-trough ratio and average percentage n-day procyclicality measure for a calendar spread and a butterfly spread portfolios.

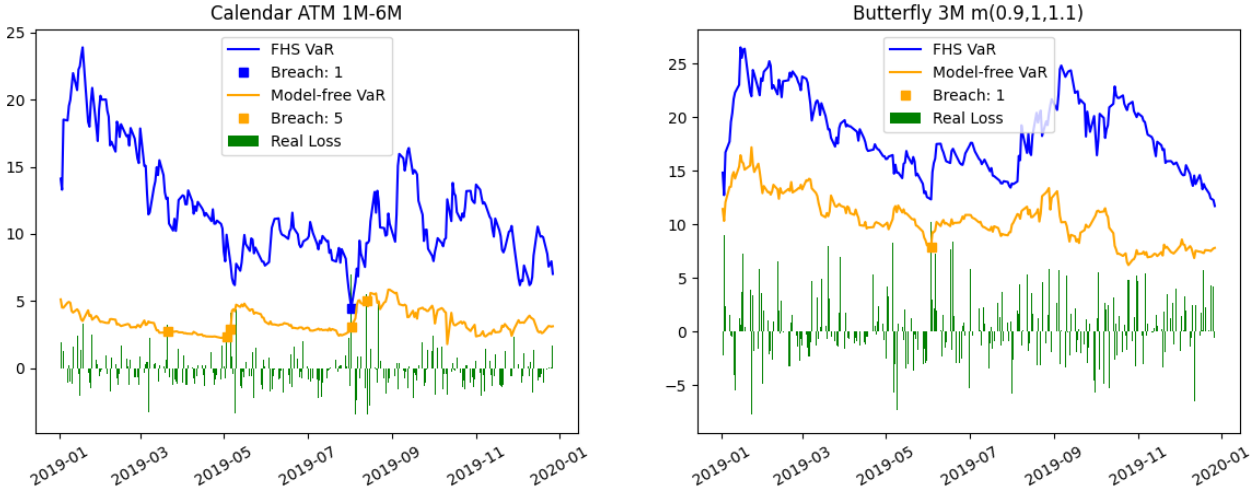

Figure 1: Margins obtained with the FHS algorithm (blue) and the short-term model-free VaR (orange), for a calendar spread ATM 1M-6M portfolio (left) and a butterfly spread 3M with moneyness $(0.9, 1, 1.1)$ *portfolio (right).*

6.4 Practical implementation of the neural-SDE model

In this section we consider the neural-SDE model for normalized option prices, and in particular we compare VaR estimations obtained as empirical quantiles on simulations (see section 3.4.1) and VaR values resulting from the approximated closed formula of theorem 5.6.

We use the same data as in section 6.3 and compute forward and discount factors similarly.

Before calibrating the G factors on the time-to-maturity, log-forward moneyness grid identified in section 6.3, historical prices must be interpolated on such a grid. With this aim, we use the same interpolation/extrapolation algorithm consisting in implied volatility's linear interpolation on the space dimension and total variances' linear interpolation in the time direction.

Now, for a fixed date t, the past 5 years historical data (1260 observations) is used to calibrate the G and ξ factors. We choose to calibrate the factors ξ_s for $s \leq t$ in the most efficient way, only looking at the statistical accuracy. In particular, for the fixed time-to-maturity and log-forward moneyness grid, we choose G_0 as the average historical prices and the remaining G_i as the principal components of the residuals between prices and values of G_0. See [7] for a detailed description of the calibration algorithm. The calibration code that

we use is the one implemented in the Github repository of the cited article.[5] Based on calibration accuracy and process time, we decide to take 2 statistical accuracy factors ξ_s.

At this point we have the constant factors G on the fixed grid and an history of factors ξ_s. Furthermore, the calibration uses neural networks to estimate the distributions of S and ξ as in eq. (3). In particular, we have today's parameters α_t, β_t, μ_t and σ_t, and for any value of S and ξ, the neural network can predict the corresponding parameters. Observe that we take the covariance matrix $P_t = I_d$ as in [7].

For future computations, the matrix G has to be interpolated outside the fixed time-to-maturity and log-forward moneyness grid. To do so, we firstly interpolate normalized call prices on the target couple (τ, k) for every historical past day. The interpolation is performed as in the preparation of the initial database, computing implied volatilities and interpolating them linearly on the space direction and linearly in total variance on the time direction. Once all history on (τ, k) is retrieved, the $G(\tau, k)$ factors are the intercept and the coefficients of the linear regression of prices along the history of ξs.

At present, the two VaR calculation methodologies can be implemented. We test the same two portfolios of section 6.3 consisting of an ATM calendar spread 1M-6M, and a butterfly spread on maturity 3M and moneyness 0.9, 1, 1.1. We consider an MPOR of 1 day and a VaR confidence level $\theta = 0.99$.

For both VaR methodologies, we work under the assumption of constant risk-free rates in the MPOR horizon. Then, the discount factor $\mathrm{DF}_{t+h}(\tau)$ in h days on a time-to-maturity τ is equal to today's discount factor $\mathrm{DF}_t(\tau)$, while the forward value $F_{t+h}(\tau)$ in h days on a time-to-maturity τ becomes $\frac{S_{t+h}}{S_t} F_t(\tau)$.

For the empirical VaR, simulations are performed under the hypothesis that parameters α, β, μ and σ are constant between t and $t+h$, following Euler's scheme in eq. (16). Starting with the estimation of today's parameters, values of S_{t+h} and ξ_{t+h} are simulated 10000 times. Future normalized prices are computed using the model relation eq. (2) with the G factors evaluated on time-to-maturity $\tau - h$ and log-forward moneyness $k + \log \frac{F_t(\tau)}{F_{t+h}(\tau-h)}$, and the estimated values of ξ_{t+h}. Once the simulated normalized call prices are computed, they are re-denormalized multiplying by $\mathrm{DF}_t(\tau - h)$ and $F_{t+h}(\tau - h)$. The final VaR is the $1 - \theta$ empirical quantile of P&Ls obtained as difference of simulated future prices and today price.

In the case of VaR obtained via approximated closed formula, in order to be consistent with the empirical VaR, the distribution of S_{t+h} is taken to be a log-normal distribution, so that the used closed formula coincides with eq. (21). Derivatives of the components of G with respect to k are computed as the average between backward and forward finite differences.

We plot the percentage ratio between the absolute difference between the two VaR estimations and the portfolio value along year 2019 in fig. 2. We can see that the empirical VaR and the approximated formula eq. (21) for the VaR generally have a very small error (about 4% for the calendar portfolio and 1% for the butterfly portfolio), with some higher picks which could reach the 10% of the portfolio. This is due to the fact that the approximated formula is less procyclical than the empirical one and reacts slower to market changes. All in all, the results confirm the consistency of hypothesis in theorem 5.6.

[5] https://github.com/vicaws/neuralSDE-marketmodel

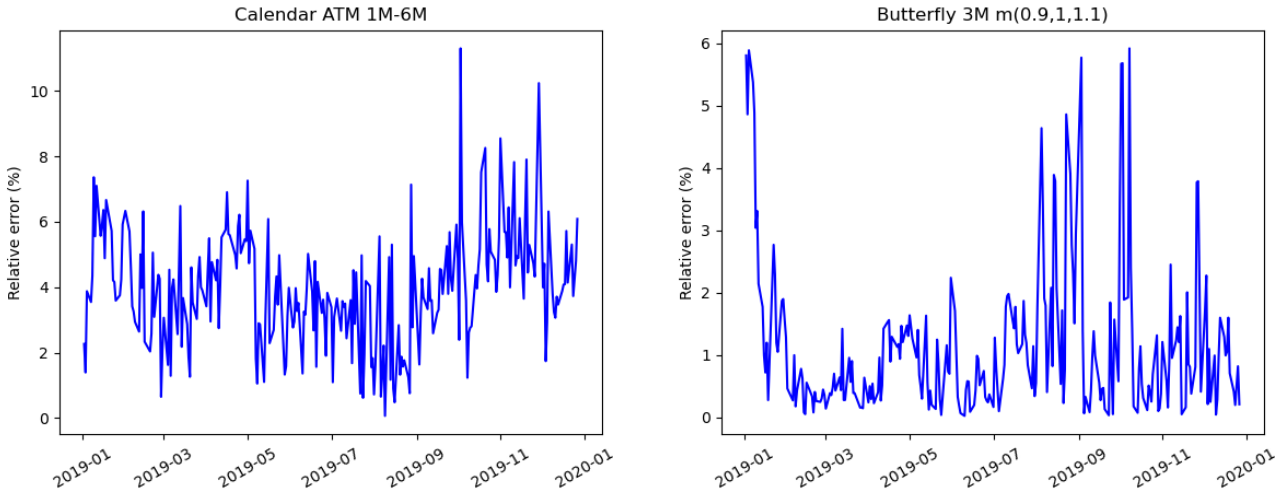

Figure 2: Percentage relative error between the empirical neural-SDE VaR calculated as in section 3.4.1 and the approximated closed formula VaR in theorem 5.6, for a calendar spread ATM 1M-6M portfolio (left) and a butterfly spread 3M with moneyness $(0.9, 1, 1.1)$ *portfolio (right).*

7. Conclusion

We summarize and analyze the methodologies that CCPs currently use for the initial margin of option portfolios. In particular, we compute a quasi-explicit formula for the VaR of option portfolios in the neural-SDE model of [7], and propose a closed asymptotic short-term model-free formula for the VaR at small time horizons.

Based on the numerical experiments that we conduct, we are confident that this new short-term model-free formula could be considered as a candidate for the core component of IM methodologies for option portfolios, duly complemented by Short Option Minimum and Stress Historical VaR components.

A. Proof of lemma 5.4

Let us consider the distribution of $\frac{\text{P\&L}}{\sqrt{h}}$ given in eqs. (14) and (16). We write $W_{0,t+h} - W_{0,t} = \sqrt{h}Y$ and $W_{t+h} - W_t = \sqrt{h}X$ where Y is a standard Gaussian random variable and X is a d-dimensional Gaussian random variable with correlation matrix P_t not depending on h.

We first consider the term $\frac{A(h,S_{t+h})}{\sqrt{h}}$ and look at its limit for h going to 0. Since S_{t+h} goes to S_t, both the numerator and the denominator go to 0. We then use L'Hôpital's rule to develop the limit. The derivative of S_{t+h} with respect to h is

$$S_{t+h}\left(\alpha_t - \frac{\beta_t^2}{2} + \frac{\beta_t}{2\sqrt{h}}Y\right)$$

so that the derivative of $A(h, S_{t+h})$ with respect to h is

$$\begin{aligned}
&S_{t+h}\Bigg[\left(\alpha_t - \frac{\beta_t^2}{2} + \frac{\beta_t}{2\sqrt{h}}Y\right)\sum_i \pi_i \mathrm{DF}_t(T_i-(t+h))f(T_i-(t+h))\Big(G_0\Big(T_i-(t+h), \log\frac{K_i}{f(T_i-(t+h))S_{t+h}}\Big)+\\
&\quad + G\Big(T_i-(t+h), \log\frac{K_i}{f(T_i-(t+h))S_{t+h}}\Big)\cdot\xi_t - \mathbb{1}_{\text{Puts}}(i)\Big)+\\
&+\sum_i \pi_i \frac{d}{dh}\Big(\mathrm{DF}_t(T_i-(t+h))f(T_i-(t+h))\Big)\Big(G_0\Big(T_i-(t+h), \log\frac{K_i}{f(T_i-(t+h))S_{t+h}}\Big)+\\
&\quad + G\Big(T_i-(t+h), \log\frac{K_i}{f(T_i-(t+h))S_{t+h}}\Big)\cdot\xi_t - \mathbb{1}_{\text{Puts}}(i)\Big)+\\
&-\sum_i \pi_i \mathrm{DF}_t(T_i-(t+h))f(T_i-(t+h))\Big(\partial_\tau G_0\Big(T_i-(t+h), \log\frac{K_i}{f(T_i-(t+h))S_{t+h}}\Big)+\\
&\quad + \partial_\tau G\Big(T_i-(t+h), \log\frac{K_i}{f(T_i-(t+h))S_{t+h}}\Big)\cdot\xi_t\Big)+\\
&-\sum_i \pi_i \mathrm{DF}_t(T_i-(t+h))f(T_i-(t+h))\Big(\partial_k G_0\Big(T_i-(t+h), \log\frac{K_i}{f(T_i-(t+h))S_{t+h}}\Big)+\\
&\quad + \partial_k G\Big(T_i-(t+h), \log\frac{K_i}{f(T_i-(t+h))S_{t+h}}\Big)\cdot\xi_t\Big)\times\\
&\quad\times\left(-\frac{\partial_\tau f(T_i-(t+h))}{f(T_i-(t+h))} + \alpha_t - \frac{\beta_t^2}{2} + \frac{\beta_t}{2\sqrt{h}}Y\right)\Bigg]
\end{aligned}$$

where $\mathbb{1}_{\text{Puts}}(i)$ is 1 if the index i refers to a put, otherwise it is null. This quantity explodes for h going to 0 with speed $\frac{\gamma Y}{2\sqrt{h}}$ where

$$\begin{aligned}
\gamma = S_t\beta_t \sum_i \pi_i \mathrm{DF}_t(T_i-t)f(T_i-t)\times\\
\times\Big[G_0\Big(T_i-t, \log\frac{K_i}{f(T_i-t)S_t}\Big) + G\Big(T_i-t, \log\frac{K_i}{f(T_i-t)S_t}\Big)\cdot\xi_t - \mathbb{1}_{\text{Puts}}(i)+\\
-\partial_k\Big(G_0\Big(T_i-t, \log\frac{K_i}{f(T_i-t)S_t}\Big) + G\Big(T_i-t, \log\frac{K_i}{f(T_i-t)S_t}\Big)\cdot\xi_t\Big)\Big].
\end{aligned}$$

This means that the ratio $\frac{A(h,S_{t+h})}{\sqrt{h}}$ tends to γY, where γ does not depend on h.

We now look at the term $B(h, S_{t+h}) \cdot \frac{\xi_{t+h}-\xi_t}{\sqrt{h}}$. Firstly, the limit of $B(h, S_{t+h})$ for h going to 0 is

$$S_t \sum_i \pi_i \mathrm{DF}_t(T_i - t) f(T_i - t) G\Big(T_i - t, \log \frac{K_i}{f(T_i - t)S_t}\Big)$$

which is simply a sum of call surfaces and, in general, is different from 0. Secondly, the ratio $\frac{\xi_{t+h}-\xi_t}{\sqrt{h}}$ is equal to $\mu_t\sqrt{h} + \sigma_t \cdot X$ and goes to $\sigma_t \cdot X$ when h tends to 0.

All in all, $\frac{\mathrm{P\&L}}{\sqrt{h}}$ tends to $\gamma Y + B(0, S_t) \cdot \sigma_t \cdot X$ almost surely, hence in law. Then, since the cumulative density function of the random variable $\gamma Y + B(0, S_t) \cdot \sigma_t \cdot X$ has a continuous inverse, all the quantiles of $\frac{\mathrm{P\&L}}{\sqrt{h}}$ converge to the corresponding quantiles of $\gamma Y + B(0, S_t) \cdot \sigma_t \cdot X$ as h tends to 0.

The h-days VaR with confidence level θ is the value of $v(\theta, h)$ solving

$$\begin{aligned} 1 - \theta &= P(\mathrm{P\&L} \leq v(\theta, h)) \\ &= P\bigg(\frac{\mathrm{P\&L}}{\sqrt{h}} \leq \frac{v(\theta, h)}{\sqrt{h}}\bigg). \end{aligned}$$

From the above discussion, we have $v(\theta, h) = u(\theta)\sqrt{h} + o(\sqrt{h})$, where $u(\theta) = F^{-1}_{\gamma Y + B(0,S_t)\cdot\sigma_t \cdot X}(1 - \theta)$.

B. Proof of theorem 5.6

From the definition in eq. (16), S_{t+h} has a log-normal distribution with density

$$p_{S_{t+h}}(s) = \frac{1}{s\beta_t\sqrt{2\pi h}} \exp\left(-\left(\frac{\log\frac{s}{S_t} - \left(\alpha_t - \frac{\beta_t^2}{2}\right)h}{\beta_t\sqrt{2h}}\right)^2\right).$$

We look at the RHS of eq. (19) when h goes to 0.

We can re-write the integral using a change of variable $y = \frac{\log\frac{s}{S_t} - \left(\alpha_t - \frac{\beta_t^2}{2}\right)h}{\beta_t\sqrt{h}}$ as

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) \Phi\left(\frac{v(\theta,h) - \hat{A}(h, s(h,y))}{\|\hat{B}(h, s(h,y))\|_2}\right) dy$$

where $s(h,y) = S_t \exp\left(y\beta_t\sqrt{h} + \left(\alpha_t - \frac{\beta_t^2}{2}\right)h\right)$. Observe that the integrand is dominated by the integrable function $\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{y^2}{2}\right)$.

If we prove that for h going to 0 the integrand converges pointwise to a function of the form $\Phi\left(\frac{-c_t y + u(\theta)}{q_t}\right)$ where c_t and q_t do not depend on y, for the Lebesgue's dominated convergence theorem the whole integral converges to

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) \Phi\left(\frac{-c_t y + u(\theta)}{q_t}\right) dy. \tag{23}$$

Assuming the proof of the convergence is done (see appendix B.1), we can pick-up a pair of independent standard Gaussian random variables X, Y and write the latter expression as

$$E\left[\mathbb{1}\left(X \leq \frac{-c_t Y + u(\theta)}{q_t}\right)\right].$$

The random variable $Z = \frac{q_t X + c_t Y}{\sqrt{c_t^2 + q_t^2}}$ has a standard normal distribution, so that the latter quantity is equal to

$$E\left[\Phi\left(\frac{u(\theta)}{\sqrt{c_t^2 + q_t^2}}\right)\right] = \Phi\left(\frac{u(\theta)}{\sqrt{c_t^2 + q_t^2}}\right).$$

Then, we can finally recover the expression of the initial margin $\mathrm{VaR}_{\theta,t}(h)$ as in eq. (21).

B.1 Proof of the pointwise convergence

The pointwise convergence of

$$\Phi\left(\frac{v(\theta,h) - \hat{A}(h, s(h,y))}{\|\hat{B}(h, s(h,y))\|_2}\right)$$

to a function of the form $\Phi\left(\frac{-c_t y + u(\theta)}{q_t}\right)$ can be proved firstly observing that we can work under hypothesis of continuous functions, given that normalized prices can be considered to be continuous in time-to-maturity and log-forward moneyness. Also, since we are looking at the limit when h is small, we can use theorem 5.4 and substitute $v(\theta,h)$ with $u(\theta)\sqrt{h} + o(\sqrt{h})$.

Firstly observe that the matrix b_t does not depend on h, and $\|\hat{B}(h, s(h,y))\|_2 = \sqrt{h}\|B(h, s(h,y)) \cdot \sigma_t \cdot b_t\|_2$ where

$$B(h, s(h,y)) = s(h,y) \sum_i \pi_i \mathrm{DF}_t(T_i - (t+h)) f(T_i - (t+h)) G\Big(T_i - (t+h), \log \frac{K_i}{f(T_i - (t+h))s(h,y)}\Big).$$

As shown in appendix A, the limit of $B(h, s(h,y))$ for h going to 0 is

$$S_t \sum_i \pi_i \mathrm{DF}_t(T_i - t) f(T_i - t) G\Big(T_i - t, \log \frac{K_i}{f(T_i - t)S_t}\Big)$$

which is different from 0. Then, the ratio $\frac{u(\theta)\sqrt{h} + o(\sqrt{h})}{\|\hat{B}(h, s(h,y))\|_2}$ goes to

$$\frac{u(\theta)}{\|B(0, S_t) \cdot \sigma_t \cdot b_t\|_2} = \frac{u(\theta)}{q_t}$$

in 0.

We now consider the ratio $\frac{\hat{A}(h, s(h,y))}{\sqrt{h}\|B(h, s(h,y)) \cdot \sigma_t \cdot b_t\|_2}$. The function $\hat{A}(h, s(h,y))$ is in turn the sum between $A(h, s(h,y))$ and $B(h, s(h,y)) \cdot m_t(s(h,y))$. The latter term is equal to $B(h, s(h,y)) \cdot (\mu_t h + y\sigma_t \cdot P_{S,\xi,t}\sqrt{h})$, so that its ratio with $\sqrt{h}\|B(h, s(h,y)) \cdot \sigma_t \cdot b_t\|_2$ goes to $\frac{B(0,S_t) \cdot \sigma_t \cdot P_{S,\xi,t}}{q_t} y$.

We shall now focus on the ratio $\frac{A(h, s(h,y))}{\sqrt{h}\|B(h, s(h,y)) \cdot \sigma_t \cdot b_t\|_2}$. Since both the numerator and denominator go to 0 with h, we use L'Hôpital's rule to develop the limit.

The derivative of $B(h, s(h,y))$ with respect to h is

$$\begin{aligned}
&s(h,y)\Bigg[\left(\frac{y\beta_t}{2\sqrt{h}} + \alpha_t - \frac{\beta_t^2}{2}\right) \sum_i \pi_i \mathrm{DF}_t(T_i - (t+h)) f(T_i - (t+h)) G\Big(T_i - (t+h), \log \frac{K_i}{f(T_i - (t+h))s(h,y)}\Big) + \\
&+ \sum_i \pi_i \frac{d}{dh}\Big(\mathrm{DF}_t(T_i - (t+h)) f(T_i - (t+h))\Big) G\Big(T_i - (t+h), \log \frac{K_i}{f(T_i - (t+h))s(h,y)}\Big) + \\
&- \sum_i \pi_i \mathrm{DF}_t(T_i - (t+h)) f(T_i - (t+h)) \partial_\tau G\Big(T_i - (t+h), \log \frac{K_i}{f(T_i - (t+h))s(h,y)}\Big) + \\
&- \sum_i \pi_i \mathrm{DF}_t(T_i - (t+h)) f(T_i - (t+h)) \partial_k G\Big(T_i - (t+h), \log \frac{K_i}{f(T_i - (t+h))s(h,y)}\Big) \times \\
&\times \left(-\frac{\partial_\tau f(T_i - (t+h))}{f(T_i - (t+h))} + \frac{y\beta_t}{2\sqrt{h}} + \alpha_t - \frac{\beta_t^2}{2}\right)\Bigg]
\end{aligned}$$

and for h going to 0, it explodes with a speed of

$$\frac{1}{2\sqrt{h}} S_t y \beta_t \sum_i \pi_i \mathrm{DF}_t(T_i - t) f(T_i - t)\Big(G\Big(T_i - t, \log \frac{K_i}{f(T_i - t)S_t}\Big) - \partial_k G\Big(T_i - t, \log \frac{K_i}{f(T_i - t)S_t}\Big)\Big).$$

Similarly, the derivative of $A(h, s(h, y))$ with respect to h explodes with a speed of

$$\begin{aligned}
&\frac{1}{2\sqrt{h}} S_t y \beta_t \sum_i \pi_i \mathrm{DF}_t(T_i - t) f(T_i - t) \times \\
&\times \Big[G_0\Big(T_i - t, \log \frac{K_i}{f(T_i - t)S_t}\Big) + G\Big(T_i - t, \log \frac{K_i}{f(T_i - t)S_t}\Big) \cdot \xi_t - \mathbb{1}_{\mathrm{Puts}}(i) + \\
&- \partial_k \Big(G_0\Big(T_i - t, \log \frac{K_i}{f(T_i - t)S_t}\Big) + G\Big(T_i - t, \log \frac{K_i}{f(T_i - t)S_t}\Big) \cdot \xi_t \Big) \Big].
\end{aligned}$$

Doing the derivative of $\sqrt{h}\|B(h, s(h, y)) \cdot \sigma_t \cdot b_t\|_2$, we find

$$\frac{1}{2\sqrt{h}} \|B(h, s(h, y)) \cdot \sigma_t \cdot b_t\|_2 + \sqrt{h} \frac{\big(\frac{d}{dh} B(h, s(h, y)) \cdot \sigma_t \cdot b_t\big)^T \cdot \big(B(h, s(h, y)) \cdot \sigma_t \cdot b_t\big)}{\|B(h, s(h, y)) \cdot \sigma_t \cdot b_t\|_2},$$

and this explodes with the first term.

All in all, L'Hôpital's rule shows that the limit of $\frac{\hat{A}(h, s(h, y))}{\sqrt{h}\|B(h, s(h, y)) \cdot \sigma_t \cdot b_t\|_2}$ for h going to 0 is $\frac{c_t}{q_t} y$.

References

[1] Giovanni Barone-Adesi and Kostas Giannopoulos. Non parametric VaR techniques. Myths and realities. *Economic Notes*, 30(2):167–181, 2001.

[2] Giovanni Barone-Adesi, Kostas Giannopoulos, and Les Vosper. Var without correlations for nonlinear portfolios. *Journal of Futures Markets*, 1997.

[3] Christian Berg and Christophe Vignat. On the density of the sum of two independent Student t-random vectors. *Statistics & probability letters*, 80(13-14):1043–1055, 2010.

[4] Maxime Bergeron, Nicholas Fung, John Hull, Zissis Poulos, and Andreas Veneris. Variational autoencoders: A hands-off approach to volatility. *The Journal of Financial Data Science*, 4(2):125–138, 2022.

[5] Jiling Cao, Jeong-Hoon Kim, and Wenjun Zhang. Pricing variance swaps under hybrid CEV and stochastic volatility. *Journal of Computational and Applied Mathematics*, 386:113220, 2021.

[6] Samuel N Cohen, Christoph Reisinger, and Sheng Wang. Detecting and repairing arbitrage in traded option prices. *Applied Mathematical Finance*, 27(5):345–373, 2020.

[7] Samuel N Cohen, Christoph Reisinger, and Sheng Wang. Arbitrage-free neural-SDE market models. *arXiv preprint arXiv:2105.11053*, 2021.

[8] Samuel N Cohen, Christoph Reisinger, and Sheng Wang. Estimating risks of option books using neural-SDE market models. *arXiv preprint arXiv:2202.07148*, 2022.

[9] European Commission. Regulation (EU) 648/2012 of 4 July 2012 of the European Parliament and Council on OTC derivatives, central counterparties and trade repositories. *Official Journal of the European Union*, 2012.

[10] European Commission. Commission Delegated Regulation (EU) No 153/2013 of 19 December 2012, supplementing Regulation (EU) No 48/2012 of the European Parliament and of the Council with regard to regulatory technical standards on requirements for central counterparties. *Official Journal of the European Union*, 2013.

[11] Rama Cont and Milena Vuletić. Simulation of arbitrage-free implied volatility surfaces. *Available at SSRN*, 2022.

[12] Paul Glasserman and Dan Pirjol. W-shaped implied volatility curves and the Gaussian mixture model. *Available at SSRN 3951426*, 2021.

[13] Fredrik Gunnarsson. Filtered Historical Simulation Value at Risk for Options: A Dimension Reduction Approach to Model the Volatility Surface Shifts, 2019.

[14] Pedro Gurrola-Perez and David Murphy. Filtered historical simulation Value-at-Risk models and their competitors. 2015.

[15] Christian Kahl and Peter Jäckel. Not-so-complex logarithms in the Heston model. *Wilmott magazine*, 19(9):94–103, 2005.

[16] Brian Ning, Sebastian Jaimungal, Xiaorong Zhang, and Maxime Bergeron. Arbitrage-free implied volatility surface generation with variational autoencoders. *arXiv preprint arXiv:2108.04941*, 2021.

[17] Lauren W Wong and Yang Zhang. Procyclicality control in risk-based margin models. *Journal of Risk*, 23(5), 2021.

[18] Jinglun Yao, Sabine Laurent, and Brice Bénaben. Managing Volatility Risk: An Application of Karhunen-Loève Decomposition and Filtered Historical Simulation. *arXiv preprint arXiv:1710.00859*, 2017.