A Proposed Risk Modeling Shift from the Approach of Stochastic Differential Equation towards Machine Learning Clustering:

Illustration with the concepts of Anticipative & Responsible VaR

(working paper)

Babak Mahdavi-Damghani[1] and Stephen Roberts[2]

Oxford-Man Institute of Quantitative Finance

Abstract— **The aim of this technical document is threefold with the bigger picture being to contribute, within the challenging regulatory environment, to bring closer together traditional conflicting practices such as trading vs risk as well as risk responsiveness vs stability. In order to achieve this goal, we first expose some of the complexity associated to the risk factors and arbitrage constraints associated with the options and the high frequency markets by re-introducing the Implied Volatility Parametrization (IVP) [3], [11] and the High Frequency Trading Ecosystem (HFTE) [41]. The exposed complexity is then contrasted with the current obsolete Risk Methodologies which are based on simplistic SDEs which we extent using the Cointelation model [10], [13] in order to partially address some of the complexity introduced by the challenging regulatory environment such as scenario coherence. We then present a simple Machine Learning clustering methodology which is designed to address and mirror the enhancements of these SDEs in a simpler fashion. We illustrate our findings by introducing few new risk concepts such as the Anticipatible VaR which aims at being a leading as opposed to a lagging (Responsive) risk measure to a market regime change.**

Keywords: Stochastic Differential Equation, Gaussian Process, Cointelation, Value at Risk (VaR), Responsive VaR, Stable VaR, Responsible VaR, Anticipative VaR, Anticipatible VaR, Stochastic Differential Equations (SDE), Implied Volatility Parametrization (IVP), High Frequency Trading Ecosystem (HFTE), Variance reduction, Volatility surface, SVI, gSVI, Arbitrage Free Volatility Surface, Ornstein-Uhlenbeck, Cox-Ingersoll-Ross, Hull-White, FRTB, Fundamental Review of the Trading Book

I. Scope

A. Market context

The financial crisis of 2009 and the resulting social uproar pushed the Basel Committee on Banking Supervision to revisit its policy on capital requirement [49]. Although quite verbose in form the new capital requirement introduced meant, in spirit, and in the context of this paper:

- Capital Requirement of each financial institution will be linked to its VaR[1] and the latter must be calculated with historical data.
- VaR must take into account the procyclicality of the market. Having from roots, the observation that a big market move is likely to follow another big market move, risk models must adapt quickly and adjust to the sudden increase of volatility of the market.
- In order to eliminate the risk associated to liquidity shortage and the resulting systemic risk, VaR should on top of being Responsive remain as stable as possible.
- P&L associated to trading should be mapped to appropriate risk factors.
- Simulated scenarios must be coherent (example no arbitrage allowed).

B. A few definitions

Before introducing the context of this paper, let us first introduce few definitions which may not be necessarily known by the reader even for the quantitative analytics community as they are relatively new, Risk specific, practitioners focused definitions.

Remark The concept of VaR and Expected shortfall will be interchangeably used in this document as the use of either of these risk measures can be interchangeably used in the context of risk *anticipativity* and risk *responsibility*. Also this technical document has been written for practitioners and therefore the mathematics is tailored in such a way that intuition is conserved at the cost of sometimes, unfortunately, rigor.

Definition (Responsive VaR): A VaR model that will be able to adapt, *a posteriori*, to increased volatility conditions will be referred to Responsive VaR.

Definition (Anticipative VaR): A VaR model that will be able to adapt, *a priori*, to increased volatility conditions will be referred to Anticipative VaR.

Definition (Stable VaR): A VaR model that will be able to remain robust will be referred to Stable VaR.

Definition (Responsible VaR): A VaR model that is both Stable and Responsive will be referred to as Responsible.

Definition (Anticipatible VaR): A VaR model that is both Stable and Anticipative will be referred to as Anticipatible.

[1]babak.mahdavidamghani@oxford-man.ox.ac.uk

[2]steve.roberts@oxford-man.ox.ac.uk

[1]Later into Expected Shortfall, but this change is irrelevant in the context of this paper as going from one to the other when one has the simulated scenarios is relatively easy.

C. Problem formulation

Besides exposing the contrast between market complexity to risk models absurd simplicity, the objective of this paper is really twofold:

- The current risk models available to practitioners are at best Responsive and therefore lagging with respect to regime changes which means that one (or few depending on the quantile level) risk breaches is needed for the mathematical model to be able to adjust to the changing market condition. The first objective for this paper is to lay down the mathematical specification for a risk system that would be leading as opposed to lagging.
- It seems quite obvious a VaR model cannot be Stable and Responsive at the same time. The second objective of this paper is to attempt to partially reconcile these conflicting risk concepts which are, interestingly, equally desirable model features for risk managers, traders and financial mathematics practitioners despite their apparent discordant properties.

D. Structure of this technical document

We re-introduce the IVP model in section II in order to recall the complexity of the risk factors in what is commonly referred to the low frequency domain. We then delve in the complex constraints associated to non-linear product especially when it comes to coherent scenario analysis in the form of arbitrage creation in section III. Given that the stakes are equally important for both the low and the high frequency domain, we also re-introduce the HFTE model in order to expose, in section IV, some of the complexity in the latter domain as well. In section VI we introduce the proposed enhancements.

II. Complex Low Frequency Risk Factors

The objective of this section is to expose the complexity of the risk factors associated to the simplest of the non linear products, the vanilla options, which are the stepping stones of more complex derivatives. This is done in order to contribute at exposing the ridicule of the simplistic risk model specifications that are used by practitioners. Studying vanilla options can be done in couple of domains, the price domain or the implied volatility domain which has been developed to address the limitations of the Black-Scholes model. As it happens, working on the implied volatility domain offers lots of benefits that the price domain cannot replicate. The latter ones, in the context of this paper, are the ones associated to isolating the risk factors, relevant to risk management[2]. There exists many parametrization of the implied volatility surface, notably the Schonbucher and the SABR models [55], [27], have had their share of practitioners. However, we will only discuss the SVI [21], [23], [22], [24] model and its most advanced extension, the IVP as it is currently the one which has the most comprehensive number of risk factors.

[2]also relevant to pricing and trading

A. The Raw Stochastic Volatility Inspired (SVI) model

Remark In terms of notations, we use the traditional notation [24] and in the foregoing, we consider a stock price process $(S_t)_{t\geqslant 0}$ with natural filtration $(\mathcal{F}_t)_{t\geqslant 0}$, and we define the forward price process $(F_t)_{t\geqslant 0}$ by $F_t := \mathbb{E}(S_t|\mathcal{F}_0)$. For any $k \in \mathbb{R}$ and $t > 0$, $C_{\mathrm{BS}}(k, \sigma^2 t)$ denotes the Black-Scholes price of a European Call option on S with strike $F_t e^k$, maturity t and volatility $\sigma \geqslant 0$. We shall denote the Black-Scholes implied volatility by $\sigma_{BS}(k,t)$, and define the total implied variance by

$$w(k, \chi_R) = \sigma^2_{\mathrm{BS}}(k,t)t.$$

The implied variance v shall be equivalently defined as $v(k,t) = \sigma^2_{\mathrm{BS}}(k,t) = w(k,t)/t$. We shall refer to the two-dimensional map $(k,t) \mapsto w(k,t)$ as the volatility surface, and for any fixed maturity $t > 0$, the function $k \mapsto w(k,t)$ will represent a slice.

1) History: One advertised[3] advantage of the SVI is that it can be derived from Heston [31], [23], a model used by many financial institutions for both risk, pricing and sometimes statistical arbitrage purposes. One of the main advantages of this parametrization is its simplicity. Advertised as being parsimonious, its parametrization assumed linearity in the wings (which yields a poor fit in the wings) because of its inability to handle variance swaps, leading it to become decommissioned couple of years after its birth. Another limitation of the SVI became apparent after the subprime crisis and the subsequent call for mathematical models that would incorporate liquidity which the SVI did not incorporate [11].

2) Formula: For a given maturity slice, we shall use the notation $w(k, \chi_R)$ where $\chi_R = \{a, b, \rho, m, \sigma\}$ represents a set of parameters, and the t-dependence is dropped.

Remark Note that the term "parameters" and "risk factors" can be used interchangeably in this section.

For a given parameter set. Then the raw SVI parameterization of implied variance reads:

$$w(k, \chi_R) = a + b[\rho(k-m) + \sqrt{(k-m)^2 + \sigma^2}] \qquad (1)$$

with k being the log-moneyness ($\log(\frac{K}{F})$ with F being the value of the forward).

Remark Note that there exist several other forms of the SVI model which are equivalent to each other through a set of transform functions [24]. The motivation of their existence and the details of the transforms are out of scope but we refer to the original papers [24] for the motivated reader.

The advantage of Gatheral's model was that it was a parametric model that was easy to use, yet had enough complexity to properly model the volatility surface and its dynamic. Figure 3 illustrates the change in the ρ parameter (the skew risk), Figure 2 illustrates the change in the b parameter (the vol of vol risk), Figure 1 illustrates the change in the a parameter

[3]One of the main point of this paper is to expose a small mistake that was done in one particular paper [21] but for the sake of the introduction, we will make this remark as a footnote.

(the general volatility level risk), Figure 5 illustrates the change in the σ parameter (the ATM volatility risk) and finally Figure 4 illustrates the change in the m parameter (the horizontal displacement risk).

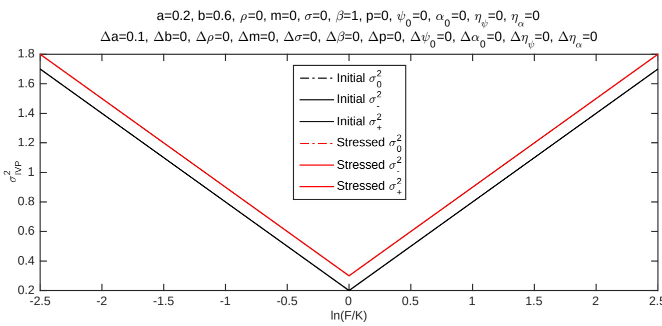

Fig. 1. Change in the a parameter in the rawSVI/gSVI/IVP model

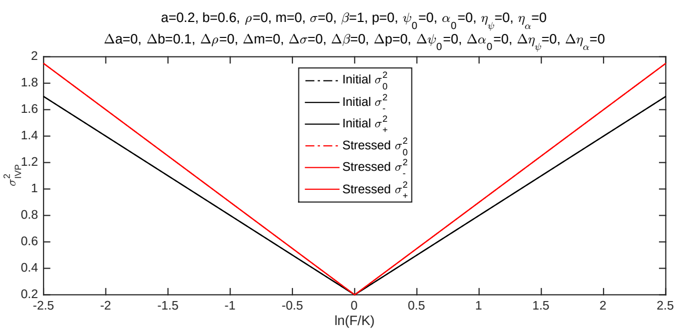

Fig. 2. Change in the b parameter in the rawSVI/gSVI/IVP model

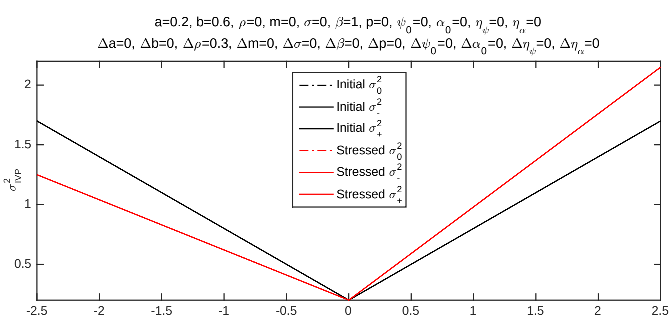

Fig. 3. Change in the ρ parameter in the rawSVI/gSVI/IVP model

B. Relation between IVP and raw SVI

Jim Gatheral developed the SVI model at Merrill Lynch in 1999 and implemented in 2005. The SVI was subsequently decommissioned in 2010 because of its limitations in accurately pricing out of the money variance swaps (for example short maturity Var Swaps on the Eurostoxx are overpriced when using the SVI). This is because the wings of the SVI are linear and have a tendency to overestimate the out of the money (OTM) variance swaps. Benaim, Friz and Lee [5] gave a mathematical justification for this market observation. The paper suggests that the implied volatility cannot grow asymptotically faster than $\sqrt{k}$ but may grow slower than $\sqrt{k}$ when the distribution of the underlier does not have finite moments (eg: has heavy tails). This suggest that the linear wings of the SVI model may overvalue really deeply OTM options which is observable in the markets. In order to address the limitations of the SVI model in the wings, while keeping its core skeleton intact, Mahdavi-Damghani [3] proposed a change of variable which purpose was to penalize the wings's linearity. The additional relevant parameter was called β and was later extended in order to also address the liquidity constraints of the model [11] especially given the challenging regulatory environment[4]. Mahdavi-Damghani initially named the model "generalized SVI" (gSVI) [3] but renamed it "Implied Volatility Parametrization" (IVP) [11] once the liquidity parameters were incorporated. In order to keep the number of factor limited, this β penalization functions was made symmetrical on each wing[5]. The function needed to be increasing as it gets further away from m and majored by a linear function increasing in $[m; +\infty[$ and decreasing in $]-\infty; m]$ and increasing in concavity the further away it gets from the center. Equation (2) summarizes the gSVI[6]. The penalization was given by equation (2b). Figure 6 illustrates the change

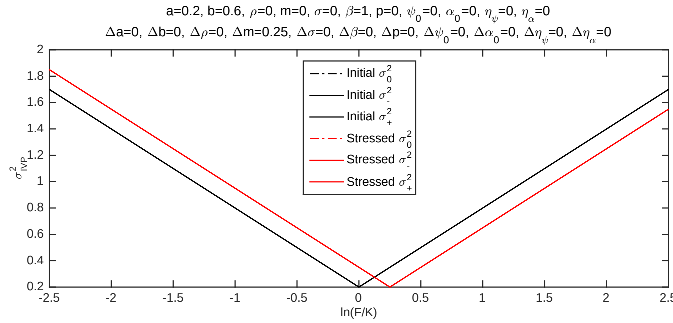

Fig. 4. Change in the m parameter in the rawSVI/gSVI/IVP model

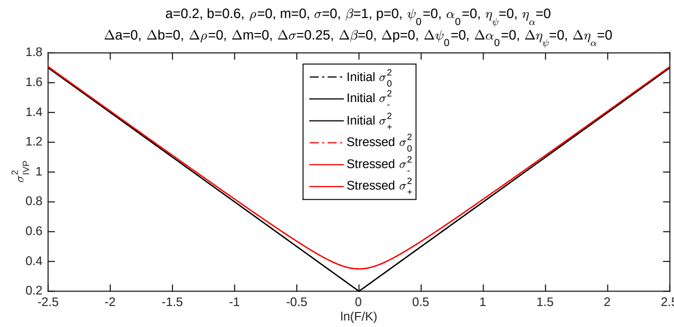

Fig. 5. Change in the σ parameter in the rawSVI/gSVI/IVP model

[4]e.g. Fundamental Review of the Trading Book (FRTB)

[5]But induced geometrically more significant on the steepest wing: for e.g. more significant on the left wing in the Equities market and more significant on the right wing of the Commodities (excluding oil) market

[6]or alternatively IVP's mid, model

in the β parameter.

$$\sigma_{gSVI}^{2}(k) = a + b\left[\rho(z-m) + \sqrt{(z-m)^{2} + \sigma^{2}}\right] \tag{2a}$$

$$z = \frac{k}{\beta^{|k-m|}}, 1 \leqslant \beta \leqslant 1.4 \tag{2b}$$

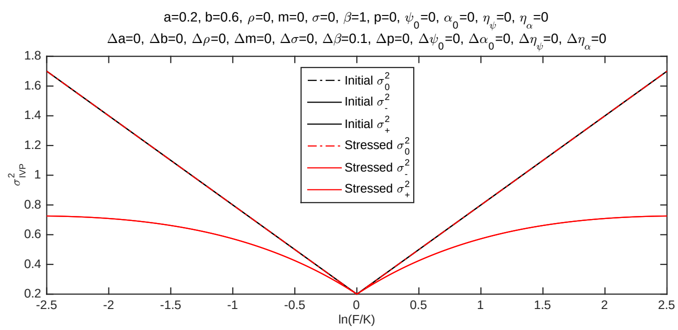

Fig. 6. Change in the β parameter in the gSVI/IVP model

Remark The downside transform in the gSVI [3] was arbitrarily given by $z = \frac{k}{\beta^{|k-m|}}, 1 \leqslant \beta \leqslant 1.4$. It is however, important to note, that there are many ways of defining the downside transform. One general approach would be to define μ and η like it is done in equation (3a). That idea can be prolonged to $\exp$ like function such as the one in equation (3b). The idea is always the same: the further away you are from the ATM, the bigger the necessary adjustment on the wings.

$$z = \frac{k}{\beta^{\mu+\eta|k-m|}} \tag{3a}$$

$$z = e^{-\beta|k-m|}(k-m) \tag{3b}$$

Mahdavi-Damghani, in introducing the IVP model [11] picked in equation (3a) a $\mu = 1$ and $\eta = 4$ and have the transformation in the form $z = \frac{k}{\beta^{1+4|k-m|}}$ because it yields better optimization results on the FX markets and also because it relaxes the constraint on β but our intuition is that the $\exp$ like function may work better when it comes to showing convergence between the modified Heston and the IVP model.

C. Risk factors associated to Liquidity

By incorporating the information on the gSVI, the ATM Bid Ask spread and the curvature adjustment of the wings Mahdavi-Damghani [3], [11] defines what he labeled the Implied Volatility surface Parametrization (IVP) below:

$$\begin{aligned}
\sigma_{IVP,o,\tau}^{2}(k) &= \left[\rho_\tau(z_{o,\tau}-m_\tau) + \sqrt{(z_{o,\tau}-m_\tau)^{2} + \sigma_\tau^{2}}\right] \\
&\quad \times b_\tau + a_\tau \\
z_{o,\tau} &= \frac{k}{\beta_{o,\tau}^{1+4|k-m|}} \\
\sigma_{IVP,+,\tau}^{2}(k) &= \left[\rho_\tau(z_{+,\tau}-m_\tau) + \sqrt{(z_{+,\tau}-m_\tau)^{2} + \sigma_\tau^{2}}\right] \\
&\quad \times b_\tau + a_\tau + \alpha_\tau(p) \\
z_{+,\tau} &= z_{o,\tau}[1+\psi_\tau(p)] \\
\sigma_{IVP,-,\tau}^{2}(k) &= \left[\rho_\tau(z_{-,\tau}-m_\tau) + \sqrt{(z_{-,\tau}-m_\tau)^{2} + \sigma_\tau^{2}}\right] \\
&\quad \times b_\tau + a_\tau - \alpha_\tau(p) \\
z_{-,\tau} &= z_{o,\tau}[1-\psi_\tau(p)] \\
\alpha_\tau(p) &= \alpha_{0,\tau} + (a_\tau - \alpha_{0,\tau})(1-e^{-\eta_{\alpha_\tau}p}) \\
\psi_\tau(p) &= \psi_{0,\tau} + (1-\psi_{0,\tau})(1-e^{-\eta_{\psi_\tau}p})
\end{aligned}$$

The functions $\alpha(p)$ (figure 7) and ψp (figure 8) model the ATM and wing curvature of the Bid-Ask keeping in mind the idea that the bigger the position size the bigger the market impact and hence the wider the Bid-Ask. This market impact parameter is controlled by p (figure 9). Finally, couple of additional parameters model the elasticity of the liquidity: η_ψ (figure 10) and η_α (figure 11).

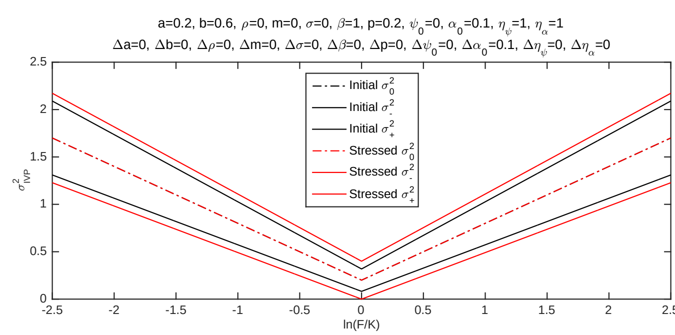

Fig. 7. Change in the α parameter in the IVP model

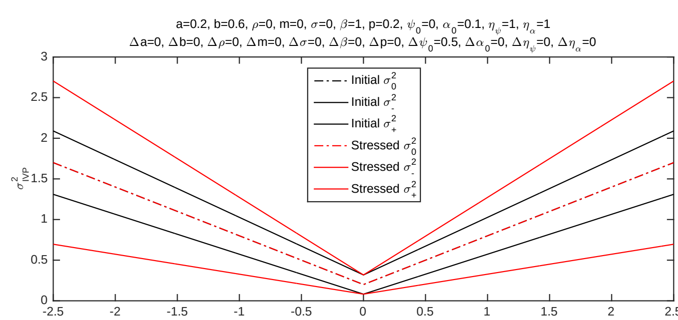

Fig. 8. Change in the ψ parameter in the IVP model

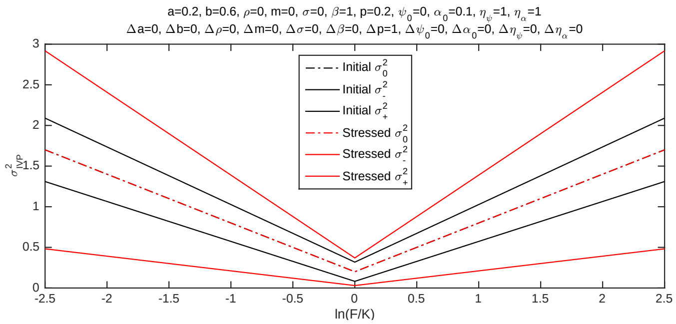

Fig. 9. Change in the p parameter in the IVP model

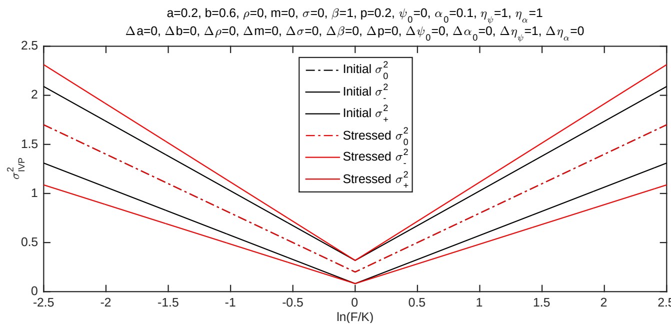

Fig. 10. Change in the η_ψ parameter in the IVP model

III. Arbitrage & the Options Market

The way stress testing is assessed for the options market is usually threefold. First, the performance as defined by the difference between the number of exceptions as returned from the back-testing exercise and the quantile level of our VaR, is of central importance at the first glance. Having a poor risk engine that does not take into account arbitrage creation may distort many scenarios especially when the shape of the implied volatility surface is highly skewed or/and high. Second, many of the risk engines uses numerical methods which break if an arbitrage is created on the implied volatility surface. Finally, many of the risk engines whether presented internally in the financial institution or outside with the regulators is scrutinized and if arbitrage is not seriously considered the reputation of the managers/bank

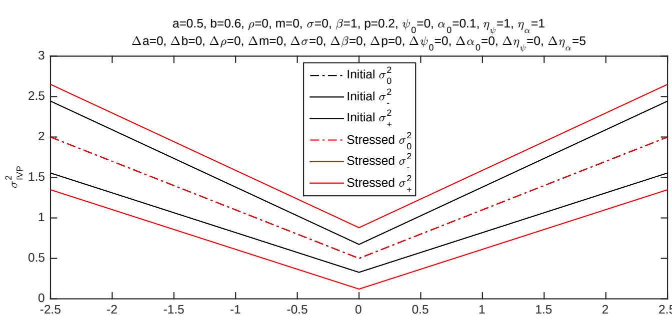

Fig. 11. Change in the η_α parameter in the IVP model

is compromised and the likelihood of acceptance of the corresponding risk model decreases as a result. We will see in this section the constraints around the arbitrage frontiers given by the conditions on the strike (section III-A) and tenor (section III-B) spaces.

A. Condition on the strike

The model set up is the usual. Let us set up the probability space $(\Omega, (f)_{(t\geqslant 0)}, \mathbb{Q})$, with $(f)_{(t\geqslant 0)}$ generated by the $T+1$ dimensional Brownian motion and $\mathbb{Q}$ is the risk neutral probability measure under which the discounted price of the underlier, rS, is a martingale. We also assume that the underlier can be represented as a stochastic volatility lognormal Brownian motion as represented by 4.

$$dS_t = rS_t dt + \sigma_t S_t dW_t \tag{4}$$

In order to prevent arbitrages on the volatility surface we will start from basic principles and derive the constraints relevant to the strike and tenor.

1) Theoretical form: Using Dupire's work [16], [17], we can write the price of a call the following way: $C(S_0, K, T) = e^{-rT}\mathbb{E}^{\mathbb{Q}}[S_T - K]^+ = e^{-rT}\int_K^{+\infty}(S_T - K)\phi(S_T, T)dS_T$ with $\phi(S_T, T)$ being the final probability density of the call. Differentiating twice we find equation (5).

$$\frac{\partial^2 C}{\partial K^2} = \phi(S_T, T) > 0. \tag{5}$$

Proof: We write our call price $C(S_0, K, T) = e^{-rT}\mathbb{E}^{\mathbb{Q}}[S_T - K]^+$ which, using integration gives $e^{-rT}\int_K^{+\infty}(S_T - K)\phi(S_T, T)dS_T \frac{\partial C}{\partial K}$ which we simplify to $-e^{-rT}\int_K^{+\infty}\phi(S_T, T)dS_T = -e^{-rT}\mathbb{E}(S_T > K)$. Also we know that $0 \leqslant -e^{-rT}\frac{\partial C}{\partial K} \leqslant 1$. Differentiating a second time and setting $r = 0$ we find $\phi(S_T, T) = \frac{\partial^2 C}{\partial K^2}$. ■

Using numerical approximation we get equation (6) which is known in the industry as the arbitrage constraint of the positivity of the butterfly spread [60].

$$\forall \Delta, C(K - \Delta) - 2C(K) + C(K + \Delta) > 0 \tag{6}$$

Proof: Given that the probability density must be positive we have $\frac{\partial^2 C}{\partial K^2} \geqslant 0$, using numerical approximation, we get

$$\begin{aligned}\frac{\partial^2 C}{\partial K^2} &= \lim_{\Delta\to 0}\frac{[C(K-\Delta) - C(K)] - [C(K) - C(K+\Delta)]}{\Delta^2} \\ &= \lim_{\Delta\to 0}\frac{C(K-\Delta) - 2C(K) + C(K+\Delta)}{\Delta^2}\end{aligned}$$

therefore $C(K - \Delta) - 2C(K) + C(K + \Delta) \geqslant 0$ ■

Gatheral and Jacquier [24] proved that the positivity of the butterfly condition comes back to making sure that the function $g()$ below is strictly positive.

$$g(k) := \left(1 - \frac{Kw'(k)}{2w(k)}\right)^2 - \frac{w'(k)^2}{4}\left(\frac{1}{w(k)} + \frac{1}{4} + \frac{w''(k)}{2}\right)$$

Proof: We have shown in equation (5) that $\frac{\partial^2 C}{\partial K^2} = \phi()$. Applying this formula to the Black-Scholes equation gives for a given tenor $\phi(k) = \frac{g(k)}{\sqrt{2\pi w(k)}}\exp\left(-\frac{d_2(k)^2}{2}\right)$ where

$w(k,t) = \sigma_{BS}^2(k,t)t$ is the implied volatility at strike K and where $d_2(k) := \frac{-k}{\sqrt{w(k)}} - \sqrt{w(k)}$. ■

Function $g(k)$ yields a polynomial of the second degree with a negative highest order which suggest that the function is inverse bell curve like and potentially only positive given **two constraints** which may appear as contradicting some of the initial slides Gatheral presented back in 2004. If g_1^e and g_2^e happens to be the exact roots of $g(k) = 0$ with $g_2^e \geqslant g_1^e$ then the volatility surface is arbitrage free with respect to the butterfly constraint if $w(k) \leqslant g_2^e$ and $w(k) \geqslant g_1^e$.

2) Necessary but not sufficient Practical form: There exists another version of this butterfly (equation (5)) condition that is a necessary but **not sufficient** condition to make a volatility surface arbitrage free but remains useful when one has a more practical objective which will be illustrated with an example in section II. This condition is given by equation (7).

$$\forall K, \forall T, |T\partial_K \sigma^2(K,T)| \leqslant 4 \tag{7}$$

Proof: The intuition behind the proof is taken from Rogers and Tehranchi [53] but is somewhat simplified for practitioners. Assuming $r = 0$, let us define the Black-Scholes call function $f : \mathbb{R} \times [0,\infty) \longrightarrow [0,1)$ in terms of the tail of the standard Gaussian distribution $\Phi(x) = \frac{1}{\sqrt{2\Pi}}\int_x^{+\infty} \exp(\frac{-y^2}{2})dy$ and given by:

$$f(k,\nu) = \begin{cases} \Phi(\frac{k}{\sqrt{\nu}} - \frac{\sqrt{\nu}}{2}) - e^k\Phi(\frac{k}{\sqrt{\nu}} + \frac{\sqrt{\nu}}{2}) & \text{if } \nu > 0 \\ (1+e^k)^+ & \text{if } \nu = 0 \end{cases}$$

Let us call $V_t(k,\tau)$ the implied variance at time $t \geqslant 0$ for log-moneyness k and time to maturity $\tau \geqslant 0$. Let's now label our Kappa and Vega, with the convention that $\phi(x) = \frac{1}{\sqrt{2\Pi}}\exp(\frac{-x^2}{2})$.

$$f_k(k,\nu) = -e^k\Phi(\frac{k}{\sqrt{\nu}} + \frac{\sqrt{\nu}}{2})$$
$$f_\nu(k,\nu) = \phi(\frac{k}{\sqrt{\nu}} + \frac{\sqrt{\nu}}{2})/2\sqrt{\nu}$$

Now define the function $I : \{(k,c) \in \mathbb{R}\times[0,\infty) : (1+e^k)^+ \leqslant c < 1\} \longrightarrow [0,1)$ implicitly by the formula:

$$f(k, I(k,c)) = c$$

Calculus gives $I_c = \frac{1}{f_\nu}$ and $I_k = -\frac{f_k}{f_\nu}$, from here using the chain rule, designating $\partial_{k+}V$ as the right derivative. We have

$$\partial_{k+}V = I_k + I_c\partial_k\mathbb{E}[(S_\tau - e^k)^+]$$

$$\begin{aligned}\partial_{k+}V &= -\frac{f_k}{f_\nu} - \frac{\mathbb{P}(S_\tau > e^k)}{f_\nu} \\ &< -\frac{f_k}{f_\nu} = 2\sqrt{\nu}\frac{\Phi(\frac{k}{\sqrt{\nu}} + \frac{\sqrt{\nu}}{2})}{\phi(\frac{k}{\sqrt{\nu}} + \frac{\sqrt{\nu}}{2})}\end{aligned}$$

Now using the bounds of the Mills' ratio $0 \leqslant 1 - \frac{x\Phi(x)}{\phi(x)} \equiv \varepsilon(x) \leqslant \frac{1}{1+x^2}$, we have:

$$\partial_{k+}V \leqslant \frac{4}{k/V+1} < 4$$

Similarly we can show [53] that $\partial_{k-}V > -4$, therefore we have $|\partial_k V| < 4$ ■

One can think of the boundaries of the volatility surface, as extrapolated by equation (7), as more relaxed boundaries (but still "close") in the strike space compared to the exact solution from $g(k)$ set to 0 which are both necessary and sufficient conditions for the volatility surface to be arbitrage free for the butterfly condition. Formally if g_1^a and g_2^a happens to be the exact roots of $|T\partial_K\sigma^2(K,T)| - 4 = 0$, with $g_2^a \geqslant g_1^a$ then we have $g_1^a \leqslant g_1^e \leqslant w(k) \leqslant g_2^e \leqslant g_2^a$. The reason why equation (7) is practical is because in de-arbitraging methodologies (as we will see more in details in section II), there exist for the pricers, a component of tolerance anyways (the pricers are stable if the volatility surface is slightly away of its arbitrage frontier). This suggests that finding a close enough solution but building on top of that an iterative methodology to get closer and closer to the practical arbitrage frontier is almost equally fast, but with less computing trouble, than having the exact theoretical solution (and building an error tolerance finder on top of it anyways). This is because there is less probability to make a typo mistakes in typing the exact solution of $g(k)$ (or its numerical approximation) especially if your parametrized version of the volatility surface is complex which is the case in most banks ($\{g_1^a, g_2^a\}$ are easier to find than $\{g_1^e, g_2^e\}$). Also as we will see in section II that given that we would like a liquidity component around a mid price, having a simple "close enough" constraint on the mid becomes very useful especially if we are happy to allow the mid to have arbitrages on it, something which happens to be the case from time to time on the mid vol of the market anyways. Figure 12 represents a counter example of $|T\partial_K\sigma^2(K,T)| \leqslant 4$ applied to the Raw SVI parametrisation[7] in which $(a,b,m,\rho,\sigma) = (0.0410, 0.1331, 0.3586, 0.3060, 0.4153)$ respect the $b(1+|\rho|) \leqslant \frac{4}{T}$ inequality but for which the probability density function at expiry in negative around moneyness of 0.8 yielding a butterfly arbitrage.

B. Condition on the tenor

The model setup is the same as in section III-A, that is let us set up the probability space $(\Omega, (f)_{(t\geqslant 0)}, \mathbb{Q})$, with $(f)_{(t\geqslant 0)}$ generated by the $T+1$ dimensional Brownian motion and $\mathbb{Q}$ is the risk neutral probability measure under which the discounted price of the underlier, rS, is a martingale. We also assume that the underlier can be represented as a stochastic volatility lognormal Brownian motion as represented by equation (4). In order to prevent arbitrages on the volatility surface on the tenor space we will split this subsection in its theoretical form in section III-B.1 and III-B.2 for its practical form.

[7] which we discuss more in details in section II.

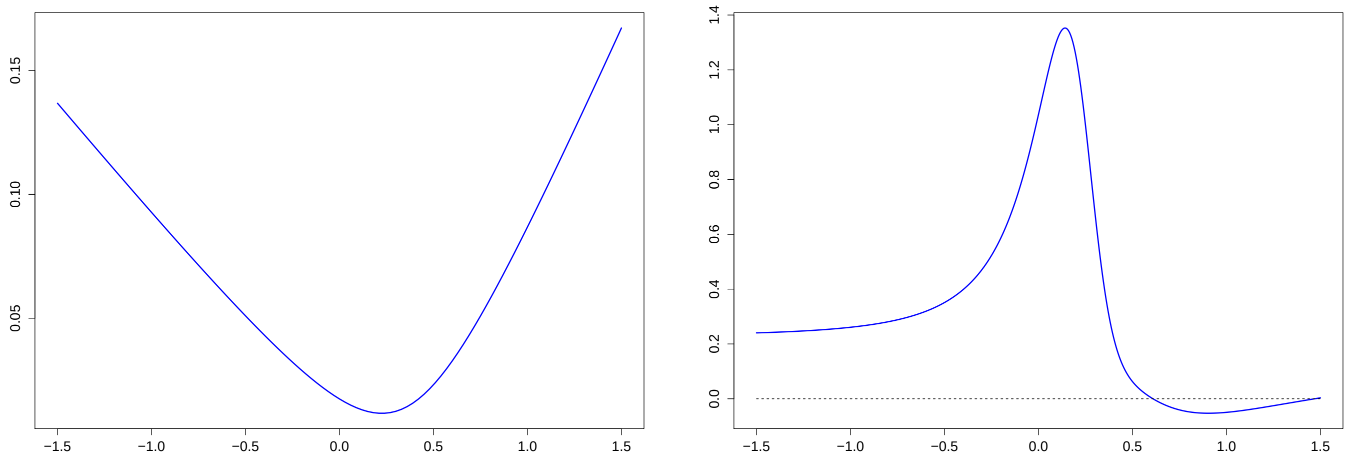

Fig. 12. Axel Vogt [62] counter-example for $b(1+|\rho|) \leqslant \frac{4}{T}$ being arbitrage free

1) Theoretical form: The condition on the tenor axis which insures the volatility surface to be arbitrage free is that the calendar spread should be positive:

$$C(K, T+\Delta) - C(Ke^{-r\Delta}, T) \geqslant 0 \tag{8}$$

Proof: One application of Dupire's formula [16], [17] is that the pseudo-probability density must satisfy the Fokker-Planck [20], [52] equation. This proof is taken from El Karoui [34]. Let us apply Itô to the semi-martingale . This is formally done by introducing the local time Λ_T^K: $e^{-r(T+\varepsilon)}\left(S_{T+\varepsilon}-K\right)^+ - e^{-r(T)}\left(S_T-K\right)^+ = \int_T^{T+\varepsilon} re^{-ru}\left(S_u-K\right)^+ du + \int_T^{T+\varepsilon} e^{-ru} 1_{\{S_u \geqslant K\}} dS_u + \frac{1}{2}\int_T^{T+\varepsilon} e^{-ru} d\Lambda_u^K$. Local times are introduced in mathematics when the integrand is not smooth enough. Here the call price is not smooth enough around the strike level at expiry. Now we have: $E\left(e^{-ru} 1_{\{S_u \geqslant K\}} S_u\right) = C(u,K) + Ke^{-ru} P\left(S_u \geqslant K\right) = C(u,K) - K\frac{\partial C}{\partial K}(u,K)$. The term of the form $E\left(\int_T^{T+\varepsilon} e^{-ru} d\Lambda_u^K\right)$ is found due to the formula of local times, that is:

$$\begin{aligned}
E\left(\int_T^{T+\varepsilon} e^{-ru} d\Lambda_u^K\right) &= \int_T^{T+\varepsilon} e^{-ru} du E\left(\Lambda_u^K\right) \\
&= \int_T^{T+\varepsilon} e^{-ru} du \sigma^2(u,K) K^2 \phi(u,K) \\
&= \int_T^{T+\varepsilon} \sigma^2(u,K) K^2 \frac{\partial^2 C}{\partial K^2}(u,K)\, du
\end{aligned}$$

Plugging these results back into the first equation we get:

$$\begin{aligned}
C(T+\varepsilon, K) =& C(T,K) - \int_T^{T+\varepsilon} rC(u,K)\, du + (r-q) \\
&\times \int_T^{T+\varepsilon} \left(C(u,K) - K\frac{\partial C}{\partial K}(u,K)\right) du \\
&+ \frac{1}{2}\int_T^{T+\varepsilon} \sigma^2(u,K) K^2 \frac{\partial^2 C}{\partial K^2}(u,K)\, du
\end{aligned}$$

If we want to give a PDE point of view of this problem we can notice that $\phi(T,K) = e^{-rT}\frac{\partial^2 C}{\partial K^2}(T,K)$ verifies the dual forward equation:

$$\begin{aligned}
\phi_T'(T,K) =& \frac{1}{2}\frac{\partial^2\left(\sigma^2(T,K) K^2 \phi(T,K)\right)}{\partial K^2} \\
&- \frac{\partial^2\left((r-q) K\phi(T,K)\right)}{\partial K}
\end{aligned}$$

Integrating twice by part, we find:

$$\begin{aligned}
\frac{\partial e^{-rT} C(T,K)}{\partial T} =& \frac{1}{2}\sigma^2(T,K) K^2 e^{rT}\frac{\partial^2 C(T,K)}{\partial K^2} \\
&- \int_K^{+\infty} (r-q) Ke^{rT} \\
&\times \frac{\partial^2 C(u,K)}{\partial K^2} \partial K(T,K)\, du
\end{aligned}$$

Now integrating by part again and setting dividends to 0 we find the generally admitted relationship:

$$\frac{\partial C}{\partial t} = \frac{\sigma^2}{2} K^2 \frac{\partial^2 C}{\partial K^2} - rK\frac{\partial C}{\partial K}$$

and therefore we have:

$$\sigma = \sqrt{2\frac{\frac{\partial C}{\partial t} + rK\frac{\partial C}{\partial K}}{K^2\frac{\partial^2 C}{\partial K}}}$$

From this formula and from the positivity constraint on equation (5) we find that $\frac{\partial C}{\partial t} + rK\frac{\partial C}{\partial K} \geqslant 0$. Note that for very small Δ, we have $C(Ke^{-r\Delta}, T) \approx C(K - Kr\Delta, T)$. Using Taylor expansion we get $C(K - Kr\Delta, T) = C(K,T) - Kr\Delta\frac{\partial C}{\partial K} + \ldots$ and therefore $rK\frac{\partial C}{\partial K} \approx \frac{C(K,T) - C(Ke^{-r\Delta},T)}{\Delta}$. Using forward difference approximation we also have $\frac{\partial C}{\partial K} = \frac{C(K,T+\Delta) - C(K,T)}{\Delta}$ and from Fokker-Planck we have $\frac{\partial C}{\partial t} + rK\frac{\partial C}{\partial K} \geqslant 0$. Substituting, we obtain $\frac{C(K,T+\Delta) - C(K,T)}{\Delta} + \frac{C(K,T) - C(Ke^{-r\Delta},T)}{\Delta} \geqslant 0$. Simplifying further we find $C(K, T+\Delta) - C(Ke^{-r\Delta}, T) \geqslant 0$. ■

2) Practical form: Similarly to section III-A there exists a more practical equivalent to the calendar spread criteria. This equivalent criteria is known as the falling variance criteria which states that if S is a martingale under the risk neutral probability measure $\mathbb{Q}$,

$$\forall t > s,\ e^{-rt}\mathbb{E}^{\mathbb{Q}}(S_t - K)^+ \geqslant e^{-st}\mathbb{E}^{\mathbb{Q}}(S_s - K)^+ \quad (9)$$

Proof: $e^{-rt}\mathbb{E}^{\mathbb{Q}}(S_t - K)^+ \geqslant e^{-rs}\mathbb{E}^{\mathbb{Q}}(S_s - K)^+ \Rightarrow e^{-rt}\mathbb{E}^{\mathbb{Q}}(S_t - K)^+ - e^{-rs}\mathbb{E}^{\mathbb{Q}}(S_s - K)^+ \geqslant 0 \Rightarrow$ Calendar Spread $\geqslant 0 \Rightarrow C(K, T + \Delta) - C(Ke^{-r\Delta}, T) \geqslant 0$ ∎

C. Arbitrage Frontiers and de-arbitraging

As we have seen from equations (6) and (8) there are couple of arbitrages types, the calendar and butterfly arbitrage as summarized my equation (10b).

$$\forall\Delta, C(K - \Delta) - 2C(K) + C(K + \Delta) > 0 \quad (10a)$$

$$\forall\Delta, \forall T, C(K, T + \Delta) - C(Ke^{-r\Delta}, T) \geqslant 0 \quad (10b)$$

A new wave of risk methodologies with the objective of making incoherent scenarios like the ones allowing an arbitrage is currently being developed [3], [24], and though promissing few questions remain to be addressed [11].

Remark Note that once Bid Ask has been incorporated, we care a bit less about the mid in the context of vanilla options market making. Though the mid may have arbitrages at the portfolio level, the Bid-Ask relaxes the butterfly spread equations. We get, in the context of the IVP mode described in section II: $\forall\Delta, C(K - \Delta, \sigma_{IVP,+,t}(k)) - 2C(K, \sigma_{IVP,-,t}(k)) + C(K + \Delta, \sigma_{IVP,+,t}(k)) > 0$ which gives: $C(K, T + \Delta, \sigma_{IVP,+,t}(k)) - C(Ke^{-r\Delta}, T, \sigma_{IVP,-,t}(k)) \geqslant 0$.

IV. Complex High Frequency Risk Factors

In this section we will expose the complexity of risk factors at the high frequency domain, which is by definition a highly fast market in which the decisions are taken by rule based methods which then go on impacting the order books which other robots read and act upon, in a systematic fashion and at a lightning speed. To illustrate this latter point we summarize the HFTE model, recently introduced [41].

A. Market Observation

1) Introduction: After the subprime crisis of 2008 and the resulting social uproar, governments strongly pushed the regulators to develop more efficient risk monitoring systems[8]. The new candidate sector under question was that of algorithmic systematic trading which led to the flash crash of May 6, 2010, in which the Dow Jones Industrial Average lost almost 10% of its value in matter of minutes. However, the current state of the art risk models are the ones inspired by the last subprime crisis and are essentially models of networks in which each node can be impacted by the connected nodes through contagion [29] and is better suited to lower frequency, linear models. Indeed, on 06/08/2011 a seemingly relatively unnoticed event occurred on the natural gas commodities market. We say "relatively unnoticed" simply because the monetary impact was limited and finance is unfortunately an industry in which warning signs are usually dismissed until it is too late. We can see from Figure 1 in [41] that clearly something non-random is occurring. This feeling is exacerbated by the strong intuition that only interacting agents falling into some sort of quagmire could yield such series of increasing oscillations followed by a mini crash. Indeed, commodities has historically been seen as a physical market, this in turn meaning that the prices are driven by supply and demand of commodities which can be consumed, stored and/or produced. This particular point is a unique feature compared to the other markets (Equities, FX, or Rate). Also this suggests that the common, though perhaps a bit lazy view, that crashes occur through totally unpredictable [59] events may not be true for algorithmic trading.

2) Rational: The HFTE is currently a model under construction which formalism and conclusions need to be ironed out in a more rigorous fashion [41]. Despite its seemingly unfinished aspect, it still exposes through the bridge-fields it connects with, to the complexity of doing proper risk management at the high frequency level. In order to understand the latter model though, we propose to go over a relevant literature review of theoretical biology (in section IV-B), more specifically Predator/Prey models, as the latter can be seen as a very rough deterministic and perfectly visible simplification of what would be required in order to study an ecosystem of strategies. We will also see in subsection IV-C an Optimal Control Theory review and in subsection IV-D a relevant Game Theoretical review. As the reader will notice the added relative complexity of these fields is to be understood as only a lower deterministic band of what it would be required to study risk at the stochastic level.

B. Theoretical Biology & Predator/Prey models

1) Review: To bring context it was discussed in the 1960s [28] that complexity in an ecosystem insures its stability or to keep the same jargon "communities not being sufficiently complex to damp out oscillations" [19], [32] have a higher likelihood of vanishing. It also is widely accepted, in the context of ecosystem simulation, that complexity should always arise from simplicity [44], [8]. The diversity-stability debate in the context of ecosystem modeling has been ongoing since the 1950s [45] with no consensus being ever reached. It was initially believed [45], [40], [18] nature was infinitely complex and therefor more diverse ecosystem should insure more stability. This assertion was however ultimately challenged through rigorous mathematical specification [44], [64], [50] in the 1970s and 1980s by using Lotka-Volterra's Predator/Prey model initially published in the 1920's [63], [39] with similar "non-intuitive" results. More recently the work has been extended to small ecosystems of three-species food chain [7]. The intuitive 3 species example we have chosen to discuss is the one containing Sharks (chosen to be the z parameter), Tuna (chosen to be the y parameter) and Small Fishes (chosen to be the x parameter), the idea

[8] In this context risk is viewed as a mixture or Market and Reputation.

being that tunas eat small fishes which in turn are eaten by sharks. Without loss of generality sharks are assumed to die of natural causes and their decomposing bodies go on to feed the small fishes. The set of differential equations has been summarized in equation (11).

$$\begin{cases} \frac{dx}{dt} &= ax - bxy \\ \frac{dy}{dt} &= -cy + dxy - eyz \\ \frac{dz}{dt} &= -fz + gyz \end{cases} \tag{11}$$

where a is the natural growth rate of species x in the absence of predator, d the one of y in the absence of z. We also have b representing the negative predation effect of y on a and e the one of z on y. We also have g which mirrors the efficiency of reproduction of z in the presence of prey y. Note that we assume that x never dies of natural causes (if it's too old then it can't run fast enough to outrun predator y) but this is not the case for z since it is an alpha predator and therefore needs some natural death rate which is symbolized by f. This relatively simple system of three equations has been studied extensively [45] for stability. For example figure 14 represents a particular instance in which the system is unstable. Indeed, we can notice that the oscillations between the 3 species increases to the point, here not shown, where the amplitudes are so big that z goes instinct and at which point x and y start oscillating, with however a constant amplitude. We refer the motivated reader back to the original papers [45] for the other cases and interesting idiosyncratic properties. One interesting point to notice is that in cases of "relative best stability", in which $a = b = c = d = e = f = g = 1\%$ from figure 13, we have oscillation which are stable through time with the highest peek from the ultimate prey (x) coming first with the highest peek and the the one of the ultimate predator (z) coming last but with the smallest amplitude. This suggest that sophisticated working trading strategies[9] need enough prey like strategies[10] in the same ecosystem to get them to be profitable. One other interesting observation is that the total ecosystem population as depicted in the thick black line from the same figure suggest that it itself oscillates which may not necessarily be intuitive. Indeed one could have speculated that the loss of a species directly benefits the other and that therefore the total population should stay constant. This interesting observation suggest that the oscillations of a financial market may likewise be subject of similar dynamics: a financial ecosystem may go through periods in which it thrives followed by period in which it declines, the economy itself is cyclical with, some may argue oscillations which are more and more important like one depicted by the unstable ecosystem from figure 14. The stunning similarities of the competitive resource driven biological ecosystem along with some compelling similarities in some of its cyclical behavior makes the Lotka-Volterra n-species food chain equation an interesting candidate when it comes to studying the stability of the financial market in the context of the HFTE because

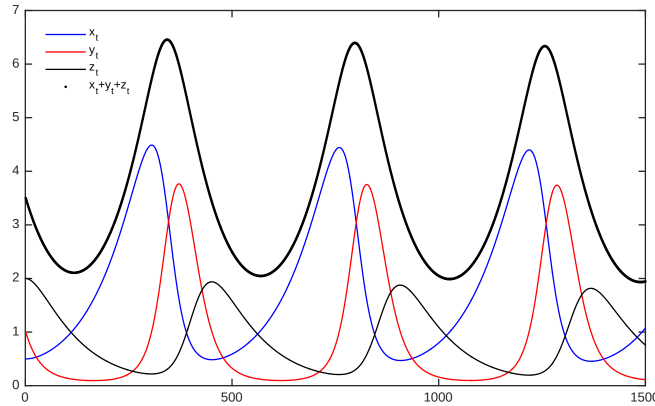

Fig. 13. The Lotka-Volterra three-species food chain equation 11 with $x_1 = 0.5$, $y_1 = 1$, $z_1 = 2$ and $a = b = c = d = e = f = g = 1\%$

of its systematic rule based approach and zero sum game like roots. However, these are hypothesis that we need to check more rigorously.

2) Regulatory Implications: The second and last immediate application we will take a look at in the context of this paper is the one of systemic risk. Given that this paper proposes that the fluctuations of the markets are linked to the frequency of the strategies composing the ecosystem of the market, we propose a model which would take advantage of this assumptions to build original high level regulations. The exercise would consist of monitoring these strategies interactions and flag the market when necessary. Suppose now that we label strategies of figure 17, 18 and 19 by respectively x, y and z and we use equation (11). If we can somehow infer what the frequency of x, y and z are in the ecosystem, then we can study whether or not the ecosystem is stable [7]. Returning to the actual mathematical study of the stability of the financial market, determining a market composed of 3 strategies is stable requires studying the Jacobian matrix J from equation (12).

$$J(x,y,z) = \begin{bmatrix} a - by & -xb & 0 \\ yd & -c + dx - ez & -ye \\ 0 & -zg & -f + gy \end{bmatrix} \tag{12}$$

By examining the eigenvalues of $J(x,y,z)$ we can indirectly gain information around the equilibrium of our financial system at the regulatory level[11]. More specifically if all eigenvalues of $J(x,y,z)$ have negative real parts then our system is asymptotically stable. Figure 14 gives an illustration of a situation in which one of the eigenvalues is negative. Many questions could be raised here: how can the regulators gain information on the parameters composing systems of equation (11)? Also the market has surely more than 3 types of strategies, how many exactly? Are these strategies easily classifiable in terms of prey, predator and super predator or can you find more subtle instances? It is very likely that trading desks especially in the high frequency domain refuse

[9] perhaps from top algorithmic desks in top tier investment banks?
[10] perhaps the retail clients of the world?
[11] we assume for the sake of this example that we only have 3 strategies

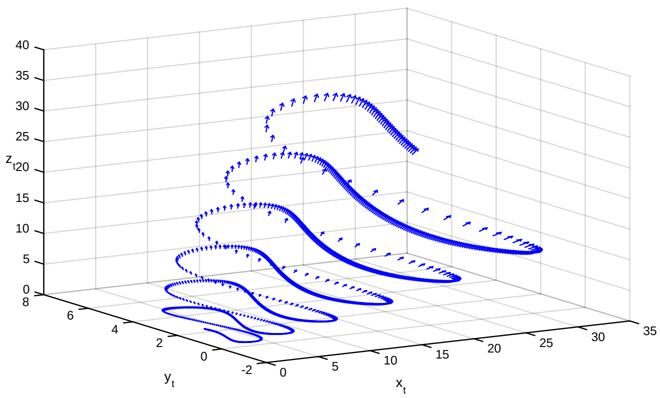

Fig. 14. The Lotka-Volterra three-species food Chain equation 11 with $x_1 = 0.5$, $y_1 = 1$, $z_1 = 2$, $a = b = c = d = e = f = 1\%$ and $g = 1.6\%$

to provide their sets of strategies for the regulators to study the Jacobian matrix in order to take the relevant actions[12].

C. Optimal Control Theory

The Hamilton-Jacobi-Bellman (HJB) partial differential equation [4] was developed in 1954 and is widely considered as a central theme of optimal control theory. Its solutions is the value function giving the minimum cost for a given dynamical system and its associated cost function. Solved locally, the HJB is a necessary condition, but when over the entire of state space, it is referred to as necessary and sufficient for an optimum. Its method can be generalized to stochastic systems. Its discrete version is referred to as the Bellman equation and its continuous version, the Hamilton-Jacobi equation. Formally we consider the problem in deterministic optimal control over the time period $[0, T]$:

$$V(x(0), 0) = \min_u \left\{ \int_0^T C[x(t), u(t)]\, dt + D[x(T)] \right\} \quad (13)$$

where $C[]$ is the scalar cost rate function, $D[]$ is the utility at the final state, $x(t)$ the system state vector with $x(0)$ usually given, and finally $u(t)$ where $0tT$ is called the control vector we aim at finding. The system of equation is also subject to $\dot{x}(t) = F[x(t), u(t)]$ where $F[]$ is a deterministic vector describing the evolution of the state vector over time. The HJB partial differential equation is given by:

$$\dot{V}(x, t) + \min_u \{\nabla V(x, t) \cdot F(x, u) + C(x, u)\} = 0 \quad (14)$$

subject to the terminal condition $V(x, T) = D(x)$. $V(x, t)$, commonly known as the Bellman value function (our unknown scalar) represents the cost incurred from starting in x at time t and controlling the system optimally until T. $V(x(t), t)$ is the optimal cost-to-go function, then by Bellman's principle of optimality from time t to $t + dt$, we have $V(x(t), t) = \min_u \left\{ V(x(t + dt), t + dt) + \int_t^{t+dt} C(x(s), u(s))\, ds \right\}$. The Taylor expansion of the first term is $V(x(t + dt), t + dt) = V(x(t), t) + \dot{V}(x(t), t)\, dt + \nabla V(x(t), t) \cdot \dot{x}(t)\, dt + o(dt)$ where $(o)(dt)$ denotes the higher order terms of the Taylor expansion. Canceling $V(x(t), t)$ on both sides and dividing by dt, and taking the limit as dt approaches zero, we obtain the HJB equation. Its resolutions is done backwards in time which can be extended to its stochastic version. In this latter case we have $\min_u \mathbb{E} \left\{ \int_0^T C(t, X_t, u_t)\, dt + D(X_T) \right\}$, with this time $(X_t)_{t\in[0,T]}$ being stochastic and needing optimization and $(u_t)_{t\in[0,T]}$ the control process. By first using Bellman and then expanding $V(X_t, t)$ with Ito's rule, one finds the stochastic HJB equation $\min_u \{\mathcal{A}V(x, t) + C(t, x, u)\} = 0$ where $\mathcal{A}$ represents the stochastic differentiation operator, and subject to the terminal condition $V(x, T) = D(x)$[13].

D. Game Theoretical Review

Another area of investigation is the one of Game Theory. Broadly speaking the prisoner's dilemma (PD) can be formalized into a matrix[14] of 2 by 2 with CC, CD, DC and DD with respective payoffs (2,2), (0,3), (3,0) and (1,1). The reason why this game theory concept is within the family of dilemmas is because although the prisoners clearly should cooperate here, given that they do not know what the other is going to do, by expectation (with equal probability for a C and a D) any user should deceit given that the expectation of the payoff for a deceit is 2 as opposed to a 1 for a cooperation.

1) Axelrod's computer tournament: however this dilemma presented in the previous subsection proved to shuffle the rules of payoff strategy optimality when the game became iterative, Robert Axelrod main contribution to the field. Indeed Axelrod [1], [2] designed a computer tournament which aim was to take a look at what strategy would prevail in an iterative format. In that occasion he invited few Mathematicians, Computer Scientists, Economists and Political Scientists to code a strategy they believed could win such tournament with the constraints of a PD rules in which it is not known when the tournament will stop[15]. Many strategies were thrown into this ecosystem in form of a tournament ranging from being simplistic like "Always Deceit" (AD) strategy[16] to many other more complicated strategies which generic representation can be looked at in Figure 15b). Surprisingly the Tit For Tat (TFT) strategy came at the top of this tournament. The TFT is considered in the literature to be a nice strategy, meaning that it is never the first to deceit (its first move is by design to be a C), but it is also a strategy that is able to retaliate in situation in which it was previously deceited. Finally, it is a strategy that is

[12]instruct the trading desks to increase or decrease their notionals so as to enforce a manual intervention for the sake of the market's stability

[13]the randomness has disappeared.

[14]Figure 15a)

[15]eg: it is by expectation best to deceit if one plays the PD only once. By iteration he should always deceit on the last move, but knowing this, the adversary should also deceit. Using this logic each player should deceit on the next to the last move and the same logic kicks in and very quickly one is led to arrive to the conclusion that he/she should deceit from the very first move.

[16]or its mirror: the AC "Always Cooperate" (AC) strategy

able to forgive meaning that if it sees that the adversary has decided to cooperate after a deceit, then he switches back to a C.

2) Evolutionary Dynamics: Martin Nowak [48] recently enhanced some of Axelrod's work by introducing new strategies and further developing the concepts of invasion/dominance[17] within a competitive strategic ecosystem. For instance as we can see from Figure 15d) that some strategies invade others but these latter strategies can be in turn invaded by other ones which in turn can be invaded by the very first strategy mentioned and induce cycles[18]. Indeed an ecosystem composed of a set of unbiased random strategies (that would randomly C or D) would invite the invasion of an ALLD (always defect) kind. In term the frequency of ALLD would take the ecosystem which would invite the TFT strategy which would benefit from the mutual cooperation when within the same proximity etc ... Figure (15) exposes how some of these strategies may interact with each other. The following additional information may help

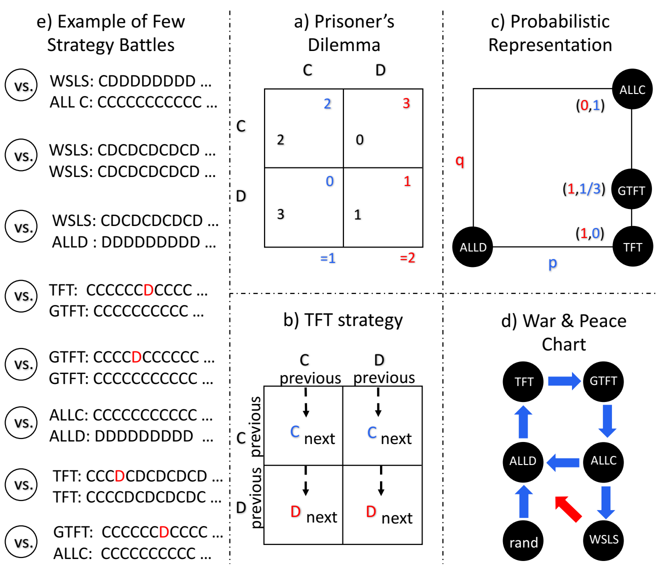

Fig. 15. Some classic Game theory representations [48].

in refreshing what some of these acronyms mean:

- TFT (*Tit of Tat*) developed in the previous section
- GTFT (*Generous Tit of Tat*) which makes it slightly less grudge prone compared to the TFT as it only deceits for 2 successive D's from the opponent.
- WSLS (*Win-Stay, Lose-Shift*) that outperforms tit-for-tat in the Prisoner's Dilemma game [48], [57]
- ALLD (*Always Deceits*) which is self explanatory
- ALLC (*Always Cooperates*) which is also self explanatory
- rand (*Random Strategy*) which outputs a C or a D with equal probability.

The main takeaway from this parallel was to expose how the rise and fall of strategies can easily be engineered through simple systematic rules based on an ecosystem and how complexity can be induced from simple rules.

E. Review of the HFTE model

Recently, the concept of ecosystems of strategies [35] was introduced. Though the idea had great potential, the paper assumes a set of static strategies which does offer to some extent an interesting snapshot of the market but does not offer:

- a history for this snapshot,
- an inspiring future for the field,
- a topology for these strategies (in the form of a DNA),
- a sense of how to study the stability of the ecosystem,
- insight about how this should impact the regulatory horizon,
- a connection to other fields[19] with concepts and properties that could be used to increase our mathematical weaponry.

Definition (HFTE): We call HFTE the High Frequency Trading Ecosystem model which attempt is to answer the points raised.

F. Network & learning potential

Two important milestones in Machine Learning are worth remembering, as they shed light on why the core building blocks of our HFTE model is a certain way. First, Warren McCulloch and Walter Pitts [51] introduced their threshold logic model in 1943 which is agreed to have guided the research in network topology as it relates to artificial intelligence for more or less a decade. Second, Rosenblatt [54], formally introduced the perceptron concept in 1962 though some early stage work had started in the 1950s. The idea of the perceptron was one in which the inputs x_1 and x_2 could act as separators[20] and therefore a direct equivalence could be made to the multi-linear regression which we will elaborate on more in details is section IV-G.2. One observed limitation of the perceptron as described by Rosenblatt, in 1969, was that a simple yet critical well known functions such as the XOR function could not be modeled [46]. This resulted in a loss of interest in the field until it was shown that a Feedforward Artificial Neural Network (ANN) with two or more layers could in fact model these functions. Added, to this we have the well known overfitting [58] problems when it comes to supervised learning which has been there since inception though regular progress is being made in that domain without real breakthrough.

G. The Funnel

The Funnel, introduced by Martin Nowak [48], represents the simplest possible network to model (therefore which minimizes overfitting) the key functions of our application. The area of evolutionary graph theory is quite rich, and

[17] by extension when applied to finance some strategies may dominate and invade others.

[18] economical cycles for example when applied to our primary problem

[19] eg: Game Theory, Mathematical Biology, Signal Processing

[20] the exact research was one in which the methodology acted as a $1,0$ through a logistic activation function $f(x) = \frac{1}{1+e^{-x}}$ as opposed to a linear one. However that small distinction is not significant enough in the context to delve too much into it but deserved a clarification in the footnotes.

graphs provide interesting properties. We can formalize the learning process from all of our strategies using the topology of Figure 16 by providing a set $\mathcal{T}$, as described by equation (15) of weights corresponding to all the possible weights of this particular figure.

$$\mathcal{T} \triangleq \left\{ \begin{array}{cc} \cup_{j\in[1,9]} w^{i}_{\bar{s},j} & \cup_{j\in[1,9]} w^{i}_{s,j}, \\ \cup_{j\in[1,9],i\in[1,3]} w^{h_1}_{\bar{s},i,j} & \cup_{j\in[1,9],i\in[1,3]} w^{h_1}_{s,i,j}, \\ \cup_{j\in[1,3]} w^{h_2}_{\bar{s},j} & \cup_{j\in[1,3]} w^{h_2}_{s,j}, \\ w^{o}_{\bar{s},j\in[1,9]} & w^{o}_{s,j\in[1,9]} \end{array} \right\} \tag{15}$$

with w^i, w^h and w^o, respectively the weights associated to the input, hidden and output layers.

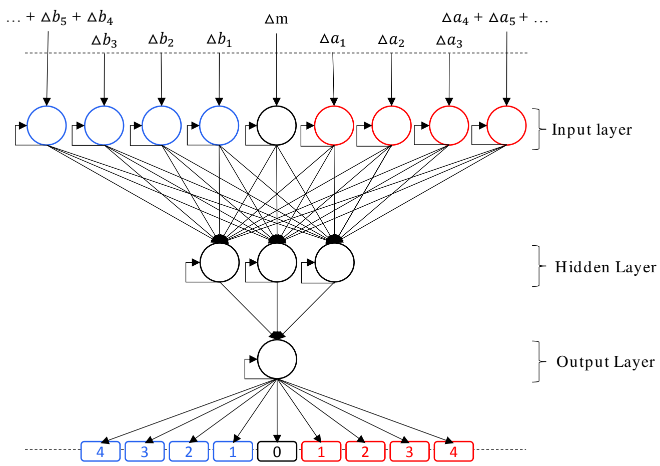

Fig. 16. The High Frequency Financial Funnel

Remark Note that in the context of this paper we have chosen to work with Martin Nowak's [48] funnel, as Figure 16. This topological structure offers the advantage of making some interesting bridges between the worlds of:

- information theory since it also resembles the classic structure of a Neural Network and can therefore easily accommodate the mapping of classic and less classic financial strategies,
- evolutionary dynamics since Moran-like Processes [47] can easily be formalized,
- biology since it is a potent amplifier of selection [48][21].

We will conclude this subsection by providing a definition of the High Frequency Financial Funnel.

Definition We define the **H**igh **F**requency **F**inancial **F**unnel (HFFF) to be a network structure with 9 inputs, 3 hidden layers and 1 output layer. Each node connects to the next layer and to itself. Each self connection will be labelled by w_s and the others by $w_{\bar{s}}$. We will admit that $w_{\bar{s}} \sim \mathcal{U}[-1, 1]$ and that $w_s \sim \mathcal{U}[0, 1]$ hence:

$$w_x \sim \mathcal{U}[-1_{x=\bar{s}}, 1] \tag{16}$$

[21] Indeed, as we will see in section IV-G.1 its simplest structure (the EWMA) serves as pillar to the section IV-G.2 (MLR) which itself does the same for the XOR strategy. So we have this incremental complexity in the network that corresponds to an incremental complexity in information processed.

1) The Trend Following Topology: a very common trading strategy is trend following (TF). The idea of TF is that if the price has been going a certain way (eg: up or down) in the recent past, then it is more likely to follow the same trend in the immediate future.

Definition The mathematical formulation of TF can be diverse but in the context of this paper we use an exponentially weighted moving average (EWMA), formally described by equation (17),

$$\hat{x}_t = (1-\lambda)x_t + \lambda\hat{x}_{t-1}, \ \lambda \in [0, 1] \tag{17}$$

in which λ represents the smoothness parameter with $\lambda \in [0, 1]$.

Remark The lower the magnitude of λ, the more the next value will be conditional to the previous value. Conversely, the higher λ, the more the future value will be function to the long term trend. The idea being that through a simple filtering process, the noise is extracted from the signal which then returns a clean time series $\hat{x}_t$.

Proposition The HFFF can model trend following strategies.

Proof: Simply set $\cup_{j\in[1,4]} w^{i}_{\bar{s},j} = 0$, $\cup_{j\in[1,4]} w^{i}_{s,j} = 0$, $\cup_{j\in[6,9]} w^{i}_{\bar{s},j} = 0$, $\cup_{j\in[6,9]} w^{i}_{s,j} = 0$ $\cup_{j\in[1,4],i\in[1,3]} w^{h_1}_{\bar{s},i,j} = 0$, $\cup_{j\in[1,4],i\in[1,3]} w^{h_1}_{s,i,j}$ $\cup_{j\in[6,9],i\in[1,3]} w^{h_1}_{\bar{s},i,j} = 0$, $\cup_{j\in[6,9],i\in[1,3]} w^{h_1}_{s,i,j}$, $w^{h}_{\bar{s},3} = 0$, $w^{h}_{s,1} = 0$ and $w^{h}_{s,3} = 0$. ■

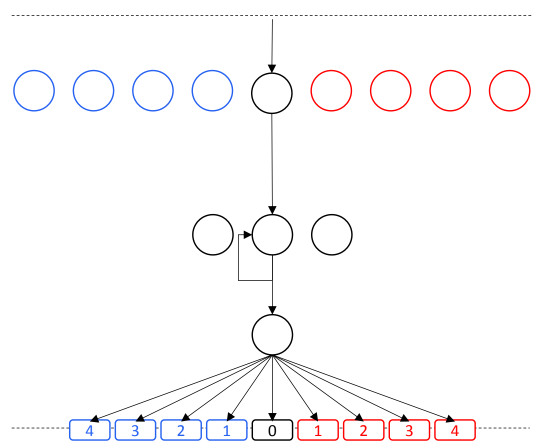

Fig. 17. The EWMA strategy translated in terms of network topology (the weights equal to 0 have not been represented)

Remark The proof is visually illustrated by Figure 17 (the weight equal to 0 have not been represented).

2) Multi Linear Regression Topology: the Multi Linear Regression (MLR) is another well known strategy traders have been using for a time in the industry.

Definition Given a data set $\{y_i, x_{i-1,1}, \ldots, x_{i-1,9}\}_{i=1}^{n}$, where n is the sample size, $\{\beta_i\}_{i=1}^{9}$, the weight of the explanatory variables and y_i the output, then our MLR is formalized by

$$\begin{aligned} y_i &= \beta_1 x_{i-1,1} + \cdots + \beta_9 x_{i-1,9} + \varepsilon_i \\ &= \mathbf{x}_{i-1}^{\mathrm{T}}\beta + \varepsilon_i, \ i = 1, \ldots, n \end{aligned} \tag{18}$$

Proposition The HFFF can model multi linear regression like strategies.

Proof: Simply set $\cup_{j\in[1,4]} w^{i}_{\bar{s},j} = 0$, $\cup_{j\in[1,4]} w^{i}_{s,j} = 0$, $\cup_{j\in[6,9]} w^{i}_{\bar{s},j} = 0$, $\cup_{j\in[6,9]} w^{i}_{s,j} = 0$, $w^{h}_{\bar{s},1} = 0$, $w^{h}_{\bar{s},3} = 0$, $w^{h}_{s,1} = 0$, $w^{h}_{s,3} = 0$. ■

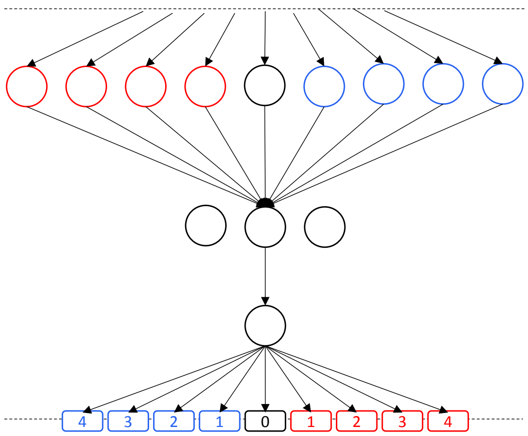

Fig. 18. The MLR strategy translated in terms of network topology

Remark We will make 4 remarks:

- MLR is illustrated in Figure 18 (weights equal to 0 have not been represented).
- Logistic or weighted MLR can be modeled through the topology of Figure 18 by simply changing respectively the activation function (from linear to logistic) and the weights.

3) XOR Topology: How is the XOR function relevant to HFT? Let's look at the following known HF rational.

Definition If we define the Open Interest (OI) as being the total volume left on the order book then it is known that when:

- the price and the OI are rising then the market is bullish,
- the price is rising but the OI is falling then the market is bearish,
- the price is falling but the OI is rising then the market is bearish,
- the price and OI are both falling then the market is bullish.

Remark These 4 market situations can be summarized by table IV-G.3.

Proposition The HFFF can model XOR like strategies.

Proof: Simply set $\cup_{j\in[1,4]} w^{i}_{\bar{s},j} = 0$, $\cup_{j\in[1,4]} w^{i}_{s,j} = 0$, $\cup_{j\in[6,9]} w^{i}_{\bar{s},j} = 0$, $\cup_{j\in[6,9]} w^{i}_{s,j} = 0$, $w^{h}_{\bar{s},1} = 0$, $w^{h}_{\bar{s},3} = 0$, $w^{h}_{s,1} = 0$, $w^{h}_{s,3} = 0$. ■

Remark We will make the following 2 observations:

- The preceding proof is visually illustrated by Figure 19 (the weights equal to 0 have not been represented).
- The XOR topology can be designed in various ways.

Open Interest	Price	Combined Symbol	Signal
Rising	Rising	⇈	Buy
Rising	Falling	⇅	Sell
Falling	Rising	⇵	Sell
Falling	Falling	⇊	Buy

TABLE I
THE RELATIONSHIP BETWEEN OPEN INTEREST (OI), PRICE (I) & SIGNAL FOR XOR STRATEGY [41]

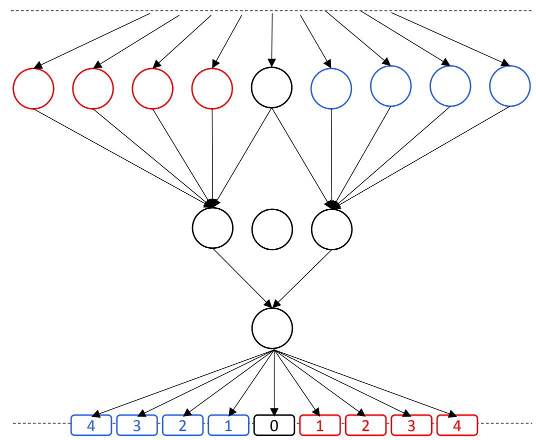

Fig. 19. The XOR strategy translated in terms of network topology

4) Execution strategy: to make the problem more realistic, one needs to formalize an execution strategy which would apply to all strategies, but still be rule based and function of its topology. In this paper we will take the simple approach in which all strategies have that same execution strategy.The idea of this algorithm will be that:

- the execution strategy will be subject to a certainty-like function,
- certainty will be decided by the historical returns from the relevant topology split into intervals,
- since the decision needs to be made and that data comes regularly a rolling percentile function should be used.

In this context our algorithm returns a value between 0 and 9, the 9 explanatory variables of our HFFF and corresponding to all of the admissible actions in our order book. The tested input is compared against the current output as it compares to the historical outputs and returns the corresponding percentile which then goes on populating the order-book. Given that no history exists in the first iteration and that the first few iterations are not significant, we will randomize the first R_n iterations.

H. Genetic Algorithm as a means to study the market through time

We will take a look at a couple of methods to study the market through time. We first take, in this subsection, an approach with the objective to gain intuition in order to strategize with respect to future research and then a second method which is mathematically more optimal in lieu of the RJ-MCMC mentioned in the literature review from the “Report Format Document”.In this “intuitive” section we specify the genetic algorithm which we have used to study our problem with intuition in mind as opposed to optimality.

1) Looping & Fitting Function: Throughout this subsection we will refer to Micro and Macro increments.

Definition We will define two types of iterations:

- the first type being **Micro** corresponding to an infinitesimal increment in our environment, namely an increment in which a strategy S analyses and in turn changes the order book by placing a order itself.
- the second type being **Macro**, corresponding to a generational increment in our environment, namely *a certain equal number* of Micro increments per strategy leading to a calculation of a Profit and Loss (P&L) and a survival process[22] based on this P&L.

We will label as N_k the number of total live strategies, N_k^e the number of tr**e**nd following like strategies, N_k^m the number of **m**ulti-linear regretion like strategies, N_k^r the number of xo**r** like strategies and N_k^o the number of ***o**ther unclassified* strategies[23]. The relationship between these entities can be summarized by equation (19).

$$N_k = N_k^e + N_k^m + N_k^r + N_k^o \tag{19}$$

A strategy $\mathcal{S}$ will consist of a topology $\mathcal{T}$, a rolling P&L $\mathcal{P}$ and a common orderbook $\mathcal{O}$ as shown by equation (20).

$$\mathcal{S} \triangleq \{\mathcal{P}, \mathcal{T}, \mathcal{O}\}\,. \tag{20}$$

2) Survival & birth processes: the survival, death & birth processes are a set of processes which impact the number of live strategies N_k at any time k the following way. If we call $\mathcal{S}_{N_k} = S_{(1)}, S_{(2)}, \ldots, S_{(n)}, S_{(n+p)}, \ldots, S_{(N_k)}$, the strategies ranked with respect to their P&L from highest to lowest, we will admit the following definitions:

Definition The Survivor set[24] is the set of strategies with a positive P&L. Namely if $\mathcal{S}_a = S_{(1)}, S_{(2)}, \ldots, S_{(s)}$ with $S_{(s)} \geqslant 0$ and $S_{(s+1)} < 0$. We will subdivide this set by distinguishing:

- secondary survivors set with cardinality $a_2 = \lfloor \frac{s}{2} \rfloor$, survive without reproducing
- primary survivors set with cardinality $a_1 = s - a_2$, survive and have one offspring with a “slightly different DNA” in form of a conditional resampling of their topology.

Definition We will call the Birth process, the set of rules conducting the selection of top strategies and their reproduction with mutations. The protocol starts by selecting the ranked first half of survived strategies. Namely, if $a_1 = b = \lfloor \frac{s}{2} \rfloor$ the strategies $\mathcal{S}_1 \ldots \mathcal{S}_{a_1}$ will both survive and reproduce and create a set of equal size but with a slightly different topology and with cardinality $b = a_1$.

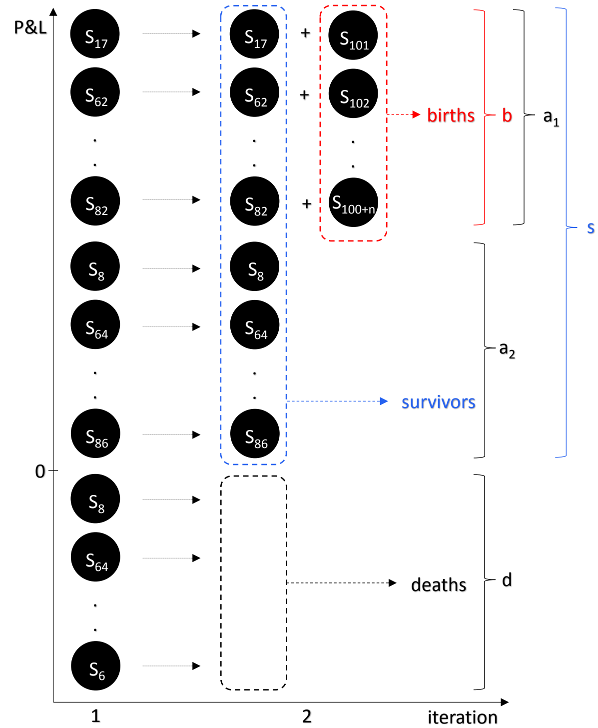

Fig. 20. Illustration for the Death and Birth processes in our GA

Definition We define the Death process, the set of protocols aiming at eliminating part of the strategies from our ecosystem, more specifically, the set of strategies with a negative P&L. Namely if $\mathcal{S}_d = S_{(s+1)}, S_{(s+2)}, \ldots, S_{(N_k)}$ will disappear from the market at Macro iteration $k+1$.

Remark We can easily see that $s = a_1 + a_2$, $a_1 \geqslant a_2$, $a_1 = b$. Figure 20 illustrates these definitions.

3) Inheritance with Mutations : the intuition about the mutation process is that each birth is a function of a successful strategy (the positive P&L of parents $\mathcal{S}_1 \ldots \mathcal{S}_{a_1}$) and should resemble the single parent[25] which produced it. We have seen in section IV that the DNA of our strategies is essentially their topology $\mathcal{T}$ (which is itself a combination of weights). We will therefore concentrate on performing the re-sampling on the weights of the offspring. The reason why this distribution is interesting[26] is that:

- is defined in a closed interval [0,1] and can therefore be rescaled easily through a change of variable to [-1,1], an interval which is a basic way of formalizing a normalized importance of each node in the topology decision making of Figure 16.
- on the contrary to the uniform distribution, it is more flexible and offers a broad range of interesting shapes allowing the possibility to code a conditional resampling

[22] explained next
[23] This label will be the same in section IV-B.
[24] or alternatively alive process
[25] no crossover in this model
[26] though, again not optimal

model and therefore make clever proximity changes around the symbolic levels: -1, 0 and 1. This way we can prevent too large deviations and rather select small incremental changes and intuitively follow the principles of selection. We can see that the $\mathcal{B}eta(x, 1, 7)$ or $\mathcal{B}eta(1-x, 1, 7)$ both concentrate a great deal of the distribution towards 0 and 1 respectively. Likewise $\mathcal{B}eta(x, 3, 7)$ and $\mathcal{B}eta(x, 5, 7)$ provide a more Gaussian like distribution towards in between zones which is what we want.

I. Preliminary results

There are issues related to handling classification scopes as well as bespoke simulation issues but those are outside the scope of this paper. However, few of the simulations seemed to indicated a positive correlation between TF and MLR strategies on strictly increasing or strictly decreasing markets. When the market trend happened to be less clear the correlation between their growth rate seemed less significant and perhaps even negative. It was difficult to make a proper quantification of these observations due to the low speed of each simulation and also because the observation conditional to market tendencies came a posteriori of the simulations. Not enough simulations were performed to really be able to assert the mentioned relationship definitively. Similar results were found as for the relationship between XOR and MLR strategies, though with even less significance. Few simulations were actually such that, the results remind us to the Lotka-Volterra 3-predator-prey model with however a great deal of noise and unclassified strategies. These limitations are currently motivations for additional work associated to the HFTE model. Finally our first paper [41] ended with the conjecture below:

Conjecture Diversity in financial strategies in the market leads to its instability.

J. Tracking the High Frequency Ecosystem

To be able to appreciate the complexity of tracking the high frequency market, which simplification is perhaps the HFTE, Mahdavi-Damghani [42] recently proposed a methodology to track the ecosystem through time. In order to appreciate the complexity of the tracking task let us first recall some results from Sequential Monte Carlo (SMC) methods.

1) Sequential Monte Carlo Methods: SMC methods [15], [38] known alternatively as Particle Filters (PF) [26], [36] or also seldom CONDENSATION [33], are statistical model estimation techniques based on simulation. They are the sequential (or 'on-line') analogue of Markov Chain Monte Carlo (MCMC) methods and similar to importance sampling methods. If they are elegantly designed they can be much faster than MCMC. Because of their non linear quality they are often an alternative to the Extended Kalman Filter (EKF) or Unscented Kalman Filter (UKF). They however have the advantage of being able to approach the Bayesian optimal estimate with sufficient samples. They are technically more accurate than the EKF or UKF. The aims of the PF is to estimate the sequence of hidden parameters, x_k for $k = 1, 2, 3, \ldots$, based on the observations y_k. The estimates of x_k are done via the posterior distribution $p(x_k|y_1, y_2, \ldots, y_k)$. PF do not care about the full posterior $p(x_1, x_2, \ldots, x_k|y_1, y_2, \ldots, y_k)$ like it is the case for the MCMC or importance sampling (IS) approach. Let's assume x_k and the observations y_k can be modeled in the following way:

- $x_k|x_{k-1} \sim p_{x_k|x_{k-1}}(x|x_{k-1})$ and with given initial distribution $p(x_1)$.
- $y_k|x_k \sim p_{y|x}(y|x_k)$.
- equations (21) and (22) gives an example of such system.

$$x_k = f(x_{k-1}) + w_k \tag{21}$$

$$y_k = h(x_k) + v_k \tag{22}$$

It is also assumed that $cov(w_k, v_k) = 0$ or w_k and v_k mutually independent and iid with known probability density functions. $f(\cdot)$ and $h(\cdot)$ are also assumed known functions. Equations (21) and (22) are our state space equations. If we define $f(\cdot)$ and $h(\cdot)$ as linear functions, with w_k and v_k both Gaussian, the KF is the best tool to find the exact sought distribution. If $f(\cdot)$ and $h(\cdot)$ are non linear then the Kalman filter (KF) is an approximation. PF are also approximations, but convergence can be improved with additional particles. PF methods generate a set of samples that approximate the filtering distribution $p(x_k|y_1, \ldots, y_k)$. If N_P in the number of samples, expectations under the probability measure are approximated by equation (23).

$$\int f(x_k)p(x_k|y_1, \ldots, y_k)dx_k \approx \frac{1}{N_P}\sum_{L=1}^{N_P} f(x_k^{(L)}) \tag{23}$$

Sampling Importance Resampling (SIR) is the most commonly used PF algorithm, which approximates the probability measure $p(x_k|y_1, \ldots, y_k)$ via a weighted set of N_P particles

$$\left(w_k^{(L)}, x_k^{(L)}\right) \;:\; L = 1, \ldots, N_P \tag{24}$$

Algorithm 1 RESAMPLE(w)

Require: array of weights w_1^N
Ensure: array of weights w_1^M resampled

```
u^0 ~ U[0, 1/M]
for m = 1 to N do
    i^(m) <- ⌊(w_n^(m) - u^(m-1) m)⌋ + 1
    u^(m) = u^(m) + i^(m)/M - w_n^(m)
end for
```

The importance weights $w_k^{(L)}$ are approximations to the relative posterior probability measure of the particles such that $\sum_{L=1}^{P} w_k^{(L)} = 1$. SIR is a essentially a recursive version of importance sampling. Like in IS, the expectation of a function $f(\cdot)$ can be approximated like described in equation

(25).

$$\int f(x_k)p(x_k|y_1,\ldots,y_k)dx_k \approx \sum_{L=1}^{N_P} w^{(L)} f(x_k^{(L)}) \tag{25}$$

The algorithm performance is dependent on the choice of the proposal distribution $\pi(x_k|x_{1:k-1}, y_{1:k})$ with the optimal proposal distribution being $\pi(x_k|x_{0:k-1}, y_{0:k})$ in equation (26).

$$\pi(x_k|x_{1:k-1}, y_{1:k}) = p(x_k|x_{k-1}, y_k) \tag{26}$$

Because it is easier to draw samples and update the weight calculations the transition prior is often used as importance function.

$$\pi(x_k|x_{1:k-1}, y_{1:k}) = p(x_k|x_{k-1})$$

The technique of using transition prior as importance function is commonly known as Bootstrap Filter and Condensation Algorithm. Figure 21 gives an illustration of the algorithm just described. Note that on line 5 of algorithm 2, $\hat{w}_k^{(L)}$, simplifies to $w_{k-1}^{(L)} p(y_k|x_k^{(L)})$, when $\pi(x_k^{(L)}|x_{1:k-1}^{(L)}, y_{1:k}) = p(x_k^{(L)}|x_{k-1}^{(L)})$.

Algorithm 2 SMC(w)

Require: array of weights w_p^N, $\pi(x_k|x_{1:k-1}^{(L)}, y_{1:k})$
Ensure: array of weights w_p^N resampled

1: **for** $L = 1$ to N_P **do**
2: $\quad x_k^{(L)} \sim \pi(x_k|x_{1:k-1}^{(L)}, y_{1:k})$
3: **end for**
4: **for** $L = 1$ to N_P **do**
5: $\quad \hat{w}_k^{(L)} = w_{k-1}^{(L)} \frac{p(y_k|x_k^{(L)})p(x_k^{(L)}|x_{k-1}^{(L)})}{\pi(x_k^{(L)}|x_{1:k-1}^{(L)}, y_{1:k})}$
6: **end for**
7: **for** $L = 1$ to N_P **do**
8: $\quad w_k^{(L)} = \frac{\hat{w}_k^{(L)}}{\sum_{J=1}^{P} \hat{w}_k^{(J)}}$
9: **end for**
10: $\hat{N}_{\mathit{eff}} = \frac{1}{\sum_{L=1}^{P}\left(w_k^{(L)}\right)^2}$
11: **if** $\hat{N}_{\mathit{eff}} < N_{thr}$ **then**
12: resample: draw N_P particles from the current particle set with probabilities proportional to their weights. Replace the current particle set with this new one.
13: **for** $L = 1$ to N_P **do**
14: $\quad w_k^{(L)} = 1/N_P$.
15: **end for**
16: **end if**

2) Resampling Methods: Resampling methods are usually used to avoid the problem of weight degeneracy in our algorithm. Avoiding situations where our trained probability measure tends towards the Dirac distribution must be avoided because it really does not give much information on all the possibilities of our state. There exists many different resampling methods, Rejection Sampling, Sampling-Importance Resampling, Multinomial Resampling, Residual Resampling, Stratified Sampling, and the performance of our algorithm can be affected by the choice of our resampling method. The stratified resampling proposed by Kitagawa [37] is optimal in terms of variance. Figure 21 gives an illustration of the Stratified Sampling and the corresponding algorithm is described in algorithm 1. We see at the top of the figure 21 the discrepancy between the estimated pdf at time t with the real pdf, the corresponding CDF of our estimated PDF, random numbers from $[0, 1]$ are drawn, depending on the importance of these particles they are moved to more useful places.

3) Importance Sampling : Importance sampling (IS) was first introduced in [43] and was further discussed in several books including in [30]. The objective of importance sampling is to sample the distribution in the region of importance in order to achieve computational efficiency via lowering the variance. The idea of importance sampling is to choose a proposal distribution $q(x)$ in place of the true, harder to sample probability distribution $p(x)$. The main constraint is related to the support of $q(x)$ which is assumed to cover that of $p(x)$. In equation (27) we write the integration problem in the more appropriate form.

$$\int f(x)p(x)dx = \int f(x)\frac{p(x)}{q(x)}q(x)dx \tag{27}$$

In IS the number, N_p, usually describes the number of independent samples drawn from $q(x)$ to obtain a weighted sum to approximate $\hat{f}$ in equation (28).

$$\hat{f} = \frac{1}{N_p}\sum_{i=1}^{N_p} W\left(x^{(i)}\right) f\left(x^{(i)}\right) \tag{28}$$

where $W(x^{(i)})$ is the Radon-Nikodym derivative of $p(x)$ with respect to $q(x)$ or called in engineering the importance weights (equation (29)).

$$W\left(x^{(i)}\right) = \frac{p\left(x^{(i)}\right)}{q\left(x^{(i)}\right)} \tag{29}$$

If the normalizing factor for $p(x)$ is not known, the importance weights can only be evaluated up to a normalizing constant, as equation (30).

$$W\left(x^{(i)}\right) \propto p\left(x^{(i)}\right) q\left(x^{(i)}\right) \tag{30}$$

To ensure that $\sum_{i=1}^{N_p} W(x^{(i)}) = 1$, we normalize the importance weights to obtain equation (31).

$$\hat{f} = \frac{\frac{1}{N_p}\sum_{i=1}^{N_p} W\left(x^{(i)}\right) f\left(x^{(i)}\right)}{\frac{1}{N_p}\sum_{i=1}^{N_p} W\left(x^{(i)}\right)} = \frac{1}{N_p}\sum_{i=1}^{N_p} \tilde{W}\left(x^{(i)}\right) f\left(x^{(i)}\right) \tag{31}$$

where $\tilde{W}\left(x^{(i)}\right) = \frac{W\left(x^{(i)}\right)}{\sum_{i=1}^{N_p} W\left(x^{(i)}\right)}$ are called the normalized importance weights. The variance of importance sampler estimate [6] in equation (31) is given by $Var_q[\hat{f}] = \frac{1}{N_p} Var_q[f(x)W(x)] = \frac{1}{N_p} Var_q[f(x)p(x)/q(x)]$. Using the integration format we

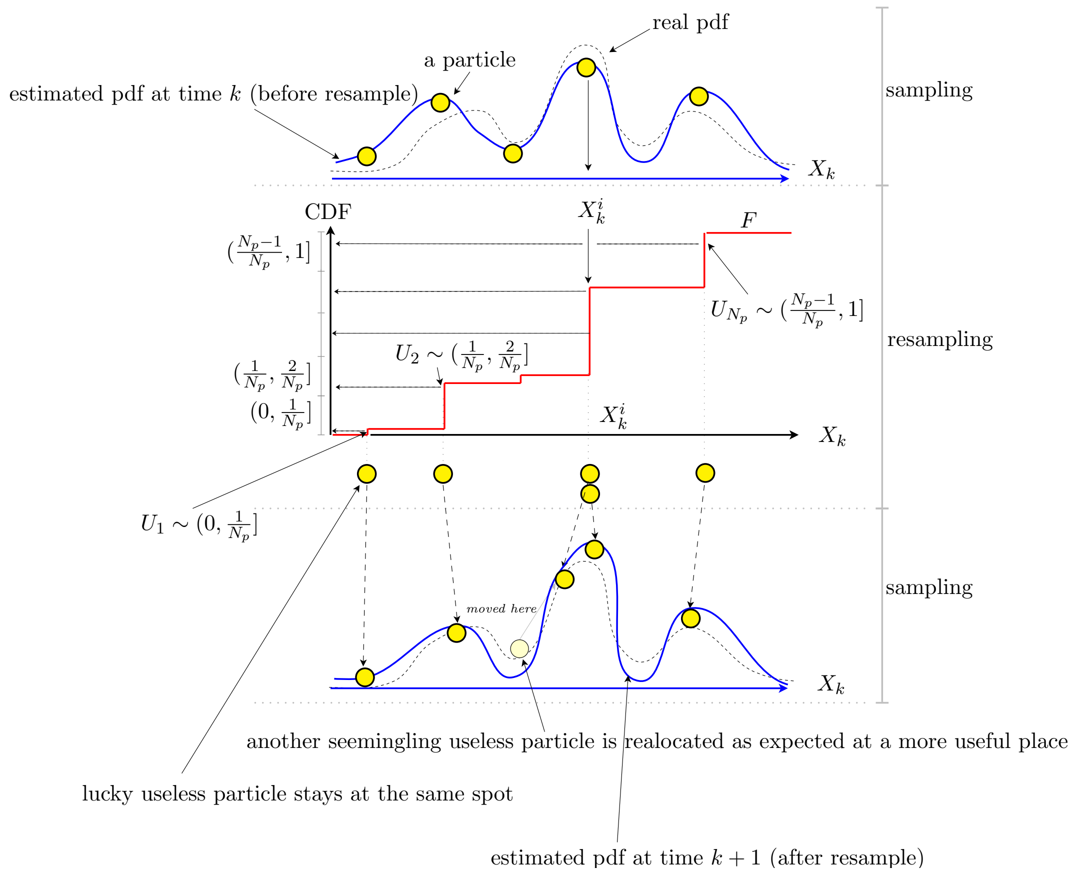

Fig. 21. Stratified Sampling illustration

have $\frac{1}{N_p}\int\left[\frac{f(x)p(x)}{q(x)}-\mathbb{E}_p[f(x)]\right]^2 q(x)dx$ which simplifies to $\frac{1}{N_p}\int\left[\left(\frac{(f(x)p(x))^2}{q(x)}\right)-2p(x)f(x)\mathbb{E}_p[f(x)]\right]dx + \frac{(\mathbb{E}_p[f(x)])^2}{N_p} = \frac{1}{N_p}\int[(\frac{(f(x)p(x))^2}{q(x)})]dx - \frac{(\mathbb{E}_p[f(x)])^2}{N_p}$. The variance can be reduced when an appropriate $q(x)$ is chosen to either match the shape of $p(x)$ so as to approximate the true variance; or to match the shape of $|f(x)|p(x)$ so as to further reduce the true variance.

Proof: $\frac{\partial Var_q[f]}{\partial q(x)} = -\frac{1}{N_p}\int[(\frac{(f(x)p(x))^2}{q(x)^2})]dx = -\frac{1}{N_p}\int[(\frac{(f(x)p(x))^2}{q(x)q(x)})]dx$. $q(x)$ having the constraint of being a probability measure that is $\int_{-\infty}^{+\infty} p(x)dx = 1$, we find that $q(x)$ must match the shape of $p(x)$ or of $|f(x)|p(x)$. ■

In essence each particle, in yellow (Figure 21), will represent an ecosystem in which the study of the risk and stability can be partially done using some of the methodology used in section IV-B. This latter problem is currently an open problem in quantitative finance.

V. Current market Risk model limitations

After exposing some of the complexities associated with the low frequency vanilla options[27] market and the high frequency trading systematic trading market[28], we summarize, in this section, some of the absurd simplifications and therefore limitations of the current risk models as used by practitioners. More specifically we examine stressed scenario generation using historical distributions in section V-A, the differences between an Anticipative and Responsive VaR models in section V-B and the opposition between a Responsive VaR and a Stable VaR in section V-C.

A. Current Modeling Issues with Historical Distributions

Depending on whether one wants to allow the risk factor to go below 0 or not, most responsive VaR models take for assumption drift-less diffusions like normality or log normality as specified in equations (32a) and (32b) in which we let $(\Omega,(\mathcal{F})_{(t\geqslant 0)},\mathbb{Q})$ be our probability space, with $(\mathcal{F})_{(t\geqslant 0)}$ generated by the $T+1$ dimensional Brownian motion and $\mathbb{Q}$ is the risk neutral probability measure.

$$dX_t = 0dt + \sigma_t dW_t \tag{32a}$$

$$\frac{dX_t}{X_t} = 0dt + \sigma_t dW_t \tag{32b}$$

Conceptually, we can observes that risk factors of different financial products are widely diverse:

- futures or shares seem to follow more of a traditional lognormal diffusion,
- basis risk or spread mean reverts but have not boundary constraints
- implied ATM vol mean reverts and remains positive,

[27] In sections III and II.
[28] In section IV.

- rates mean reverts and, up to recently were, bounded below by 0.
- implied vol skew mean reverts and remains bounded between $[-1, 1]$ [3], [11].

However the current traditional Risk methodologies assume consistently "symmetric" diffusions (relative or absolute). Some of the pitfall of these models are that:

- Bumping volatility when vols are high using relative bumps overestimates the upward moves,
- when vols are low using absolute moves can yield negative vols,
- relative bumps underestimate the potential upward moves,
- using either proportional moves or absolute moves on vol increases the probability of creating **arbitrages** on the stressed scenarios (or/and negative risk factors undesirable for vol).

Remark Note that in this section we take the example of interest rates as an asset class but the latter can easily be replaced by more complex financial concepts such as the implied volatility risk factor that we have mentioned in details in section II. Also we may chose the term "bump" from time to time, which is an industry jargon used to describe a stress testing methodology without more details.

The issues with proportional moves applied to interest rates (IR) when interest rates are high is that we grossly over-estimate their risk but when rates are low historically we wanted to avoid interest rates going negative so we would use proportional stress testing in these situation. Whether it makes sense to use either of these methodology to asses the risk of IR is not the point of this subsection but rather how we can reconcile practitioners culture with these 2 market observable phenomenon. Equation (33) is a proposal that reconcile partially these issues. The rational is that when $r_t >> \mu_t$ where μ_t is its historical rolling mean, then most of the contribution to the bumps should come in absolute term ($\underbrace{\lim}_{r_t \to \infty} \exp^{-\theta_t(r_t-\mu_t)} = 0$, so we get $\underbrace{\lim}_{r_t \to \infty} \lambda(r_t, \theta_t, \mu_t) = 1$). Similarly when $r_t << \mu_t$ we get $\underbrace{\lim}_{r_t \to \infty} \lambda(r_t, \theta_t, \mu_t) = 0$ and we get proportional moves. However, this idea needs to be backtested and the argument of the function studied a bit more in depth for special situations[29].

$$
\begin{array}{ll}
dr_t & = \lambda(r_t, \theta_t, \mu_t) dr_t^a + [1 - \lambda(r_t, \theta_t, \mu_t)] dr_t^p \\
\lambda(r_t, \theta_t, \mu_t) & = \frac{1}{1+\exp^{-\theta_t(r_t-\mu_t)}}
\end{array}
\tag{33}
$$

In this mixture model, μ_t is some sort of rolling mean that may want to rescale using a EWMA or not and which timsescale of relevance may be also optimized like is currently done by practitioners. Similarly, θ_t is optimized so as to get good backtesting results and may also have to go through a EWMA which decay factor we guess will have to be far less responsive than the μ[30]. Absolute and relative mixture models are therefore convoluted and not entirely convincing. We will see next that using stochastic calculus one may actually come up with a much more elegant solution. Once the underlying assumptions of normal and log-normality (or mixture) have been decided the stressed scenarios need to be adjusted in order to address the market change of volatility, so as to get the label of "responsiveness". There are few ways to conceptually address this concept and they all rely in the industry with a scaling of a long term volatility compared to a recent volatility. For instance if we define $\sigma_h = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \overline{x})^2}{N-1}}$ as being the standard deviation over our entire history of relevant data $[1, \ldots, N]$, and $\sigma_c = \sqrt{\frac{\sum_{i=p}^{N}(x_i - \overline{x})^2}{N-p-1}}$ the standard deviation over more recent history, with $N > p > 1$, the responsive VaR formula is given by:

$$
\begin{aligned}
\mathrm{RVaR}_\alpha(X) &= \frac{\sigma_c}{\sigma_h} \mathrm{VaR}_\alpha(X) \\
&= \frac{\sigma_c}{\sigma_h} \inf\{x \in \mathbb{R} : P(X + x < 0) \leqslant 1 - \alpha\} \\
&= \frac{\sigma_c}{\sigma_h} \inf\{x \in \mathbb{R} : 1 - F_X(-x) \geqslant \alpha\}
\end{aligned}
$$

where X is the underlying and α the quantile level.

Remark Note that p is an ongoing parameter for debate which is usually the result of an optimization by constraints problem in which, the financial institution calculating its VaR tries to minimizes its capital requirement[31] with the constraints being set by the regulators in order to make the relevant statistics significant[32].

Note that another elegant way to address this idea of most recent data must have the most to say about σ_c, an EWMA is often used as a solution since the latter does not have an arbitrary cutting point but rather the older data set contribution in the calculation of σ_c decreases exponentially. In this situation the unnormalized weights are calculated using $w_1 = 1$ and $w_t = \lambda w_{t-1}$ and the normalized weights are given by $\tilde{w}_t = \frac{w_t}{\sum_{i=1}^{N} w_i}$ with λ chosen in order to make $\sum_{i=1}^{N} w_i$ bigger than 2 years of data.

B. Anticipative Vs Responsive VaR

Most of the options' risk models currently used by market practitioners are drift-less. By drift-less we mean that most if not all the known used diffusion are solely split between Log-Normal[33] as described, in continuous time, by equation (32b) or alternatively Normal[34] as described by equation

[29]Note that when $\mu_t = r_t$, $\lambda(r_t, \theta_t, \mu_t) = \frac{1}{2}$ so the contribution of proportional bumps to absolute bumps is the same in terms of determining future potential IR scenarios.

[30]if the intuition is not clear, try to think of the following: θ needs a lot more data away from its mean to be calibrated and r moves slowly so it does not make sense to make it overly responsive

[31]and choses the best p to minimize that VaR.

[32]the Basel committee usually likes to see 2 years of data, therefore $N-p$ needs to be at least 2 years.

[33]we use historical "proportional" bumps to stress our scenarios.

[34]we use historical "absolute" bumps to stress our scenarios.

(32a). Whether we use either of these models, the family of responsive VaR can be described as one in which the dynamic adjustment in the VaR model is **lagging** with respect to the sudden change in market behavior[35]. The most responsive type of VaR one can define while abiding by the rules of the current challenging regulatory environment, would be one in which, the weights of your stressed scenarios be scaled according to an EWMA model, which decay factor would be tailored such that on average your returns point to a minimum of 6 months [49]. This constraint comes from the requirement that any VaR model should have at least 1 year of historical data [49]. However, this methodology on top of being unable to reconcile VaR Responsiveness to VaR stability suffers from an even bigger issue which is that it needs to endure a big move in order to adjust (a big loss which is blind with respect to the model guidelines and needs a dangerous breach in its Risk model before adjusting to its model requirements). The reader can perhaps already guess that if a Responsive VaR is lagging, an Anticipative VaR must be leading. Indeed, an Anticipative VaR is essential a conditional probability model which can be used on any VaR engine where the underlying risk process is Mean Reverting (eg: implied Vol, Rates etc ...). For instance, equations (32b) and (32a) in an Anticipative VaR are changed, in continuous time, into equations (35a) and (35b) which are derived from the celebrated Ornstein-Uhlenbeck [25]. It is perhaps interesting to note that in equation (35a), the X_t in from of the stochastic part can be replaced by $\sqrt{X_t}$ and we get the CIR model [9].

C. Responsive Vs. Stable VaR

there is a plethora of technical documents from practitioners [65], [56] available on the web which attempt to expose the conflicting properties of Responsive and Stable VaR. Youngman's [65] is a simple enough and stereotypical example of how the duality between between Responsive and Stable VaR are understood and used by practitioners. Indeed figure 22 plots 3 graphs for the 99% VaR of BBB corporate bonds (so a linear product) using 3 different lookback periods. The VaR model in this situation is in terms of complexity the introductory model used in the industry in which we assume that the underliers follows a rolling log-normal distribution in which the rolling windows are, in this example 1 (in green), 3 (in red) and 5 (in blue) years. What we can see is that the green graph which is the one with the shortest rolling window happens to be the most responsive to market events whereas the blue graph (the biggest rolling window) happen to be the most stable. The rational is that for VaR stability aficionados the green graphs fluctuates too much and can create liquidity congestion in the case where the market would get used to low VaR market environment. For VaR responsiveness aficionados, the blue line is too conservative in low vol environment and not reactive enough in situations of increased volatility. This interesting market observation is the second problem we will attempt at solving by introducing to the concept of Responsible VaR which is a portmanteau neologism in finance, designed to signify a hybrid method between Responsive and Stable VaR.

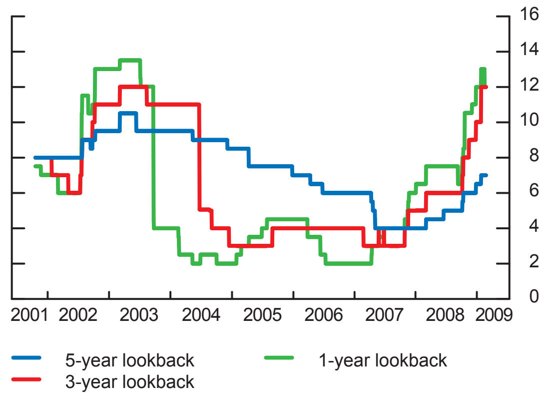

Fig. 22. Youngman's [65] 99% VaR BBB corporate bonds' example

VI. Proposed Enhancements

In this section we try to address the dual limitations associated to model distribution selection[36] and the model reactivity[37] with one stone. In doing so we first examine the problem using the traditional SDE approach and finally propose a Machine Learning equivalent.

A. Anticipative VaR

1) The Stochastic Differential Approach: A better bumping methodology would be to apply a conditional scenario cleanser in which the stressed scenario distributions of our risk factors (eg: implied vol, short term interest rates) going up or down are as much influenced by historical data than by a metaphoric elastic which would pull that historical distribution back to its historical mean and which would prevent the risk factor going too high or going too far below 0. Couple of models that would capture this idea are the OU process [25] or alternatively the Cox Ingersoll Ross model [9][38]. Equations (35b) and (35a) are the mirrors of equations(32a) and (32b) in an Anticipative VaR settings. Equation (35b) is essential the OU process [25], and equation (35a) a modified CIR model[9]. Equation (35c) in a mean reverting SDE bounded by $[-1, +1]$ [12].

$$dX_t = \theta(\mu - X_t)dt + X_t\sigma_t dW_t \tag{35a}$$
$$dX_t = \theta(\mu - X_t)dt + \sigma_t dW_t \tag{35b}$$
$$dX_t = \theta(\mu - X_t)dt + \sigma_t(1 - X_t^2)dW_t \tag{35c}$$

Equation (36) represents the cointelation model [13], which is the most complex of the mentioned mean reverting SDE is the core inspiration of the Anticipative VaR model as it

[35]the reason we have added the enigmatic $0dt$ is more a tactical pedagogic strategy to break some of the market practice misconception that we will introduce in sectionVI-A.3.

[36]selecting normal vs log-normal assumptions

[37]selecting an enhancement that would be anticipative rather than responsiveness.

[38]this latter option would enforce positive IR while still capturing the essence of mean reversion

exposes like we can see in figure 23 the dangers of working with correlation when dealing with mean reverting SDEs [10].

$$\frac{dS_t}{S_t} = \sigma dW_t^1 \tag{36a}$$
$$dS_{g,t} = \theta(S_t - S_{g,t})dt + \sigma S_{g,t} dW_t^2 \tag{36b}$$
$$d < W_t^1, W_t^2 >= \rho dt \tag{36c}$$

In the generalized bumping methodology in which we

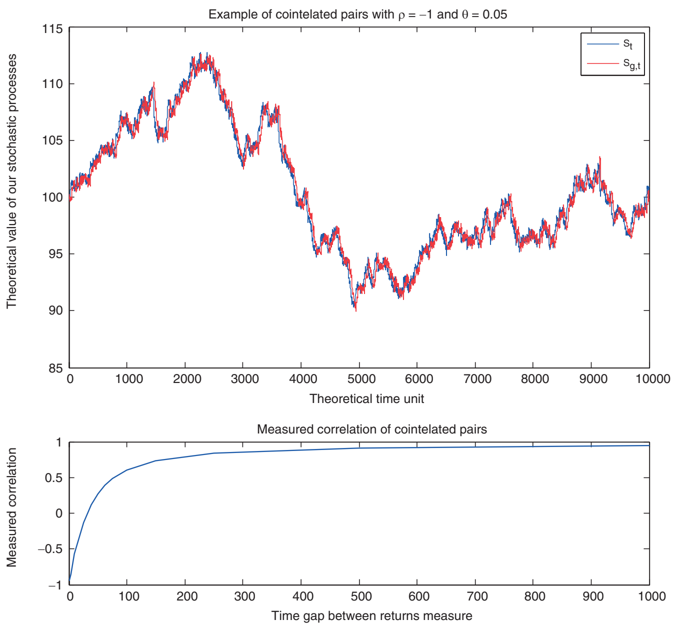

Fig. 23. Example of Cointelation model with a $\rho = -1$ and it's resulting mirror measured correlation [10]

disregard the secondary parameters[39], we assume that any risk factor X_t follows the modified cointelation model [10], [13], [12] given by equation (37). This stochastic process essential can model:

- Proportional bump (log-normal diffusion). Simply enforce $\theta = 0$, $\alpha = 1$, $\beta = 0$ and we get equation (32b).
- Absolute bumps (normal diffusion). Simply enforce $\theta = 0$, $\alpha = 0$, $\beta = 0$ and we get equation (32a),
- Mean reverting bumps where we enforce positivity (like in the case of the CIR [9] diffusion),
- Mean reverting bumps where we do not enforce positivity (like in the case of the OU [25] diffusion),
- Mean reverting bumps bounded in $[-1, 1]$. For example the dynamics of the ρ parameter in the SVI/gSVI/IVP [24], [12], [11] implied volatility parametrization.

$$dX_t = \theta_{t,\tau}(\mu_{t,\tau} - X_t)dt + \sigma X_t^\alpha (1 - X_t^2)^\beta dW_t \tag{37}$$

We will call the following parameters the primary ones:

- θ_t, the rolling speed of mean reversion,
- μ_t, the long term rolling mean,

[39]explained next.

- α the positivity flag enforcer,
- β, the $[-1, +1]$ boundary flag enforcer.
- and $\{\bigcup dW_i\}_{i=t-\tau}^t$, the set of historical deviation of your assumed model's distribution (eg: all the historical absolute bumps in the context of a normal diffusion).

We will call the following parameters the secondary ones:

- τ, the rolling window length (eg: 3 years), meaning that all of the primary parameters will now be function of this rolling window,
- κ_θ, the speed of mean reversion dampener,
- κ_W, the variance enhancer.

Incorporating the secondary parameters into the equation (37) we get the final generalized bumping in the form of equation (38).

$$dX_t = \frac{\theta_t}{\kappa_\theta}(\mu_t - X_t)dt + \kappa_W \sigma X_t^\alpha (1 - X_t^2)^\beta dW_t^\perp \tag{38}$$

with μ_t function of a constant drift μ and stochastic part dW_t with $< dW_t, dW_t^\perp >= \rho dt$. There are multiple benefits in choosing such generalized bumping formula:

- It is versatile: it models all the known risk models on top of new ones,
- It is deployable and robust: once the calibration has been performed the same code works for every risk factor,
- It is leading: it allows for anticipation in the regime change as opposed to waiting passively for responding to a regime change,
- It is more realistic: when Vols (or interest rates) are high applying relative shifts overestimates the moves on the upside but underestimate the moves on the downside,
- It deceases arbitrages scenarios: since the diffusion of equation (37) is more realistic with respect to market observable phenomenon which are in turn function of arbitrage conditions that can occur at the portfolio level (implied vol mean reverts), the number of arbitrage opportunities in the stressed scenario generations decreases drastically when the generalized methodology is used as opposed to relative shifts especially when it comes to skew like strategies (eg: butterfly, call spreads etc ...).

2) The Machine Learning Approach: In section VI-A.1 we have seen that the selection of the parameters, and it will be detailed in section VI-B, their calibration and the culture associated to the this way of doing Risk Management presents challenges on multi level which benefits to complexity ratio is often such that these methodologies are rejected by practitioners of average to low quantitative knowledge.

Lemma 1a Let $R = \{x_1, \ldots, x_n\}$ be a set of empirical random variables taken from equation 37 with cumulative distribution function $F(x)$ and density $f(x)$. Let's call $O = \{x_{(1)}, \ldots, x_{(n)}\}$ the ordered set of R such that $x_{(1)} < x_{(2)} <$

$\ldots < x_{(n)}$ and $O_p^i = \{x_{(\lceil n((i-1)+1)/p \rceil)}, \ldots, x_{(\lfloor n(i)/p \rfloor)}\}$. Then the empirical distribution function for an SDE of the form of equation (37) can be approximated by a union of band-wise Bernoulli process given by:

$$\hat{F}_n(x_i|\mathcal{F}_t) = \frac{1}{n}\sum_{j=1}^{p}\sum_{i=\eta}^{\zeta}\mathbf{1}_{x_i \in O_p^j} \tag{39}$$

with $\eta = \lceil n((i-1)+1)/p \rceil$ and $\zeta = \lfloor n(i)/p \rfloor$.

Remark In the case $p = 3$, using a a Gaussian Mixture such that $\hat{F}_n(x_i|\mathcal{F}_t) = \mathcal{N}(-3,1)\mathbf{1}_{x_t \in O_3^1} + \mathcal{N}(0,1)\mathbf{1}_{x_t \in O_3^2} + \mathcal{N}(3,1)\mathbf{1}_{x_t \in O_3^3}$, we get the approximate stratification of figure 26. The stratification in our case being made so that the carnality in each O_p^j region remains approximately the same, as opposed to being the result of a geometrical separation function of $x_{(1)}$ and $x_{(n)}$. Figure 24 illustrates this remark.

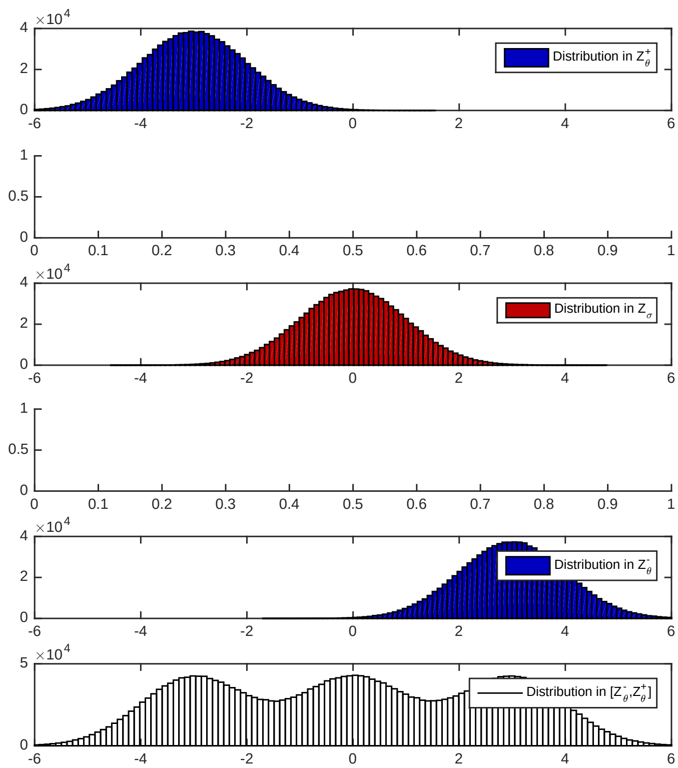

Fig. 24. Gaussian distribution in 3 different size homogeneous zones mimicking approximately figure 26.

Lemma 1b The distribution given by equation 39 converges towards a p-Gaussian Mixture.

Proof: $\mathbf{1}_{x_i \in O_p^j}$ is a Bernoulli random variable with parameter p, and since the sum of Bernoulli random variable is also Bernoulli, $\hat{F}_n(x_i|\mathcal{F}_t) = \frac{1}{n}\sum_{j=1}^{p}\sum_{i=\eta}^{\zeta}\mathbf{1}_{x_i \in O_p^j}$ is Bernoulli distributed. We can also see that in equation (37) $\lim_{n\to\infty, p\to\infty}(\mu_{t,\tau} - X_t) = \lambda_{t,\tau}$ and therefore $dX_t - \lambda_{t,\tau} = \sigma X_t^\alpha(1 - X_t^2)^\beta dW_t$ becomes a locale martingale. Using Glivenko-Cantelli theorem [61], [14], $\|F_n - F\|_\infty = \sup_{x\in\mathbb{R}} |F_n(x) - F(x)| \xrightarrow{a.s.} 0$. Distribution from equation (37) can therefore be approximated by $\cup_{i=1}^{p}\mathcal{N}(\lambda_i, \sigma_i)$. ■

Remark We can see how increasing p can lead to a smoothing probability distribution function by looking at the difference between figure 24 and figure 25.

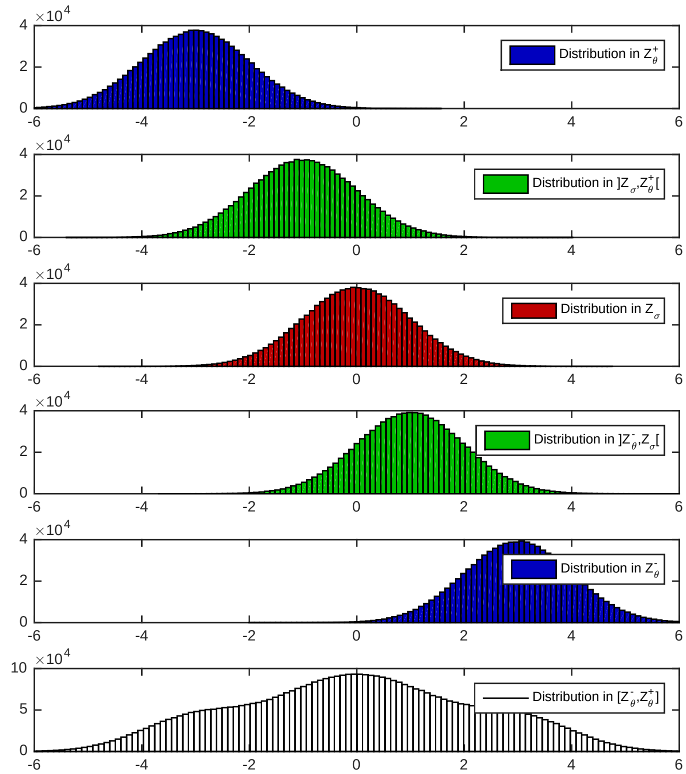

Fig. 25. Gaussian distribution in 5 different zones

Theorem 1 Let $(\Omega, (\mathcal{F})_{(t\geqslant 0)}, \mathbb{Q})$ be our probability space, with $(\mathcal{F})_{(t\geqslant 0)}$ generated by the $T+1$ dimensional Brownian motion and $\mathbb{Q}$ is the risk neutral probability measure. The probability distribution $f(x|\mathcal{F}_t)$ induced by the Stochastic Differential Equation $dX_t = \theta_{t,\tau}(\mu_{t,\tau} - X_t)dt + \sigma X_t^\alpha(1 - X_t^2)^\beta dW_t$ converges almost surely towards a p-Gaussian mixture as n and p converge towards ∞.

Proof: The proof can be split in 2 steps using Lemma 1a and Lemma 1b. ■

3) Margining under the classic and the Anticipative methodology: Market participant call full revaluation methodology the following. If we call N the total number of risk factors relevant to a portfolio P and R_i each of the relevant risk factors (eg: exchange rate, interest rate, skew, vol of vol, ATM vol etc ...) of this portfolio, then we will define f, the function that takes all the bumped historical scenarios set $\cup_{i=1}^{N}\cup_{\tau=1}^{T} R_{i,t} + \Delta_{t-\tau}^{R_{i,\tau}}$ as input into f and would revalue the portfolio[40] and would return for example the worst scenario or the average of the worst w scenarios in the case of the expected shortfall. The best

[40]Note that $\tau \in [0, T]$ with T being the length of the available relevant data (eg: 10 years)

way to describe the margining methodology is to see how it works with the currently understood models in the market changing a bit the format the models are presented and use this new format to expose the way the margining in done under Anticipative VaR. Table II is a good visual aid in understanding how the scenarios are generated taking into account the "co-movement" of spot (assuming a log-normal diffusion[41]), with, for the lack of a better tools "normal" diffusion[42] chosen for the relevant vol point risk factors. The resulting returns at the portfolio level would be given by $\text{P\&L}_{t,\tau} = f\Big(S_t(1+\Delta S_\tau), \Sigma_{E,K,t} + \Delta\Sigma_{E,K,\tau}, \cdots\Big)$. Note that in the function f, we have incorporated the symbol "$\cdots$" to signify that the same methodology is used for all the relevant vol points for all the relevant tenors. The stressed

Date τ	$dS_t = S_t d\hat{W}^a_t$	$d\Sigma_{E,K,\tau} = d\hat{W}^b_t$	$\cdots$	$\text{P\&L}_{t,\tau}$
$t-1$	$\frac{\Delta S_{t-1}}{S_{t-1}}$	$\Delta\Sigma_{E,K,t-1}$	$\cdots$	-1.7%
$t-2$	$\frac{\Delta S_{t-2}}{S_{t-2}}$	$\Delta\Sigma_{E,K,t-2}$	$\cdots$	+ 0.7%
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$t-750$	$\frac{\Delta S_{t-750}}{S_{t-750}}$	$\Delta\Sigma_{E,K,t-750}$	$\cdots$	-1.4%

TABLE II
CLASSIC RISK MODEL FULL REVALUATION EXAMPLE

scenarios may be cleansed prior the revaluation using de-arbitraging methodologies [12] but, in spirit table II would be an intuitive enough representation to expose the way Anticipative VaR is used as we will see next. Table III is the corresponding visual aid to table II under the assumption of mean reversion capability for ATM vol using the generalized bumping methodology from equation (38). Usually the part which is most misunderstood in this methodology is how $\{\bigcup dW_i\}^t_{i=t-\tau}$ behaves in either of the models. The key in understanding this part is to realize that $\{\bigcup dW_i\}^t_{i=t-\tau}$ represent the "deviations/error" from an assumed model which errors maybe correlated to the other risk factors which may or may not assume the same Stochastic Differential Equation as when it comes to their diffusion. For instance you could mix a model in which you try to capture the risk of a call and try to define the corresponding risk factor. You assume that the underlier follows a log normal distribution (you enforce $\theta = 0$ but $\alpha = 1$), you assume that the ATM vol is mean reverting and positive (you enforce $\theta \neq 0$ but set $\alpha = 1$), you assume interest rates are mean reverting but could go be negative (you enforce $\theta \neq 0$ but $\alpha = 0$). The correlation at the infinitesimal level of these errors from these models $\{\bigcup dW_i\}^t_{i=t-\tau}$ will be concerved if the time stamp is conserved. The resulting returns at the portfolio level would be given by $\text{P\&L}_{t,\tau} = f\Big(S_t(1+\Delta S_\tau), \Sigma_{E,K,t-1} + \theta(\hat{\mu} - \Sigma_{E,K,t}) + \Delta S_\tau, \cdots\Big)$. We invite the reader to take a look at the anecdotal column labeled $\text{P\&L}_{t,\tau}$ in both table II and III. The numbers filled in table II one have been chosen randomly but the ones from table III are chosen in terms of what one would expect in terms of approximate difference had the vol been different from its historical mean. So the P&L is shifted because the risk is now asymmetric. If that last sentence sounded convoluted, the reader should go back to figure 27 and think about what the VaR would look like depending on whether one is long or short a straddle under Anticipative VaR and depending on which zone we are in the figure 27.

[41]"we use proportional bumps"
[42]"we use absolute bumps"

B. Calibration

1) The SDE approach: As we will see in this section the calibration of the parameters of our bumping model from equation (38) comes in 3 steps:

- The choice of our assumption. This is where we decide whether the risk factor ought to be mean reverting or not, enforce positivity or not, enforce being in $[-1, +1]$ or not,
- Sequential estimation of the primary parameters,
- Intelligent fine tuning of the backtest thanks to the secondary parameters.
- Adjustment of the λ parameter in the Responsible VaR equation (43a) to fit the manager's risk appetite.

We recommend the following flags to be enforced according to the risk factor:

- For any, assumed, non mean reverting random process which has to stay positive (eg: spot), simply enforce $\theta = 0$, $\alpha = 1$, $\beta = 0$ and we get equation (32b).
- For any volatility related mean reverting bumps where we enforce positivity, simply enforce $\theta \neq 0$, $\alpha = 1$, $\beta = 0$,
- For interest rate assume, mean reverting bumps where we do not enforce positivity (like in the case of the OU [25] diffusion), simply enforce $\theta \neq 0$, $\alpha = 0$, $\beta = 0$.
- For assigning a diffusion on correlation itself or the ρ parameter in the SVI/gSVI/IVP [24], [12], [11], that is mean reverting bumps bounded in $[-1, 1]$, simply enforce $\theta \neq 0$, $\alpha = 0$, $\beta = 1$,
- For assigning a diffusion on the minimum of the moneyness axis in an implied Vol, we recommend a mean reverting process which can go negative and positive, simply enforce $\theta = 0$, $\alpha = 0$, $\beta = 0$ and we get equation (32a)[43].

2) Calibration of the primary parameters: The calibration of the model in the situation in which we do not enforce mean reversion is trivial. If this is not clear we invite the reader to go back to equation (37) and think a bit more about what each parameter does. Once we are in the context of mean reversion, we need to calibrate in sequence $\hat{\mu}$, $\hat{\theta}$

[43]Note that no suggest of product assuming absolute bumps (normal diffusion) is disregarded. The reason is because this is the diffusion that is most likely to be made obsolete as a result of this new proposed methodology. However if you are still attached to this methodology (for STIR's for example)

Date τ	$dS_t = S_t d\hat{W}_t^a$	$d\Sigma_{E,K,\tau} = \hat{\theta}(\hat{\mu} - \Sigma_{E,K,t})dt + \Sigma_{E,K,\tau} d\hat{W}_t^b$	$\cdots$	P&L$_{t,\tau}$
t-1	$\frac{\Delta S_{t-1}}{S_{t-1}}$	$\Delta\Sigma_{E,K,t-1}$	$\cdots$	-1.9%
t-2	$\frac{\Delta S_{t-2}}{S_{t-2}}$	$\Delta\Sigma_{E,K,t-2}$	$\cdots$	+0.8%
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$t-750$	$\frac{\Delta S_{t-750}}{S_{t-750}}$	$\Delta\Sigma_{E,K,t-750}$	$\cdots$	-1.6%

TABLE III
EXAMPLE OF FULL REVALUATION TABLE UNDER ANTICIPATIVE RISK ENGINES

and finally the errors from the model $\{\bigcup d\hat{W}_i\}_{i=t-\tau}^{t}$. The calibration of μ is straight forward, done by equation (40).

$$\hat{\mu}_{t,\tau} = \mathbb{E}[X_{t,\tau}|\mathcal{F}_{t,\tau}] = \frac{1}{N}\sum_{i=t-\tau}^{t} X_i \tag{40}$$

where $N = \text{card}\{\tau, \tau+1, \cdots, t-1, t\}$[44]. The calibration of θ happens to be tricky, indeed if one rearranges equation (37), we get $\theta_{t,\tau} = \frac{dX_t - \sigma X_t^\alpha (1-X_t^2)^\beta dW_t}{(\mu_{t,\tau} - X_t)dt}$. However if one was to take all the available samples for θ and perform $\mathbb{E}[\theta_{t,\tau}|\mathcal{F}_{t,\tau}]$ like in equations (40), the estimation would quickly be dominated by instances where $\mu_{t,\tau}$ is very close to X_t and where θ "explodes" as a consequence and creates a random bias in the estimation of $\theta_{t,\tau}$. The variance reduction idea comes from noticing the explosion effect described and deliberately choosing to neglect zones in which the explosion is highly likely. In the original paper [13] in which this technique was first introduced, the relevant zones were $B_+ = |\frac{\max(X_i, i\in[t-\tau,t])}{2}|$, and $B_- = |\frac{\inf(X_i, i\in[t-\tau,t])}{2}|$[45]. These zones can be visualized in figure 26. In this figure, Z_θ represents situation in which sampling θ is very likely to be of high quality. When we apply the idea of the variance

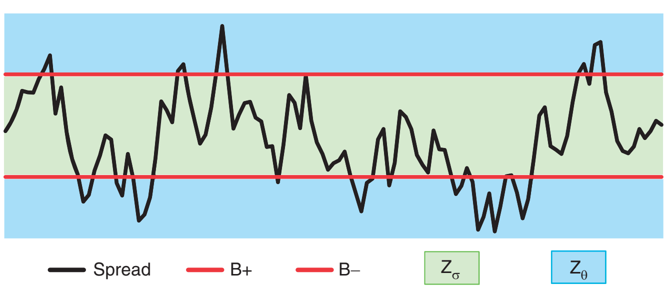

Fig. 26. Visual representation for the sampling zones for θ [13].

reduction technique we get equation (41) as for a proper estimation of θ.

$$\hat{\theta}_{t,\tau} = \frac{1}{n}\sum_{i\in Z_\theta} \frac{\Delta X_i - \sigma X_i^\alpha (1-X_i^2)^\beta \Delta W_i}{(\hat{\mu}_{t,\tau} - X_i)} 1_{X_i \in Z_\theta} \tag{41}$$

[44] which is essentially the number of sample used to estimate μ.

[45] We note that the estimation of θ is noised when $Z_\sigma = B_+ > |X_i, i \in [t-\tau, t]| > B_-$. The reverse is true when $Z_\theta = |X_i, i \in [t-\tau, t]| > B_+ \bigcup |X_i, i \in [t-\tau, t]| < B_-$, sampling θ is a good idea.We will therefore sample θ in Z_θ.

where n is the cardinality of the set of all instances in which we sampled in the Z_θ zone. In practice doing this average over the 40th and 60th percentiles provides enough data and filters out enough outliers to make a quality estimator. Once μ and θ have been calibrated, we need to calibrate the deviation from the model, this is done by isolating the dW_t's in equation (37). The estimation for the errors becomes equation (42).

$$\Delta\hat{W}_i = \frac{\Delta X_i - \hat{\theta}_{t,\tau}(\hat{\mu}_{t,\tau} - X_i)}{X_i^\alpha (1-X_i^2)^\beta} \tag{42}$$

3) Calibration of the secondary parameters in the Anticipative VaR context : As we have seen there are 3 secondary parameters. The calibration of these secondary parameters are, as we will see, very much qualitative in approach and geared towards practitioners rather than pure probabilists.

- The first secondary parameter, τ, represents the "rolling window of interest". This window can be chosen so as to either satisfy the regulatory constraints on model selection or/and based on how well your backtest performs with respect to your risk appetite.
- The second secondary parameter as it can be seen from equation (38) and (37) happens to be κ_θ which can be understood as the "elastic aging factor". The bigger κ_θ happens to be, the more the speed of mean reversion calibrated by the assumed model would get weaker and the more we converge towards a drift-less model. This is particularly useful in situation in which the risk manager believes that the long term mean fluctuates faster than assumed by the model and with respect to the rolling window.
- The third and last secondary parameter, κ_W, can be thought of the "returns beef-upper". It was created so as to more or less replicate the current, in my opinion not ideal, market standard which re-levels the returns so that the historical VaR matches the desired risk appetite.

We can observe how the historical deviations of equation (37) behaves as a function of where the stochastic process stands with respect to its long term mean. Indeed, in a situation where one assign $\beta = 0$, $\alpha = 1$ for the ATM implied vol for the 2 years expiry we get figure 27. What this figure attempts at exposing is the asymmetric/skewed behavior for the distribution of the generated stressed scenarios as a function of where the risk factor stands with respect to its long term mean[46]. Indeed, we can see in zones 3 and 4, that the risk factor is biased towards the upside whereas in zones 1 and 5 in which the risk factor is above its historical rolling mean, the simulated distribution is biased on the downside. Note here that in a situation in which you are below your historical mean, you can still have simulations that take you below your current value (the reverse is true for the symmetric situation in which you are above your historical mean). A second interesting point to note is that the more the risk factor is significantly above its historical mean, the more skewed is the resulting distribution (for example the distribution of zone 5 is more skewed towards the downside compared to the one of zone 2 because the deterministic side of the stochastic differential equation pressures the simulations more on the downside, even though it still allows moves on the upside). From these few zones, the reader can now understand the terminology chosen ("Anticipative") for this new risk concept. For instance, depending on where one stands with respect to the equilibrium point (eg: rolling mean) a historically model calibrated "view" is incorporated on top of the traditional deviations from the model to adjust a bit the distribution. This can be opposed to the concept of Responsive VaR in which one always takes a symmetric approach and rescales the risk factors returns once a big market move has already occurred (so the VaR is Responsive as opposed to Anticipative).

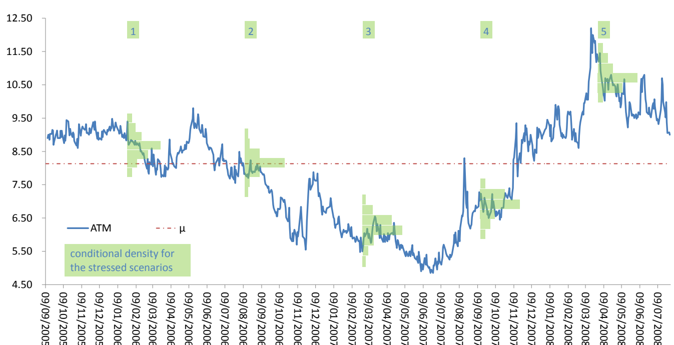

Fig. 27. Example of distributions for the stressed scenarios under the Anticipative VaR and statistical arbitrage hypothesis

4) The Machine Learning Approach:

A straightforward learning algorithm: The idea behind the calibration of the Band-Wise Gaussian Mixture is similar, though not exactly the same as to the variance reduction technique we saw earlier, more specifically of figure 26. Namely, depending on the selected zone, the resulting approximated distribution of the samples differ. The calibration algorithm will then consist of creating as many zones as possible trying to converge to the results from the theorem 1 page 21. Algorithm 3 is a rough pseudo-code for the calibration process.

[46]We have assumed for display purposes that the mean will not be rolling in this graph but fixed.

Algorithm 3 BAND-WISE GAUSSIAN MIXTURE(X, p)

Require: array $X_{1:n}$ and number of bands p
Ensure: $\Omega^{(1:p)}$, $[B^+_{(1:p)}, B^-_{(1:p)}]$ are returned.

Sorting state:
$X_{(1:n)} \leftarrow$ QuickSort$(X_{1:n})$
$[B^+_{(1:p)}, B^-_{(1:p)}] \leftarrow$ FindPercentileBands$(X_{(1:n)},\, p)$
$\Omega^{(1:\lceil n/p \rceil)} \leftarrow []$

Allocation state:
for $j = 1$ to p **do**
  for $i = 1$ to n **do**
    if $B^-_{(j)} \leqslant X_{(i)} < B^+_{(j)}$ **then**
      Amend$(\Omega^{(j)}, X_{(i)})$
    end if
  end for
end for

Checking Approximation state:
$\hat{\mu}_{1:p} \leftarrow$ mean$(\Omega^{(1:p)})$
$\hat{\sigma}_{1:p} \leftarrow$ stdev$(\Omega^{(1:p)})$
Print$(\cup_{i=1}^{p} \mathcal{N}(\hat{\mu}_i, \hat{\sigma}_i))$

Return state:
$\Omega^{(1:p)}$, $[B^+_{(1:p)}, B^-_{(1:p)}]$

Remark Note that in algorithm 3, we have used a QuickSort which can be substituted by other sorting algorithm. We invite the motivated readers to investigate on their own this idiosyncratic issue. Also note that this algorithm has neither been optimized nor checked for data quality (eg: the combination of n and p should be such that each band has enough data (eg: minimum 30) for the statistical estimators to be significant.

5) A model that can dynamically accommodate regime changes: One important point to note about the additional benefits of the Machine Learning approach over the SDE approach[47] is to take a look at the example associated to the bizarre world of interest rates. Indeed up to 2014, it was assumed that interest rates could never become negative[48]. It is reasonable to assume there a risk manager attracted to prowess of equation 37 would have chosen a $\beta = 0$ and an $\alpha = 1$, the latter enforcing positivity for the simulated scenarios of our risk factor. This very reasonable assumption would have crashed the whole risk engine. The Machine Learning approach would have however been able to continue its dynamical learning scenario without any problem.

C. Responsible VaR

1) Theory: In this section we introduce the concept of Responsible VaR which is, the second new Risk term after

[47]beyond the obvious benefits associated to achieving the same results though through a simpler channel and also bypassing convoluted SDE calibration issues in the process.

[48]why would you pay to put your money in the bank? That would essentially be the physical question one may ask oneself.

”Anticipative”, that we aim at introducing in this paper. Responsible is a portmanteau term that aims at signifying a VaR model that is Responsive on the upside and Stable on the downside. If we assume ν_t represents the VaR at the α level at time t, then mathematically the concept of Responsible VaR is jointly defined by the stochastic processes $\tilde{\nu}_t^+$ and $\tilde{\nu}_t^-$ summarized in equation (43a).

$$\alpha = \int_{-\infty}^{\nu_t^+} p_t(x)dx \tag{43a}$$

$$\tilde{\nu}_0^+ = \nu_0^+ \tag{43b}$$

$$\tilde{\nu}_t^+ = \max\left(\nu_t^+, \lambda\tilde{\nu}_{t-1}^+ + (1-\lambda)\nu_t^+\right) \tag{43c}$$

$$1-\alpha = \int_{\nu_t^-}^{+\infty} p_t(x)dx \tag{43d}$$

$$\tilde{\nu}_0^- = \nu_0^- \tag{43e}$$

$$\tilde{\nu}_t^- = \min\left(\nu_t^-, \lambda\tilde{\nu}_{t-1} + (1-\lambda)\nu_t^-\right) \tag{43f}$$

Figures 28, 29, 30 and 31 expose how λ, which can be intuitively understood as the stability coefficient impacts how the Responsible VaR level for a portfolio, changes as a function of λ. This methodology can in fact be used independently of whether we are in a Responsive or an Anticipative VaR context.

2) Practice: As the reader can see the concept of Responsible VaR can be summarized by a system of two controlled Snell envelopes of the VaR level at their respective quantile level, which control is fined tuned by the λ parameter to match essentially the practitioner risk appetite. In a practical point of view one must record the VaR of a specific portfolio[49] in time and record the instantaneous Anticipative or Responsive VaR and adjust it based on equation (43a). That particular last point may be deterrent in direct use if the IT constraints are not flexible or too slow. Figures 28, 29, 30 and 31 illustrate how Responsible VaR changes as a function of λ. The model can be thought as, the more λ is close to 0, the more the model is purely Responsive both on the up and downside, and the more λ is close to , the more the model is stable on the downside while being equally responsive on the upside.

3) Calibrating the responsible VaR parameters: Finally, the λ parameter can be chosen so as to fit the risk appetite of the risk manager and the financial institution he/she represents. Figures 28, 29, 30 and 31 provides an illustration of how the risk measure changes as a function of different values of λ.

D. Simulations

In this section we have chosen to only show the backtest using the SDE approach, though the backtest using the Machine Learning approach was similar but more straightforward. Also showing the performances of both approaches would have increased unnecessarily the number of graphs without really exposing performance issues which are more associated calibration issue and parameter model limitation. We have chosen in this section to perform a backtest under the margining context of the generalized bumping methodology of section VI-A.1. We have chosen, a straddle of 2 year expiry assuming spot follows a log normal diffusion ($\theta = 0$, $\alpha = 1$, $\beta = 0$) with implied vol points in the 10, 25, 50, 75 and 90 delta for the relevant expiry are bumped following a mean reverting assumption in which positivity is enforced ($\theta \neq 0$, $\alpha = 1$, $\beta = 0$). Figures 28, 29, 30 and 31 exposes how the backtest performs as a function of λ. Figure 28 show few interesting points. First point to notice is that, in 2008, the market experienced the kind of turmoil it never experienced up to then. Given that we are in a context of historical VaR, it is not surprising then to have few breaches in 2008. The second point to notice is that when λ from the Responsible VaR formula of equation (43a), the Anticipative VaRs superpose the Anticipative Responsible VaRs of the same quantile. Figure 29 exposes how increasing λ by not a significant amount creates time series which are still more stable on the downside compared to figure 28 but still too responsive on the downside. It can be speculated from figure 29 that the VaR level although more stable on the downside compared to figure 28, might still be overly responsive on the downside. Adjusting the λ in equation (43a) allows to improve a bit stability on the downside and avoid additional breaches in 2009 and 2014 while still allowing the VaR level to provide relief to market participants. Finally figure 31 exposes the relation between equation (43a) and the $\max$ and $\min$ functions for the VaR at the relevant quantile level when $\lambda = 1$.

VII. Conclusion

We first exposed some of the complexity associated to the risk factors and arbitrage constraints associated with the options and the high frequency markets by re-introducing the Implied Volatility Parametrization (IVP) [3], [11] and the High Frequency Trading Ecosystem (HFTE) [41]. The complexity was then contrasted with the current obsolete Risk Methodologies which are based on simplistic SDEs. We first extended the latter SDEs using the Cointelation model [10], [13] in order to partially address some of the complexity introduced by the challenging regulatory environment such as scenario coherence. We then presented a simple Machine Learning clustering methodology which is designed to address and mirror the enhancements of these SDEs in a simpler fashion. We have laid out the benefits of such methodology which we organized in concepts such as versatility, deploy-ability, robustness, leading as opposed to lagging, realistic and partially non arbitrage-able. We finally illustrated our findings by introducing few new risk concepts such as the Anticipative VaR which aims at being a leading as opposed to a lagging (Responsive) risk measure to a market regime change, as well as found a way to reconcile the latter to the concept of Stable VaR to formalize the concepts of Anticipatible and Responsible VaR.

[49] A portfolio of option’s keeping the same weights but decaying by one day would not constitute a constant portfolio as defined in the context of Responsible VaR

REFERENCES

[1] Robert Axelrod. The evolution of cooperation. *Basic Books*, 1984.

[2] Robert Axelrod. The complexity of cooperation: Agent-based models of competition and collaboration. *Princeton University Press*, 1997.

[3] Andrew Kos Babak Mahdavi Damghani. De-arbitraging with a weak smile: Application to skew risk. pages 40–49, 2013.

[4] Richard Bellman. Dynamic programming and a new formalism in the calculus of variations. *Proceedings of the National Academy of Sciences of the United States of America*, 40(4):231–235, 1954.

[5] S. Benaim, P. Friz, and Roger Lee. On the black-scholes implied volatility at extreme strikes. page 2, 2008.

[6] J.A. Bucklew. Large deviation techniques in decision, simulations, and estimation. *Wiley*, 1990.

[7] Erica Chauvet, Joseph E. Paullet, Joseph P. Previte, and Zac Walls. A Lotka-Volterra three-species food chain. *Mathematics Magazine*, 75:243–255, 2002.

[8] John Horton Conway. *On Numbers and Games*. 1976.

[9] J.E. Ingersoll Cox, J.C. and S.A. Ross. A theory of the term structure of interest rates. pages 385–407, 1985.

[10] Babak Mahdavi Damghani. Introduction to the cointelation model and inferred correlation. 2013.

[11] Babak Mahdavi Damghani. Introducing the implied volatility surface parametrisation (IVP):application to the fx market. 2015.

[12] Babak Mahdavi Damghani and Andrew Kos. De-arbitraging with a weak smile. 2013.

[13] Babak Mahdavi Damghani, Daniella Welch, Ciaran O'Malley, and Stephen Knights. The misleading value of measured correlation. 2012.

[14] A. W. Van der Vaart. Asymptotic statistics. page 265, 1998.

[15] A. Doucet, S. Godsill, and C. Andrieu. On sequential monte carlo sampling methods for bayesian filtering. *Statistics and Computing*, 10(9):197–208, 2000.

[16] Bruno Dupire. Pricing with a smile. pages 17–20, 1994.

[17] Bruno Dupire. Pricing and hedging with a smile. pages 103–111, 1997. In M.A.H. Dempster and S.R. Pliska, editors, Mathematics of Derivative Securities, Isaac Newton Institute.

[18] C. S. Elton. Ecology of invasions by animals and plants. *Chapman & Hall, London*, pages 228–233.

[19] C. S. Elton. The ecology of invasions by animals and plants. *Methuen*, 1958.

[20] A.D. Fokker. Die mittlere energie rotierender elektrischer dipole im strahlungsfeld. pages 810–820, 1914.

[21] J. Gatheral and A. Jacquier. Convergence of Heston to SVI. *Quantitative Finance*, 11:1129–1132, 2011.

[22] Jim Gatheral. The volatility surface.

[23] Jim Gatheral and Antoine Jacquier. Convergence of heston to svi. 11:1129–1132, 2011.

[24] Jim Gatheral and Antoine Jacquier. Arbitrage-free svi volatility surfaces. page 8, 2012.

[25] G.E.Uhlenbeck and L.S.Ornstein. On the theory of brownian motion. 36:823–841, 1930.

[26] N. Gordon, D. Salmond, and A. Smith. Novel approach to nonlinear/non-gaussian bayesian state estimation. *IEE Proceedings-F*, 140(2):107–113, 1993.

[27] Patrick S. Hagan, Deep Kumar, Andrew S. Lesniewski, and Diana E. Woodward. Managing smile risk. *Wilmott*, pages 84–108, 2002.

[28] Andrew G. Haldane and Robert M. May. Systemic risk in banking ecosystems. *Nature*, 469:351–355, 2011.

[29] Rama Cont Hamed Amini and Andrea Minca. Resilience to contagion in financial networks. *Mathematical Finance*, 26:329–365, 2016.

[30] J.E. Handschin and D.Q. Mayne. Monte carlo techniques to estimate conditional expectation in multi-state non-linear filtering. *Int. J. Contr.*, 9(5):547–559, 1969.

[31] S. L. Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies*, 6:327–343, 1993.

[32] G. E. Hutchinson. Homage to santa rosalia, or why are there so many kinds of animals? *Nature*, 93:145–159, 1959.

[33] M. Isard and A. Blake. Condensation conditional density propagation for visual tracking. *International Journal of Computer Vision*, 29(1):5–28, 1998.

[34] Nicole El Karoui. Couverture des risques dans les marches financiers.

[35] S. Mankad; G. Michailidis; A. Kirilenko. Discovering the ecosystem of an electronic financial market with a dynamic machine-learning method. *Algorithmic Finance*, 2:2:151–165, 2013.

[36] G. Kitagawa. Monte carlo filter and smoother for non-gaussian nonlinear state space models. *Journal of Computational and Graphical Statistics*, 5(1):1–25, 1996.

[37] Genshiro Kitagawa. Monte carlo filter and smoother for non-gaussian nonlinear state space models. *Journal of Computational and Graphical Statistics*, 5(1):1–25, 1996.

[38] J.S. Liu and R. Chen. Sequential monte carlo methods for dynamic systems. *Journal of the American Statistical Association*, 93:1032–1044, 1998.

[39] Alfred J. Lotka. Elements of physical biology. *Williams & Wilkins Co.*, 1925.

[40] R. H. MacArthur. Fluctuations of animal populations and a measure of community stability. *Ecology*, 36:533536, 1955.

[41] Babak Mahdavi-Damghani. Introducing the HFTE model: a multi-species predator prey ecosystem for high frequency quantitative financial strategies. 2016.

[42] Babak Mahdavi-Damghani. Deciphering price formation in the high frequency domain: Systems & evolutionary dynamics as keys for construction of the high frequency trading ecosystem. *working paper*, 2017.

[43] A. Marshall. The use of multi-stage sampling schemes in monte carlo computations. *in Symposium on Monte Carlo Methods, M. Meyer Ed. New York: Wiley*, pages 123–140, 1956.

[44] Robert M. May. Simple mathematical models with very complicated dynamics. *Nature*, pages 459–467, 1976.

[45] Kevin Shear McCann. The diversitystability debate. *Nature*, pages 228–233.

[46] Marvin Minsky and Seymour Papert. *Perceptrons: An Introduction to Computational Geometry*. 2nd edition with corrections, first edition 1969 edition, 1972.

[47] Patrick Alfred Pierce Moran. The statistical processes of evolutionary theory. *Oxford: Clarendon Press*, 1962.

[48] Martin Nowak. *Evolutionary dynamics : exploring the equations of life*. 2006.

[49] Basel Committee on Banking Supervision. Basel 3: A global regulatory framework for more resilient banks and banking systems. page 11, 2011.

[50] S. L. Pimm and J. H. Lawton. On feeding on more than one trophic level. *Nature*, 275:542–544, 1978.

[51] Warren McCulloch; Walter Pitts. A logical calculus of ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5:115–133, 1943.

[52] M. Planck. Sitz. ber. preu. akad. 1917.

[53] Chris Rogers and Mike Tehranchi. The implied volatility surface does not move by parallel shifts. 2009.

[54] Frank Rosenblatt. *Principles of neurodynamics: perceptrons and the theory of brain mechanisms*. 1962.

[55] P. J. Schonbucher. A market model for stochastic implied volatility.

[56] Meera Sharma. What is wrong with quantitative standard d.?

[57] Marton Nowak; Karl Sigmund. A strategy of win-stay, lose-shift that outperforms tit-for-tat in the prisoner's dilemma game. *Nature*, 364:56–58, 1993.

[58] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, and Ruslan Salakhutdinov Ilya Sutskever and. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15:1929–1958, 2014.

[59] Nassim Nicholas Taleb. *The Black Swan*. Allen Lane, 2007.

[60] Peter Tankov and Nizar Touzi. Calcul stochastique en finance. 2010.

[61] Howard Tucker. A generalization of the glivenko-cantelli theorem. page 828830, 1959.

[62] Axel Vogt.

[63] Vito Volterra. Variazioni e fluttuazioni del numero dindividui in specie animali conviventi. *Mem. Acad. Lincei Roma*, pages 31–113, 1926.

[64] P. Yodzis. The stability of real ecosystems. *Nature*, 289:674–676, 1981.

[65] Peter Youngman. Procyclicality and value at risk. available on line

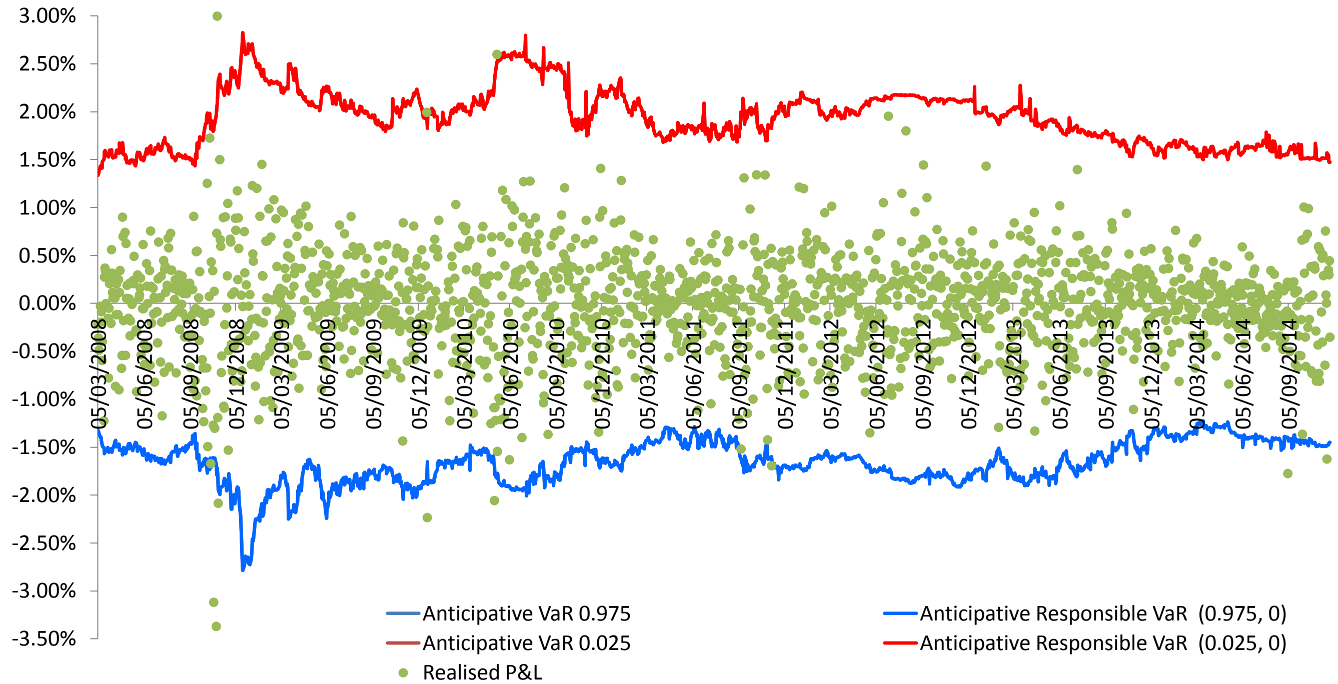

Fig. 28. USD/EUR 2 years expiry straddle strategy backtest under Anticipative Responsible VaR with $\lambda = 0.000$.

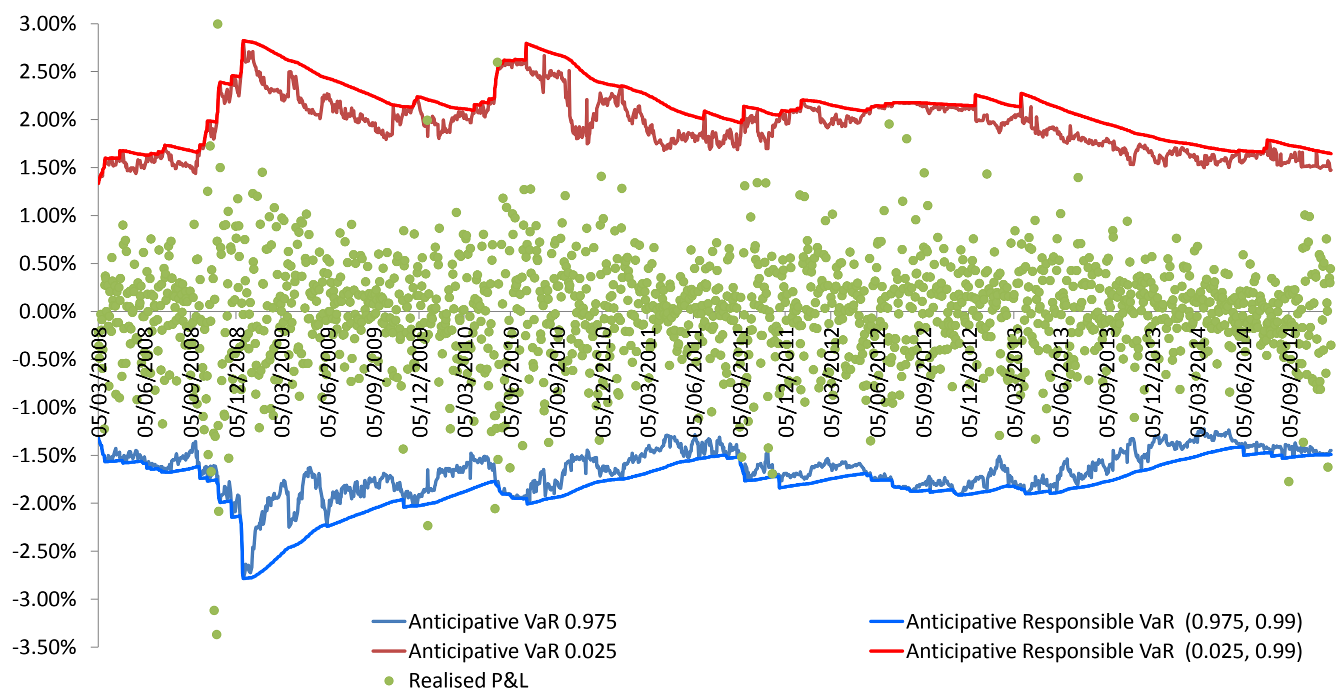

Fig. 29. USD/EUR 2 years expiry straddle strategy backtest under Anticipative Responsible VaR of with $\lambda = 0.990$.

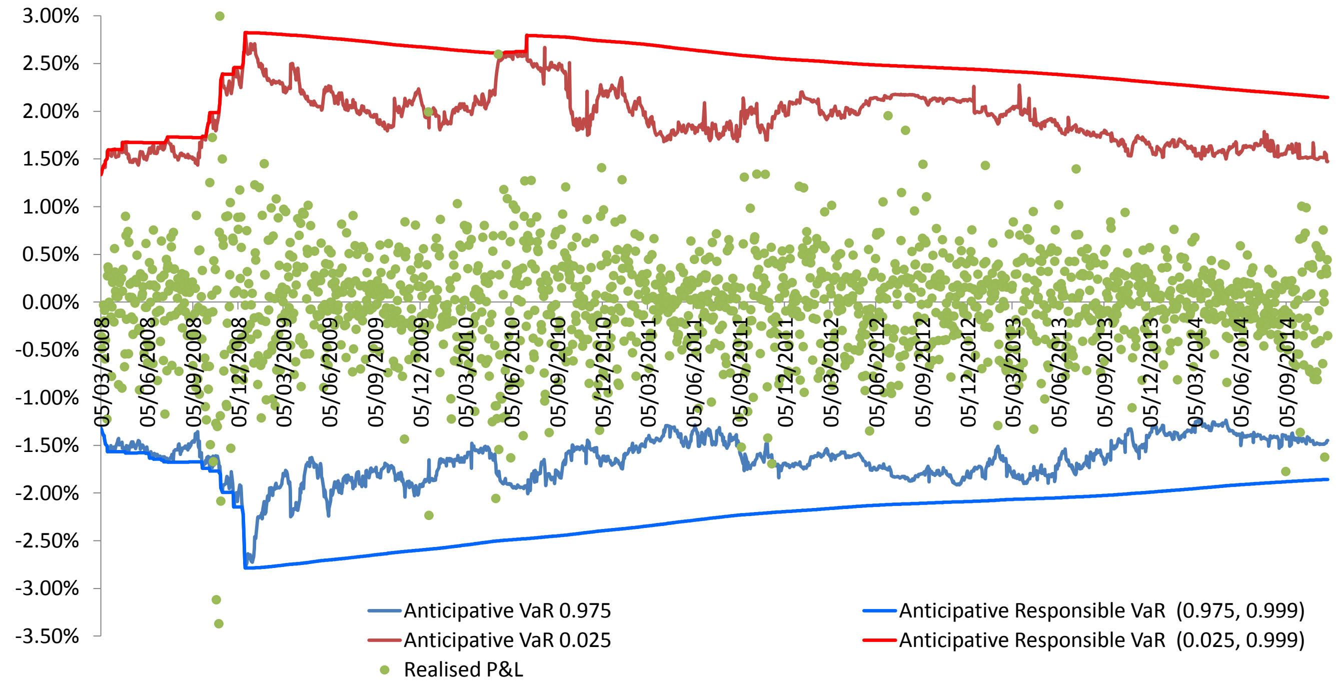

Fig. 30. USD/EUR 2 years expiry straddle strategy backtest under Anticipative Responsible VaR of with $\lambda = 0.999$.

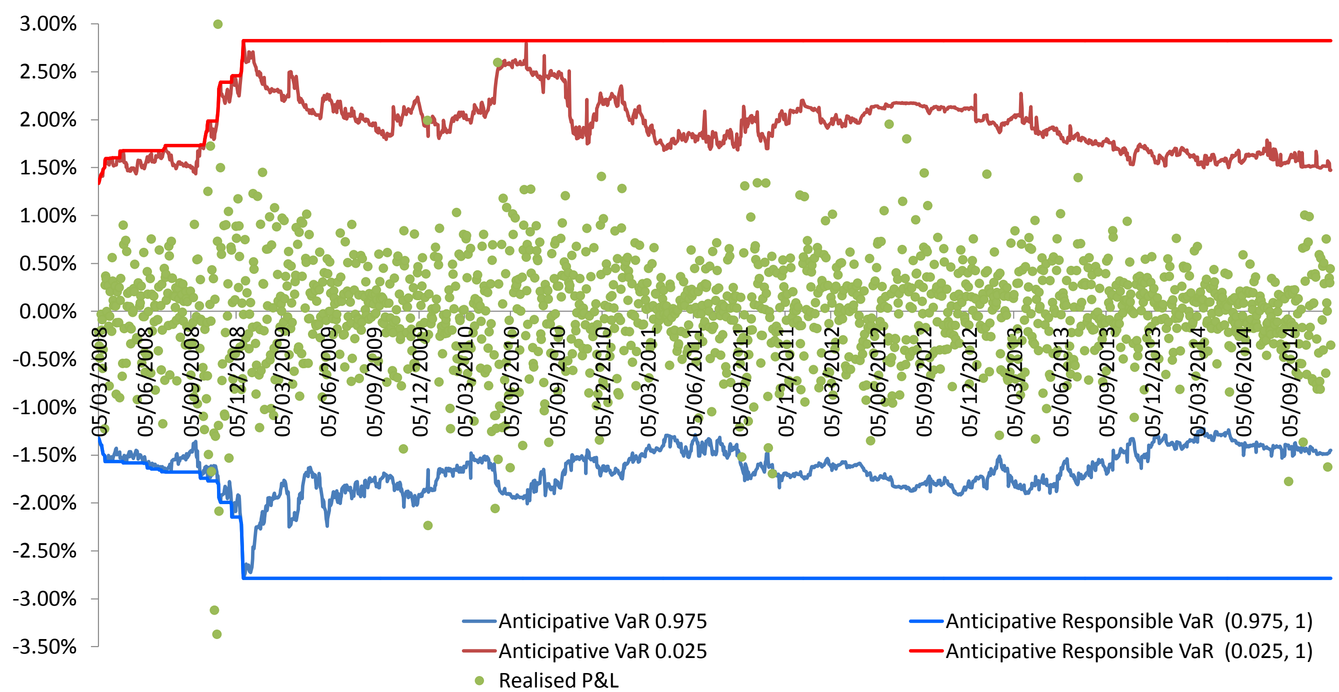

Fig. 31. USD/EUR 2 years expiry straddle strategy backtest under Anticipative Responsible VaR with $\lambda = 1.000$.