Machine Learning Techniques for Deciphering Implied Volatility Surface Data in a Hostile Environment:

Scenario Based Particle Filter, Risk Factor Decomposition & Arbitrage Constraint Sampling

(working paper: version of January 24, 2019)

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

Oxford-Man Institute of Quantitative Finance

***Abstract*— The change subsequent to the sub-prime crisis pushed pressure on decreased financial products complexity, going from exotics to vanilla options but increase in pricing efficiency. We introduce in this paper a more efficient methodology for vanilla option pricing using a scenario based particle filter in a hostile data environment. In doing so we capitalize on the risk factor decomposition of the the Implied Volatility surface Parametrization (IVP) recently introduced [70] in order to define our likelihood function and therefore our sampling methodology taking into consideration arbitrage constraints.**

Keywords: Implied Volatility Parametrization (IVP), Volatility surface, SVI, gSVI, Arbitrage Free Volatility Surface, Fundamental Review of the Trading Book (FRTB).

I. Scope

A. Market context

The financial crisis of 2009 and the resulting social uproar led to an ethical malaise [51], [69], [74] that grew in the scientific community as well as within practitioners which changed the market in many ways including the following:

- convoluted financial products with high volatility or/and low liquidity and/or without any societal need, other than as speculative tool, such as exotic products were chastised [49] and many desks were closed as a result,
- the product class that took the niche of exotics became simpler vanilla products, which hedging property has still utilitarian value[1], more liquid, less volatile and therefore more in-line with the role of derivatives at their inception,
- traditional financial mathematical programs focused on derivatives in which highest likelihood and mathematical convenience prevailed over data supported by the market were chastised and rethought [61] which popularized Machine Learning (ML) and more specifically Gaussian processes (GP) within them because they provided a flexible non-parametric framework to which, one could incorporate growing data. The latter academic scheme is already making good progress [91] at modeling the options market but it seems there are some room for improvement especially when coherence as defined by arbitrage constraints is taken into consideration,
- liquidity modeling became of central focus for government led initiatives [84] such as for example within the Fundamental Review of the Trading Book (FRTB),
- risk models increased in sophistication more specifically in the context of coherent non-arbitrageable scenarios [71] and risk factor to P&L mapping [84].

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

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

[1]For example a farmer would use a put options in order to hedge himself against the prices of its crop going down few months before maturity.

B. Problem Formulation

In this paper we expose some of the challenges associated to the process of price discovery in the context of vanilla options market making, more specifically resulting to its asynchronous and multi-space[2] properties and with the arbitrage and liquidity constraints.

1) The problem of normalizing rolling contracts: Though sometimes including small idiosyncratic differences most derivatives listed markets typically offer new contracts once a month with a two year expiry on a fixed date. This means that once, two years have elapse from the first issuance of the listed contracts, we have every months the contracts which were issues 2 years back that expire. The day an issuance occurs:

- we have a new information about what the information stored in implied volatility surface is worth with a two year expiry,
- a new information about short dated options (which expired the day of issuance)
- as well as information about the surface every month in between these two time-lapses.

It is usually agreed that there exists 9 important pillars in which linear interpolation in variance space gives reasonable results[3]. These pillars are Over Night (ON), 1 Week (1W), 2W, 1M, 2M, 3M, 6M, 1 Year (1Y) and 2Y. Figure I-B.1 gives an illustration of these pillars the instant of simultaneous issuance of the longest expiry options with the expiry of the shortest expiry contracts. Figure I-B.1 illustrates the case in which the last expiry was more than a day away associated with the challenges in estimating these invisible points.

[2]to be understood as the strike and tenor space

[3]For example on the FX markets.

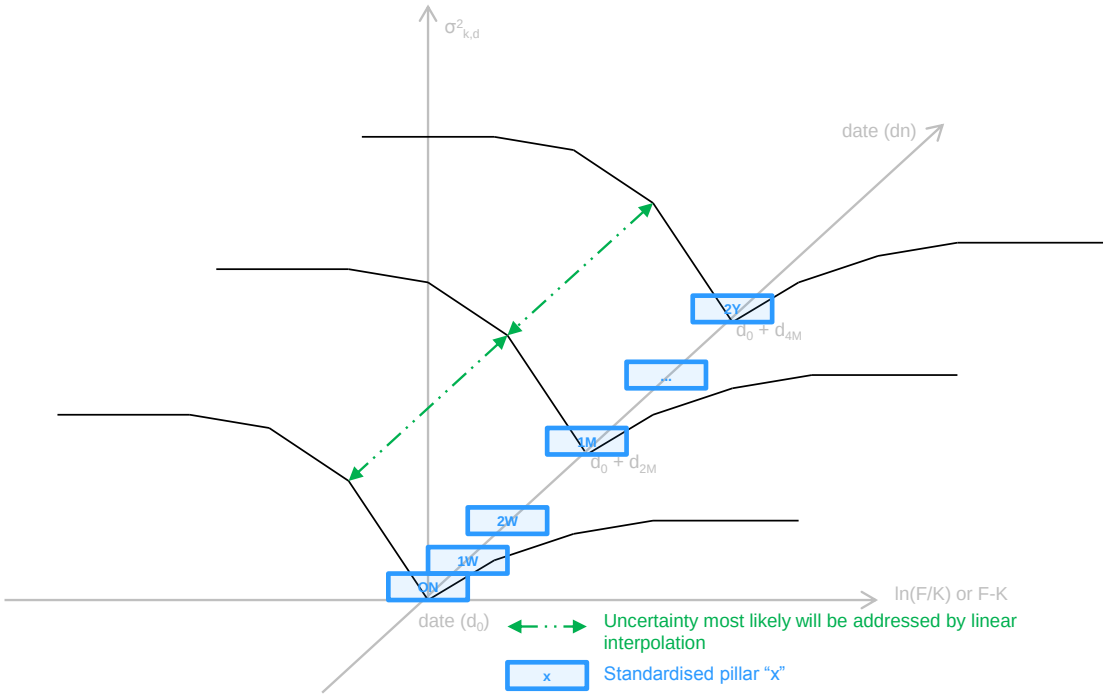

Fig. 1. When contracts roll we have a simple solution

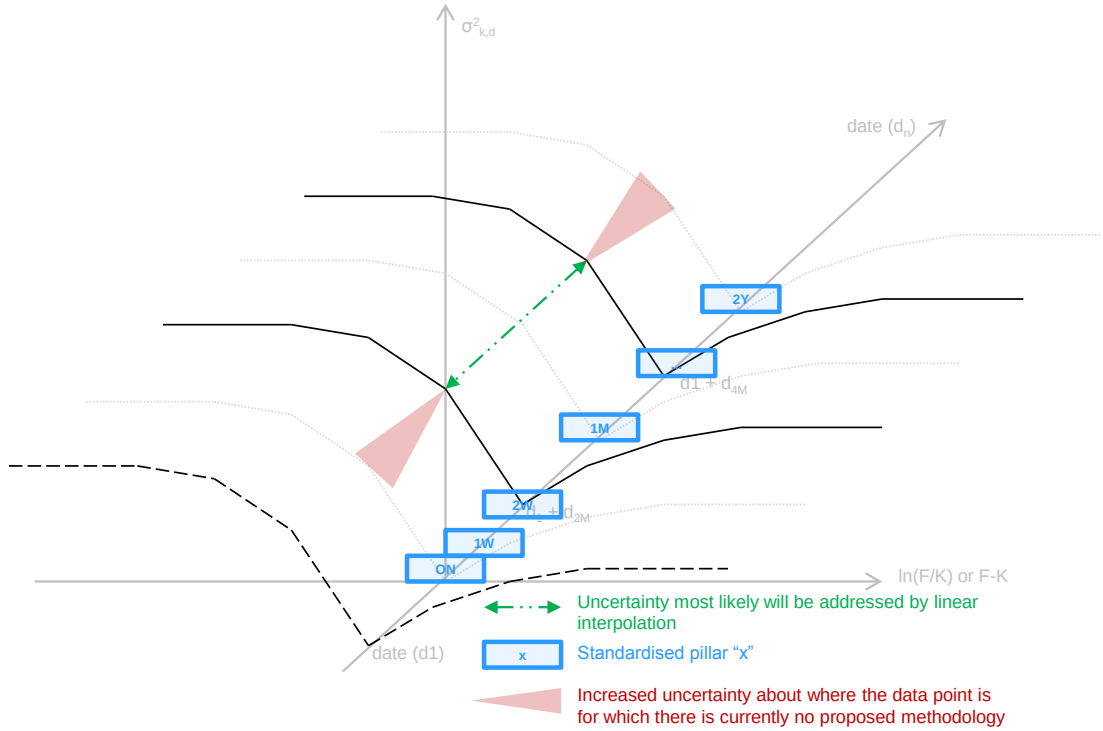

Fig. 2. When there is no roll, how do we populate the red zones without jumps?

2) The problem of asynchronous data from the OTC market: The second problem we will address in the one associated to the question of marked implied volatility surface (IVS) update in cases of non listed volatilities. For instance figure I-B.2 represents a tranche of our IVS. We can see that a random arrival at a specific point of the IVS may diffuse throughout the tranche and beyond but the way this is done is critical in the context of Market Making. for instance in figure I-B.2 the arrival of information on a specific moneyness (the arrow) could mean many things:

- the specific moneyness has increased without any change on the remainder of the IVS,
- the overall IVS price has increased but we only observe one specific point,
- a change of skew,
- a change in the total variance structure,
- a localized change in the implied, volatility with small impact in the direct vicinity without change in the distant points,
- an increase in that specific point with a decrease in the rest of the IVS.
- other less likely change in various risk factors.

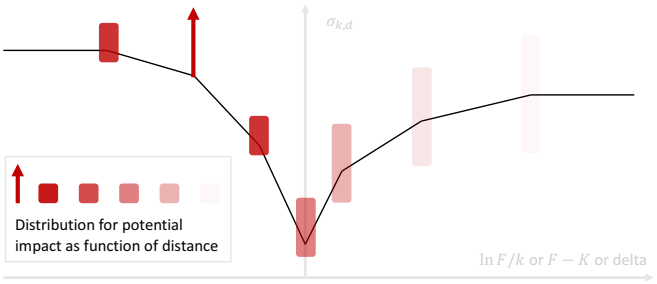

Fig. 3. Asynchronous information arrival on specific moneyness and the intuitive representation for the impact on other the other strikes of the same tenor

These few changes and their mix will have to be adequately addressed in our proposed methodologies at the distribution level in which the latter is of various scenarios of IVS changes.

C. Structure of this technical document

We first explore the science of fetching the raw, available but sparse data from the markets in section II. In section III we present the risk factors associated to the volatility surface by re-introducing the ones of the Implied Volatility Surface Parametrization (IVP) [70]. We will recall in section IV the arbitrage constraints inherent to the IVS, a necessary step to the resampling methodology we will introduce in section VI. A literature review for scenario tracking will however be summarized prior to that in section V.

II. The Science of Fetching the Raw Data

A. Black-Scholes related models

The celebrated Black-Scholes-Merton (BSM) model is the most respected closed form equation that provides the mathematical weaponry to price European options model. It can take 3 main form depending on the underlying diffusion:

- Log-Normal diffusion: the most well known form,
- Normal Assumption: which has become more useful in the recent past especially on the interest rate market in which we have seen in 2016 the bizarre economic state of negative ones,
- Garman Kohlhagen model on the FX market which formalizes the log-normal diffusion as a ratio of log-normal diffusion.

1) Log-Normal Assumption: The BSM formula using the log-normal diffusion is given equation (1).

$$C(S_t,T) = e^{-r(T-t)}[FN(d_1) - KN(d_2)] \quad (1)$$

with $d_1 = \frac{1}{\sigma\sqrt{T-t}}\left[\ln\left(\frac{S_0}{K}\right) + (r - q + \frac{1}{2}\sigma^2)(T-t)\right]$, $d_2 = d_1 - \sigma\sqrt{T-t}$, $N(\cdot)$ the cumulative distribution function of the standard normal distribution, $T-t$ the time to maturity, S_t the spot price of the underlying asset, F_t the forward price, K is the strike price, r the risk free rate, q the dividend yield and σ the volatility of returns of the underlying asset.

Proof: If we take the BSM diffusion $dS_t/S_t = (r-q)dt + \sigma dW_t$, using Ito's lemma, we get $S_T = S_t e^{(r-q+\frac{\sigma^2}{2})(T-t)+\sigma W_{T-t}}$. The price of a European Call is given by $C(S_t,T) = e^{-r(T-t)}\mathrm{E}^Q[S_T - K]^+ = e^{-r(T-t)}\frac{1}{\sqrt{2\pi\sigma}}\int_{-\infty}^{\infty}(S_T - K)1_{S_T>K}e^{\frac{-x^2}{2\sigma}}dx$. We can also note that $S_T > K \Leftrightarrow x < \frac{1}{\sigma\sqrt{T-t}}\left[\ln\left(\frac{S_0}{K}\right) + (r - q + \frac{1}{2}\sigma^2)(T-t)\right]$. We can get rid of the indicator function and adjust the born of the integral function and we get equation (1). ■

2) Normal Assumption: The BSM pricing methodology using the normal diffusion is given equation (2).

$$C(S_0,t) = e^{-r(T-t)}[(F-K)N(d) + \sigma\sqrt{T-t}N'(d)] \quad (2)$$

with $d = \frac{F-K}{\sigma\sqrt{T-t}}$.

Proof: If we take the BSM normal diffusion $dS_t = (r-q)dt + \sigma dW_t$, using the proof methodology of II-A.1 we get equation (2). ■

3) Garman Kohlhagen model: An adjustment in the FX market is necessary compared to the other markets. This is done with the Garman Kohlhagen (GK) [27] model instead of the BSM model to account for the presence of two interest rates relevant to pricing: r_d the domestic risk free simple interest rate and r_f the foreign risk free simple interest rate. Equation (3), with the usual BSM naming conventions, provides the pricing method laid down by GK.

$$C = S_0 e^{-r_f T}N(d_1) - Ke^{-r_d T}N(d_2) \quad (3)$$

with $d_1 = \frac{\ln(S_0/K) + (r_d - r_f + \sigma^2/2)T}{\sigma\sqrt{T}}$ and $d_2 = d_1 - \sigma\sqrt{T}$.

Proof: If we take the BSM normal diffusion $dS_t/S_t = (r_d - r_f)dt + \sigma dW_t$, using the proof methodology of II-A.1 we get equation (2). ■

B. Ensemble learning and the Brent Algorithm

The BSM assumes constant volatility but the vanilla options' prices, as seen on the market, suggest that the BSM is wrong.

1) Motivation: There are many reasons why the BSM is still used but for the sake of making the reasons brief, we can put forward the argument of the Greeks being critical across all main banking functions:

- Front Office: where pricing and hedging is used on daily basis on options' desks,
- Middle Office: risk management in which VaR methodologies using sensitivities are quite common and
- Back Office: product control in which clearing methodologies require live P&L and often sensitivities are used.

To reconcile the BSM equation to the market observable prices, the only non-observable value is the volatility input value. We call IVS the geometrical 3D structure which takes as input a tenor and a moneyness[4] and returns the volatility value which reconciles the BSM equation to the market observable price (as we can see from figure 4). If we call the function $Pricer$, the closed form formula returning the BSM of equation (1), (2) or (3) with $\mathcal{F} \in \{\mathcal{N}, \mathcal{L}, \mathcal{G}\}$ acting as a flag to the pricing methodology. There are 3 main algorithms used going from price risk to implied vol, namely the Bisection, the Newton-Raphson and their "ensemble learning", Brent algorithm.

2) Bisection: The Bisection method described in algorithm 1 has the property of always converging but can be a bit slow. This root-finding method repeatedly bisects the interval of study and subsequently selects the subinterval in which the root is contained. The process is repeated until the root is found within an arbitrary error.

Algorithm 1 Bisection method returns IVS

Require: P, S_t, K, r_d, r_f, T
Ensure: $P \approx C(\mathcal{F}, S_t, K, \sigma_i, T, r_d, r_f, q, r, T)$
$\epsilon \leftarrow 0.01$; $N = 50$; $\sigma_+ \leftarrow 3.0$; $\sigma_- \leftarrow 0.01$;
for $i = 1$ to N **do**
  $\sigma_i \leftarrow \frac{\sigma_+ + \sigma_-}{2}$
  if $P > C(S_t, K, \sigma, T, r_d, r_f, q, r)$ **then**
    $\sigma_+ \leftarrow \sigma_i$
  else
    $\sigma_- \leftarrow \sigma$
  end if
  if $|P - C(S_t, K, \sigma, T, r_d, r_f, q, r)| < \epsilon$ **then**
    $i \leftarrow N$
  end if
end for
Return σ_i

3) Newton-Raphson: The Newton-Raphson described in algorithm 2 is faster than the bisection method but does not always converge. The idea of this root finding algorithm starts with an initial guess, assumed reasonably close to the the solution. Using basic calculus, the tangent line is then calculated as a mean to approximate the next guess. The method is iterated until a stopping criteria such as an approximate error is enforced.

[4]or log-moneyness or delta space depending on which form of the BSM and which asset class we are dealing with.

Algorithm 2 Newton method returns IVS σ_i

Require: P, S_t, K, r_d, r_f, T
Ensure: $P \approx C(\mathcal{F}, S_t, K, \sigma_i, T, r_d, r_f, q, r, T)$
$\sigma_0 \leftarrow .5$,
$\varepsilon \leftarrow 10^{-5}$,
$\epsilon \leftarrow 10^{-14}$,
$M = 20$
$\mathcal{S} = false$
$y' \leftarrow Ke^{-r\tau}\phi(d_2)\sqrt{\tau}$
for $i = 1$ to M **do**
  $y \leftarrow C(\mathcal{F}, S_t, K, \sigma_0, T, r_d, r_f, q, r, T)$
  $d_2 \leftarrow \frac{\ln(S/K)+(r-q-\sigma_0^2/2)\tau}{\sigma_0\sqrt{T}}$
  $y' = Ke^{-r\tau}\phi(d_2)\sqrt{\tau}$
  if $|y'| < \epsilon$ **then**
    break;
  end if
  $\sigma_1 = \sigma_0 - y/y'$
  if $|\sigma_0 - \sigma_1| < \varepsilon \times |\sigma_1|$ **then**
    $\mathcal{F} =$ true
    break;
  end if
  $\sigma_0 = \sigma_1$
end for
if $\mathcal{S} =$ true **then**
  Return σ_i
else
  "Algorithm did not converge"
end if

4) Brent: The Brent algorithm is a mixture of the Bisection described in algorithm 1 and the Newton methods described in algorithm 2. The Brent algorithm essentially tried the Newton method and switches to the Biscetion method if the algorithm has hard time converging. If both speed and accuracy matter, we recommend this algorithm, otherwise the simplicity of the Bisection method works in most financial applications.

Remark We will call $C(.)$ the pricing function that gets as input an implied volatility $\sigma(t, k)$ and returns the relevant price. $\mathcal{B}$ will therefore be the normal, log-normal or the Garman Kohlhagen formula depending on the asset class in which one works.

III. Implied Volatility Surface Risk Factors

The objective of this section is to discuss the risk factors associated to the simplest of the non linear products, the vanilla options, which are the stepping stones of more complex derivative strategies[5]. Studying vanilla options can be done in couple of domains, the price domain or the IVS domain, developed to address the limitations of the Black-Scholes model. As it happens, working on the IVS domain offers lots of benefits that the price domain cannot replicate. There exists many parametrization of the IVS, notably the

[5]eg: straddle, strangle, butterfly, call & calendar spread, condor, etc ...

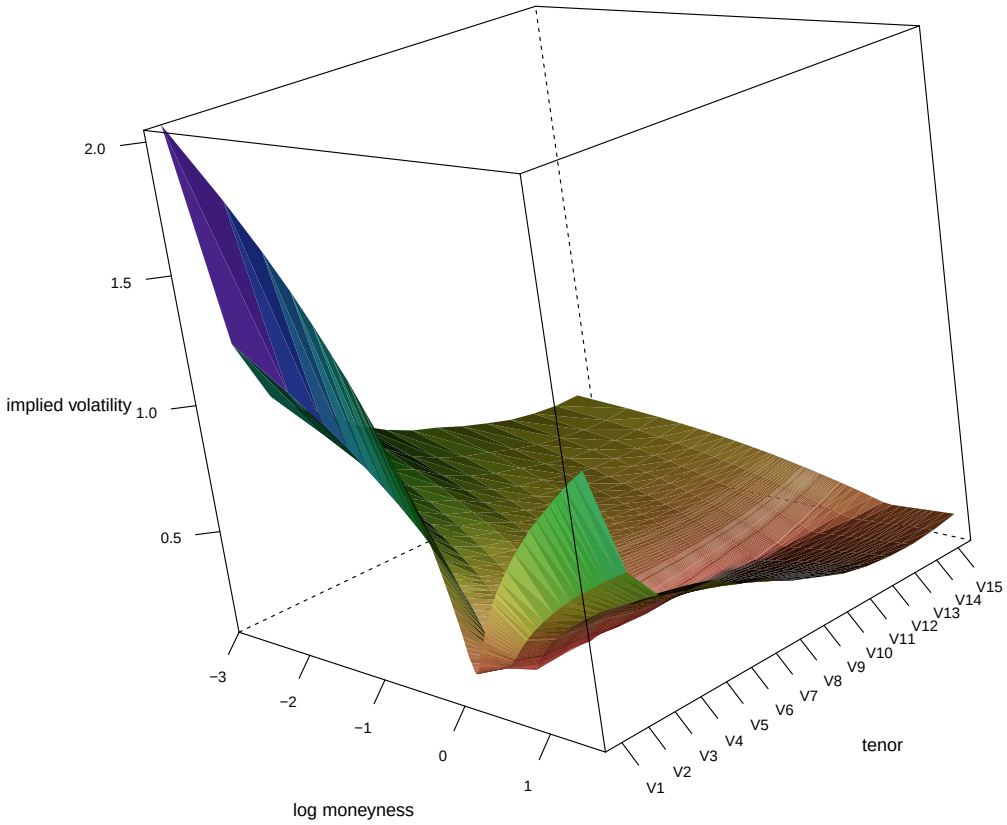

Fig. 4. A possible arbitrage free IVS plot

Schonbucher and the SABR models [98], [40], [4], have had their share of practitioners enthusiasts. However, we will only discuss the SVI [28], [30], [29], [31] model and its most advanced extension, the IVP as it is currently the one which has the most comprehensive number of risk factors.

A. The Raw Stochastic Volatility Inspired (SVI) model

1) History: One advertised[6] advantage of the SVI is that it can be derived from Heston [44], [30], 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 (in which it yields a poor fit) 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 [70].

Remark In terms of notations, we use the traditional notation [31] 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 $t > 0$, $C_{\text{BS}}(k, \sigma^2 t)$ 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 IVS by $\sigma_{BS}(k,t)$, and define the total implied variance by

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

The implied variance v shall be equivalently defined as $v(k,t) = \sigma^2_{\text{BS}}(k,t) = w(k,t)/t$. 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.

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 in the context of an implied volatility parametrization, "parameters" and "risk factors" can be used interchangeably.

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

$$w(k, \chi_R) = a + b[\rho(k-m) + \sqrt{(k-m)^2 + \sigma^2}] \qquad (4)$$

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

[6]One of the main point of this paper is to expose a small mistake that was done in one particular paper [28] but for the sake of the introduction, we will make this remark as a footnote.

Remark Note that there exist several other forms of the SVI model which are equivalent to each other through a set of transform functions [31]. The motivation of their existence and the details of the transforms are out of scope but we refer to the original papers [31] 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 model with great accuracy a great portion of the volatility surface and its dynamic. Figure 5 illustrates the change in the a parameter (the general volatility level risk), Figure 6 illustrates the change in the b parameter (the vol of vol risk), Figure 7 illustrates the change in the ρ parameter (the skew risk), Figure 9 illustrates the change in the σ parameter (the ATM volatility risk) and finally Figure 8 illustrates the change in the m parameter (the horizontal displacement risk).

B. Relation between IVP and raw SVI

Jim Gatheral developed the SVI model at Merrill Lynch in 1999 and implemented it 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 [6] gave a mathematical justification for this market observation. The paper suggests that the IVS 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 [4] 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 [70] especially given the challenging regulatory environment[7]. Mahdavi-Damghani initially named the model "generalized SVI" (gSVI) [4] but renamed it "Implied Volatility Parametrization" (IVP) [70] 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[8]. The function needed to be increasing as it gets further away from m, majored by a linear function increasing in $[m;+\infty[$, decreasing in $]-\infty;m]$ and increasing in concavity the further away it gets from the center. Equation (5) summarizes the gSVI[9]. The penalization was initially given by equation (5b). Figure 10 illustrates the change in the β parameter.

$$\sigma^2_{gSVI}(k) = a + b\left[\rho(z-m) + \sqrt{(z-m)^2+\sigma^2}\right] \quad (5a)$$

$$z = \frac{k}{\beta^{|k-m|}}, 1 \leqslant \beta \leqslant 1.4 \quad (5b)$$

Remark The downside transform in the gSVI [4] 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 (6a). That idea can be prolonged to exponential like function such as the one in Equations (6b) or (6c). The idea is always the same: the further away you are from the ATM, the bigger the necessary adjustment on the wings.

$$z = \frac{k}{\beta^{\mu+\eta|k-m|}} \quad (6a)$$

$$z = e^{-\beta|k-m|}(k-m) \quad (6b)$$

$$z = \log(\beta|k-m|) \quad (6c)$$

Mahdavi-Damghani, in introducing the IVP model [70] picked in equation (6a) 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 exponential like function may work better when it comes to showing convergence between the modified Heston and the IVP model [72].

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 [4], [70] defines what he labeled the Implied Volatility surface Parametrization (IVP) below:

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

The functions $\alpha(p)$ (figure 11) and $\psi(p)$ (figure 12) 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

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

[8] 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

[9] or alternatively IVP's mid, model

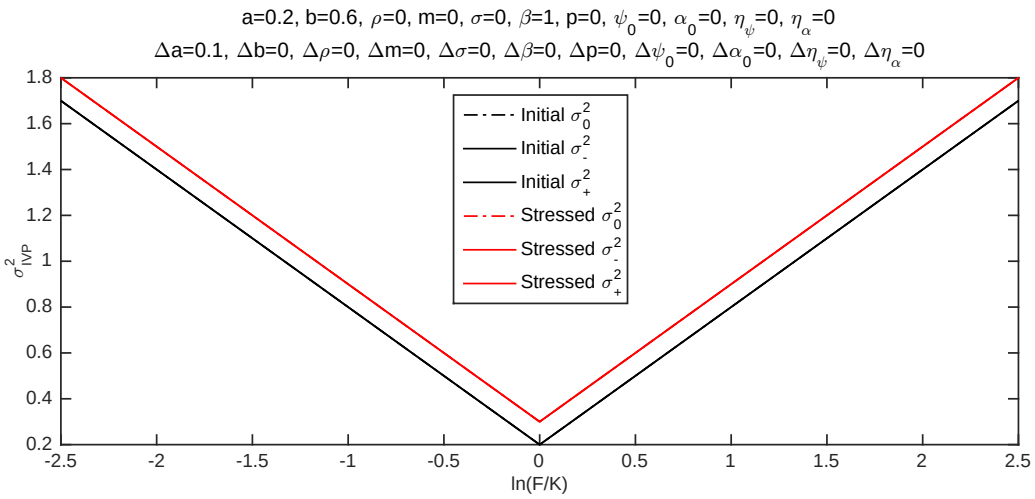

Fig. 5. Change in the a parameter in the rawSVI/gSVI/IVP model

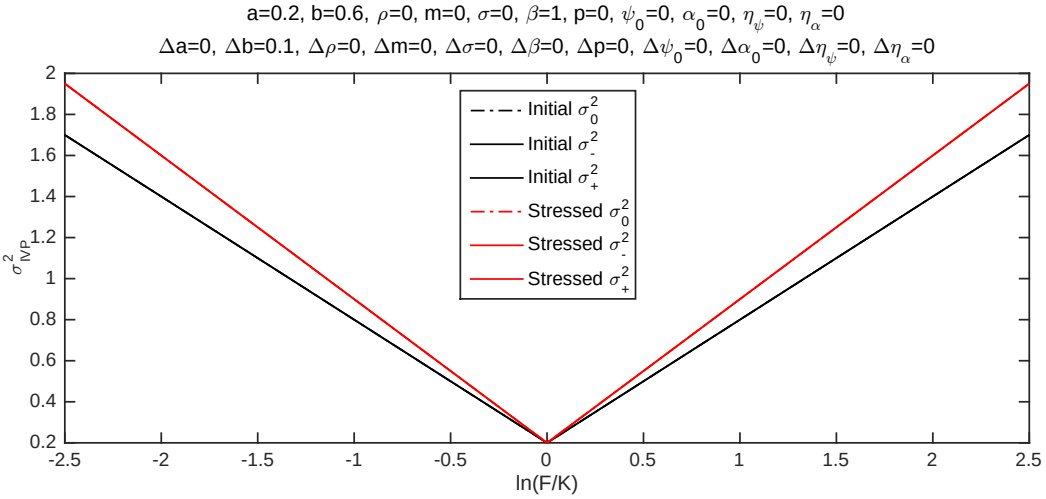

Fig. 6. Change in the b parameter in the rawSVI/gSVI/IVP model

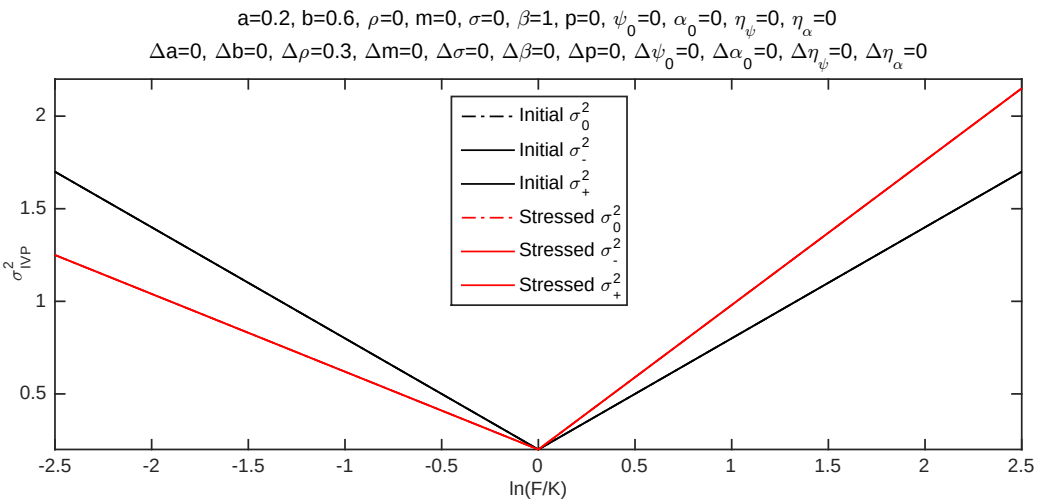

Fig. 7. Change in the ρ parameter in the rawSVI/gSVI/IVP model

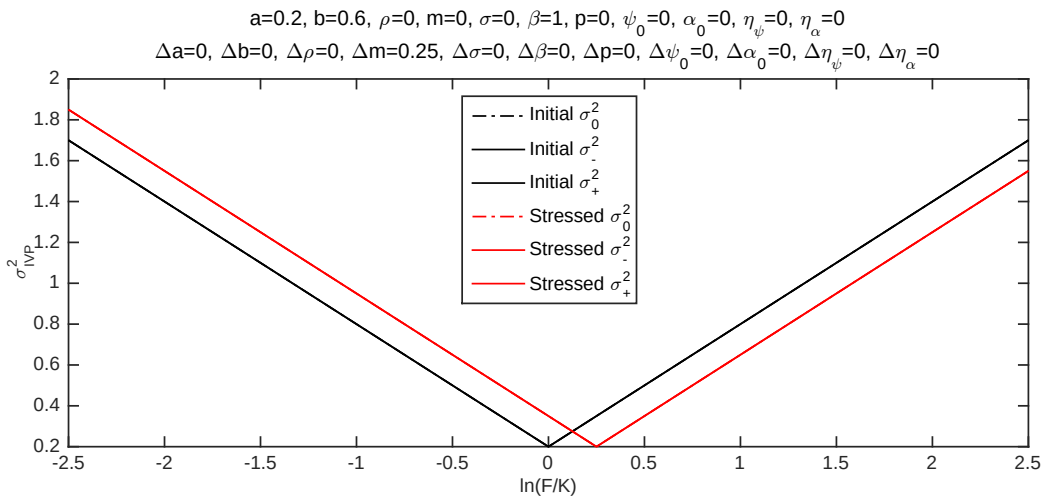

Fig. 8. Change in the m parameter in the rawSVI/gSVI/IVP model

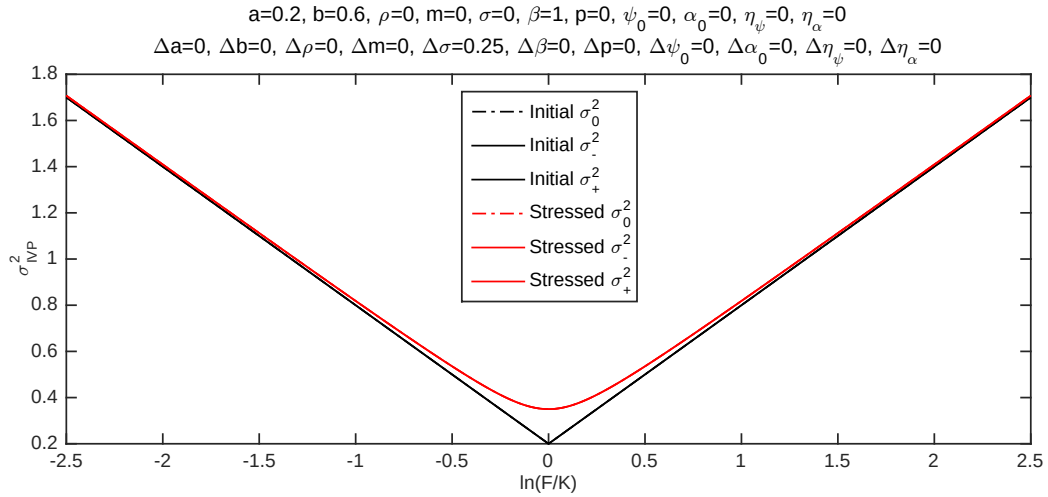

Fig. 9. Change in the σ parameter in the rawSVI/gSVI/IVP model

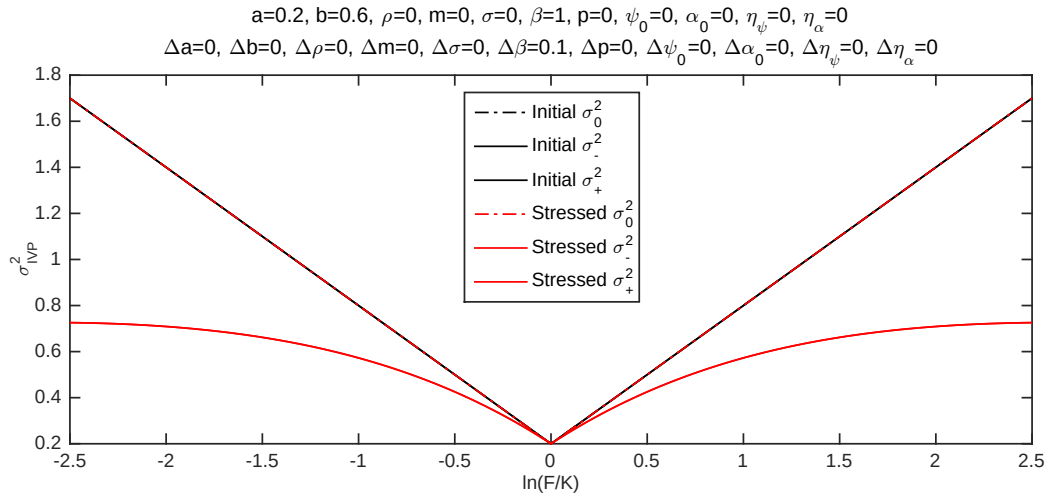

Fig. 10. Change in the β parameter in the gSVI/IVP model

impact and hence the wider the Bid-Ask. This market impact parameter is controlled by p (figure 13). Finally, couple of additional parameters model the elasticity of the liquidity: η_ψ (figure 14) and η_α (figure 15).

IV. Arbitrage & the Options Market

As we have seen in section I-A, having coherent risk scenarios has become of central importance in the last few years. 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 IVS is highly skewed or/and high. Second, many of the risk engines uses numerical methods which break if an arbitrage is created on the IVS. 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 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 IV-A) and tenor (section IV-B) spaces.

A. Condition on the strike

The model set up is the usual. Let us set up the probability space $(\Omega, (\mathcal{f})_{(t\geqslant 0)}, \mathbb{Q})$, with $(\mathcal{f})_{(t\geqslant 0)}$ 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 7.

$$dS_t = rS_t dt + \sigma_t S_t dW_t \tag{7}$$

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 results [20], [21], we can write the price of a call: $C(S_0,K,T) = e^{-rT}\mathbb{E}^{\mathbb{Q}}[S_T - K]^+ = e^{-rT}\int_K^{+\infty}(S_T-K)\phi(S_T,T)dS_T$ with $\phi(S_T,T)$ being the final probability density of the call. Differentiating twice we find equation (8).

$$\frac{\partial^2 C}{\partial K^2} = \phi(S_T,T) > 0. \tag{8}$$

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 (9) which is known in the industry as the arbitrage constraint of the positivity of the butterfly spread [107].

$$\forall \Delta, C(K-\Delta) - 2C(K) + C(K+\Delta) > 0 \tag{9}$$

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 [31] 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 (8) 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 condition, in equation (8), 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 III. This condition is given by equation (10).

$$\forall K, \forall T, |T\partial_K \sigma^2(K,T)| \leqslant 4 \tag{10}$$

Proof: The intuition behind the proof is taken from Rogers and Tehranchi [93] 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

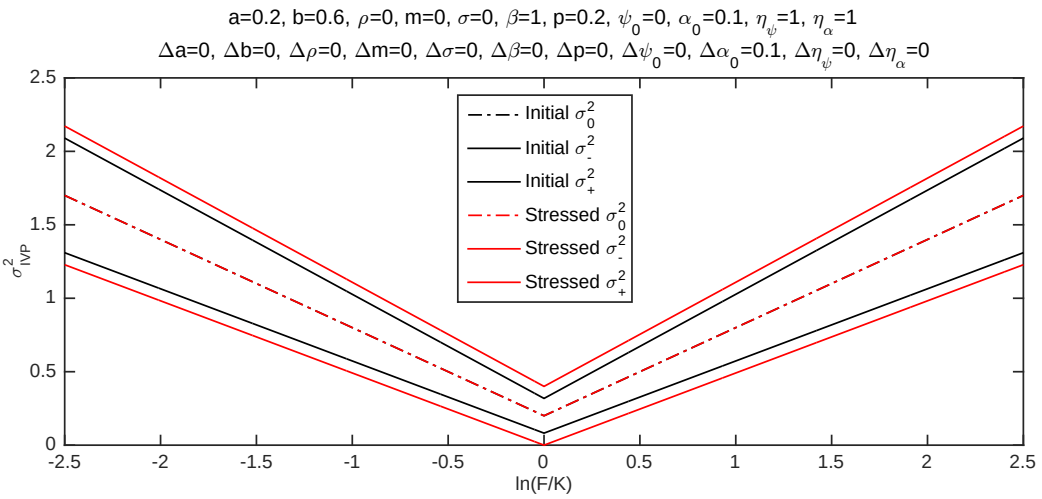

Fig. 11. Change in the α parameter in the IVP model

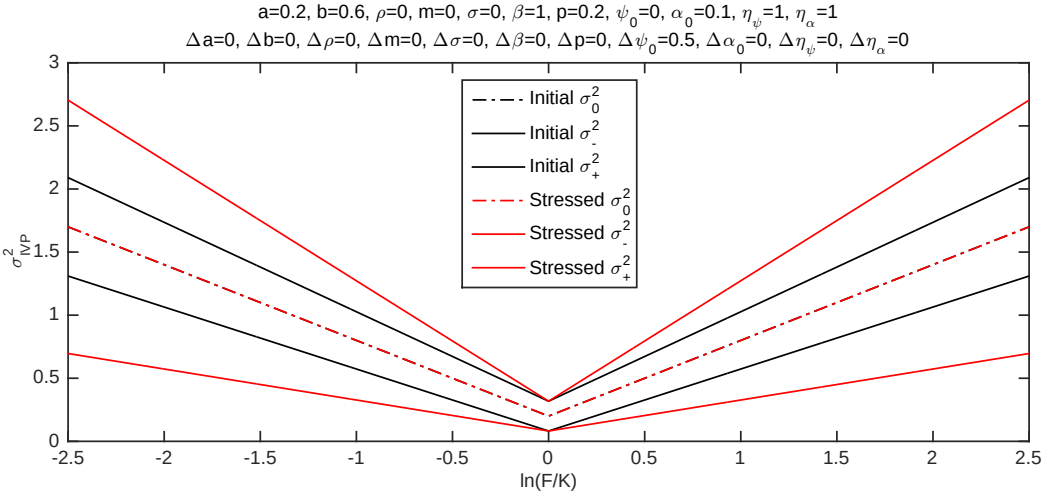

Fig. 12. Change in the ψ parameter in the IVP model

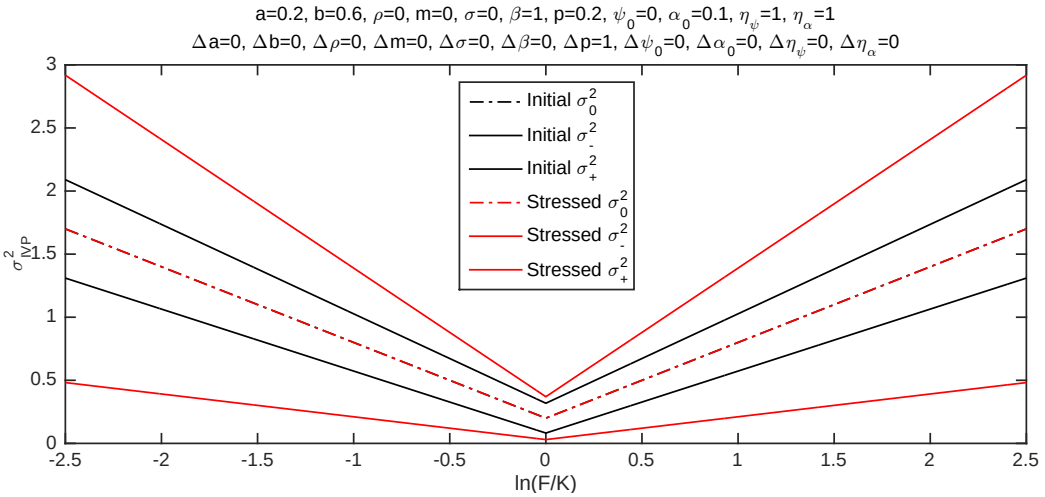

Fig. 13. Change in the p parameter in the IVP model

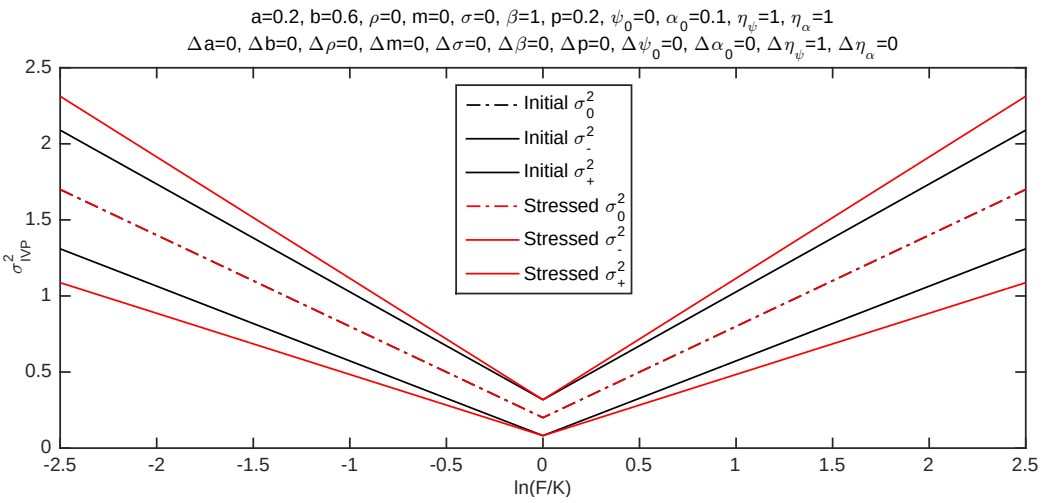

Fig. 14. Change in the η_ψ parameter in the IVP model

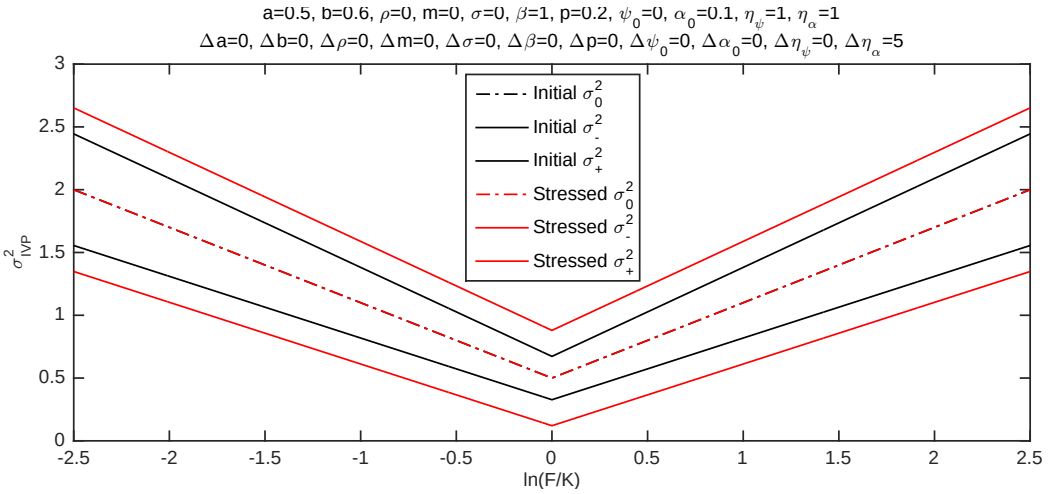

Fig. 15. Change in the η_α parameter in the IVP model

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 [93] 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 (10), 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 (10) is practical is because in de-arbitraging methodologies (as we will see more in details in section III), 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 III 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 16 represents a counter example of $|T\partial_K\sigma^2(K,T)| \leqslant 4$ applied to the Raw SVI parametrisation[10] 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 IV-A, that is let us set up the probability space $(\Omega, (\mathcal{f})_{(t\geqslant 0)}, \mathbb{Q})$, with $(\mathcal{f})_{(t\geqslant 0)}$ 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 (7). In order to prevent arbitrages on the volatility surface on the tenor space we will split this subsection in its theoretical form in section IV-B.1 and IV-B.2 for its practical form.

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:

$$C(K,T+\Delta) - C(Ke^{-r\Delta},T) \geqslant 0 \tag{11}$$

Proof: One application of Dupire's formula [20], [21] is that the pseudo-probability density must satisfy the Fokker-Planck [24], [88] equation. This proof is taken from El Karoui [56]. Let us apply Itô to the semi-martingale . This is formally done by introducing the local time Λ_T^K: $e^{-r(T+\varepsilon)}(S_{T+\varepsilon}-K)^+ - e^{-r(T)}(S_T-K)^+ = \int_T^{T+\varepsilon} re^{-ru}(S_u-K)^+ du + \int_T^{T+\varepsilon} e^{-ru}1_{\{S_u\geqslant K\}}dS_u + \frac{1}{2}\int_T^{T+\varepsilon} e^{-ru}d\Lambda_u^K$. 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: $E\left(e^{-ru}1_{\{S_u\geqslant K\}}S_u\right) = C(u,K) + Ke^{-ru}P(S_u \geqslant K) = C(u,K) - K\frac{\partial C}{\partial K}(u,K)$. 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:

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

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:

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

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 (8) 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 $\frac{\partial C}{\partial t} +$

[10] which we discuss more in details in section III.

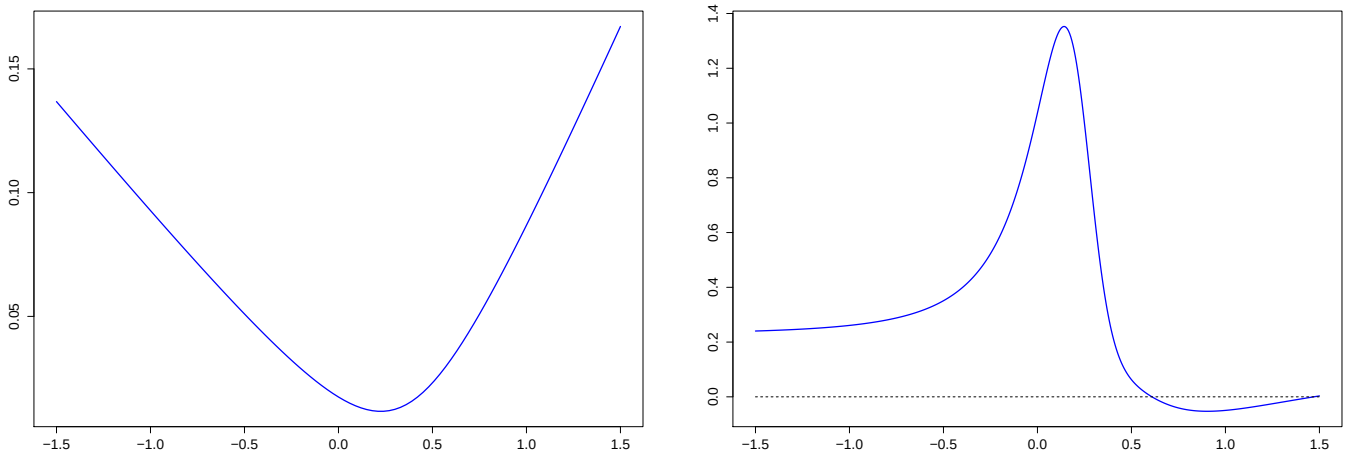

Fig. 16. Axel Vogt [111] counter-example for $b(1+|\rho|) \leqslant \frac{4}{T}$ being arbitrage free

$rK\frac{\partial C}{\partial K} \geqslant 0$. 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 IV-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 $\mathbb{Q}$,

$$\forall t > s,\ e^{-rt}\mathbb{E}^{\mathbb{Q}}(S_t - K)^+ \geqslant e^{-st}\mathbb{E}^{\mathbb{Q}}(S_s - K)^+ \qquad (12)$$

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

1) General picture: As we have seen from equations (9) and (11) there are couple of arbitrages types, the calendar and butterfly arbitrage as summarized my equation (13b).

$$\forall \Delta, C(K-\Delta) - 2C(K) + C(K+\Delta) > 0 \qquad (13a)$$
$$\forall \Delta, \forall T, C(K,T+\Delta) - C(Ke^{-r\Delta},T) \geqslant 0 \qquad (13b)$$

A new wave of risk methodologies with the objective of making incoherent scenarios like the ones allowing an arbitrage is currently being developed [4], [31], and though promissing few questions remain to be addressed [70].

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 III: $\forall \Delta, C(K-\Delta, \sigma_{IVP,+,t}(k)) - 2C(K, \sigma_{IVP,-,t}(k)) + C(K+\Delta, \sigma_{IVP,+,t}(k)) > 0$ which gives: $C(K,T+\Delta,\sigma_{IVP,+,t}(k)) - C(Ke^{-r\Delta},T,\sigma_{IVP,-,t}(k)) \geqslant 0$.

2) Intuitive Mathematical Specification: If we were to take an intuitive representation of the IVS at market observable pillars with a double array (figure 17), then if we disregard macro-economical, asset specific factors[11], then the relative co-movements of these pillars as driven by the pure arbitrage opportunities as explained by section IV-A and IV-B would be best described by figure 17.

Remark This intuitive representation of figure 17 no longer works with the FX pillars (figure 18). This is because the data in FX is listed in delta space but the classic de-arbing algorithms assumes that the data is conveniently aligned in log-moneyness space[12]. Indeed the market delta space pillars are the 10, 25, 50, 75, 90 delta[13]. The delta to log-moneyness conversion creates increasing mis-alignments as the tenor increases (figure 18).

3) Few Definitions:

Definition Let C^τ be the set of standardized pillars, C^k the set of standardized strikes[14]. Let's call C^d the set of live contract expiries. We will call $\sigma_t(C_i^\tau, C_j^k)$ the implied volatility, as observed from the price space and $\hat{\sigma}_t(C_i^\tau, C_j^k)$ the closest implied volatility spanned by the IVP parameters for $\sigma_t(C_i^\tau, C_j^k)$ with:

- "i"th observed element of C^τ where $1 < i < |C^\tau|$ and
- "j"th element of C^k of C^d where $1 < j < |C^d|$ and

Definition Let's call $\tilde{C}^\tau \in C^\tau$ and $\tilde{C}^k \in C^k$ the set of incomplete data taken within the standardized strikes.

Remark For most cleared asset classes, we usually have:

[11] In the Equities market the skew is such that it reflect the market participants fear of bankruptcies and hence a premium towards a price decrease of the underlier. In the commodities market we see the reverse (with an exception for oil), people are afraid of prices going up and hence the observation of a reverse skew.

[12] $\Delta_f = \phi e^{-rft} N(\phi \frac{1}{2}\sigma\sqrt{t})$

[13] On the tenor axis the pillars are usually ON, 1W, 2W, 1M, 2M, 3M, 6M, 1Y, 18M, 2Y.

[14] which can depending on the market be expressed in moneyness, log-moneyness or delta.

$C^\tau = \{ON, 1W, 2W, 1M, 2M, 3M, 6M, 1Y, 18M, 2Y\}$.

Remark If we are dealing with the FX market, we have $C^k = \{10, 25, 50, 75, 90\}$.

Remark In general we have $|C^k| < |C^\tau|$.

4) Optimization by constraint specification: De-arbing is a convoluted mathematical optimization which perfect solution falls outside the scope of what, people in the industry, especially within the risk space usually define to be below the threshold for the pragmatic benefits to complexity ratio, so for this section of the practitioners, we propose, a partial de-arbing process for which a simplified de-arbitraging methodology has been illustrated in figure 17 and for which the optimization by constraint algorithm is described below.

solve:
$\hat{\sigma}_t(\tau, d) = \underset{\tilde{\sigma}_t(\tau,d)}{\arg\min} \sum_\tau \sum_d [C(\sigma_{i,t}(\tau, d)) - C(\tilde{\sigma}_t(\tau, d))]^2$
subject to:
$\forall d \in C^k,\ C_1^{\mathcal{B},d,\tau} < C_2^{\mathcal{B},d,\tau}$ and
$\forall \tau \in C^\tau,\ C_1^{\mathcal{C},d,\tau} < C_1^{\mathcal{C},d,\tau}$

Where we call $\mathcal{B}$ the call spread[15] arbitrage flag and $CS_1^{d,\tau}$ its impact in price given by $CS_1^{d,\tau} = \left| C_1^{\mathcal{B},d,\tau} - C_2^{\mathcal{B},d,\tau} \right| 1_{\mathcal{B}}$ where $C_1^{\mathcal{B},d,\tau} = C\Big(K - \Delta, \sigma_0(K - \Delta, \tau)\Big)$ and $C_2^{\mathcal{B},d,\tau} = C\Big(K, \sigma_0(K, \tau)\Big)$. Let $\mathcal{C}$ be the Calendar spread arbitrage flag and $CS_2^{d,\tau}$ its price impact given by $CS_2^{d,\tau} = \left| C_1^{\mathcal{C},d,\tau} - C_2^{\mathcal{C},d,\tau} \right| 1_{\mathcal{C}}$ where $C_1^{\mathcal{C},d,\tau} = C\Big(K, \tau + \Delta, \sigma_0(K, \tau + \Delta)\Big)$ and $C_2^{\mathcal{C},d,\tau} = C\Big(Ke^{-r\Delta}, \sigma_0(Ke^{-r\Delta}, \tau)\Big)$. In this algorithm, we make sure that for every pillar tenor and every pillar strikes the relevant points are mutually arbitrage free[16]. We try to find the shortest distances between the input vol and its closest arbitrage free mirror subject to the Call spread (equivalent to butterfly) and Calendar spread Conditions. In order to use the usual optimization tools, we need to adjust the objective function to take in the constraints of the problem. Now adjusting the the objective function as described in equation (14).

$$\begin{aligned} \hat{\sigma}_t(\tau, d) &= \arg\min_{\tilde{\sigma}_t(\tau,d)} \sum_\tau \sum_d [\sigma_{i,t}(\tau, d) - \tilde{\sigma}_t(\tau, d)]^2 \\ &\quad + \mathcal{K}(CS_1^{d,\tau} + CS_2^{d,\tau}) \end{aligned} \tag{14}$$

where $\mathcal{K}$ is the constraint scalar[17].

5) Economical argument around the closest arbitrage frontier: The Mean Square Error (MSE) methodology of the pointwise de-arbitraging methodology of algorithm 14 gives an intuitive representation in figures 17 and 18 of how the closest geometrical implied vol would be adjusted like. Though these optimization algorithms in the L2-norm appear intuitive for a Mathematician, they do not make sense in a market point of view. A better approach would be to do the optimization in L2-norm but rather on the price space instead of the IVS space like proposed by equation (15) with $BS(.)$ representing any of the 3 methodologies of section II-A and $\tilde{\mathcal{K}}$ representing a normalized scalar chosen[18] to benefit the optimization process on the price space, namely

$$\begin{aligned} \hat{\sigma}_t(\tau, d) &= \arg\min_{\tilde{\sigma}_t(\tau,d)} \sum_\tau \sum_d \tilde{\mathcal{K}}(CS_1^{d,\tau} + CS_2^{d,\tau}) \\ &\quad + [C(\sigma_{i,t}(\tau, d)) - C(\tilde{\sigma}_t(\tau, d))]^2 \end{aligned} \tag{15}$$

Though closer, this latter approach on the mid still does not reflect the critical liquidity aspect which creates on regular basis a mid volatility surface which is itself not arbitrage free but, however, not arbitrage-able when liquidity is taken into account. Moreover, the strong correlation between the different tenors of the volatility surface may influence the convenient substitution in situation of impossibility of perfect hedging. Also the view in the better methodology would be to interpret the movement as an economical argument rather than a MSE argument on the implied volatility in which the movement have higher variances on the lower tenor without much price impact.

V. Review of Inference models

A. Markov Chain Monte Carlo

Markov Chain Monte Carlo (MCMC) algorithms [78] sample from a probability distribution based on a Markov chain that has a desired equilibrium distribution, the quality of the sample improving at each additional iteration. We will see next few version of the MCMC algorithm.

1) Metropolis-Hastings algorithm: The Metropolis-Hastings algorithm is a MCMC method that aims at obtaining a sequence of random samples from a probability distribution for which direct sampling is difficult [79] and initially advertised of high dimensions. We will see in the next few algorithm examples that the methodology is now classified as useful for low dimensional problems. At each iteration x_t, the proposal next point x' is sampled through a proposed distribution $g(x'|x_t)$. We then calculate:

- with $a_1 = \frac{P(x')}{P(x_t)}$ is the the probability ratio between the proposed sample and the previous sample,
- and $a_2 = \frac{g(x_t|x')}{g(x'|x_t)}$, the ratio of the proposal density in both directions[19].

and set $a = \max(a_1 a_2, 1)$, we then accept $x_{t+1} = x'$ if $r \sim U[0, 1] \geqslant a$ which essentially means that if $a = 1$, accept is always true otherwise you accept with a probability $a_1 a_2$. The algorithm works best if the proposal distribution is similar to the real distribution. Note that the seed is slowly forgotten as the number of iterations increases.

[15] equivalent to the Butterfly condition

[16] without any guaranty that the in between pillars are arbitrage free. However we will see that this latter technical malaise can be neglected when add a bid/ask spread.

[17] a big enough number to make sure the constraints are respected but not too big to create numerical instabilities

[18] different from $\mathcal{K}$

[19] equal to 1 is the proposal density is symmetric

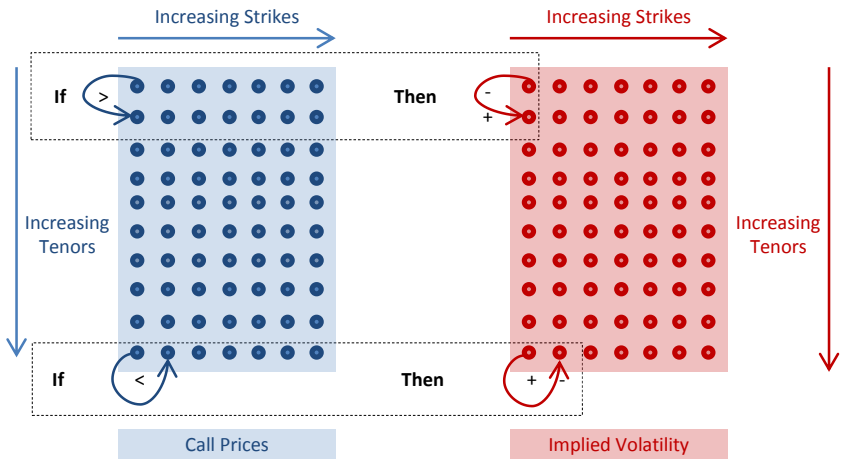

Fig. 17. Visualization for the core simple de-arbing idea

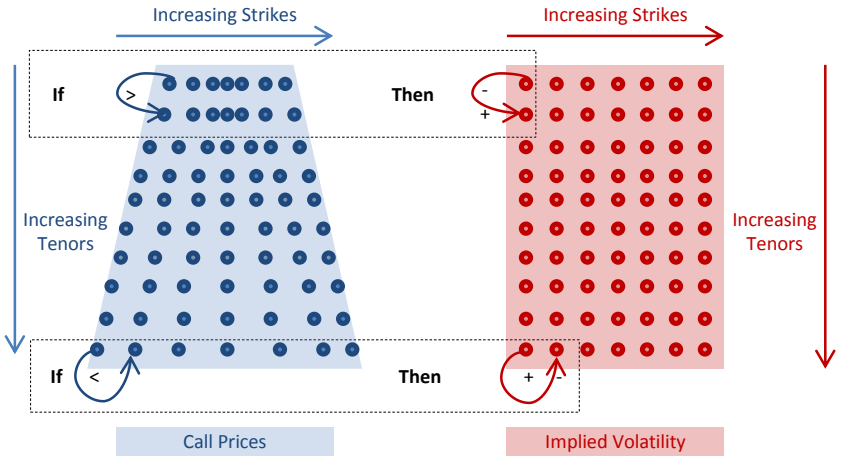

Fig. 18. Visualization for the core simple de-arbing idea approximation

2) Gibbs sampling: Perhaps one of the simplest MCMC algorithms, the Gibbs Sampling (GS) algorithm was introduced in by Geman & Geman [33] with the application of image processing. Later it was discussed in the context of missing data problems [108]. The benefice of the Gibbs algorithm for Bayesian analysis was demonstrated in Tanner and Wong [108]. To define the Gibbs sampling algorithm, let the set of full conditional distributions be: $\pi(\boldsymbol{\psi}_1|\boldsymbol{\psi}_2,\ldots,\boldsymbol{\psi}_p)$, $\ldots$, $\pi(\boldsymbol{\psi}_d|\boldsymbol{\psi}_1,\boldsymbol{\psi}_2,\ldots,\boldsymbol{\psi}_{d-1},\boldsymbol{\psi}_{d+1},\ldots,\boldsymbol{\psi}_p)$, $\ldots$, $\pi(\boldsymbol{\psi}_p|\boldsymbol{\psi}_1,\ldots,\boldsymbol{\psi}_{p-1})$. One cycle of the GS, described in algorithm 3, is completed by sampling $\{\boldsymbol{\psi}_k\}_{k=1}^{p}$ from the mentioned distributions, in sequence and refreshing the conditioning variables. When d is set to 2 we obtain the two block Gibbs sampler described by Tanner & Wong [108]. If we take general conditions, the chain generated by the GS converges to the target density as the number of iterations goes towards infinity. The main drawback with this method however is its relative computational heavy aspect because of the burn-in period.

3) Hamiltonian Monte Carlo: Hamiltonian Monte Carlo [19], sometimes also referred to[20] as hybrid Monte Carlo is an MCMC method for obtaining a sequence of random samples from a probability distribution for which direct sampling is difficult. It serves to address the limitations of the Metropolis-Hastings algorithm by adding few more parameters that aim is to reduce the correlation between successive samples using a Hamiltonian evolution process and also by targeting states with a higher acceptance rate.

4) Ordered Overrelaxation: Overrelaxation is usually a term associated with a Gibbs Sampler but in the context of this subsection we discuss Ordered Overrelaxation. The methodology aims at addressing the slowness associated in performing a random walk with inappropriately selected step sizes. The latter problem was addressed by incorporating a momentum parameter which consist of sampling n random variables (20 is considered a good [68] number for n), sorting them from biggest to smallest, looking where x_t ranks, say at p's position, amongst the n variables and the picking $n-p$

[20]though more in the past.

Algorithm 3 GIBBS-SAMPLING($\boldsymbol{\psi}_1^{(0)}, \ldots, \boldsymbol{\psi}_p^{(0)}$)

Require: Specify an initial value $\boldsymbol{\psi}^{(0)} = \left(\boldsymbol{\psi}_1^{(0)}, \ldots, \boldsymbol{\psi}_p^{(0)}\right)$

Ensure: $\left\{\boldsymbol{\psi}^{(1)}, \boldsymbol{\psi}^{(2)}, \ldots, \boldsymbol{\psi}^{(M)}\right\}$

1: **for** $j = 1, 2, \ldots, M$ **do**
2: Generate $\boldsymbol{\psi}_1^{(j+1)}$ from $\pi\left(\boldsymbol{\psi}_1 | \boldsymbol{\psi}_2^{(j)}, \boldsymbol{\psi}_3^{(j)}, \ldots, \boldsymbol{\psi}_p^{(j)}\right)$
3: Generate $\boldsymbol{\psi}_2^{(j+1)}$ from $\pi\left(\boldsymbol{\psi}_2 | \boldsymbol{\psi}_1^{(j+1)}, \boldsymbol{\psi}_3^{(j)}, \ldots, \boldsymbol{\psi}_p^{(j)}\right)$
4: $\vdots$
5: Generate $\boldsymbol{\psi}_d^{(j+1)}$ from $\pi(\boldsymbol{\psi}_d | \boldsymbol{\psi}_1, \boldsymbol{\psi}_2, \ldots, \boldsymbol{\psi}_{d-1}, \boldsymbol{\psi}_{d+1}, \ldots, \boldsymbol{\psi}_p)$.
6: $\vdots$
7: Generate $\boldsymbol{\psi}_p^{(j+1)}$ from $\pi\left(\boldsymbol{\psi}_p | \boldsymbol{\psi}_1^{(j+1)}, \ldots, \boldsymbol{\psi}_{p-1}^{(j+1)}\right)$
8: Return the values $\left\{\boldsymbol{\psi}^{(1)}, \boldsymbol{\psi}^{(2)}, \ldots, \boldsymbol{\psi}^{(M)}\right\}$
9: **end for**

for the subsequent sample x_{t+1} [80]. This form of optimal "momentum" parameter design is a central pillar of research in MCMC.

5) Slice sampling: Slice sampling is one of the remarkably simple methodologies [80] of MCMC which can be considered as a mix of Gibbs sampling, Metropolis-Hastings and rejection sampling methods. It assumes that the target density $P^*(x)$ can be evaluated at any point x but is more robust compared to the Metropolis-Hastings especially when if comes to step size. Like rejection sampling it draws samples from the volume under the curve. The idea of the algorithm is that it switches vertical and horizontal uniform sampling by starting horizontally, then vertically performing "slices" based on the current vertical position. MacKay made good contributions in its visual [68] representation.

6) Multiple-try Metropolis: One way to address the curse of dimensionality is the Multiple-try Metropolis which can be though of as a enhancement of the Metropolis-Hastings algorithm. The former allows multiple trials at each point instead of one by the latter. By increasing both the step size and the acceptance rate, the algorithm helps the convergence rate of the sampling trajectory [67]. The curse of dimensionality is another central area of research for MCMCs.

7) Reversible-Jump: Another variant of the Metropolis-Hastings is the Reversible-jump MCMC (RJ-MCMC) developed by Green [39]. One key factor or RJ-MCMC is that it is designed to address changes of dimensionality issues. In our case, as we saw in section III of "Paper Format" document, we face a dual type issues around change of dimensionality. The first being the frequency of each strategy in an ecosystem and the second element being the HFFF[21] which branching structure and size changes as a function of the strategy[22]. More formally. Let us define $n_m \in N_m = \{1, 2, \ldots, I\}$, as our model indicator and $M = \bigcup_{n_m=1}^{I} \mathbb{R}^{d_m}$ the parameter space whose number of dimensions d_m is function of model n_m (with our model indicators not needing to be finite). The stationary distribution is the joint posterior distribution of (M, N_m) that takes the values (m, n_m). The proposal m' can be constructed with a mapping $g_{1mm'}$ of m and u, where u is drawn from a random component U with density q on $\mathbb{R}^{d_{mm'}}$. The move to state (m', n'_m) can thus be formulated as $(m', n'_m) = (g_{1mm'}(m, u), n'_m)$. The function $g_{mm'} := (m, u) \mapsto (m', u')$, with $(m', u') = \left(g_{1mm'}(m, u), g_{2mm'}(m, u)\right)$ must be one to one and differentiable, and have a non-zero support: $\text{supp}(g_{mm'}) \neq \varnothing$, in order to enforce the existence of the inverse function $g_{mm'}^{-1} = g_{m'm}$, that is differentiable. Consequently (m, u) and (m', u') must have the same dimension, which is enforced if the dimension criterion $d_m + d_{mm'} = d_{m'} + d_{m'm}$ is verified ($d_{mm'}$ is the dimension of u). This criterion is commonly referred to as dimension matching. Note that if $\mathbb{R}^{d_m} \subset \mathbb{R}^{d_{m'}}$ then the dimensional matching condition can be reduced to $d_m + d_{mm'} = d_{m'}$, with $(m, u) = g_{m'm}(m)$. The acceptance probability is given by $a(m, m') = \min\left(1, \frac{p_{m'm} p_{m'} f_{m'}(m')}{p_{mm'} q_{mm'}(m,u) p_m f_m(m)} \left|\det\left(\frac{\partial g_{mm'}(m,u)}{\partial(m,u)}\right)\right|\right)$, where $p_m f_m$, the posterior probability is given by $c^{-1} p(y|m, n_m) p(m|n_m) p(n_m)$ with c being the normalising constant. Many problems in data analysis require the unsupervised partitioning. Roberts, Holmes and Denison [92] re-considered the issue of data partitioning from an information-theoretic viewpoint and shown that minimisation of partition entropy may be used to evaluate the most probable set of data generators which can be employed using a RJ-MCMC.

B. Dynamical Linear Methods

Multi-Target Tracking (MTT) which deals with state space estimation of moving targets has applications in different fields [5], [64], [105], the most intuitive ones being perhaps of radar and sonar function.

1) Kalman Filter: The Kalman Filter (KF) is a mathematical tool which purpose is to make the best estimation in a Mean Square Error (MSE) sense of some dynamical process, (x_k), perturbed by some noise and influenced by a controlled process. For the sake of our project we will assume that the controlled process is null but will still incorporate it in the general state in order to fully understand the model. The estimation is done via observations which are functions of these dynamics (y_k). Roweis and Ghahramani made a quality review [94] of the topic. The dynamics of the KF is usually referred in the literature as x_k and given by equation (16).

$$x_k = F_k x_{k-1} + B_k u_k + w_k \tag{16}$$

with F_k is the state transition model which is applied to the previous state x_{k-1}; B_k is the control-input model which is applied to the vector u_k (often taken as the null vector); w_k is the process noise which is assumed to be drawn from a zero mean multivariate normal distribution with covariance

[21] See figure 6 of "Paper Format" document for more information.
[22] See in section III and Figures 6, 9, 10 and 11.

Q_k and $w_k \sim N(0, Q_k)$. At time k an observation of x_k, y_k is made according to equation (17).

$$y_k = H_k x_k + v_k \tag{17}$$

where H_k is the observation model which maps the true state space into the observed space. v_k is the observation noise which is assumed to be zero mean Gaussian white noise with $v_k \sim N(0, R_k)$. We also assume that the noise vectors $(\{w_1, \dots, w_k\}, \{v_1 \dots v_k\})$ at each step are all assumed to be mutually independent ($cov(v_k, w_k) = 0$ for all k). The KF being a recursive estimator, we only need the estimated state from the previous time step and the current measurement to compute the estimate for the current state. $\hat{x_k}$ will represent the estimation of our state x_k at time up to k. The state of our filter is represented by two variables: $\hat{x}_{k|k}$, the estimate of the state at time k given observations up to and including time k; $P_{k|k}$, the error covariance matrix (a measure of the estimated accuracy of the state estimate). The KF has two distinct phases: Predict and Update. The predict phase uses the state estimate from the previous timestep to produce an estimate of the state at the current timestep. In the update phase, measurement information at the current timestep is used to refine this prediction to arrive at a new, more accurate state estimate, again for the current timestep. The formula for the updated estimate covariance

Algorithm 4 KALMAN-FILTER(w)

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

1: **//Predicted state**:
2: $\hat{x}_{k|k-1} = F_k \hat{x}_{k-1|k-1} + B_{k-1} u_{k-1}$
3: $P_{k|k-1} = F_k P_{k-1|k-1} F_k^{\mathrm{T}} + Q_{k-1}$
4: **//Update state**:
5: //Innovation (or residual)
6: $\tilde{y}_k = y_k - H_k \hat{x}_{k|k-1}$
7: //Covariance
8: $S_k = H_k P_{k|k-1} H_k^T + R_k$
9: //Optimal Kalman gain
10: $K_k = P_{k|k-1} H_k^{\mathrm{T}} S_k^{-1}$
11: //Updated state estimate
12: $\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k \tilde{y}_k$
13: //Updated estimate covariance
14: $P_{k|k} = (I - K_k H_k) P_{k|k-1}$

above is only valid for the optimal Kalman gain. Usage of other gain values require a more complex formula. Below we present a partial proof of the KF algorithm [53], [54].

Proof: The second line of the algorithm is derived the following way: $\hat{x}_{k|k-1} = \mathrm{E}[x_k] = \mathrm{E}[F_k x_{k-1} + B_k u_k + w_k] = F_k \hat{x}_{k-1|k-1} + B_{k-1} u_{k-1}$. The third line of the algorithm is derived the following way: $P_{k|k-1} = \mathrm{E}[x_k x_k] = \mathrm{E}\left[F_k \underbrace{x_{k-1|k-1} x_{k-1|k-1}^{\mathrm{T}}}_{\mathrm{E}\left[x_{k-1|k-1} x_{k-1|k-1}^{\mathrm{T}}\right] = P_{k|k-1}} F_k^{\mathrm{T}} \right] + 2\underbrace{\mathrm{E}\left[F_k x_{k-1|k-1} B_k u_k\right]}_{0} + 2\underbrace{\mathrm{E}\left[F_k x_{k-1|k-1} w_k\right]}_{0} + 2\underbrace{\mathrm{E}\left[B_k u_k w_k\right]}_{0} + \underbrace{\mathrm{E}\left[w_k w_k^{\mathrm{T}}\right]}_{Q_k} = F_k P_{k-1|k-1} F_k^{\mathrm{T}} + Q_{k-1}$. The 8th line is derived the following way: $S_k = \mathrm{E}[y_k y_k] = \mathrm{E}\left[H_k \underbrace{x_k x_k^{\mathrm{T}}}_{\mathrm{E}\left[x_{k-1|k-1} x_{k-1|k-1}^{\mathrm{T}}\right] = P_{k|k-1}} H_k^{\mathrm{T}} \right] + 2\underbrace{\mathrm{E}\left[H_k x_k v_k\right]}_{0} + \underbrace{\mathrm{E}\left[v_k v_k\right]}_{R_k} = H_k P_{k|k-1} H_k^{\mathrm{T}} + R_k$.

As for the Kalman Gain, we first rearrange some of the equations in a more useful form. First, with the error covariance $P_{k|k}$ as above $P_{k|k} = \mathrm{cov}(x_k - \hat{x}_{k|k})$ and substitute in the definition of $\hat{x}_{k|k}$ $P_{k|k} = \mathrm{cov}(x_k - (\hat{x}_{k|k-1} + K_k \tilde{y}_k))$ and substitute $\tilde{y}_k$. $P_{k|k} = \mathrm{cov}(x_k - (\hat{x}_{k|k-1} + K_k(y_k - H_k \hat{x}_{k|k-1})))$. $P_{k|k} = \mathrm{cov}(x_k - (\hat{x}_{k|k-1} + K_k(H_k x_k + v_k - H_k \hat{x}_{k|k-1})))$, now by collecting the error vectors we get $P_{k|k} = \mathrm{cov}((I - K_k H_k)(x_k - \hat{x}_{k|k-1}) - K_k v_k)$. Given that the measurement error v_k is uncorrelated with the other terms, we have $P_{k|k} = \mathrm{cov}((I - K_k H_k)(x_k - \hat{x}_{k|k-1})) + \mathrm{cov}(K_k v_k)$, now by the properties of vector covariance this becomes $P_{k|k} = (I - K_k H_k)\mathrm{cov}(x_k - \hat{x}_{k|k-1})(I - K_k H_k)^{\mathrm{T}} + K_k \mathrm{cov}(v_k) K_k^{\mathrm{T}}$ which, using our invariance on $P_{k|k-1}$ and the definition of R_k becomes $P_{k|k} = (I - K_k H_k) P_{k|k-1} (I - K_k H_k)^{\mathrm{T}} + K_k R_k K_k^{\mathrm{T}}$. This rearrangement is known in the literature as the Joseph form of the covariance equation, which is true independently of K_k. Now if K_k is the optimal Kalman gain, we can simplify further. The Kalman filter is a minimum MSE estimator. The error is $x_k - \hat{x}_{k|k}$. We would like to minimize the expected value of the square of the magnitude of this vector, $\mathrm{E}[|x_k - \hat{x}_{k|k}|^2]$. This idea is equivalent to minimizing the trace of the posterior estimate covariance matrix $P_{k|k}$. By expanding out the terms in the equation above and rearranging, we get: $P_{k|k} = P_{k|k-1} - K_k H_k P_{k|k-1} - P_{k|k-1} H_k^{\mathrm{T}} K_k^{\mathrm{T}} + K_k (H_k P_{k|k-1} H_k^{\mathrm{T}} + R_k) K_k^{\mathrm{T}} = P_{k|k-1} - K_k H_k P_{k|k-1} - P_{k|k-1} H_k^{\mathrm{T}} K_k^{\mathrm{T}} + K_k S_k K_k^{\mathrm{T}}$. The trace is minimized when the matrix derivative is zero: $\frac{\partial \, \mathrm{tr}(P_{k|k})}{\partial \, K_k} = -2(H_k P_{k|k-1})^{\mathrm{T}} + 2 K_k S_k = 0$. Solving this for K_k yields the Kalman gain: $K_k S_k = (H_k P_{k|k-1})^{\mathrm{T}} = P_{k|k-1} H_k^{\mathrm{T}}$ $K_k = P_{k|k-1} H_k^{\mathrm{T}} S_k^{-1}$. This optimal Kalman gain, is the one that yields the best estimates when used. The formula used to calculate the posterior error covariance can be simplified when the Kalman gain equals the optimal value derived above. Multiplying both sides of our Kalman gain formula on the right by $S_k K_k^T$, it follows that $K_k S_k K_k^T = P_{k|k-1} H_k^T K_k^T$. Referring back to our expanded formula for the posterior error covariance, $P_{k|k} = P_{k|k-1} - K_k H_k P_{k|k-1} - P_{k|k-1} H_k^T K_k^T + K_k S_k K_k^T$ we find that the last two terms cancel out, giving $P_{k|k} = P_{k|k-1} - K_k H_k P_{k|k-1} = (I - K_k H_k) P_{k|k-1}$. This formula is low latency and thus usually used. One should keep in mind that it is only correct for the optimal gain though. ■

2) Extended Kalman Filter: The EKF is essentially an approximation of the KF for non-severely-non-linear models which linearises about the current mean and covariance, so that the state transition and observation models need not be linear functions of the state but may instead be differentiable functions. The dynamics and measurements of this equation is presented in (18).

$$\begin{cases} x_k &= f(x_{k-1}, u_k) + w_k \\ y_k &= h(x_k) + v_k \end{cases} \tag{18}$$

The algorithm is very similar to the one described in algorithm 4 but with couple of modifications highlighted below algorithm 5[23].

Algorithm 5 EXTENDED-KALMAN-FILTER(w)

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

1: **//Predicted state**:
2: $\hat{x}_{k|k-1} = f(\hat{x}_{k-1|k-1}, u_k)$
3: $P_{k|k-1} = F_k P_{k-1|k-1} F_k^{\mathrm{T}} + Q_{k-1}$
4: **//Update state**:
5: //Innovation (or residual)
6: $\tilde{y}_k = y_k - h(\hat{x}_{k|k-1})$
7: //Covariance
8: $S_k = H_k P_{k|k-1} H_k^{\mathrm{T}} + R_k$
9: //Optimal Kalman gain
10: $K_k = P_{k|k-1} H_k^{\mathrm{T}} S_k^{-1}$
11: //Updated state estimate
12: $\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k \tilde{y}_k$
13: //Updated estimate covariance
14: $P_{k|k} = (I - K_k H_k) P_{k|k-1}$

Proof: The proof for algorithm 5 is very similar to the proof of algorithm 4 with couple of exceptions. First, F_k and H_k are approximations of first order of F_k and H_k. Second, we get a truncation error which can be bounded and satisfies the inequality known as Cauchy's estimate: $|R_n(x)| \leqslant M_n \frac{r^{n+1}}{(n+1)!}$, here $(a-r, a+r)$ is the interval where the variable x is assumed to take its values and M_n positive real constant such that $|f^{(n+1)}(x)| \leqslant M_n$ for all $x \in (a-r, a+r)$. M_n gets bigger as the curvature or non-linearity gets more severe. When this error increases it is possible to improve our approximation at the cost of complexity by increasing by one degree our Taylor approximation, i.e: $F_k = \frac{\partial f}{\partial x}\Big|_{f(\hat{x}_{k-1|k-1}, u_k)} + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}\Big|_{f(\hat{x}_{k-1|k-1}, u_k)^2}$ and $H_k = \frac{\partial h}{\partial x}\Big|_{f(\hat{x}_{k|k-1})} + \frac{1}{2}\frac{\partial h}{\partial x}\Big|_{f(\hat{x}_{k|k-1})^2}$. ∎

Remark Though the EKF tries to address some of the limitations of the KF by relaxing some of the linearity constraints it still needs to assume that the underlying function dynamics are both known and derivable. This particular point is not at all desirable in many applications.

[23]Note that here $F_k = \frac{\partial f}{\partial x}\Big|_{\hat{x}_{k-1|k-1}, u_k}$ and $H_k = \frac{\partial h}{\partial x}\Big|_{\hat{x}_{k|k-1}}$.

C. Dynamical Non-linear methods

1) Sequential Monte Carlo methods: Sequential Monte Carlo methods (SMC) [16], [66] known alternatively as Particle Filters (PF) [35], [58] or also seldom CONDENSATION [50], 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 (19) and (20) gives an example of such system.

$$x_k = f(x_{k-1}) + w_k \tag{19}$$
$$y_k = h(x_k) + v_k \tag{20}$$

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 (19) and (20) 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 (21).

$$\int f(x_k) p(x_k|y_1, \ldots, y_k) dx_k \approx \frac{1}{N_P} \sum_{L=1}^{N_P} f(x_k^{(L)}) \tag{21}$$

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

$$\left(w_k^{(L)}, x_k^{(L)}\right) \ : \ L = 1, \ldots, N_P \tag{22}$$

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

Algorithm 6 RESAMPLE(w)

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

1: $u^0 \sim \mathcal{U}[0, 1/M]$
2: **for** $m = 1$ to N **do**
3: $\quad i^{(m)} \leftarrow \left\lfloor (w_n^{(m)} - u^{(m-1)}m) \right\rfloor + 1$
4: $\quad u^{(m)} = u^{(m)} + \frac{i^{(m)}}{M} - w_n^{(m)}$
5: **end for**

of importance sampling. Like in IS, the expectation of a function $f(\cdot)$ can be approximated like described in equation (23).

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

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 (24).

$$\pi(x_k|x_{1:k-1}, y_{1:k}) = p(x_k|x_{k-1}, y_k) \quad (24)$$

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 19 gives an illustration of the algorithm just described. Note that on line 5 of algorithm 7, $\hat{w}_k^{(L)}$, simplifies to $w_{k-1}^{(L)} p(y_k|x_k^{(L)})$, when $\pi(x_k^{(L)}|x_{1:k-1}^{(L)}, y_{1:k}) = p(x_k^{(L)}|x_{k-1}^{(L)})$. Because it is in general difficult to design a proposal distributions with the ability of approximating the posterior distribution well, a past methodology was to sample from the transition prior which the latter can fail in situation in which new measurements happen to be in the tail of the prior or if the likelihood is too peaked in comparison to the prior [109]. These kind of situations happen often in Finance since data exhibits jump like behavior. More information around this topic can be found in [87]. This naturally invited the use of the EKF and then the UKF as the proposal distribution for the PF [109].

Proposition The latter method converges.

Proof: This proof is taken from [109], [12]. Let $B(\mathbb{R}^n)$ be the space of bounded, Borel measurable functions on $\mathbb{R}^n$. We denote $||f|| \triangleq \sup_{x \in \mathbb{R}^n} |f(x)|$. If the importance weight given by $\frac{p(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{x}_{t-1})}{q(\mathbf{x}_t|\mathbf{x}_{0:t-1},)\mathbf{y}_{1:t}}$ is an upper bound for any $(\mathbf{x}_{t-1}, \mathbf{y}_t)$, then for all $t \geqslant 0$, there exists c_t independent of N, such that for any $f_t \in B(\mathbb{R}^{n_x \times (t+1)})$ we get $c_t \frac{||f_t||^2}{N_p} \geqslant \mathbb{E}\left[\left(\frac{1}{N_P}\sum_{L=1}^{N_P} f(x_k^{(L)}) - f(x_k)p(x_k|y_{1:k})dx_k\right)^2\right]$. ■

Remark Though naturally more robust and more accommodating of fatter tails it also naturally yields bigger variance.

Algorithm 7 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: $\quad$ 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: $\quad$ **for** $L = 1$ to N_P **do**
14: $\quad\quad w_k^{(L)} = 1/N_P$.
15: $\quad$ **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 [59] is optimal in terms of variance. Figure 19 gives an illustration of the Stratified Sampling and the corresponding algorithm is described in algorithm 6. We see at the top of the figure 19 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 [77] and was further discussed in several books including in [41]. 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

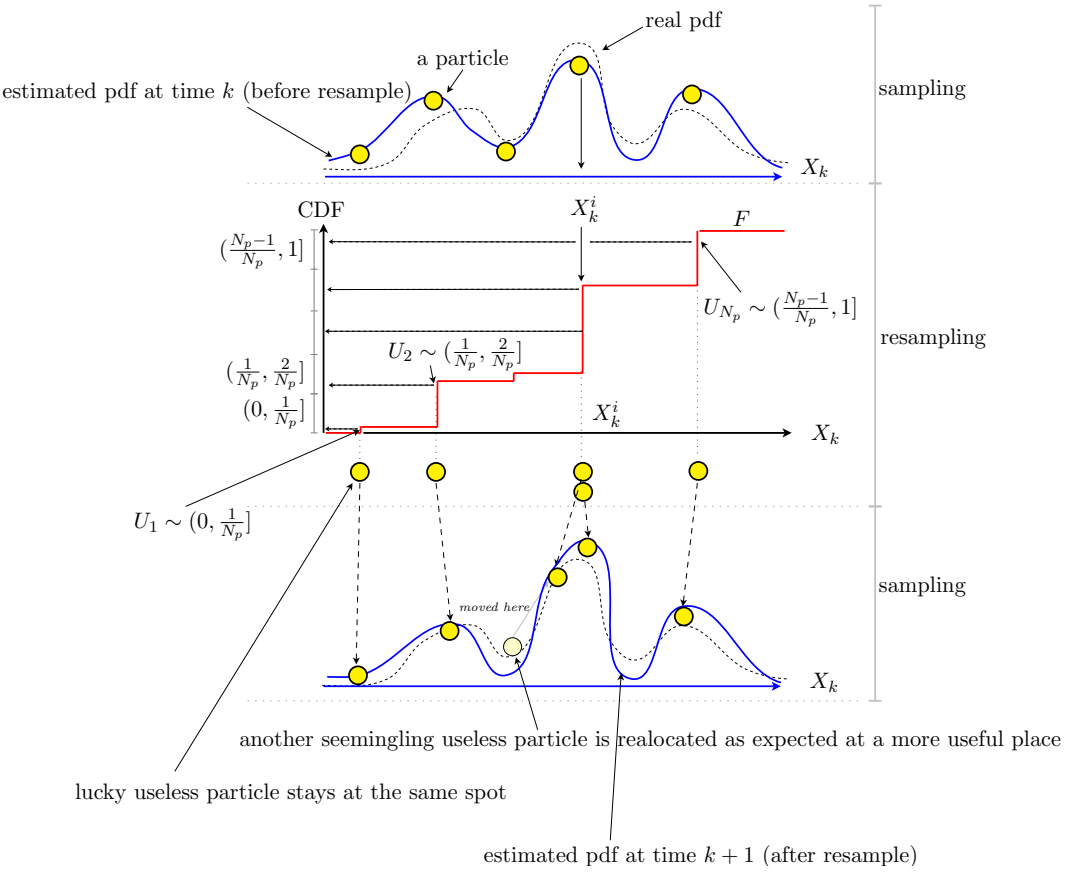

Fig. 19. Stratified Sampling illustration

of $p(x)$. In equation (25) we write the integration problem in the more appropriate form.

$$\int f(x)p(x)dx = \int f(x)\frac{p(x)}{q(x)}q(x)dx \tag{25}$$

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 (26).

$$\hat{f} = \frac{1}{N_p}\sum_{i=1}^{N_p} W\left(x^{(i)}\right) f\left(x^{(i)}\right) \tag{26}$$

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 (27)).

$$W\left(x^{(i)}\right) = \frac{p\left(x^{(i)}\right)}{q\left(x^{(i)}\right)} \tag{27}$$

If the normalizing factor for $p(x)$ is not known, the importance weights can only be evaluated up to a normalizing constant: $W\left(x^{(i)}\right) \propto p\left(x^{(i)}\right) q\left(x^{(i)}\right)$. To ensure that $\sum_{i=1}^{N_p} W(x^{(i)}) = 1$, we normalize the importance weights to obtain equation (28).

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

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 [11] in equation (28) is given:

$$\begin{aligned}
Var_q[\hat{f}] &= \frac{1}{N_p} Var_q[f(x)W(x)] \\
&= \frac{1}{N_p} Var_q[f(x)p(x)/q(x)] \\
&= \frac{1}{N_p}\int\left[\frac{f(x)p(x)}{q(x)} - \mathbb{E}_p[f(x)]\right]^2 q(x)dx \\
&= \int\left[(\frac{(f(x)p(x))^2}{q(x)}) - 2p(x)f(x)\mathbb{E}_p[f(x)]\right]dx \\
&\quad \times\frac{1}{N_p} + \frac{(\mathbb{E}_p[f(x)])^2}{N_p} \\
&= \frac{1}{N_p}\int\left[(\frac{(f(x)p(x))^2}{q(x)})\right]dx - \frac{(\mathbb{E}_p[f(x)])^2}{N_p}
\end{aligned}$$

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: $\frac{\partial Var_q[\hat{f}]}{\partial q(x)} = -\frac{1}{N_p}\int[(\frac{(f(x)p(x))^2}{q(x)^2})]dx = -\frac{1}{N_p}\int[(\frac{(f(x)p(x))^2}{q(x)q(x)})]dx$. $q(x)$ 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 $|f(x)|p(x)$. ■

D. Scenario Tracking Algorithm

1) Context: Recently, SMC methods [15], [16], [65], especially when it comes to the data association issue, have been developed. Particle Filters (PF) [35], [58], have recently become a popular framework for Multi Target Tracking (MTT), because able to perform well even when the data models are nonlinear and non-Gaussian, as opposed to linear methods used by the classical methods like the KF/EKF [42]. Given the observations and the previous target state information SMC can employ sequential importance sampling recursively and update the posterior distribution of our target state. The Probability Hypothesis Density (PHD) filter [101], [103], [75], which combines the Finite Set Statistics (FISST), an extension of Bayesian analysis to incorporate comparisons between different dimensional state-spaces, and the SMC methods, was also proposed for joint target detection and estimation [82]. The M-best feasible solutions is also a new useful finding in SMC [82], [60], [7], [62], [8]. Articles [102], [104] were proposed to cope with both the multitarget detection and tracking scenario but according to Ng, Li, Godsill, and Vermaak [81] they are not robust if the environment becomes more noisy and hostile, such as having a higher clutter density and a low probability of target detection. To cope with these problems a hybrid approach and it extensions [81] were implemented. The aim of these methods is to stochastically estimate the number of targets and therefore the multitarget state. The soft-gating approach described in [83] is an attempt to address the complex measurement-to-target association problem. To solve this issue of detection in the presence of spurious objects a new SMC algorithm is presented in [63]. That method provided a solution to deal with both time-varying number of targets, and measurement-to-target association issues.

2) Time-varying number of targets & measurement-to-target association: Currently, tracking for multiple targets has a couple of major challenges that are yet to be answered efficiently. The first of these two main challenges is the modelling of the time-varying number of targets in an environment high in clutter density and low in detection probability (hostile environment). To some extend the PHD filter [76], [102], [104], based on the FISST , has proved ability in dealing with this problem with unfortunately a significant degradation of its performance when the environment is hostile [81]. The second main challenge is the measurement-to-target association problem. Because there is an ambiguity between whether the observation consists of measurements originating from a true targets or a clutter point, it becomes obviously essential to identify which one is which. The typical and popular approach to solve this issue is the Joint Probabilistic Data Association (JPDA) [5], [25]. Its major drawback though is that its tracks tend to coalesce when targets are closely spaced [22] or intertwined. This problem has been, however, partially studied. Indeed the sensitivity of the track coalescence may be reduced if we use a hypothesis pruning strategy [9], [43]. Unfortunately the track swap problems still remain. Also performance of the EKF [42] is known to be limited by the linearity of the data model on the contrary to SMC based tracking algorithms developed by [48], [38], [37], [46]. This issue of data association can also be sampled via Gibbs sampling [46]. Also because target detection and initialization were not covered by this framework algorithms developed in [110] were suggested in order to improve detection and tracking performance. The algorithm suggested in [110] combines a deterministic clustering algorithm for the target detection issue. This clustering algorithm enabled to detect the number of targets by continuously monitoring the changes in the regions of interest where the moving targets are most likely located. Another approach in [95] combines the track-before-detect (TBD) and the SMC methods to perform joint target detection and estimation, where the observation noise is Rayleigh distributed but, according to [95], this algorithm is currently applicable only to single target scenario. Solutions to the data association problem arising in unlabelled measurements in a hostile environment and the curse of dimensionality arising because of the increased size of the state-space associated with multiple targets were given in [110]. In [110], a couple of extensions to the standard known particle filtering methodology for MTT was presented. The first extension was referred to as the Sequential Sampling Particle Filter (SSPF), sampled each target sequentially by using a factorisation of the importance weights. The second extension was referred by the Independent Partition Particle Filter (IPPF), makes the hypothesis that the associations are independent. Real world MTT problems are usually made more difficult because of couple of main issues. First realistic models have usually a very non-linear and non-Gaussian target dynamics and measurement processes therefore no closed-form expression can be derived for the tracking recursions. The most famous closed form recursion leads to the KF [2] and arises when both the dynamic and the likelihood model are chosen to be linear and Gaussian. The second issue with real world problem is due to the poor sensors targets measurements labeling which leads to a combinatorial data association problem that is challenging in a hostile environment. The complexity of the data association problem may be enhanced by the increase in probability of clutter measurements in lieu of a target in areas rich in multi-path effects. We have seen that the KF is limited in modeling non linearity because of its linear properties but it is still an interesting tool as an approximation mean like it has been done with the EKF [2] which capitalizes on linearity around the current state in non linear models. Logically the performance of the EKF decreases as the non-linearity increases. The Unscented Kalman Filter (UKF) [52] was created to answer this problem. The method maintains the second order statistics of the target distribution by recursively propagating a set of carefully selected sigma points. The advantage of this method is that it does not require linearisation as well as usually yields more robust estimates. Models with non-Gaussian state and/or observation noise were initially studied and partially solved by the Gaussian Sum Filter (GSF) [1] . That method approximates the non-Gaussian

target distribution with a mixture of Gaussians but suffers when linear approximations are required similarly to the EKF. Also, over time we experience a combinatorial growth in the number of mixture components which ultimately leads to eliminate branches to keep control of an exponential explosion as iterations go forward. Another option that does not require any linear approximations like it is the case with the EKF or the GSF was proposed [57]. In this case the non-Gaussian state is approximated numerically with a fixed grid, using Bayes' rule, the prediction step is integrated numerically. Unfortunately because the computational cost of the integration explodes with the dimension of the state-space the method becomes useless for dimensions larger than four [110]. For non-linear and non-Gaussian models, generally speaking SMC's have become popular user friendly numerical techniques that approximate Bayesian recursions for MTT. Its popularity is mainly due to flexibility, relative simplicity as well as efficiency. The method models the posterior distribution with a set of particles with an associated weights more or less big relative to the particle importance and are propagated and adjusted throughout iterations. The very big advantage with SMC method is that the computational complexity does not become exorbitant with an increase in the dimension of the state-space [57]. It has been defined in [110] that there exists numerous strategies available to solve the data association problem but they could be categorized as either single frame assignment methods, or multi-frame assignment methods. The multi-frame assignment problem can be solved using Lagrangian relaxation [89]. Another algorithm the Multiple Hypotheses Tracker (MHT) [90] tries to keep track of all the possible association hypotheses over time which makes it awkward as the number of associations hypotheses grows exponentially with each iteration.

3) The problem of pruning: The Nearest Neighbor Standard Filter (NNSF) [5] links each target with the closest measurement in the target space. This simplistic method has the flaws that one may assume it has, that is the method suppresses many feasible hypotheses. The Joint Probabilistic Data Association Filter (JPDAF) [5], [25] is more interesting in this respect as it does not do as much pruning or pruning only infeasible hypotheses. The parallel filtering algorithm goes through the remaining hypotheses and adjusts the corresponding posterior distribution. Its principal deficiency is that the final estimate looses information because, to maintain tractability, the corresponding estimate is distorted to a single Gaussian. This problem however has been identified and strategies have been suggested to address this shortcoming. For example [86], [96] proposed strategies to instead reduce the number of mixture components in the original mixture to a tractable level. This algorithm unfortunately only partially solved the problem as many feasible hypotheses may still be pruned away. The Probabilistic Multiple Hypotheses Tracker (PMHT) [32], [106] takes as hypothesis that the association variables to be independent and avoids the problems of reducing our state space. This leads to an incomplete data problem that, however may be solved using the Expectation Maximisation (EM) algorithm [14]. Unfortunately the PMHT is not suitable for sequential applications because considered a batch strategy. Moreover [112] has shown that the JPDA filter outperforms the PMHT and we have seen earlier the shortcomings of the JPDAF. Recently strategies have been proposed to combine the JPDAF with particle techniques to address the general non-linear and non-Gaussian models [100], [99], [26], [55] issue of approximation of linearity failing when the dynamic of measurement functions are severly non-linear. The feasibility of multi-target tracking with SMC has first been described in [3], [36] but the simulations dealt only with a single target. In the article [47] the distribution and the hypotheses of the association is computed using a Gibbs sampler, [33] at each iterations. This method, similar to the one described in [13], uses MCMC [34] to compute the associations between image points within the framework of stereo reconstruction. Because they are iterative in nature and take an unknown number of iterations to converge. These MCMC strategies though, are not always suitable for online applications. Doucet [37] presents a method where the associations are sampled from a well chosen importance distribution. Although intuitivly appealing it is, however, reserved to Jump Markov Linear Systems (JMLS) [17]. The follow up of this strategy, based on the UKF and the Auxiliary Particle Filter (APF) [87], so that applicable to Jump Markov Systems (JMS) is presented in [18]. Similar in [48], particles of the association hypotheses are generated via an optimal proposal distribution. SMC have also been applied to the problem of MTT based on raw measurements [10], [97]. We have seen that the MTT algorithms suffers from exponential explosion that is as the number of targets increases, the size of our state spaces increases exponentially. Because pruning is not always efficient it may commonly occur that particles contain a mixture of good estimates for some target states, and bad estimates for other target states. This problem has been first acknowledged in [85], and where a selection strategy is addressed to solve this problem. In [110] a number of particle filter based strategies for MTT and data association for general non-linear and non-Gaussian models is presented. The first, is referred to as the Monte Carlo Joint Probabilistic Data Association Filter (MC-JPDAF) and presented by the authors as a generalization of the strategy proposed in [100], [99] to multiple observers and arbitrary proposal distributions. Two extensions to the standard particle filtering methodology for MTT is developed. The first strategy is presented by the authors as an exact methodology that samples the individual targets sequentially by utilizing a factorization of the importance weights, called the Sequential Sampling Particle Filter (SSPF). The second strategy presented in [110] assumes the associations to be independent over the individual target, similar to the approximation made in the PMHT, and implies that measurements can be assigned to more than one target. This assumption claims that it effectively removes all dependencies between the individual targets, leading to an efficient component-wise sampling strategy to construct new particles. This approach was named Independent Partition Particle Filter (IPPF). Their main benefit is that as opposed to the JPDAF, neither

approach requires a gating procedure like in [48].

VI. Generating the Implied Volatility Scenarios

A. Overview

We have seen in section III the IVP as a function of its risk factors, in section IV the conditions which make a volatility surface coherent and in section V the classic multi target tracking methodologies. We will see that these sections are very useful in helping us define the core components of our SMCM that is the formalization of the likelihood function in subsection VI-C and the sampling processes in subsection VI-D. Prior to going through these subsections we will first go through few definitions in subsection VI-B that will help us navigate through the formalization.

B. Pillar Normalization and Few Definitions

Definition Let $\mathbb{T} = \{t_0, t_1 \dots, t_N\}$ be the ordered set of arrival times such that $t_i < t_{i+1}$.

Remark t_N represents the most recent time-stamp.

Definition Let $\mathbf{w} = \{w_0, w_1 \dots, w_N\}$ be the set of weights associated to our arrival times process such that:

$$w_i = \lambda w_{i+1},\ 0 < \lambda < 1 \tag{29}$$

Definition We call $\lambda^{d,\tau}$ the λ weight defined above associated to the time the implied volatility d and tenor τ, $\sigma_{d,\tau}$, arrived in our dataset.

Definition We will call $\hat{\sigma}_t(\vartheta, K)$ the linear interpolation in variance space of the implied volatility. Equation (30) gives its formula.

$$\begin{aligned}\hat{\sigma}_t^2(\vartheta, C_j^k) =& \frac{(C_{i+1}^\tau - \vartheta)\sigma_t^2(C_{i+1}^\tau, C_j^k)}{C_{i+1}^\tau - C_i^\tau} \\ &+ \frac{(\vartheta - C_i^\tau)\sigma_t^2(C_i^\tau, C_j^k)}{C_{i+1}^\tau - C_i^\tau}\end{aligned} \tag{30}$$

where $\vartheta \in [C_i^\tau, C_{i+1}^\tau]$ and $1 < i < |C^\tau|$.

Remark The above definition does not include a definition of the edges of our IVS and also assume a perfect interpolation and extrapolation methodology already exist on the strike space .

As we have seen and illustrated by figure I-B.1, when there is no roll, how do we populate the red zones without creating spurious jumps? For this specific exercise we need to create a proxy. If we call Υ the stopping[24] time at which the contract rolls as in equation (31)

$$\Upsilon = \inf\{t | \sigma_t^2(C_1^\tau, C_j^k) = \sigma_t^2(C_1^d, C_j^k)\} \tag{31}$$

then the longest tenor proxy can be better approximated by equation (32a) and the shortest by equation (32b):

$$\hat{\sigma}_t^2(\vartheta, C_j^k) = \frac{\sigma_\Upsilon^2(C_{i+1}^\tau, C_j^k)}{\sigma_\Upsilon^2(C_i^\tau, C_j^k)} \hat{\sigma}_t^2(C_i^\tau, C_j^k) \tag{32a}$$

$$\hat{\sigma}_t^2(\vartheta, C_j^k) = \frac{\sigma_\Upsilon^2(C_i^\tau, C_j^k)}{\sigma_\Upsilon^2(C_{i+1}^\tau, C_j^k)} \hat{\sigma}_t^2(C_{i+1}^\tau, C_j^k) \tag{32b}$$

with $i = |C^\tau| - 1$ for equation (32a) and $i = 1$ for equation 32b.

Definition Without loss of generality we will refrain from using the various symbolics defined in this subsection and will assume, unless otherwise specified, that throughout the paper the implied volatility tenors are the one of the normalized tenors as opposed to specific contracts.

C. Bespoke Likelihood Function

In order to define our likelihood function we must first define the arrival and the weighting process associated to the arrival time. We must also rank the risk factors from most likely to least likely. There are several methods that can be used to perform the latter task, we recommend a stepwise regression [45] as it is simple enough but recognize that other methods potentially better, may be used as well [23]. Let us call:

- λ_0 r-squared contribution of a pointwise change of the IVS,
- $\lambda^1 = \{\lambda_a, \lambda_b, \lambda_\rho, \lambda_m, \lambda_\sigma, \lambda_\beta\}$ the set of ranked r-squared contribution of each of the 6 parameters $\in \chi_t^i = \{a_t^i, b_t^i, \rho_t^i, m_t^i, \sigma_t^i, \beta_t^i\}$ where $i \in C^\tau$ representing the mid and liquidity parameters of each pillar tenors.
- $\lambda^2 = \{\lambda_{(a,b)}, \lambda_{(a,\rho)}, \dots, \lambda_{(m,\beta)}\}$ the set of unique pairwise parameter changes but there really is one set of pairwise parameters that interests us (the combine change in skew and vol of vol),
- the subset of $\lambda^1 \cup \lambda_{(b,\rho)}$ of all accepted scenarios:

$$\lambda = \{\lambda_0, \lambda_a, \lambda_b, \lambda_\rho, \lambda_m, \lambda_\sigma, \lambda_\beta, \lambda_{(b,\rho)}\} \tag{33}$$

- $\mathcal{H}(.)$ the function taking a uniform random variable $u \sim U[0,1]$ and returning the set of parameter(s) associated to λ: see algorithm (8).
- N_p^t the number of particles at time t but the number of particles only change as a function of the rounding functions used. It always mean reverts towards N_p so we will chose N_p instead for convenience sake.

D. Bespoke Sampling Algorithm

1) General Idea: When information about a specific point, $\sigma_{d,\tau}$, of the IVS has arrived we can assume that this change:

- is isolated ***b**ut this specific point **p**ropagates arbitrage free **c**onstraints on the IVS which would otherwise remain constant* (BPC)[25],
- corresponds to a change in "vol of vol" (the b parameter in the IVP model introduced in section III more specifically figure 6) BPC,
- corresponds to a change in "skew" (the ρ parameter in the IVP model introduced in section III more specifically figure 7) BPC,
- corresponds to an individual change of any other of IVP parameter introduced in section III,

[24] to take on probabilistic jargon

[25] We substitute this sentence in italic by BPC from now on

Algorithm 8 $\mathcal{H}(\lambda)$

Require: λ

Ensure: a set of parameter(s) is returned

$u \sim \mathcal{U}\left[0, \lambda_0 + \lambda_a + \lambda_b + \lambda_\rho + \lambda_m + \lambda_\sigma + \lambda_\beta + \lambda_{(b,\rho)}\right]$

if $0 \leqslant u < \lambda_0$ **then**
    return (0)
else if $\lambda_0 \leqslant u < \lambda_a$ **then**
    return (a)
else if $\lambda_a \leqslant u < \lambda_b$ **then**
    return (b)
else if $\lambda_b \leqslant u < \lambda_\rho$ **then**
    return (ρ)
else if $\lambda_\rho \leqslant u < \lambda_m$ **then**
    return (m)
else if $\lambda_m \leqslant u < \lambda_\sigma$ **then**
    return (σ)
else if $\lambda_\sigma \leqslant u < \lambda_\beta$ **then**
    return (β)
else if $\lambda_\beta \leqslant u < \lambda_{b,\rho}$ **then**
    return (b, ρ)
else
    return ERROR
end if

- corresponds to a multiple change of any of the IVP parameters introduced in section III: more specifically "spot-vol".

The reason why mapping onto the IVP parameter is a good idea is because the parameters of the IVP not only fit better than any model the market observed prices but also its parameters map to easily understandable economical risk factor like we have seen before. The sampling methodology will consist of changing each parameters of the 3 sets of types of sampling:

- the first type of sampling is one in which we only move that new point that just arrived. We call this **Sample 0P**[26].
- the second type will consists of sampling 1 parameter of the IVP in order to explain the new data to map the change of price by a change of economical climate. For example, the ATM is the point which arrived the most recently, then our PF will assume with one of its scenarios that this is due of a change in Vol of Vol (and therefore) all the point of the implied volatility should be adjusted accordingly. We will call this sampling **Sample 1P**.
- the third type will consists of assuming that the point change is the result of two economical factor happening at the same time. For example, imagine you are assigned the task to mark an in the money call for wheat and the economical climate is that there are political tension with Russia[27] (therefore vol of vol increases) and that at the same time we have information that there are possible droughts that are incoming (there is a change of skew). This leads our PF to assume with one of its scenarios that this is due to a change in Vol of Vol and Skew at the same time and that all the point of the implied volatility should be adjusted accordingly. We will call this sampling **Sample 2P**.

Remark Note that we could do the same methodology with:

- Three or more of the parameters but usually higher order greeks beyond the second-order are considered negligible on the market and therefore it is not worth adding complexity as the ratio of the benefits over the latter does not invite such extension. However, the motivated student may wish to apply higher order sampling as an exercise.
- we could very well imagine scenarios in which the parameters of different tenors react together at the same extent but we thought that the de-arbitraging methodology would partially take care of this specific additional type of sampling and we also thought that the methodology was complex enough the way it is currently proposed. We may address that specific limitation in a subsequent paper if needed.

2) Formalization: Let's call the set of IVP parameters which best fit our IVS information by:

$$\chi_t^i = \{\chi_t^{M,i} \cup \chi_t^{L,i}\} \tag{34}$$

where $i \in C^\tau$, $\chi_t^i = \{a_t^i, b_t^i, \rho_t^i, m_t^i, \sigma_t^i, \beta_t^i\}$ and $\chi_t^{M,i} = \{\psi_{0,t}^i, \alpha_{0,t}^i, \eta_{\psi,t}^i, \eta_{\alpha,t}^i\}$ representing the mid and liquidity parameters of each pillar tenors. In addition we will call $\dot{\sigma}(p,q)$ (where $q \in C^k$ and $p \in C^\tau$) the most recent data and the array of weights w_p^N. We sample our scenarios according to the following optimization by constraints is which $x_k^{(i)}$ of algorithm (7) is replaced by:

$$\begin{aligned}
&x_k^{(i)} = \underset{\tilde{\sigma}_t(\tau,d)}{\arg\min} \textstyle\sum_\tau \sum_d [C(\sigma_t(\tau,d)) - C(\tilde{\sigma}_{\mathcal{H}(\lambda),t}(\tau,d))]^2 \\
&\text{subject to:} \\
&\forall d \in C^k,\ C_1^{\mathcal{B},d,\tau} < C_2^{\mathcal{B},d,\tau} \\
&\forall \tau \in C^\tau,\ C_1^{\mathcal{C},d,\tau} < C_1^{\mathcal{C},d,\tau} \\
&\hat{\sigma}(p,q) = \dot{\sigma}(p,q) \\
&1_{|a_t^i - \tilde{a}_t^i|>0} + 1_{|b_t^i - \tilde{b}_t^i|>0} + 1_{|\rho_t^i - \tilde{\rho}_t^i|>0} + 1_{|m_t^i - \tilde{m}_t^i|>0} + \\
&1_{|\sigma_t^i - \tilde{\sigma}_t^i|>0} + 1_{|\beta_t^i - \tilde{\beta}_t^i|>0} = \text{card}(\mathcal{H}(\lambda))
\end{aligned}$$

where $\text{card}(\mathcal{H}(\lambda))$ represents the cardinality of our hash function.

Remark Note that Sample0P will not return any solution if the point updated induces arbitrages on the IVS. It also implies that IVS will have to be saved in grid form instead of parametric form.

[26]Note that we could have called this one Sample 1P with the "P" meaning "point" but this could be also interpreted as "parameter" which we use for the second type of sampling. Therefore in order to limit confusion we preferred calling it Sample 0P for 0 parameters

[27]Russia is one of the first exporters of wheat

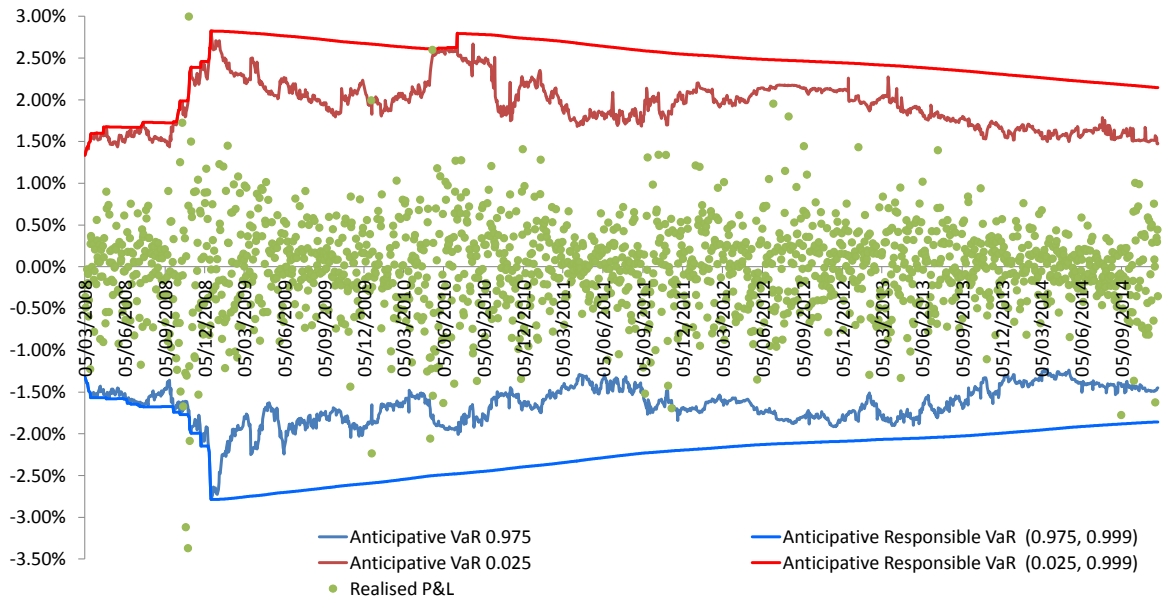

Fig. 20. USD/EUR 2 years expiry straddle strategy backtest under Anticipative Responsible VaR of with $\lambda = 0.999$.

VII. Backtesting

In terms of data, it would be good to have a time series of the implied volatility surface (IVS) in grid format as well as the specific points which led to the update of the IVS of the time series. We are ideally interested here to see how the non visible points are updated in the internal system as a result of a single visible additional information. This kind of data is unfortunately not available on the market and this for a good reason. Indeed, the market becomes visible upon the exchange of a contract or a series of contracts. For example, let us assume that we are long a 2 year expiry straddle on the USD/EUR, we only know about the change of the volatility a posteriori and through the observed price of the latter straddle. We can for example use our scenario based particle filter for risk purposes. More specifically, we can point to our related study [73] in which we also express the Risk mirror problem stress scenario generation in terms of a clustering problem and also introduce the concept of *Responsible VaR*, responsive on the upside and stable on the downside. We point to our results and its discussion in the original paper [73]. Figure 20 represent a backtest we have performed using our methodology with a $\lambda = 0.999$.

VIII. Conclusion

A. Summary

We first discussed the science of fetching the raw available sparse data from the markets in section II. In section III we explore the volatility surface risk factors as a premise, with the section IV on arbitrage constraints, to the resampling methodology that is needed in the framework of particle filter methods for which we have also done a literature review in section V. We finally discuss, in section VI the methodology which objective is to generate the implied volatility scenarios.

B. Future Research

We have raised several limitation to our current model. First we have limited our model to a span of stress, $i = \{(0), (a), (b), (\rho), (m), (\sigma), (\beta), (b, \rho)\}$, which most complex co-movement (b, ρ) is limited to a single pair of factors out of the 15 theoretical ones possible. The model does not even explore movements with 3, 4, 5 or 6 risk factors. It would seem unrealistic to explore all the possibilities but it seems equally plausible that more scenarios could be included.

References

[1] D.L. Alspach and H.W. Sorenson. Nonlinear bayesian estimation using gaussian sum approximation. *IEEE Transactions on Automatic Control*, 17(4):439–448, 1972.

[2] and Anderson, B.D.O. Optimal filtering. *Englewood Cliffs: Prentice-Hall*, 1979.

[3] D. Avitzour, W. Burgard, and D. Fox. Stochastic simulation bayesian approach to multitarget tracking. *IEE Proceedings on Radar and Sonar Navigation*, 142(2):41–44, 1995.

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

[5] Y. Bar-Shalom and T. E. Fortmann. Tracking and data association. *Academic Press*, 1988.

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

[7] D.P. Bertsekas. Auction algorithms for network flow problems: A tutorial introduction. *Computational Optimization and Applications*, 1:7–66, 1992.

[8] S. Blackman and R. Popoli. Design and analysis of modern tracking systems. *Norwood, MA: Artech House*, 1999.

[9] H.A.P. Blom and E.A. Bloem. Probabilistic data association avoiding track coalescence. *IEEE Tr. Automatic Control*, 45:247–259, 2000.

[10] Y. Boers, J.N. Driessen, F. Verschure, W. P. M. H. Heemels, and A. Juloski. A multi target track before detect application. *in Proceedings of the IEEE Workshop on Multi-Object Tracking*, 2003.

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

[12] D. Crisan and TITLE = Convergence of generalized particle filters journal = Technical Report, Cambridge University Engineering Department year = 2000 Doucet, A.

[13] F. Dellaert, S.M. Seitz, C. Thorpe, and S. Thrun. Em, mcmc, and chain flipping for structure from motion with unknown correspondence. *Machine Learning*, 50(1-2):45–71, 2003.

[14] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society, Series B*, 39(1):1–38, 1977.

[15] A. Doucet, N. de Freitas, and N. Gordon. Sequential monte carlo in practice. *New York: Springer-Verlag*, 2001.

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

[17] A. Doucet, N. Gordon, and V. Krishnamurthy. Particle filters for state estimation of jump markov linear systems. *IEEE Transactions on Signal Processing*, 49(3):613–624, 2001.

[18] A. Doucet, B. Vo, C. Andrieu, and M. Davy. Particle filtering for multi-target tracking and sensor management. *in Proceedings of the 5th International Conference on Information Fusion*, 2002.

[19] Simon Duane, A.D. Kennedy, Brian J. Pendleton, and Duncan Roweth. Hybrid monte carlo. *Physics Letters B*, 195(2):216–222, 1987.

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

[21] 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.

[22] R.J. Fitzgerald. Track biases and coalescence with probabilistic data association. *IEEE Tr. Aerospace*, 21:822–825, 1985.

[23] Lynda Flom and David L. Cassell. Stopping stepwise: Why stepwise and similar selection methods are bad, and what you should use. 01 2007.

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

[25] T. Fortmann, Y. Bar-Shalom, and Scheffe M. Sonar tracking of multiple targets using joint probabilistic data association. *IEEE Journal of Oceanic Engineering*, 8:173–184, 1983.

[26] O. Frank, J. Nieto, J. Guivant, and S. Scheding. Multiple target tracking using sequential monte carlo methods and statistical data association. *in Proceedings of the IEEE / RSJ International Conference on Intelligent Robots and Systems*, 2003.

[27] Garman and Kohlhagen. 1983.

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

[29] Jim Gatheral. The volatility surface.

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

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

[32] H. Gauvrit, J-P. Le Cadre, and C. Jauffret. A formulation of multitarget tracking as an incomplete data problem. *IEEE Transactions on Aerospace and Electronic Systems*, 33(4):1242–1257, 1997.

[33] S. Geman and D. Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6(6):721–741, 1984.

[34] W.R. Gilks, S. Richardson, and D.J. Spiegelhalter. Markov chain monte carlo in practice. *Chapman and Hall*, 1996.

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

[36] N. J. Gordon. A hybrid bootstrap filter for target tracking in clutter. *IEEE Transactions on Aerospace and Electronic Systems*, 33(1):353–358, 1997.

[37] N.J. Gordon and A. Doucet. Sequential monte carlo for maneuvring target tracking in clutter. *Proceedings SPIE*, pages 493–500, 1999.

[38] N.J. Gordon, D.J. Salmond, and D. Fisher. Bayesian target tracking after group pattern distortion. *O. E. Drummond, editor, Signal and Data Processing of Small Targets, SPIE 3163*, pages 238–248, 1997.

[39] Peter J. Green. Reversible jump markov chain monte carlo computation and bayesian model determination. *Biometrika*, 82(4):711–732, 1995.

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

[41] 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.

[42] S. Haykin. Adaptive filter theory. *Englewood Cliffs, NJ:Prentice Hall*, 4, 2000.

[43] A.P. Henk, H.A.P. Blom, and E.A. Bloem. Joint imm and coupled pda to track closely spaced targets and to avoid track coalescence. *Proceedings of the Seventh International Conference on Information Fusion*, 1:130–137, 2004.

[44] 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.

[45] R. R. Hocking. A biometrics invited paper. the analysis and selection of variables in linear regression. *Biometrics*, 32(1):1–49, 1976.

[46] C. Hue, J-P. Le Cadre, and P. Perez. Sequential monte carlo methods for multiple target tracking and data fusion. *IEEE Tr. Signal Processing*, Feb(50):309–325, 2002.

[47] C. Hue, J-P. Le Cadre, and P. Perez. Tracking multiple objects with particle filtering. *IEEE Transactions on Aerospace and Electronic Systems*, 3(38):791–812, 2002.

[48] N. Ikoma and S.J. Godsill. Extended object tracking with unknown association, missing observations, and clutter using particle filters. *Proceedings of the 2003 IEEE Workshop on Statistical Signal Processing*, September:485–488, 2003.

[49] Dan Immergluck. From the subprime to the exotic: Excessive mortgage market risk and foreclosures. *Journal of the American Planning Association*, 74(1):59–76, 2008.

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

[51] Timothy C. Johnson. Finance and mathematics: Where is the ethical malaise? *The Mathematical Intelligencer*, 37(4):8–11, Dec 2015.

[52] S.J. Julier and J. K. Uhlmann. A new extension of the kalman filter to nonlinear systems. *Proceedings of AeroSense: The 11th International Symposium on Aerospace / Defence Sensing, Simulation and Controls, vol. Multi Sensor Fusion, Tracking and Resource Management II*, 1997.

[53] R. E. Kalman. A new approach to linear filtering and prediction problems. *Transactions of the ASME - Journal of Basic Engineering*, 82:35–45, 1960.

[54] R. E. Kalman and Bucy R. S. New results in linear filtering and prediction theory. *Transactions of the ASME - Journal of Basic Engineering*, 83:95–107, 1961.

[55] R. Karlsson and F. Gustafsson. Monte carlo data association for multiple target tracking. *in Proceedings of the IEE Seminar Target Tracking: Algorithms and Applications*, page 13/113/5, 2001.

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

[57] G. Kitagawa. Non-gaussian state-space modeling of nonstationary time series (with discussion). *Journal of the American Statistical Association*, 82(400):1032–1063, 1987.

[58] 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.

[59] 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.

[60] J. Larsen and I. Pedersen. Experiments with the auction algorithm for the shortest path problem. *Nordic Journal of Computing*, 6(4):403–421, 1998.

[61] Clara Leonard. Aprs la crise, lenseignement de la finance repens. Sep 2013.

[62] R.D. Leone and D. Pretolani. Auction algorithms for shortest hyperpath problems. *Society for Industrial and Applied Mathematics*, 11(1):149–159, 2001.

[63] Jack Li, William Ng, Simon Godsill, and Jaco Vermaak. Online multitarget detection and tracking using sequential monte carlo methods. *Department of Engineering University of Cambridge*.

[64] X.R. Li, Y. Bar-Shalom, and T. Kirubarajan. Estimation, tracking and navigation: Theory, algorithms and software. *New York: John Wiley and Sons*, June, 2001.

[65] J. Liu and R. Chen. Sequential monte carlo methods for dynamic systems. *Journal of the American Statistical Association*.

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

[67] Jun S. Liu, Faming Liang, and Wing Hung Wong. The multiple-try method and local optimization in metropolis sampling. *Journal of the American Statistical Association*, 95(449):121–134, 2000.

[68] David MacKay. Information theory, inference, and learning algorithms. *Cambridge University Press*, 2003.

[69] Babak Mahdavi-Damghani. UTOPE-ia. pages 28–37, 2012.

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

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

[72] Babak Mahdavi-Damghani, Konul Mustafayeva, Cristin Buescu, and Stephen Roberts. *working paper*, 2017.

[73] Babak Mahdavi-Damghani and Stephen Roberts. *working paper*, 2017.

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

[75] R. P. S. Mahler. Multitarget bayes filtering via firstorder multitarget moments. *IEEE Transactions on Aerospace and Electronic Systems*, 39(4), 2003.

[76] R.P.S. Mahler. Multitarget bayes filtering via first-order multitarget moments. *IEEE Transactions on Aerospace and Electronic Systems*, 39(4), 2003.

[77] 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.

[78] Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, and Augusta H. Teller. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087, 1953.

[79] David D. L. Minh and Do Le (Paul) Minh. Understanding the hastings algorithm. *Communications in Statistics - Simulation and Computation*, 44(2):332–349, 2015.

[80] Radford M. Neal. Suppressing random walks in markov chain monte carlo using ordered overrelaxation. *University of Toronto, Department of Statistics*, 1995.

[81] W. Ng, J. Li, S. Godsill, and J. Vermaak. A hybrid approach for online joint detection and tracking for multiple targets. *University of Cambridge, Department of Engineering*, 2005.

[82] William Ng, Jack Li, Simon Godsill, and Jaco Vermaak. A hybrid approach for online joint detection and tracking for multiple targets. *IEEEAC paper,Department of Engineering, University of Cambridge*, 1, 2004.

[83] William Ng, Jack Li, Simon Godsill, and Jaco Vermaak. Multiple target tracking using a new soft-gating approach and sequential monte carlo methods. *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, 4:1049–1052, 2005.

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

[85] M. Orton and W. Fitzgerald. A bayesian approach to tracking multiple targets using sensor arrays and particle filters. *IEEE Transactions on Signal Processing*, 50(2):216–223, 2002.

[86] L.Y. Pao. Multisensor multitarget mixture reduction algorithms for target tracking. *AIAA Journal of Guidance, Control and Dynamics*, 17, 1994.

[87] M. K. Pitt and N. Shephard. Filtering via simulation: Auxiliary particle filter. *Journal of the American Statistical Association*, 94:590–599, 1999.

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

[89] R.L. Popp, T. Kirubarajan, and K.R. Pattipati. Survey of assignment techniques for multitarget tracking. *in Multitarget/Multisensor Tracking: Applications and Advances III, Y. Bar-Shalom and W. D. Blair, Eds. Artech House,*, 2000.

[90] D. Read. An algorithm for tracking multiple targets. *IEEE Transactions on Automation and Control*, 24(6):84–90, 1979.

[91] S. Roberts and TITLE = Sequential Sampling of Gaussian Latent Variable Models journal = Technical Report, University of Oxford year = 2018 Tegner, M.

[92] S. J. Roberts, C. Holmes, and D. Denison. Minimum-entropy data partitioning using reversible jump markov chain monte carlo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(8):909–914, Aug 2001.

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

[94] S. Roweis and Z. Ghahramani. A unifying review of linear gaussian models. *Neural Comput*, 11(2):305–345, 1999.

[95] M.G. Rutten, N.J. Gordon, and S. Maskell. Particle-based track-before-detect in rayleigh noise. *Proceedings of SPIE Conference on Signal Processing of Small Targets*, 2004.

[96] D.J. Salmond. Mixture reduction algorithms for target tracking in clutter. *in Signal and Data Processing of Small Targets, SPIE 1305, O. E. Drummond, Ed.*, pages 434–445, 1990.

[97] D.J. Salmond and H. Birch. A particle filter for track-before-detect. *in Proceedings of the American Control Conference*, pages 3755–3760, 2001.

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

[99] D. Schulz, W. Burgard, and D. Fox. People tracking with mobile robots using sample-based joint probabilistic data association filters. *in Proceedings of the IEEE International Conference on Robotics and Automation*, 22(2), 2003.

[100] D. Schulz, W. Burgard, D. Fox, and A.B. Cremers. Tracking multiple moving targets with a mobile robot using particle filters and statistical data association. *in Proceedings of the IEEE International Conference on Robotics and Automation*, pages 1665–1670, 2001.

[101] H. Sidenbladh. Multi-target particle filtering for the probability hypothesis density. *Proceedings 6th International Conference on Information Fusion*, pages 800–806, 2003.

[102] H. Sidenbladh. Multi-target particle filtering for the probability hypothesis density. *Proceedings 6th International Conference on Information Fusion*, pages 800–806, 2003.

[103] H. Sidenbladh and S.L. Wirkander. Particle filtering for finite random sets. *IEEE Transactions on Aerospace and Electronic Systems*, 2003.

[104] H. Sidenbladh and S.L. Wirkander. Particle filtering for finite random sets. *IEEE Transactions on Aerospace and Electronic Systems*, 2003.

[105] L.D. Stone, C.A. Barlow, and T.L. Corwin. Bayesian multiple target tracking.

[106] R.L. Streit and T. E. Luginbuhl. Maximum likelihood method for probabilistic multi-hypothesis tracking. *in Signal and Data Processing of Small Targets, SPIE 2235, O. E. Drummond, Ed.*, 1994.

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

[108] M.A. Tanner and W.H. Wong. The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, 82:528–549, 1987.

[109] Rudolph Van Der Merwe, Arnaud Doucet, Nando De Freitas, and Eric Wan. The unscented particle filter. 13:2, 01 2001.

[110] J. Vermaak, S. Godsill, and P. Perez. Monte carlo filtering for multi-target tracking and data association. *Signal Processing Engineering Depratmenet, University of Cambridge*, 2004.

[111] Axel Vogt.

[112] P. Willett, Y. Ruan, and R. Streit. Pmht: Problems and some solutions. *IEEE Transactions on Aerospace and Electronic Systems*, 38(3):738–754, 2002.