A Bottom-up Approach to the Financial Markets

Agent-Based Quantitative Algorithmic Strategies: Ecosystem, Dynamics & Detection
(working paper)

Babak Mahdavi-Damghani* and Stephen Roberts**

Oxford-Man Institute of Quantitative Finance

***Abstract*— In this paper we propose a new approach to studying the financial markets. Instead of the traditional top-down approach where a Brownian Motion is assumed as the driving force behind the market and where dynamic strategies are built as a result, we rather take the opposite point of view (the bottom-up approach) by assuming that it is the interaction of systematic strategies that induces the dynamics of the market. We achieve this shift in perspective, by re-introducing the High Frequency Trading Ecosystem (HFTE) model [86]. More specifically we specify an approach in which agents interact through a Neural Network structure designed to address the complexity demands of most common financial strategies but designed randomly at inception. This strategy ecosystem is then studied through a simplified genetic algorithm. Taking an approach in which simulation and hypothesis interact in order to improve the theory, we explore areas that are usually associated to fields orthogonal to Quantitative Finance such as Evolutionary Dynamics & predator-prey models. We introduce in that context concepts such as the Path of Interaction in order to study our Ecosystem of strategies through time. Finally a Particle Filter methodology is then proposed to track the market ecosystem through time.**

Keywords: High Frequency Trading Ecosystem (HFTE), High Frequency Financial Funnel (HFFF), Multi-Target Tracking (MTT), Stability of Financial Systems, Markov Chain Monte Carlo (MCMC), Data Analysis and Patterns in Data, Electronic Trading, Systemic Risk, High Frequency Trading, Game Theory, Machine Learning, Predator Prey Models, Sequential Monte Carlo, Particle Filter.

I. Introduction

A. The Rise of Big Data

1) Definition: The multiple industrial applications arising from the concurrent rise in information retrieval and computer storage capabilities has opened up Big Data in a spectacular fashion [21], [60], [71], [1], [34] and unique way since the scope is both deep and far reaching. But what really is Big Data? Though used sometimes loosely partly because of a lack of formal definition, the interpretation that seems to best describe Big Data is the one associated with large body of information that we could not comprehend when used only in smaller amounts [21]. This characterization seems to indicate that the realm of the definition goes fundamentally beyond simply reducing the confidence interval of a parameter whose estimation would benefit from an increase of the sample size. This latter intuition is the natural statistician point of view. In fact the term "datafication" has recently been introduced in order to replace the misleading term that is Big Data in order to make sure readers research the term instead of guess its meaning [21]. A good way to illustrate this point would be for instance to examine Figure 1[1], new data at the high frequency domain which allow us to explain the market from the Bottom-Up approach[2] rather than assume the Top-Down[3]. Big Data suggests in this situation that the point of reference of the mathematical model should be entirely shifted top to bottom. Big Data suggests real innovation as opposed to merely improvement of the status quo. Though the part of the definition that infers increase in size part is partly indicative of the definition, it only tells half of the story, and in that sense "Innovative[4] Data" would have been a more intuitive term, though perhaps arguably less marketing friendly. Taking its literal sense though, we can legitimately ask how big is Big Data?

2) How Big is Big Data?: There exists many anecdotal claims illustrating the size of Big Data. For instance its been suggested that if we were to take as reference the time where information was not stored digitally (for instance during the third century BC), where it was believed that the Library of Alexandria housed the sum of all human knowledge, then today, there are arguably 320 times the number of inhabitants worth of data available. More specifically if all this data was placed on CDs and these latter CDs were stacked up, the CDs would form five separate piles that would all reach to the moon [21]. Another interesting fact reported is that as much as 90% of current data was created in the last couple of years [1]. Though these figures are often the most cited by researchers there are legitimate questions around the quality and the usefulness of the data being stored. For instance Facebook likes which may have been bought or censured constitute a source of data equal in value to perhaps Geophysics data. There are also small disagreements with

* babak.mahdavidamghani@oriel.ox.ac.uk, Oxford-Man Institute of Quantitative Finance

**steve.roberts@oxford-man.ox.ac.uk, Oxford-Man Institute of Quantitative Finance

[1]Which we delve more into latter on.

[2]e.g. strategies interacting explain the fluctuation of the market

[3]e.g. the market is assume to be a Brownian motion which itself allows for dynamical strategies such as hedging for instance

[4]The idea that there is increase in the available data (the "Big" in Big Data) is implied in the latter formulation but on top of that the intuition that it also brings change is encompassed with this proposed terminology as well.

respect to how fast Data is growing but it is estimated to roughly double every three years [21], [71].

3) Scope: The Internet and Cloud Computing fast growth have led to the exponential growth of data in most if not all every industries. This hot topic attracted the attention of all segments of the population going from government, academia and the industry with far reaching opportunities but also grand challenges such as data, computational and system complexity with already proposed solutions [60]. No matter how one looks at these figures the rise of Big Data is real, its size and scope are changing our lives and our civilization at a very rapid speed. The industry has many applications of BD. The one we have chosen to explore is the financial industry as it is at the heart of our economy, comes under much scrutiny and can create systemic risk [50] with far reaching impacts.

B. A Market Changing Financial Crisis

We can go as far back as few centuries for the construction of the financial system with perhaps its first serious mathematization attempt occurring about a century ago [8] but the event that is most relevant to this paper happened about 10 years ago with additional posterior signs which served to remind us that the impact of this event was not over[5].

1) The Subprime Crisis as a Triggering Effect: The financial crisis of 2009, the resulting social uproar in the general population induced ethical malaise in the scientific community [61], [83], [90], [84] which changed the market in many ways. More specifically, after the subprime crisis governments strongly pushed the regulators to develop more efficient risk monitoring systems[6] and review the current modelling pillars so as to avoid similar crises in the future.

2) A Call for a Modelling Revolution: The new candidate sector under inspection quickly became the one of algorithmic systematic trading which flash crash of May 6, 2010, in which the Dow Jones Industrial Average lost almost 10% of its value in matter of minutes, exacerbated the scrutiny. However, the current state of the art risk models, are the ones inspired by the last subprime crisis and are essentially models of networks in which each node can be impacted by the connected nodes through contagion [50] and is better suited to lower frequency models. Indeed, on 06/08/2011 a seemingly relatively unnoticed event occurred on the natural gas commodities market. We say "relatively unnoticed" simply because the monetary impact was limited and finance is unfortunately an industry in which warning signs are usually dismissed until it is too late. We can see from Figure 1 that clearly something non-random is occurring. This feeling is exacerbated by the strong intuition that only interacting agents falling into some sort of quagmire could yield such series of increasing oscillations followed by a mini crash. Indeed, commodities has historically been seen as a physical market, this in turn meaning that the prices are driven by supply and demand of commodities which can be consumed, stored and/or produced. This particular point is a unique feature compared to the other markets (Equities, FX, or Rate). Also Figure 1 suggests that the common, though perhaps a bit lazy view, that crashes occur through totally unpredictable [131] events may not be true for algorithmic trading. These few examples amongst others have led the scientific community to encourage a revolutionary changes possibly in the form of agent-based modelling [15], [35], [16] in lieu of traditional financial mathematics models. It is in this fundamental opposition of views that part of the title of this paper must be understood. Indeed, traditional financial mathematics programs focused on derivatives[7] were chastised and rethought [72]. This decline 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 [114] at modelling the options market for example but it seems there are outstanding issues especially when coherence as defined by arbitrage constraints is taken into consideration. On the Risk side VaR must now take into account the procyclicality of the market. Data suggests that a big market move is likely to follow another big market move [50], risk models must adapt quickly, incorporate Bayesian statistics, adjust to the sudden increase of volatility of the market and therefore be responsive. Paradoxically, in order to eliminate the risk associated to liquidity shortage and the resulting systemic risk, VaR should on top of being Responsive remain as Stable as possible [105][8].

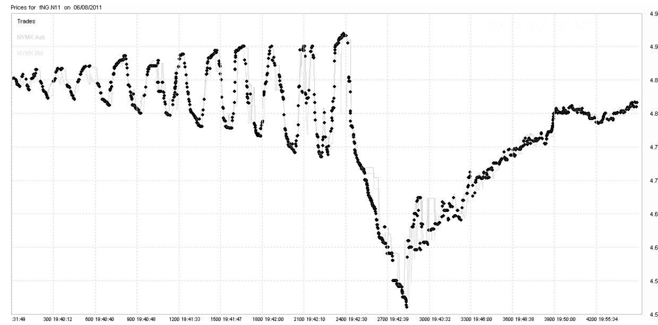


Fig. 1: Natural Gas flash crash of 06/08/2011 [99]

3) Exactitude vs Complexity: 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 [58] and many desks were closed as a result. Indeed, the last crisis led to a modelling revolution. Fueled by the lessons learnt from trading convoluted illiquid products based on wrong but mathematically convenient assumptions[9], we saw a concurrent:

[5] for example the multiple flash crashes.

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

[7] in which highest likelihood and mathematical convenience prevailed over data supported by the market.

[8] We reconcile these discordant instructions later in the document.

[9] The correlation model assumption was misleading[90] when it came to pricing complex subprime products.

- tactical step back[10] in the complexity of the products in order to gain a more focused momentum for the future Quantitative models,
- compulsory step forward in liquidity modelling since the subprime crisis was arguably a liquidity activated crisis.

The product class that took the niche of exotics became simpler vanilla products, which hedging property has still utilitarian value[11], more liquid, less volatile and therefore more in-line with the role of derivatives at their inception. Generally speaking liquidity modelling became of central focus like never before especially through government led initiatives [105] such as for example the Fundamental Review of the Trading Book (FRTB). As the risk models increased in sophistication, questions around coherence of scenarios also became of central importance [88]. The Capital Requirement of each financial institution is now linked to its VaR[12] and the latter must be calculated with historical data. Finally P&L associated to trading should be mapped to appropriate risk factors.

4) From Financial Mathematics to Machine Learning: Going all the way back to the early stages of the 20th century and Louis Bachelier pillar contribution [8] to Mathematical Finance, the world of Quantitative Finance has been gradually enhanced by the other STEM fields. Probability Theory, Physics and Statistical Mechanics, among other fields, have steadily but surely brought Quantitative Finance amongst the most challenging STEM fields as its interdisciplinary nature (Mathematics, Computer Science, Physics, Economics & Finance) makes it increasingly difficult to master in full giving rise to further granular specialization[13]. Added to these challenging historical enhancements, today the other STEM fields[14] are accelerating these changes by contributing themselves to the complexity. More specifically Bayesian Statistics, Signal Processing Statistics [89], Game Theory [86] and above all Machine Learning [21] are increasingly contributing to the interdisciplinary complexity. More specifically, with the rise of Big Data [21] and Data Science [1] and the aftermath of the financial crisis of 2009 as well as the multiple Flash Crashes of the early 2010s, resulted in social uproars in the general population and ethical malaises in the scientific community [61], [83], [90], [84] which triggered noticeable changes in Quantitative Finance. More specifically, the latter was instructed to change to the point where the highest authorities in Quantitative Finance have been calling for a modelling revolution [15], [35], [16]. To some extend the Bottom-Up approach of agent based modelling, an area of Machine Learning, as suggested by the authorities, is in total opposition to the Top-Down approach used by Financial Mathematics (as best symbolized by the use of the Brownian Motion). Added to this antagonistic change in the point of the markets models in algorithmic trading, we have been witnessing an interesting re-balancing shift between models and data. More specifically we are moving from an approach in which models assumed data towards one in which data is reassuming the models [90], [84], [88], [85] and slowly making distinguished Financial Mathematics models obsolete. To some extent the subprime crisis can be seen as the triggering effect which has seen the rise of Machine Learning and the coinciding decline of traditional Financial Mathematics models within the world of Quantitative Finance. However as we will see more in details the opposition between these two fields can sometimes be turned into an apposition. This wording needs to be understood the following way: these two field can enhance each other (appose) rather than oppose each other.

C. Problem Formulation

1) Agent-Based Intelligent System & Deep Learning: We learn about the bottom-up vs the top-down approach in introductory systems engineering classes at the undergraduate level but by the time one gets into the most advanced postgraduate financial mathematics classes, this essential beginners scientific lesson for information processing strategy, has long been forgotten and the models have become dogma. Indeed at these more advance stages of ones education it becomes much more important to be able to derive or infer meaning via these believes rather than understand the limitations of these core modelling assumptions and improve the models from inception. In fact these beliefs are so much anchored in our common academic psychs that wrong models get Nobel Prizes[15] and lead to market crashes.

Remark The latter award, is sometimes abusively called the "Nobel Prize in Economics", the same way the Fields Medal is sometimes called the "Nobel Prize in Mathematics". The Nobel Prize in Economics does not, in fact, exit. The same way Alfred Nobel left Mathematics out of his will, he also left Economics out of his will. This latter fact is less known as the wording of the Economics Prize is much closer to the wording of the other Nobel Prizes. The exact wording of the Economics Prize is the "Nobel *Memorial* Prize in Economic Sciences" which was awarded for the BSM in 1973. The Prize was in fact launched by the Sveriges Riksbank (Swedens Central Bank) in *memory* of Alfred Nobel.

In fact the embarrassment of the repeated market crashes has led the highest Quantitative Finance experts[16] to call for a modelling revolution [15] in the shape and form of an agent-based intelligent system point of view.

[10] Coming back to the basis of the derivatives markets which is to provide insurance against big market moves as opposed to create these big market moves.

[11] 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.

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

[13] We now have "Pricing Quants", "Algorithmic Trading Quants", "Risk Methodology Quants", "Structurers", "Model Validation Quants", "Quant Developpers", "Quantitative Traders" etc ...

[14] As we will see in part **??**.

[15] see: Black-Scholes model and Long-Term Capital Management history.

[16] Jean-Philippe Bouchaud was awarded the very prestigious "Quant of the Year" award the year this paper was written.

Remark Note that Quantitative Finance is often criticized as being more a social science because many[17] of its theories are wrong. This peculiarity is not exclusive to Quantitative Finance. Indeed, other STEM fields share some of this embarrassing fact. For example in Physics questions around the gravitational force are still outstanding and Newton theory only works within the confine of our planet and not beyond. In biology the individual centered view of evolution, though would explain a great deal of our surrounding was ultimately gently put aside when the gene centered view of evolution appeared. These two theories were at inception the highest ever recorded academic impact of their respective field and they were ultimately "incomplete[18]".

In any case how is this relevant to the mentioned strategy of information processing? The current modelling approach in Quantitative Finance is the lazy, though convenient, top-down approach and the one we are suggesting is the more challenging bottom-up approach. Indeed the current modelling format takes as view that financial underliers follow a random walk like process and that the latter converges towards the wiener process. Formally in the top-down approach we assume that in the "limits" a change to some price process S_t to follow a log-normal diffusion process. Recall that the log-normal assumption arises from the wiener process itself resulting in the assumption of the random walk:

Definition (Wiener Process): W_t has four main properties: $W_0 = 0$ a.s. $\forall t > 0$, $W_{t+u} - W_t$, are independent of W_s where $s < t$, $W_{t+u} - W_t \sim \mathcal{N}(0, u)$ and W_t is continuous in t.

Definition (Random Walk to Wiener Process): Let $\xi_1, \xi_2, \ldots$ be i.i.d. random variables with mean 0 and variance 1. For each n, we define $W_n(t) = \frac{1}{\sqrt{n}} \sum_{1 \leqslant k \leqslant \lfloor nt \rfloor} \xi_k$ where $t \in [0, 1]$. By the central limit theorem (and more rigorously Donsker theorem) $\lim_{n\to\infty} W_n(t) - W_n(s) \sim N(0, t-s)$.

However this top-down approach with the assumption of the increment being iid to make analysis more convenient as opposed to more exact has been criticized both in the low frequency domain [90] as well as in the higher frequency domain (see analysis associate to Figure 1) so much so that "overlay[19]" models [40] have been incorporated to the BS model to account for the mis-pricing induced by the log-normal assumption[20].

Remark Note that the random walk assumption is still useful if one takes the risk neutral approach, more specifically in the pricing branch of Quantitative Finance. However, this latter theory, in practice, often works as an approximation which is often combined with a fat "sales" fees which is really there as sometimes an exorbitant add-on model which makes you wonder about the risk neutral approach.

These later models where in turn challenged with the arrival of Big Data which exposed new limitations [88] of these overlay models but these latter models where in turn also shown to be incomplete when the question of liquidity came into play [85]. You may take two approaches in analyzing the consistent failure of these models:

- either you accept that all these models are incomplete because the core assumptions which we use to derive them, are too far from reality and therefore waiting for the next crash to improve with yet another overlay model and this is an unacceptable approach,
- or you could chose to simply assume that this is the natural course of the scientific method and building a science on precarious grounds is an acceptable alternative because we cannot do anything better yet.

The repetitive market crashes and the subsequent punitive sanctions taken by the regulators and their subsequent instructions to work with new models that directly contradict these root mathematical assumptions[21] suggests that the timing is now right to take a step back and reexamine the bottom-up approach for more consistency in the future.

Remark An interesting analogy can be made with respect to how the gene centered view of evolution (as opposed to the individual centered view of evolution) completely reshuffled our understanding of natural selection and gave the opportunity to explain altruism better. By analogy, we are trying to communicate the idea that the change of prospective from the market centered view (Top-Down) of the financial systems is the wrong way to understand the fluctuation of the market and that the strategy centered view (Bottom-Up) of the financial system provides an opportunity to explain the fluctuations of the market more effectively.

However, this bottom up approach at the intelligent agent level presents a great deal of challenges. The first one to take into account is to recall that small simple increments are the basis of any viable complex biological system. For instance what created the complexity of the eye in evolution was a slow process which went from the simple photoreceptor to the folded area (cavity) and finally[22] a complex eye.

2) Adversary Model as a Key for Enhancement: It is worthy to note that the creation of the eye was for survival purposes and that it was developed in both predators and prey as the interaction between them favored the enhancement on the complexity of the eye. The key word here to note in the scientific process is the one of interaction. This critical element of the scientific process was perhaps best exposed by Conway who also took this approach of simple rules leading to complex systems in his Game of Life research [39], [20].

[17] I am being a little politically correct here as I would rather replace "many" by "all".

[18] I have too much respect for these two contributions to really replace the term "incomplete" with "wrong".

[19] e.g. Dupire's local volatility model [30], [31].

[20] This is the whole rational behind the implied volatility surface.

[21] The regulators have recently advised the banks that their model should take into account the procyclicality of the markets [105], directly contradicting the mathematical assumption around the returns being iid.

[22] We took a bit of jump here as there are many more steps including the evolution of the cornea.

DeepMind's AlphaGo was produced with a similar idea of Adversary Model recently. A question naturally arises here. How can we apply this methodology to Quantitative Finance and more specifically market microstructure? What would be a simple mathematical structure (in the form of a DNA) that would both allow simple recognizable strategies to arise from a random swarm of strategies with the environmental pressure being the profit that they make and how can these mathematical structures, through interaction with competing strategies create fluctuations in the market as well as create a pressure for these strategies to adapt and improve their models? How can we make sure we do not incorporate the idea of foresight in our design? Can we make a parallel with other known biological systems such as ecosystems? These tasks are perhaps overly ambitious but can we come up with models and ideas that would inspire research in a new direction? These questions will fall under the Financial Mathematics & Machine Learning opposition part of the document. We however propose to study the apposition case through examples as well.

3) Stability of the Market and Multi-Target Tracking: The electronic trading market is increasingly regarded as the source of the next big market crash [86]. More specifically the fact that this type of trading is characterized by interacting algorithms, these potential crashes could come very quickly and the fact that we have not a single, even archaic, bottom-up model is frightening. It becomes of central importance to be able to decipher market movement as a result of these interacting algorithms. However the specifications of these algorithms are always hidden for proprietary reasons which makes the task of the regulators seemingly impossible. However, we can legitimately ask oneself this question. Can we come up with a solution to this regulatory quagmire? More specifically the construction of scenario based particle filter may offer us a hope in at least laying down the foundations of finding a new potential path towards handling these problems. What are these first steps towards finding a solution?

D. Agenda

We have divided this paper in 3 Sections. More specifically in Section II we express Agent-Based Intelligent Systems in Neural Network format expressing the incentive of going from Shallow to Deep Learning. This will then help us, in Section III, to study Ecosystems of strategies using tools in Evolutionary Dynamics. Finally in Section IV we study the Stability of Financial Systems using Multi-Target Tracking.

II. Agent-Based Intelligent System, from Shallow to Deep Learning

Following Bouchaud[23]'s call for a revolutionary change in economics [15], taking a bottom-up approach via Agent-Based Models, instead of the traditional top-down approach as best symbolized (the Brownian Motion assumption in Financial Mathematics). This section will expose an example in which the traditional Financial Mathematics approach can be revolutionized by tools from Machine Learning and therefore provide an argument opposing these two fields rather than apposing them. This Section will be organized in several Sections. First we make the parallel with the scientific method used in Conway's Game of Life [39], [20] in Section II-A. We will then delve into some formalization of Electronic Trading in Section II-B. In Section II-C we go through a very brief literature review of Neural Networks in order to introduce the rational of certain types of architecture and their learning potential. In Section II-D we will introduce the core DNA for our financial strategies and expose how this structure can model many of the well known financial strategies. Finally in Section II-G, in the context of Electronic Trading, we will then reflect on the relationship between architecture and meaning in order to explain the incentive for Deep Learning (DL). We use in that occasion the same methods used in adversarial algorithms in order to expose how Deep Learning can naturally result from simpler strategies in Shallow Learning.

A. The Financial Game of Life

In Sections II, III and IV of this paper we will take a methodological approach similar to the Game of Life, a well known 4 rules cellular automaton published by Conway [39] in the mid 70's[24]. In this section we inspire ourselves of

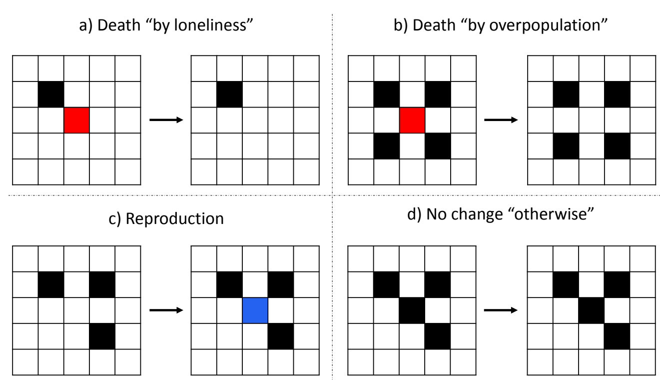


Fig. 2: Conway's Game of Life rules illustrated

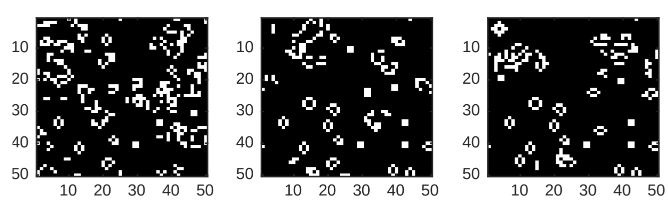


Fig. 3: Snapshots of a simulation of Conway's Game of Life

Conway's scientific method but use it instead to the world of High Frequency Trading (HFT) while adjusting some of the idiosyncratic parts of the exercise. As a reminder, Conway's

[23]Quant of the year 2017.

[24]Figure 2 provides the rules of the Automaton and Figure 3 represents 3 snapshot of one random simulation.

Game of Life assumes that complexity in an ecosystem[25] arises from simple rules. To illustrate this Simplicity-to-Complexity path, in the Game of Life, the four simple rules of Figure 2 can lead to many different families of complex automatons. As a reminder, starting with random seeds and after few iterations, the simulation leads to stable, oscillating and moving forms. For stable forms[26], the concept of financial stability may be raised through a similar methodology. For oscillating forms [27], intuitively the reader may guess that the concept of financial cycles or HF oscillations (Figure 1) may be induced through a similar methodology. Finally the moving forms[28] may have different sizes and speeds[29]. The parallel to the world of quantitative financial strategies would be the following. First interacting agents lead to market price fluctuations. More specifically their interaction determines the stability or instability of the market[30]. Second, the market will not necessarily follow the rules of a zero-sum game[31] with, however random seeds. Third agents (e.g. strategies) will follow a simple rule for their births and deaths.

B. Electronic Trading

1) Description: Traditional order book consists of a list of orders that a trading venue such as an exchanges uses to record the market participants' interests in a particular financial product. Typically a rule based algorithm records these interests taking into account, the price & the volume proposed (on either side of the Bid-Ask) as well as the time in which that interest was recorded (in situations in which interest at the same price is recorded by few different market participants, a referee decides which would win the trade: usually FIFO).

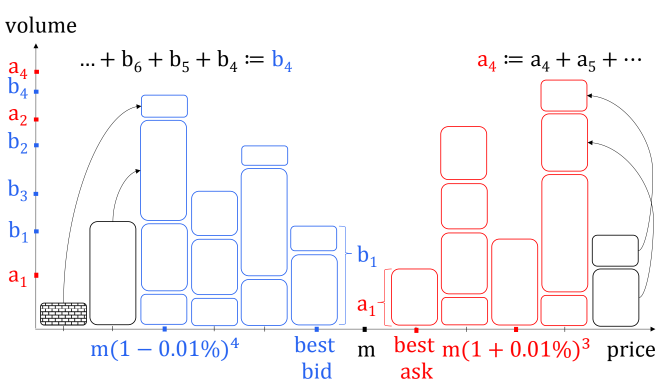


Fig. 4: Order-book visual representation

[25]We take this opportunity to mention here that in this paper Ecosystem and Market are interchangeable since the former is taken to an intuitive image of the latter.

[26]For example the "Block", the "Beehive", the "Loaf", the "Boat".

[27]For example the "Blinker (2 period iteration)", the "Toad" (2 period iteration), the "Beacon" (2 period iteration), the "Pulsar" (3 period iteration), the "Pentadecathlon" (15 period iteration).

[28]for example the "Glider" and the "Lightweight spaceship" (LWSS)

[29]This latter family is arguably conceptually not really providing a useful comparison to our problem.

[30]Depending on what the market is made of in terms of the strategies involved as well as the evolving order-book.

[31]Meaning that its evolution is determined by its initial state, requiring no further input.

2) Variable Definition:

Definition (Order-Book): In terms of naming early these different points of the order book we would label by a_1^t and b_1^t the best ask & bid total volumes at time t. By extension a_i^t, b_i^t with $i \in \{1, 2, 3, 4\}$ would correspond to total volume at the relevant depths' of the order book with the special case where $i = 4$ which then would represents the total volume at the 4th depth level in addition to all the other market depths superior in price (in the case of the Asked price and vice versa for the bid price). We will call m_t the mid price of the best bid/ask at time t. The prices at the different levels, l of the order book will be arbitrarily chosen to be 1bps[32] apart as shown by Equation (1)). Figure 4 represents our version of the order book.

$$p_l^t = m_t[1 + (-1 \times 1_{l \in b_i^t} + 1 \times 1_{l \in a_i^t}) \times 0.001\%]^i \qquad (1)$$

Definition (Leading Indicators): We will label by $\{y_i\}_{i=0}^{n-1}$ the price process of interest, $i \in [0, n]$ its discretized 500ms snapshots with $i = 0$ being the most recent snapshot and $i = n$ its most distant snapshot. Moreover we will assume here that 500ms is enough time for the trading system to take the data, reformat it, analyze it as well allow the relevant strategy to take actions[33]. Similarly we will define $\{x_{j,1}, x_{j,2}, \dots x_{j,p}\}_{j=i+1}^{n}$ the relevant, p leading indicators to the price dynamic of interest.

Remark We will assume that the Leading Indicators for the price process can only be taken from the order book which is a reasonable assumption in the higher frequencies. Some usually accepted leading indicator are the price of the underlier itself and the accumulated volume at different market depth of the order books (4 on the bid side and 4 on the ask side for a total of 9 leading indicators with the price process: see Figure 4 for visual representation).

C. Neural Net Architecture & Learning Potential

In the spirit of explaining the complex through simple logical incremental steps, this particular subsection is dedicated to how increasing simple Artificial Neural Network (ANN) architecture into more complex ones by adding hidden layers[34] can lead to complex learning potential like it can be done with Deep Learning. More specifically, taking this approach can allow us to slowly move towards Deep Learning potential and allow us to unweave[35] the black box associated to the latter perplexing concept. With this in mind, two well known, but important milestones in Machine Learning are worth reminding of. Especially for the beginners, these two milestones can shed light on why the core building blocks of our HFTE model is a certain way and also prepare intuitively the reader for sections II-D and II-G. First, Warren McCulloch and Walter Pitts [110] introduced their threshold

[32]bps stands for Basis Points or in terms of Percentage 0.01%.

[33]Last assumption we will make is that no slippage or other man made errors can bias our results.

[34]"Deep Learning" is arguably just a fancy word for a Perceptron with many hidden layers.

[35]Note that the unweaving analysis goes forward instead of backward.

logic model in 1943 which is agreed to have guided the research in network architecture as it relates to artificial intelligence for more or less a decade. Second, Rosenblatt [115], formally introduced the perceptron concept in 1962 though some early stage work had started in the 1950s. The idea of the perceptron was one in which the inputs x_1 and x_2 as depicted from Figure 5 could act as separators[36] and therefore a direct equivalence could be made to the Multi-Linear Regression (MLR) which we will elaborate more in details is Section II-E.2. One observed limitation of the perceptron as described by Rosenblatt, in 1969, was that a simple yet critical well known functions such as the XOR function could not be modeled [98]. This resulted in a loss of interest in the field until it was shown that a Feedforward Artificial Neural Network (ANN) with two or more layers could in fact model these functions (see Figure 6 for the illustration). Added, to this we have the well known over-fitting [128] problems when it comes to supervised learning which, to some extend would like to simply keep adding hidden layers when the learning potential has been absorbed. This problem of learning potential to over-fitting has been there since inception though regular progress is being made in that domain without real breakthrough[37].

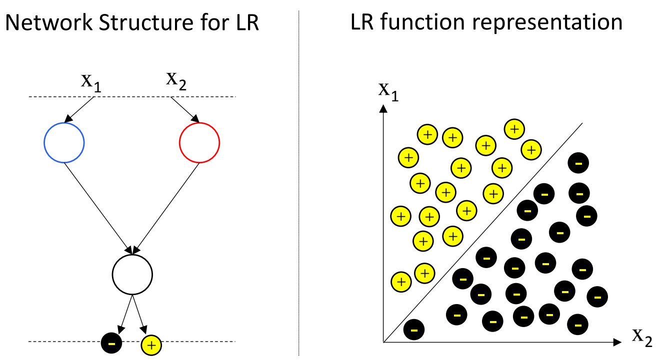


Fig. 5: Neural Network Modelling a Linear Regression

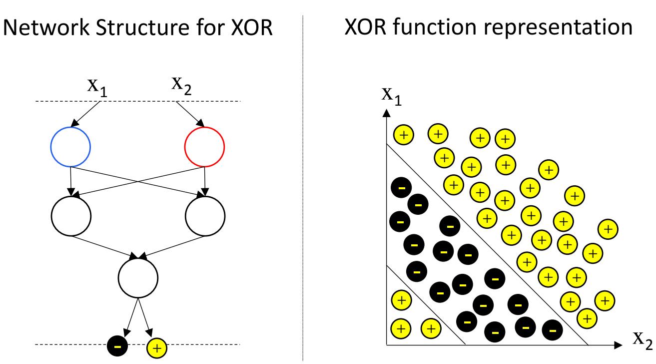


Fig. 6: ANN with 1 Hidden Layer & the XOR Function

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

[37]We refer here the reader to area of ML known as Regularization.

D. Intelligent Agents & Financial Strategies

In this subsection we will use some of the material presented in the High Frequency Trading Ecosystem (HFTE) [86] recently introduced, we will therefore summarize the main points of the referred paper for that occasion.

E. The High Frequency Financial Funnel

The pillars associated to the construction of the HFTE model has in its inspirational roots the idea that strategies in the market interact and it is their interaction that creates the fluctuations in the prices like it would be the case for cellular automatons [39]. It also assumes that strategies can invade others and therefore the study of the financial market partially comes to studying a stochastic n-species predator prey model. Another pillar is that the construction of each of these strategies must have the same DNA[38]: the financial funnel (Figure 7). Finally the financial funnel can model many of the classic financial strategies. For example it can model Trend Following (TF) strategies, Moving Average Convergence Divergence (MACD), Multi-Linear Regression (MLR) or XOR like strategies like it can be seen by figures 8,9, 10 and 13 respectively.

These few historical rationals[39] are the main drivers which have led us to propose the Funnel, introduced by Martin Nowak [104], as the simplest possible network to model (therefore which minimizes over-fitting) the key functions for our application. The area of evolutionary graph theory is quite rich. Many graphs provide interesting properties.

Definition (High Frequency Financial Funnel): We can formalize the learning process from all of our strategies using the HFFF of Figure 7 by providing a set $\mathcal{H}$, as described by Equation (2) of weights corresponding to all the possible weights of this particular figure.

$$\mathcal{H} \triangleq \left\{ \begin{array}{cc} \cup_{j\in[1,9]} w^{i}_{\bar{s},j} & \cup_{j\in[1,9]} w^{i}_{s,j}, \\ \cup_{j\in[1,9],i\in[1,3]} w^{h_1}_{\bar{s},i,j} & \cup_{j\in[1,9],i\in[1,3]} w^{h_1}_{s,i,j}, \\ \cup_{j\in[1,3]} w^{h_2}_{\bar{s},j} & \cup_{j\in[1,3]} w^{h_2}_{s,j}, \\ w^{o}_{\bar{s},j\in[1,9]} & w^{o}_{s,j\in[1,9]} \end{array} \right\} \quad (2)$$

with w^i, w^h and w^o, respectively the weights associated to the input, hidden and output layers. More formally let the **H**igh **F**requency **F**inancial **F**unnel (HFFF) [86] to be a NN of 9 inputs, 3 hidden layers and 1 output layer. Each node connects to the next layer and to itself. Each connection to itself will be label by w_s and the others by $w_{\bar{s}}$. We will admit that $w_{\bar{s}} \sim \mathcal{U}[-1, 1]$ and that $w_s \sim \mathcal{U}[0, 1]$ and therefore the results from Equation (3)).

$$w_x \sim \mathcal{U}[-1_{x=\bar{s}}, 1] \quad (3)$$

Remark Note that in the context of this paper we have chosen to work with Martin Nowak's [104] funnel, which modification is described in Figure 7. This NN structure offers the advantage of linking some interesting bridges

[38]alternatively called HFFF or Neural Network.

[39]we will discuss more in details the Bias-Variance Dilemma in Section II-G.

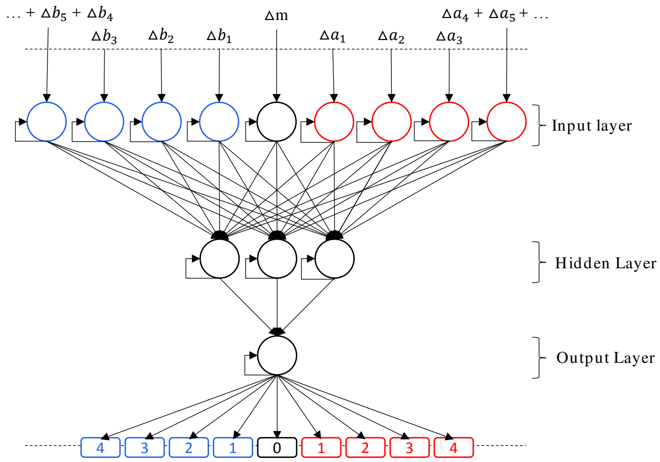


Fig. 7: The High Frequency Financial Funnel (HFFF)

between the worlds of information theory, evolutionary dynamics and biology. Indeed in information theory it also resembles the classic structure of a Neural Network and can therefore easily accommodate the mapping of classic and less classic financial strategies. In evolutionary dynamics, Moran likes Processes can easily be formalized through similar means. In biology the network structure is a potent amplifier of selection [104].

Note also that the HFFF from Figure 7 can easily be trained using a classic error back propagation algorithm like the one described in algorithm (1)[40].

Algorithm 1: Backpropagation

Input: NN $\mathcal{H}$ with unoptimized weights
Output: NN $\mathcal{H}$ with optimized weights

for *d in data* **do**

  Forwards Pass:
  Starting from the input layer, do a forward pass trough the network, computing the activities of the neurons at each layer.

  Backwards Pass
  Compute the derivatives of the error function with respect to the output layer activities **for** *layer in layers* **do**
    Compute the derivatives of the error function with respect to the inputs of the upper layer neurons Compute the derivatives of the error function with respect to the weights between the outer layer and the layer below Compute the derivatives of the error function with respect to the activities of the layer below
  end
  Updates the weights.
end

1) The EWMA NN: When we first started our research we called this subsection the Trend Following HFFF but through the simulation exercises and with increased research experience we decided to rename this subsection EWMA NN. However, in many of our simulation when we refer to the TF strategies we really mean EWMA family. We will explain this rational next. A very common trading strategy is the trend following (TF). The idea of the TF is that if the price has been going a certain way (e.g. up or down) in the recent past, then it is more likely to follow the same trend in the immediate future.

Definition (Trend Following): The mathematical formulation of a TF can be diverse but in the context of this paper we will be using an exponentially weighted moving average (EWMA) formally described by Equation (4)).

$$\hat{x}_t = (1-\lambda)x_t + \lambda \hat{x}_{t-1},\ \lambda \in [0,1] \tag{4}$$

Remark In this equation λ represents the smoothness parameter with $\lambda \in [0,1]$. The lower the λ, the more the next move will be conditional to the immediately adjacent previous move. Conversely, the higher the λ, the more the future move will be function to the long term trend. The idea being that through a simple filtering process, the noise is extracted from the signal which then return a clean time series $\hat{x}_t$ traders like to seldom use directly or sometimes by using it with couple of other similar equations with a different λ and therefore defining a signal as a difference of these various filtered time series.

Proposition The HFFF can model trend following strategies.

Proof: Simply set $\cup_{j\in[1,4]} w^i_{\bar{s},j} = 0$, $\cup_{j\in[1,4]} w^i_{s,j} = 0$, $\cup_{j\in[6,9]} w^i_{\bar{s},j} = 0$, $\cup_{j\in[6,9]} w^i_{s,j} = 0$ $\cup_{j\in[1,4],i\in[1,3]} w^{h_1}_{\bar{s},i,j} = 0$, $\cup_{j\in[1,4],i\in[1,3]} w^{h_1}_{s,i,j}$ $\cup_{j\in[6,9],i\in[1,3]} w^{h_1}_{\bar{s},i,j} = 0$, $\cup_{j\in[6,9],i\in[1,3]} w^{h_1}_{s,i,j}$, $w^h_{\bar{s},3} = 0$, $w^h_{s,1} = 0$ and $w^h_{s,3} = 0$. ∎
The proof is visually illustrated by Figure 8 (the weight equal to 0 have not been represented[41]). We will address the problem of rigorously formalizing mathematically what constitutes a trend following in a subsequent paper. However for now, in order to keep things intuitive, we will consider a trend following strategy to have a NN DNA which would look like the one from Figure 8. One of the current hurdles in our research is our classification issue and the MACD strategy is a good example as to why. Indeed the MACD strategy which is technically associated to the EWMA family has an economical meaning which can potentially be classified as an economically antithetic strategy of TF which are in the same EWMA family. This may be important for practitioners as the MACD(12,26) has for instance gained a great deal of momentum for algo traders as it can be seen on the various search results on youtube or on practitioners websites such as "investopedia".

[40]where the activation function would be linear so as to make sure the MLR strategy can be exactly replicated.

[41]Note that there is different ways to achieve the same numerical results though with a different NN architecture.

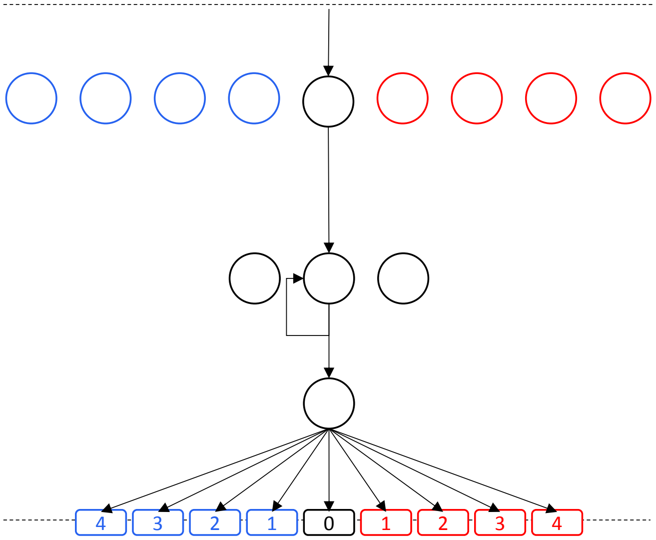


Fig. 8: The EWMA Strategy in HFFF format

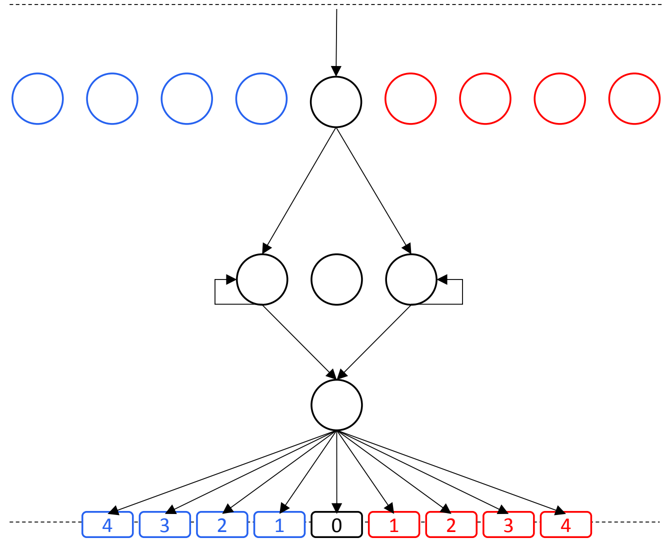


Fig. 9: MACD (difference of EWMAs) in HFFF format

Proposition The HFFF can model the MACD(12,26) strategy.

Proof: The Moving Average Convergence/Divergence (MACD) was designed to reveal changes in the direction and duration of a trend. It essentially models difference between a "fast" ($S_t^{N_f}$) EMA and another "slower" ($S_t^{N_s}$). For instance the popular MACD(12,26), $M_t^{12,26}$ is given by:

$$M_t^{N_f,N_s} = S_t^{N_f} - S_t^{N_s} \quad (5)$$

$$S_t^{\alpha} = \begin{cases} S_1, t = 1 \\ \alpha \cdot S_t + (1-\alpha) \cdot S_{t-1}^{\alpha}, t > 1 \end{cases} \quad (6)$$

$$\alpha = 2/(N_\alpha + 1) \quad (7)$$

$$N_\alpha = \{N_f, N_s\} = \{12, 26\} \quad (8)$$

Figure 9 represents a generic MACD. If one is looking specifically for a MACD(12,26), then the weights of the hidden layers must be such that $\alpha_{12} = 2/13$ and $\alpha_{26} = 2/27$ and the ones of the output layers must be a simple subtraction to abide by the above definition. ■

2) Multi Linear Regression NN: The Multi Linear Regression (MLR) is another well known simple strategy traders have been using in the industry.

Definition (Multi Linear Regression): Given a data set $\{y_i, x_{i-1,1}, \ldots, x_{i-1,9}\}_{i=1}^n$ where n is the sample size, and y_i then our MLR is formalized by the Equation (9).

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

where T denotes the transpose, so that $\mathbf{x}_{i-1}^{\mathrm{T}}\beta$ is the inner product between vectors x_i and β. The best unbiased estimator of β is given by $\hat{\beta} = (\mathbf{x}^{\mathrm{T}}\mathbf{x})^{-1}\mathbf{x}^{\mathrm{T}}y$ and sometimes also referred to β^{OLS}.

Proposition The HFFF can model multi linear regression like strategies.

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

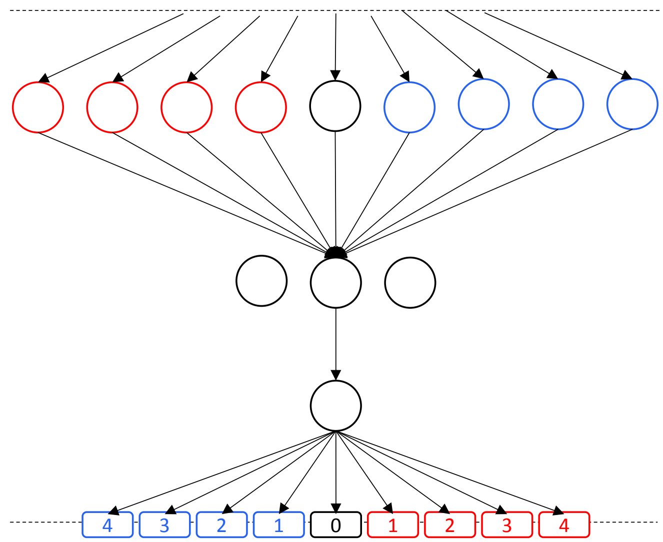


Fig. 10: MLR strategy in HFFF format

Remark Do the activation functions matter when it comes to modelling MLR strategies? The answer to this question is obviously yes. The MLR is by definition a linear strategy and not a sigmoid strategy, otherwise it would be called a MSR. We will make 4 additional remarks. First, an MLR can be illustrated by Figure 10 (the weights equal to 0 have not been represented). Second, as we have explained before, different NN architecture may lead more or less to the same strategy. Figure 11 is another example of a MLR. Third we will address the problem of rigorously formalizing mathematically what constitutes a MLR in the context of the HFFF in a subsequent paper. However for now, in order to keep things intuitive, we will consider a trend following strategy to have a NN architecture which would look like the one from Figure 10. Finally, logistic or weighted MLR can be modeled through the same HFFF of Figure 10 by simply changing respectively the activation function (from linear to logistic) and the weights.

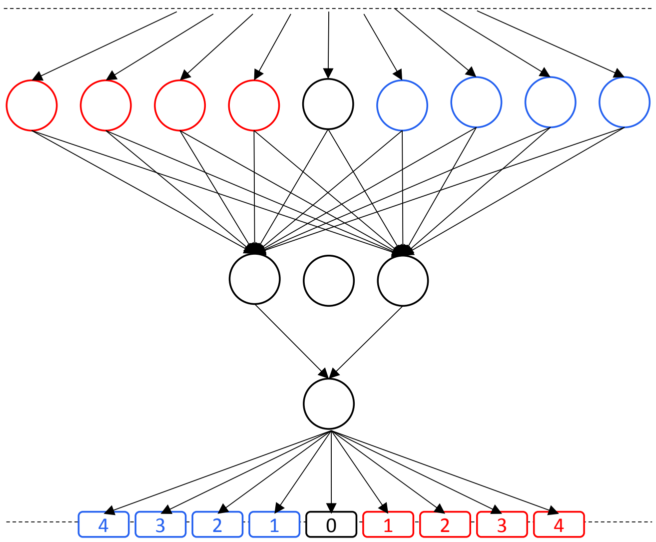


Fig. 11: Another MLR strategy in HFFF format

3) Regularized NN & Lasso Regression: The bias-variance dilemma (BVD) is a technical term representing the optimization by constraints problem which aims at simultaneously minimizing the error from erroneous assumptions (bias) in our learning algorithm or commonly called "under-fitting" and the error from the out of sample analysis (variance) or commonly called "over-fitting". One of the properties of DL is its dual ability to learn the most complicated functions but also makes it prone for over-fitting. It is therefore recommended that one applies conscious efforts in studying carefully the associated benefits to complexity ratio in the context of the BVD. Regularization is usually the term employed for the methodology that aims at finding the optimal model according to the BVD. The mathematical formalization suggests that we calibrate a function f which takes as input a potential infinite number of explanatory variable $x_1, x_2, ...x_n$ so as to minimize the distance to a target y under some cost measure V subject to a penalization, or regularization term[42] $R(f)$. Equation (10) refers to this generic Regularization.

$$\min_f \sum_{i=1}^{n} V(f(x_i), y_i) + \lambda R(f) \tag{10}$$

Within the family of Regularized methodologies the Lasso[43] methodology is the most common one and usually associated with the MLR we have seen in the previous paragraph . They have been gaining momentum in the past few years as they represent the simplest ML technique which has the reputation to work in systematic trading provided the strategy and the input variables are sound.

[42]or regularizer

[43]Short for Least Absolute Shrinkage and Selection Operator.

Definition (Lasso Regression): Given a data set $\{y_i, x_{i-1,1}, \ldots, x_{i-1,9}\}_{i=1}^{n}$ where n is the sample size, and y_i then our Lasso Regression is formalized by Equation (11).

$$y_i = \beta_1 x_{i-1,1} + \cdots + \beta_9 x_{i-1,9} + \varepsilon_i. \tag{11}$$

subject to $\sum_{j=1}^{9} |\beta_j| \leqslant t$ where t is an input parameter that determines the amount of regularisation desired.

Proposition The HFFF can model Lasso regression like strategies.

Proof: Simply set $w^h_{\bar{s},2} = 0$, make sure the regularizarion is done exclusively on one of the remaining hidden layer and finally make sure the remaining hidden layer calibrates its weight the same way at the β^{OLS}. Figure 12 gives an illustration. ■

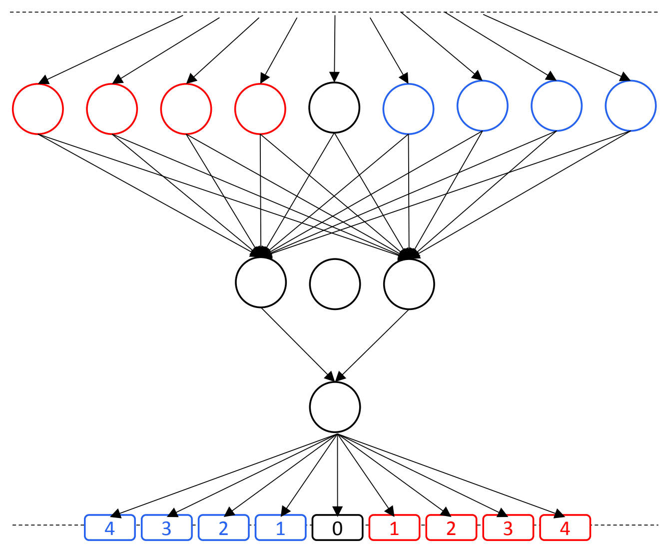


Fig. 12: Lasso regression strategy in HFFF format

Remark Figure 12 and Figure 14 look actually the same but the weights and activation functions are actually very different.

F. XOR Architecture

We recall here the truth table associated by the XOR function in table I. Let's look at the following known HF rational, which will hopefully shed light on the reason why we are discussing the XOR function.

Definition (Open Interest): If we define the Open Interest (OI) as being the total volume left on the order book then it is known that when the price and the OI are rising then the market is bullish, when the Price is rising but the Open Interest is Falling then the market is bearish, when the Price is falling but the Open Interest is rising then the market is bearish, and finally when the Price is falling and the Open

I_1	I_2	O	Price (I_1)	Open Interest (I_2)	Signal (O)
1	1	0	Rising	Rising	Buy
1	0	1	Rising	Falling	Sell
0	1	1	Falling	Rising	Sell
0	0	0	Falling	Falling	Buy

TABLE I: Relationship Between Open Interest, Price & XOR

Interest is falling then the market is bullish. These 4 market situations can be summarized by table I.

Proposition The HFFF can model XOR like strategies.

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

Remark We will make the following 2 observations. First the preceding proof is visually illustrated by Figure 13 (the weights equal to 0 have not been represented). Second The XOR HFFF can be designed in various ways. We will address the problem of rigorously formalizing mathematically what constitutes an XOR in a subsequent paper. However for now, in order to keep things intuitive, we will consider an XOR strategy to have a NN architecture which would look like the one from Figure 13. Figure 14 represents an equivalent alternative example of an XOR strategy.

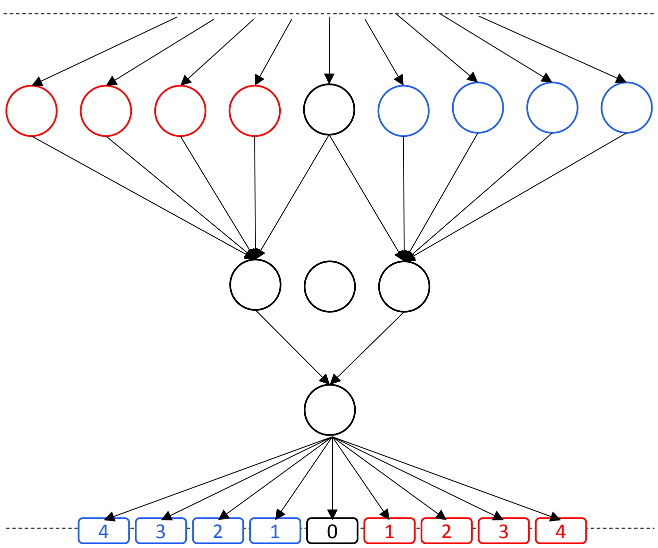


Fig. 13: XOR strategy in HFFF format

G. Network HFFF & Deep Learning Thoughts

Deep learning has been gaining a great deal of momentum, as a subbranch of Machine Learning for very good reasons. For instance Deep-Mind was created in 2011 and subsequently bought by Google in 2014 for 400M [18]. It was made famous for building an AI algorithm, AlphaGo, which outperformed the best Go master in the world. Though, an event in which a Machine Learning algorithm would beat

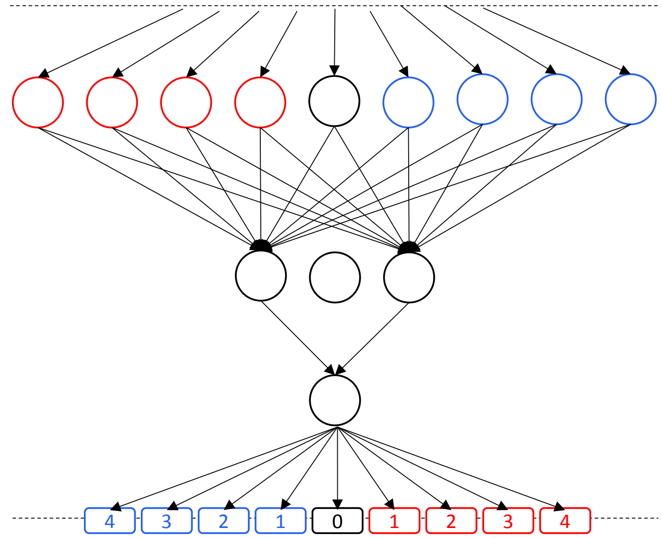


Fig. 14: Another XOR strategy in HFFF format

a master was not an original feature[44], the complexity of the game and the number of possible moves made the Deep-Mind director speculate that the algorithm worked intuitively rather than using a pure logic based approach like it was done with chess and Deep Blue. This extraordinary feature was achieved through an initial deep neural network architecture in which an initial data is used for training purposed and the algorithm, once the data exhausted, would play older version of itself and gets incrementally better this way. The ultimate goal behind this self learning A.I. is to create a general purpose Algorithm. The scientific methodology behind the construction of the game of GO is one we wish to apply to our HFFFs and create a dynamical ecosystem. For instance increasingly advanced strategies compete with each other and we eventually get an interesting portfolio of strategies as well as their co-evolution. However, the HFFF itself potentially suffers from similar kind of limitations that prevented the XOR function to be learnt without 1 hidden layer (see Figure 6 and 5 as well as paragraph II-C). A legitimate question can be asked on whether a single hidden layer is enough. The answer to this question is in fact negative as Convolutional Neural Network (CNN) have shown more potential extracting trading signal compared to shallow learning [133][45]. Some other studies reveal universal features of price formation [127] but lack a study on simpler benchmarks. For instance in [127] a logistic regression is used for a benchmark. It would have been interesting to see some more complex benchmarks[46]. We have arbitrarily taken as hypothesis the HFFF to be good enough to model few critical strategies in the domain of QF and above all proceeding this way is important in unweaving the black

[44] The computer "Deep Blue" beat Kasparov in May 1997 [4].

[45] I am however personally skeptical on the results of these published studies but I do accept the potential of CNN in trading.

[46] starting with a shallow NN and increasing in complexity in order to understand whether the universal features learnt are because the NN is deep or is it because it has a hidden layer.

box associated DL. With this in mind it is interesting to notice that the TF strategy has been designed to dominate a random swarm of strategies. In turn the MLR strategy has been designed to dominate the TF with the key point being that the MLR strategy capitalizes on areas of the orderbook the TF strategy does not have the DNA for (to perceive information of the OI). Similarly the XOR strategy seem to dominate the MLR by splitting the OI surface in additional zones that the MLR cannot understand (lacking the necessary hidden layer). Figure 15 illustrates these observations. In some way you could extrapolate this

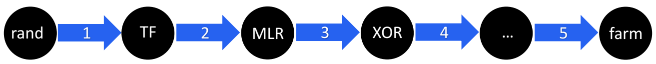


Fig. 15: Illustration for intuitive strategy invasion

"invasion" to "increased network complexity" tendencies to a system that could potentially converge towards a deep learning infrastructures. It is however true that the likelihood of overfitting increases as one adds hidden layers but we have also seen with Shallow Learning that adding hidden layers can also allow us to do regularization which removes the last hurdle argument against DL.

III. Evolutionary Dynamics & Strategy Ecosystem

In the context of the bottom-up approach for algorithmic trading in which we discussed the strategy DNA in Section II, we here formalize the interaction rules at the ecosystem level. More specifically, we go from the premise that a good theory can be simulated but that a simulation can also help bring intuition on what the theory might be, and these two research tools can go back and forth until the theory is ironed out [135]. More specifically we have organized this Section the latter way, assuming that our strategies follow the HFFF model described by Figure 7. We first let the random states of the latter HFFF strategies interact through the primitive strategies swarm via the market order-book. We show, in Section III-D, that although not necessarily optimal, the simulation provides however, a great deal of intuition that help us come up with a more polished theory in Section III-F with the concept of *Path of Interaction*[47]. In doing so we will first do a very brief review of relevant inference and dynamical models in Section III-A, Game Theory in Section III-B and Theoretical Biology in Section III-C.

A. Review of Markov Chain Monte Carlo Models

In this section we go over a brief overview of classic Inference and Dynamical models focusing on Markov Chain Monte Carlo (MCMC). MCMC algorithms [96] 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.

[47]Defined formally later in this Section

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 [97] because 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[48].

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.

2) Hamiltonian Monte Carlo: Hamiltonian Monte Carlo [29], sometimes also referred to[49] 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.

3) Gibbs Sampling: Perhaps one of the simplest MCMC algorithms, the Gibbs Sampling (GS), in pseudo-code in algorithm 2. It was first introduced in Geman & Geman [42] in the context of an application to image processing. Later it was discussed in the context of missing data problems [132]. The benefice of the Gibbs algorithm for Bayesian analysis was demonstrated in Tanner and Wong [132]. To define the Gibbs sampling algorithm, let the set of full conditional distributions be: $\pi(\psi_1|\psi_2, \ldots, \psi_p)$, $\ldots$, $\pi(\psi_d|\psi_1, \psi_2, \ldots, \psi_{d-1}, \psi_{d+1}, \ldots, \psi_p)$, $\ldots$, $\pi(\psi_p|\psi_1, \ldots, \psi_{p-1})$. One cycle of the GS, described in algorithm 2, is completed by sampling $\{\psi_k\}_{k=1}^{p}$ 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 [132]. 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.

4) Ordered Over-relaxation: Over-relaxation is usually a term associated with a Gibbs Sampler but in the context of this subsection we discuss Ordered Over-relaxation. 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

[48]equal to 1 is the proposal density is symmetric
[49]though more in the past.

Algorithm 2: Gibbs Sampling

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

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

Sample:

for $j = 1, 2, \ldots, M$ **do**

  Generate $\boldsymbol{\psi}_1^{(j+1)}$ from $\pi\left(\boldsymbol{\psi}_1|\boldsymbol{\psi}_2^{(j)}, \boldsymbol{\psi}_3^{(j)}, \ldots, \boldsymbol{\psi}_p^{(j)}\right)$

  Generate $\boldsymbol{\psi}_2^{(j+1)}$ from $\pi\left(\boldsymbol{\psi}_2|\boldsymbol{\psi}_1^{(j+1)}, \boldsymbol{\psi}_3^{(j)}, \ldots, \boldsymbol{\psi}_p^{(j)}\right)$

  $\vdots$

  Generate $\boldsymbol{\psi}_d^{(j+1)}$ from $\pi(\boldsymbol{\psi}_d|\boldsymbol{\psi}_1, \boldsymbol{\psi}_2, \ldots, \boldsymbol{\psi}_{d-1}, \boldsymbol{\psi}_{d+1}, \ldots, \boldsymbol{\psi}_p)$.

  $\vdots$

  Generate $\boldsymbol{\psi}_p^{(j+1)}$ from $\pi\left(\boldsymbol{\psi}_p|\boldsymbol{\psi}_1^{(j+1)}, \ldots, \boldsymbol{\psi}_{p-1}^{(j+1)}\right)$

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

end

---

momentum parameter which consist of sampling n random variables (20 is considered a good [82] 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 $n-p$ for the subsequent sample x_{t+1}. This form of optimal "momentum" parameter design is a central pillar of research in MCMC [100].

5) Slice Sampling: Slice sampling is one of the remarkably simple methodologies [100] 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 [82] 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 [79]. The curse of dimensionality is another central area of research for MCMCs [100].

7) Reversible-Jump: Another variant of the Metropolis-Hastings, and perhaps most promising methodology when it comes to our application is the Reversible-jump MCMC (RJ-MCMC) developed by Green [48]. One key factor or RJ-MCMC is that it is designed to address changes of dimensionality issues. 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 which branching structure and size changes as a function of the strategy[50]. 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 $(m', n'_m) = (g_{1mm'}(m,u), n'_m)$. 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: $\mathrm{supp}(g_{mm'}) \neq \varnothing$, 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 normalizing constant. Many problems in data analysis require the unsupervised partitioning. Roberts, Holmes and Denison [113] re-considered the issue of data partitioning from an information-theoretic viewpoint and shown that minimization of partition entropy may be used to evaluate the most probable set of data generators which can be employed using a RJ-MCMC.

B. Game Theory Review

In this Section we present relevant concepts from the world of Game Theory.

1) Prisoner's Dilemma: The prisoner's dilemma (PD) is a well known standard example of a game. The way it is usually explained is in the context of a situation involving 2 prisoners who have organized illegal actions for which they have been caught by a third party (the police) who however needs confessions from either of the prisoners in order to abide by the complex legal proceedings. The prosecutor wants to close the case and send someone in prison (at least one of the two suspects) so he offers a deal involving a confession against a more lenient judgment. Both captives are offered this deal independently and away from each other. If the criminals both cooperate (C), nobody goes to prison but they each get a heavy fine. If one denounces[51] (D) the

[50] See in section III and Figures 6, 9, 10 and 11.

[51] Sometimes also referred in the literature as "Deceits".

other, then he will be free without any fine, but the one being denounced has to go to prison and get a fine. If they each denounce each other they go to prison without a fine. Broadly speaking that little story can be formalized into a 2 by 2 matrix[52] with CC, CD, DC and DD with respective payoffs (2,2), (0,3), (3,0) and (1,1). Although the prisoners should clearly cooperate here, given that they do not know what the other is going to do, by expectation (with equal probability for a C and a D) any of the two users should denounce the other given that the expectation of the payoff for denouncing is 2 as opposed to a 1 for a cooperation. This is the reason why this game theory concept is referred to as a "dilemma".

2) Axelrod's Computer Tournament: However this optimal strategy in a "single interation" presented in Section III-B.1 changes when the game becomes iterative. This concept was formalized by Robert Axelrod [6], [7]. Indeed, he designed a computer tournament aiming at understanding what makes a strategy optimal in the context of an ecosystem in an iterative format. In that occasion he invited few Mathematicians, Computer Scientists, Economists and Political Scientists to code a strategy they believed could win such tournament with the constraints of a PD rules in which it is not known when the tournament will stop[53]. Many strategies were thrown into this ecosystem in this form of computer tournament. The range of strategies went from being being very simplistic like "Always Deceit" (AD)[54] to many other more complicated strategies which generic representation can be looked at in Figure 16b). Surprisingly the Tit For Tat (TFT) strategy came at the top of this tournament. The TFT is considered in the literature to be a nice strategy, meaning that it is never the first to deceit (its first move is by design to be a C), but it is also a strategy that is able to retaliate in situation in which it was previously deceived. Finally, it is a strategy that is able to forgive: meaning that if it sees that the adversary algorithm has decided to cooperate after a deceit, then he switches back to a C.

3) Evolutionary Dynamics: Martin Nowak [104] enhanced some of Axelrod's work by introducing new strategies and further developing the concepts of invasion/dominance[55] within a competitive strategic ecosystem. For instance we can see from Figure 16d) that some strategies invade others but these latter strategies can be in turn invaded by other ones which in turn can be invaded by the very first strategy mentioned and induce cycles[56]. Indeed an ecosystem composed of a set of unbiased random strategies (that would randomly C or D) would invite the invasion of an ALLD (always defect) kind. In turn the frequency of ALLD would take the ecosystem which would invite the TFT strategy which would benefit from the mutual cooperation within the same proximity. This process continues in a similar fashion. Figure 16d) exposes how some of these strategies may interact with each other. The following additional information may help in refreshing what some of these acronyms mean: The main

Acronym	Strategy	Description
TFT	Tit for Tat	C first then replicate
GTFT	Generous Tit for Tat	Less grudge prone than TFT
WSLS	Win-Stay, Lose-Shift	Outperforms TFT [104], [126]
ALLC	Always Cooperates	Self explanatory
ALLD	Always Deceits	Self explanatory
rand	Random Strategy	C or D with $p = \frac{1}{2}$

TABLE II: Evolutionary Dynamics Related Strategies

takeaway from this parallel was to expose how the rise and fall of strategies can easily be engineered through simple systematic rules based on an ecosystem and how complexity can be induced from simple rules. Figure 16 summarizes some of the main take aways from Axelrod [6], [7] and Nowak's [104], [126] work.

C. Theoretical Biology Review

It was discussed in the 1960s [49] that complexity in an ecosystem insures its stability or to keep the same jargon "communities not being sufficiently complex to damp out oscillations" [33], [56] have a higher likelihood of vanishing. It is however widely accepted, in the context of ecosystem simulation, that complexity should always arise from simplicity [94], [20]. The diversity-stability debate has in fact been ongoing since the 1950s [95] with no consensus being ever reached. It was initially believed that nature was infinitely complex and therefore more diverse ecosystem should insure more stability [95], [81], [32]. This assertion was however ultimately challenged through rigorous mathematical specification [94], [139], [108] in the 1970s and 1980s by using Lotka-Volterra's Predator/Prey model initially published in the 1920's [137], [80] with similar "non-intuitive" results. More recently the work has been extended to small ecosystems of three-species food chain [19]. The intuitive 3 species example we have chosen to discuss is the one containing Sharks (chosen to be the z parameter), Tunas (chosen to be the y parameter) and Small Fishes (chosen to be the x parameter), the idea being that tunas eat small fishes which in turn are eaten by sharks. Without loss of generality sharks are assumed to die of natural causes and their decomposing bodies go on to feed the small fishes along with other infinite supply of food for the small fishes. The set of differential equations has been summarized in Equation (12)).

Definition (Lotka-Volterra 3-Species Predator Prey): Let a be the natural growth rate of species $x(t)$ (with

[52]Figure 16a.

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

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

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

[56]one may extrapolate that economical cycles may be influenced by similar kind of processes.

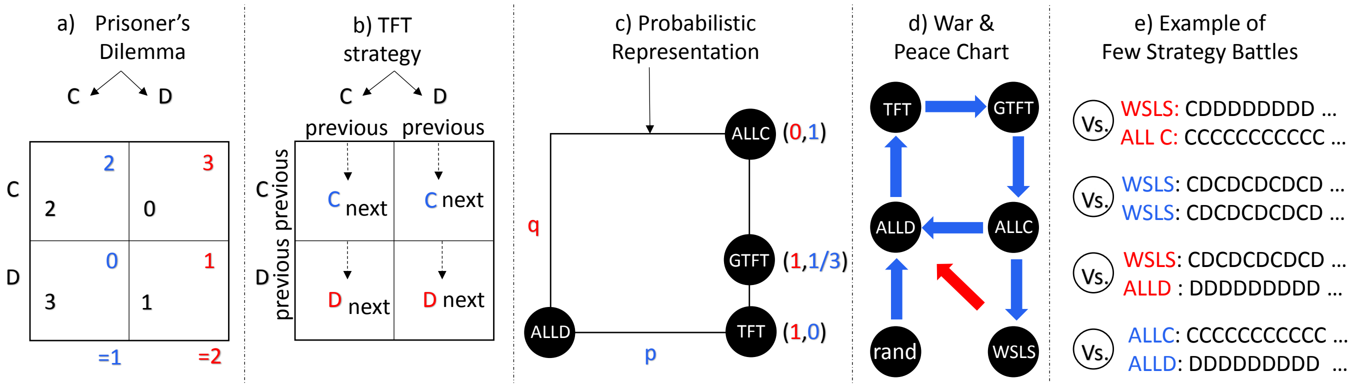


Fig. 16: Some classic Game theory representations [104].

$\mathbb{R} \xrightarrow{x} \mathbb{R}$) in the absence of predator, d the one of $y(t)$ (with $\mathbb{R} \xrightarrow{y} \mathbb{R}$) in the absence of $z(t)$ (with $\mathbb{R} \xrightarrow{z} \mathbb{R}$). We also have b representing the negative predation effect of $y(t)$ on a and e the one of $z(t)$ on $y(t)$. We also have g which mirrors the efficiency of reproduction of $z(t)$ in the presence of prey $y(t)$. The relationship between $x(t)$, $y(t)$ and $z(t)$ is given by Equation (12).

$$\begin{cases} \frac{dx(t)}{dt} &= ax(t) - bx(t)y(t) \\ \frac{dy(t)}{dt} &= -cy(t) + dx(t)y(t) - ey(t)z(t) \\ \frac{dz(t)}{dt} &= -fz(t) + gy(t)z(t) \end{cases} \tag{12}$$

Remark Note that we assume that $x(t)$ never dies of natural causes (if it's too old then it can't run fast enough to outrun predator $y(t)$) but this is not the case for $z(t)$ since it is an alpha predator and therefore needs some natural death rate which is controlled by f.

This relatively simple system of three equations has been studied extensively [95] for stability, for example via Lyapunov coefficient [76] and the eigenplane of the Jacobian matrix [19]. There are different traditional ways to represent stability or instability for these kind of equation, for example Figure 17 represents a 2D stable representation and Figure 19 represents a particular 3D unstable representation. For the latter case, we can notice that the oscillations between the 3 species increases to the point, here not shown, where the amplitudes are so big that z goes instinct and at which point x and y start oscillating, with however a constant amplitude. We refer the motivated reader back to the original papers [95] for the other cases and interesting idiosyncratic properties. One interesting point to notice is that in cases of "relative best stability", in which $a = b = c = d = e = f = g = 1\%$ from Figure 17, we have oscillation which are stable through time with the highest peek from the ultimate prey (x) coming first and the lowest peek of the ultimate predator (z) coming last. This suggests that sophisticated working trading strategies[57] need enough prey like strategies[58] in the same ecosystem to get them to be profitable. One other interesting observation is that the total ecosystem population as depicted in the thick black line from the same figure suggests that it itself oscillates which may not necessarily be intuitive. Indeed one could have speculated that the loss of a species directly benefits the other and that therefore the total population should stay constant. This interesting observation suggests that the oscillations of a financial market *may likewise be subject to similar dynamics*[59]: a financial ecosystem may go through periods in which it thrives followed by period in which it declines. The economy itself is somewhat of a noisy version of Figure 19. The stunning similarities of the competitive resource driven biological ecosystem along with some compelling similarities in some of its cyclical behavior makes the Lotka-Volterra n-species food chain equation an interesting candidate when it comes to studying the stability of the financial market especially the electronic trading markets because of its systematic rule based approach and non zero sum game like roots.

D. Formalizing the Evolutionary Process

With the aim of providing intuition with respect to the sort of interaction that may occur between strategies, we need to formalize the Evolutionary Process (EP), but first, we go through few definitions.

Definition (Evolutionary Process): In the context of our study, the set of rules that control the continuous change of an ecosystem and more specifically its agents (e.g. strategies), will be arbitrarily called Evolutionary Process.

Definition (Iteration Types): We will define two types of iterations. The first type of iteration will be called **Micro**, corresponding to an infinitesimal increment in our environment namely, an increment in which a strategy S analyses and in turn changes the order book by placing an order itself. The second type of iteration will be called **Macro**, corresponding to a generational increment in our environment namely, *a certain equal number* of Micro increment

[57] perhaps from top algorithmic desks in top tier investment banks?

[58] perhaps the retail clients of the world?

[59] It is worth to mention that the oscillations of the financial market may be due to totally different reasons. We revert back here to what we mentioned in the introduction of this Section. More specifically, the relationship between theory and simulation in the scientific method, namely that a good theory should be simulated and a good simulation should be able to help in polishing the theory [135]. At this stage of the Section, we would like to help the reader to follow the train of thought of the author.

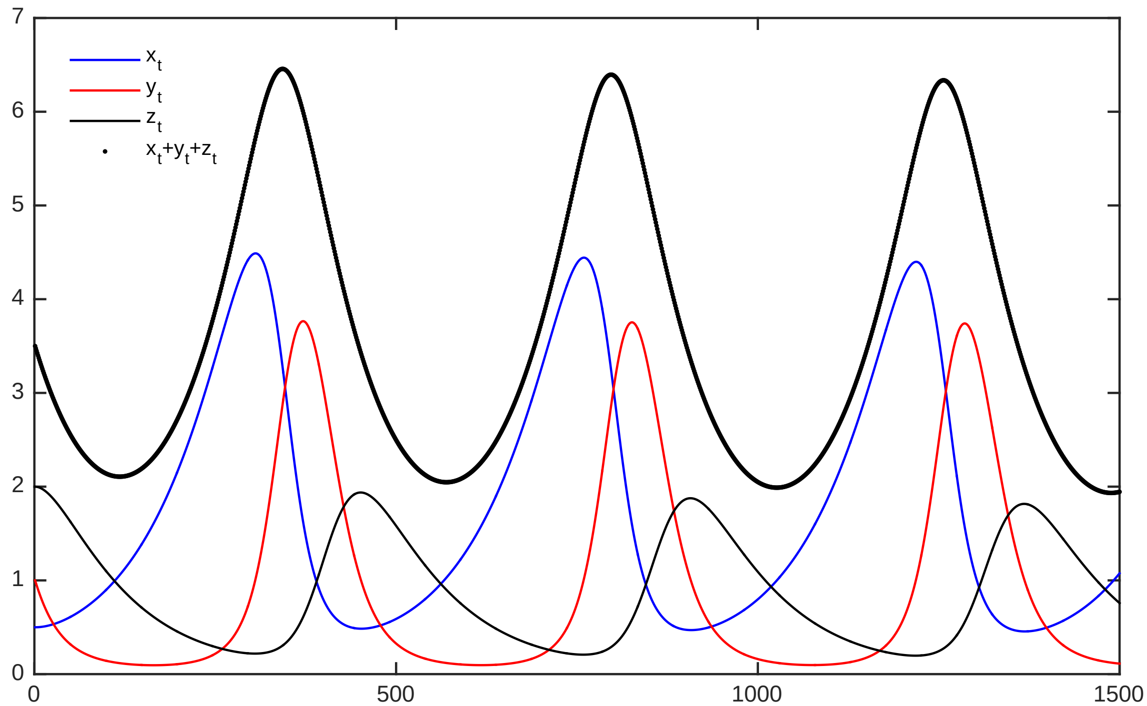


Fig. 17: Stable 3-species Lotka-Volterra Simulation

per strategy leading to a calculation of profit and loss (P&L) and a survival process[60] based on this P&L.

Strategy A strategy will consist of an HFFF $\mathcal{H}$, a rolling P&L $\mathcal{P}$ and a common orderbook $\mathcal{O}$ as shown by Equation (13)).

$$\mathcal{S} \triangleq \{\mathcal{P}, \mathcal{H}, \mathcal{O}\} \tag{13}$$

Definition (Alive/Survived Strategy): A strategy is defined as alive if it does currently take action on the EP.

Definition (Dead Strategy): A strategy is defined as dead if it no longer take action on the EP.

Definition (Born Strategy): A strategy is born if it will take action on the EP for the first time on the next iteration.

Definition (Strategy Classification): We will label N_k the number of total alive strategies, N_k^e the number of tr**e**nd following like strategies, N_k^m the number of **m**ulti-linear regression like strategies, N_k^r the number of xo**r** like strategies and N_k^o the number of ***o****ther unclassified* strategies[61]. The relationship between these entities can be summarized by Equation (14)).

$$N_k = N_k^e + N_k^m + N_k^r + N_k^o \tag{14}$$

Remark One may ask why we have not chosen the first letters of each of the strategies ("t" for trend following, "m" for multi-linear regression and "x" for XOR strategy). The reason why this has been named this way is because as we will see in Section III-C N_k^e behaves in mathematical biology like the number of pr**e**ys in a Lotka-Volterra (LV) 3 species equations [19], that N_k^m behaves in mathematical biology like the number of **m**ixed (both prey and predator) in the same system of equation and that N_k^r behaves in mathematical biology like the number of super predato**r**s as the third species of that system of equations. The different possible permutations, constraints on the first letters being different for each type of strategy and the association to the LV 3 species equation, made the choice of e, m and r at first glance the most optimal in this qualitative optimization by constraint problem.

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

Definition (Survivor Set): The Survivor set[62] is the set of strategies with a positive P&L. Namely if $\mathcal{S}_a = S_{(1)}, S_{(2)}, \ldots, S_{(s)}$ with $S_{(s)} \geqslant 0$ and $S_{(s+1)} < 0$. We will subdivide this set by distinguishing the secondary survivors set which carnality $a_2 = \left\lfloor \frac{s}{2} \right\rfloor$, survive without reproducing and the primary survivors set which carnality $a_1 = s - a_2$, which survive and have one offspring with a "slightly different DNA" in form of a conditional resampling of their NN architecture.

Definition (Birth Process): We will call the Birth process, the first half of survived strategies. Namely, if $a_1 = b = \left\lfloor \frac{s}{2} \right\rfloor$

[60]explained next.
[61]This label will be the same in Section III-C.
[62]or alternatively alive process.

the strategies $\mathcal{S}_1 \ldots \mathcal{S}_{a_1}$ will both survive and reproduce and create a set of equal size but with a slightly different HFFF and with carnality $b = a_1$.

Definition (Death Process): We will call the Death process, the set of strategies with a negative P&L. Namely if $\mathcal{S}_d = S_{(s+1)}, S_{(s+2)}, \ldots, S_{(N_k)}$ will disappear from the market at Macro iteration $k+1$.

Remark We can easily see that $s = a_1 + a_2$, $a_1 \geqslant a_2$, $a_1 = b$. Figure 18 illustrates these few definitions.

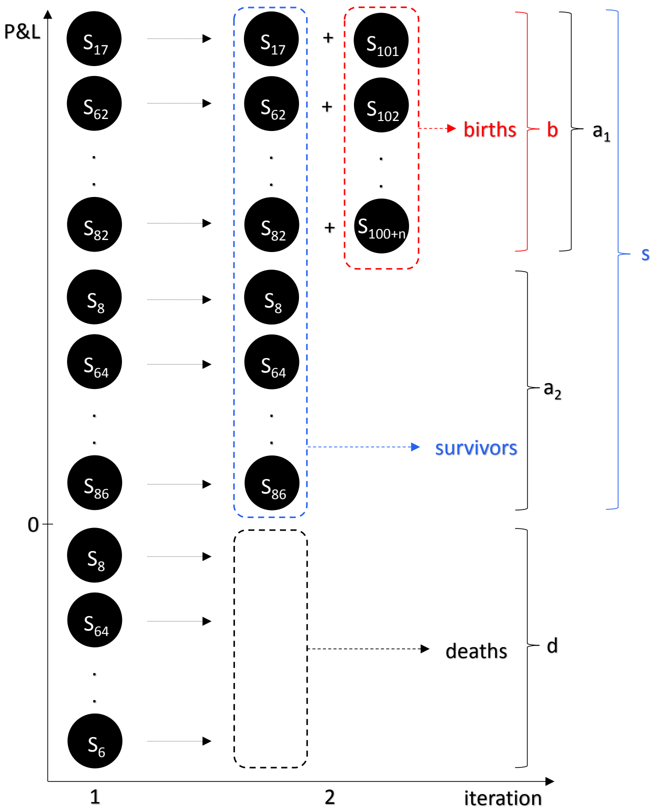


Fig. 18: Death and Birth processes in our GA

2) Inheritance with Mutations : The intuition about the mutation process is that each birth is function of a successful strategy (the positive P&L of parents $\mathcal{S}_1 \ldots \mathcal{S}_{a_1}$) should resemble a great deal to that single parent[63] which produced it but be at the same time be a bit different to allow the ecosystem to evolve. We have seen in Section II-E that the DNA of our strategies is essentially their HFFF $\mathcal{H}$ (which is itself a combination of weights). We will therefore concentrate on performing the re-sampling on the weights of the offspring. Recall that the pdf of the beta distribution, is given by

$$\mathcal{B}(x, \alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\, x^{\alpha-1}(1-x)^{\beta-1} \tag{15}$$

with $\Gamma(n) = (n-1)!$, $0 \leqslant x \leqslant 1$, and $\alpha, \beta > 0$, the shape parameters. The reason why this distribution is interesting is that it is defined in a closed interval [0,1] and can therefore be rescaled easily through a change of variable to [-1,1], an interval which is a basic way of formalizing a normalized importance of each node in the NN architecture decision making of Figure 7. It also offers a broad range of interesting shapes allowing the possibility to code a conditional resampling model and therefore make clever proximity changes around the symbolic levels: -1, 0 and 1. This way we can prevent too large deviations and rather select small incremental changes and intuitively follow the principles of selection. We can see that the $\mathcal{B}(x, 1, 7)$ or $\mathcal{B}(1-x, 1, 7)$ both concentrate a great deal of the distribution towards 0 and 1 respectively. Likewise $\mathcal{B}(x, 3, 7)$ and $\mathcal{B}(x, 5, 7)$ provide a more unbiased modification of the deviations since more symmetrical.

[63] so no crossover in this model.

Definition (Mutation Sampling): Each branch of the HFFF described by Equation (2) will be sampled according to Equation (16).

$$\mathcal{D}(\tilde{x}) = \mathcal{B}\Big(\frac{\tilde{x}+1}{2}, \alpha(\tilde{x}), \beta\Big) 1_{\tilde{x} \leqslant -\frac{1}{5}} \tag{16a}$$

$$+ \mathcal{B}\Big(1 - \frac{\tilde{x}+1}{2}, \alpha(\tilde{x}), \beta\Big) 1_{\tilde{x} > \frac{1}{5}} \tag{16b}$$

$$+ f(\tilde{x}) 1_{|\tilde{x}| \leqslant \frac{1}{5}} \tag{16c}$$

$$\text{where } \alpha(\tilde{x}) = \begin{cases} 1, & \text{if } 1 > |\tilde{x}| \geqslant \frac{3}{4} \\ 3, & \text{if } \frac{3}{4} > |\tilde{x}| \geqslant \frac{1}{2} \\ 5, & \text{if } \frac{1}{2} > |\tilde{x}| \geqslant \frac{1}{5} \end{cases} \tag{16d}$$

$$\text{and } F(k) = \begin{cases} \frac{1}{10}, & \text{if } k \leqslant -\frac{1}{5} \\ \frac{8}{10}, & \text{if } |k| < \frac{1}{5} \\ 1, & \text{if } k \geqslant \frac{1}{5} \end{cases} \tag{16e}$$

$$\text{and } \tilde{x} \in]-1, 1[\text{ and } \beta = 7 \tag{16f}$$

Remark The function $\alpha(\tilde{x})$ models the interval of condition and is arbitrary chosen, though constructed by noticing that the mode of the Beta distribution is given by $\frac{\alpha-1}{\alpha+\beta-2}$ and also so as to make the fractions easy and the intervals loosely equal.

E. Evolutionary Dynamics Simulation

1) Observations: Following Cedric Villani's [135] comment on the relationship between theory and simulation, more specifically around how simulations can give us good intuition about the theory, we lay forward the results which helped us formalize the hypothesis that we investigate latter on in the Section more specifically when it comes to the kind of interactions that may take place. These interactions will be formally addressed through the concept of Path of Interaction that we will introduce in Subsection III-F. However in the meantime, in order to discuss the matter at the intuitive level only we will call "HFFF 1" *Trend Following* (TF), "HFFF 2" *Multi-Linear Regression* (MLR) and "HFFF 3" *XOR*. We may interpret the following. First the market was bullish in the first parts of the zones then became bearish in the next parts of the zones, the TF type strategies, first increases in frequencies then diminishes suddenly in the middle of the zones, the MLR type strategies increases in a short burst right in the middle of the zone and immediately decreases. Finally the XOR strategies frequency increases suddenly in the middles of the zones and decreases slowly.

2) Possible Hypothetical Interpretation: At the early stages of research, we hypothesized the following incomplete interpretation. TF strategies are what people commonly call self fulfilling like prophecies meaning that they only work as long as everyone making up the competitive environment follow the same trend. The biological mirror as described from Section III-C would be an ultimate prey which given an environment without any predator would never die and actually grow exponentially. The XOR strategy is hypothesized as being a super predator strategy (similar to the z parameter in Section III-C) and feeds on the MLR strategies. MLR are hypothesized as being both predator and prey strategies. It feeds onto the TF strategies but are used as prey by the XOR strategies. The way the MLR dominates the TF strategy is due to the fact that it looks at additional leading information on the orderbook (the volumes at the different depth of the order book) so it is leading in the trend whereas the TF is lagging on the trend. XOR strategies can only survive if enough prey (MLRs) are present in the ecosystem otherwise it dies. The way the XOR strategy dominates the MLR strategy is due to its ability to hide its cards better and is able to better decipher spurious positions at higher depths of the orderbook. The XOR strategy cannot invade the TF strategies on its own since the sophistication of its bait (the systematic strategy built to bait the MLR) is too complex to trick the TF. An analogy could be made with a professional poker player playing with a beginner whose moves are almost random.

Remark It has been speculated that the need for a bigger brain in men is partly due to the need for human to elaborate deceitful strategies with their rivals and cooperative strategies with their allies. It is therefore not entirely ridiculous to associate increased neural network branching (to be roughly understood as increase in cranial size) with increased strategy complexity. However, increased intelligence does not necessarily equate to survival. A way to illustrate this is to observe the shark population, which is considered like an apex predator but with a relatively small brain that has not evolved for millennia. By analogy we could speculate that strategies with increased complexity may win in the short run but may not necessarily prevail in the long run.

3) Regulatory Implications: The second and last immediate application we will take a look at in the context of this paper is the one of systemic risk. Given that this paper proposes that the fluctuations of the markets are linked to the frequency of the strategies composing the ecosystem of the market, we propose a model which would take advantage of this assumptions to propose to build the first few steps of a theory that would help high level regulations. The exercise would consist of monitoring these strategies interactions and flag the market when necessary. This may sound a bit grand or overly ambitious but for the sake of opening up a discussion or at least exposing the benefits of future research let us develop a bit the argument. Suppose now that we label strategies of Figure 8, 10 and 14 by respectively x, y and z and that we use Equation (12)). If we can somehow correctly classify and guess what the frequency of x, y and z are in the ecosystem, then we can study whether or not the ecosystem is stable [19]. Now going back to the actual mathematical study of the stability of the financial market. Answering if a financial market composed of 3 strategies is stable would come to studying the Jacobian matrix J from Equation (17)).

$$J(x,y,z) = \begin{bmatrix} a-by & -xb & 0 \\ yd & -c+dx-ez & -ye \\ 0 & -zg & -f+gy \end{bmatrix} \quad (17)$$

By examining the eigenvalues of $J(x,y,z)$ we can indirectly gain information around the equilibrium of our financial system at the regulatory level[64]. More specifically if all eigenvalues of $J(x,y,z)$ have negative real parts then our system is asymptotically stable. Figure 19 gives an illustration of a situation in which one of the eigenvalues is negative. Many questions could be raised here: how can the regulators gain information on the parameters composing the system of equations (12)? Also the market has surely more than 3 types of strategies, how many exactly? Are these strategies easily classifiable in terms of prey, predator and super predator or can you find more subtle instances? It is very likely that trading desks especially in the high frequency domain refuse to provide their sets of strategies for the regulators to study the Jacobian matrix in order to take the relevant actions[65]. We take this opportunity to recall the conjecture we introduced in our last paper [86]:

Conjecture 1 (Diversity & the Financial Stability): Diversity in financial strategies in the market lead to its instability.

4) Optimal Control Theory: The Hamilton-Jacobi-Bellman (HJB) partial differential equation [10] was developed in 1954 and is widely considered as a central theme of optimal control theory. Its solutions is the value function giving the minimum cost for a given dynamical system and its associated cost function. Solved locally, the HJB is a necessary condition, but when over the entire of state space, it is referred to as necessary and sufficient for an optimum. Its method can be generalized to stochastic systems. Its discrete version is referred to as the Bellman equation and its continuous version, the Hamilton-Jacobi equation.

5) Optimal Control Formalization: Formally we consider the problem in deterministic optimal control over the time period $[0,T]$:

$$V(x(0),0) = \min_u \left\{ \int_0^T C[x(t),u(t)]\,dt + D[x(T)] \right\} \quad (18)$$

where $C[]$ is the scalar cost rate function, $D[]$ is the utility at the final state, $x(t)$ the system state vector with $x(0)$ usually given, and finally $u(t)$ where $0tT$ is called the control vector we aim at finding. The system of equation is also subject to

[64] we assume for the sake of this example that we only have 3 strategies.

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

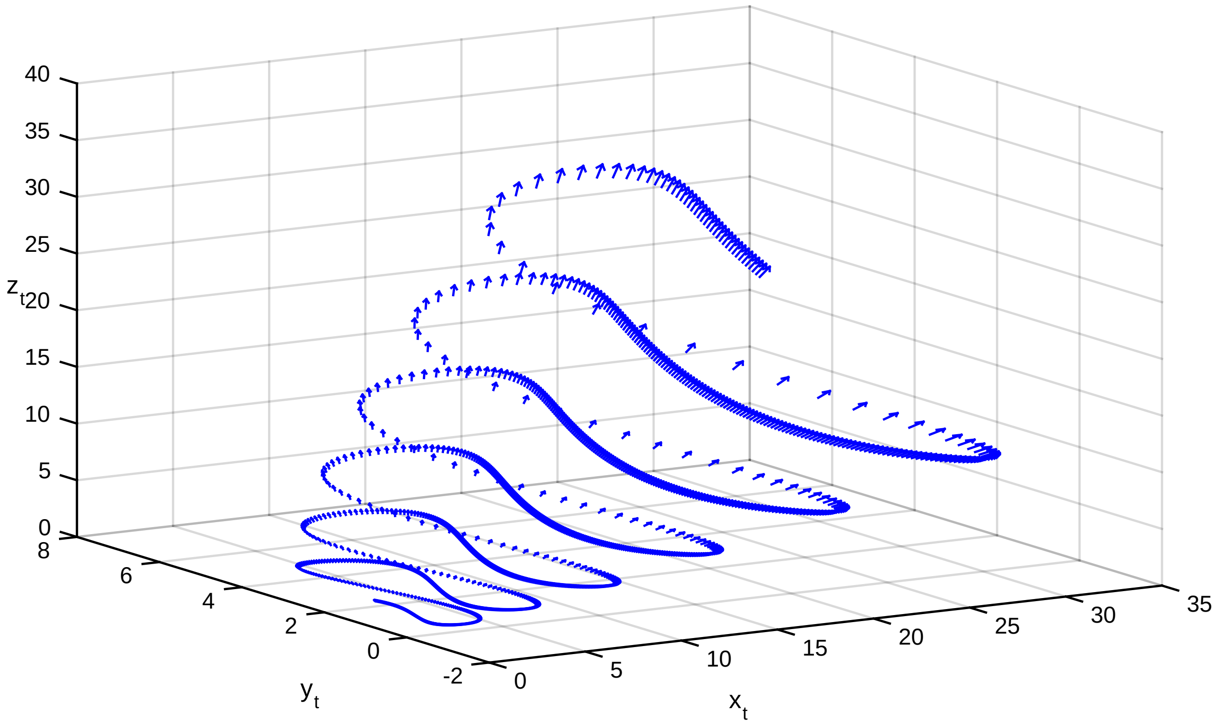


Fig. 19: Unstable 3-Species Lotka-Volterra Simulation

$\dot{x}(t) = F[x(t), u(t)]$ where $F[]$ is a deterministic vector describing the evolution of the state vector over time.

6) Partial Differential Equation Specification: The HJB partial differential equation is given by:

$$\dot{V}(x,t) + \min_u \{\nabla V(x,t) \cdot F(x,u) + C(x,u)\} = 0 \quad (19)$$

subject to the terminal condition $V(x,T) = D(x)$. $V(x,t)$, commonly known as the Bellman value function (our unknown scalar) represents the cost incurred from starting in x at time t and controlling the system optimally until T.

7) Equation derivation: $V(x(t),t)$ is the optimal cost-to-go function, then by Bellman's principle of optimality from time t to $t + dt$, we have $V(x(t),t) = \min_u \left\{ V(x(t+dt), t+dt) + \int_t^{t+dt} C(x(s), u(s))\, ds \right\}$. The Taylor expansion of the first term is $V(x(t+dt), t+dt) = V(x(t),t) + \dot{V}(x(t),t)\, dt + \nabla V(x(t),t) \cdot \dot{x}(t)\, dt + o(dt)$ where $(o)(dt)$ denotes the higher order terms of the Taylor expansion. Canceling $V(x(t),t)$ on both sides and dividing by dt, and taking the limit as dt approaches zero, we obtain the HJB equation. Its resolutions is done backwards in time which can be extended to its stochastic version. In this latter case we have $\min_u \mathbb{E} \left\{ \int_0^T C(t, X_t, u_t)\, dt + D(X_T) \right\}$, with this time $(X_t)_{t \in [0,T]}$ being stochastic and needing optimization and $(u_t)_{t \in [0,T]}$ the control process. By first using Bellman and then expanding $V(X_t, t)$ with Ito's rule, one finds the stochastic HJB equation $\min_u \{\mathcal{A}V(x,t) + C(t,x,u)\} = 0$ where $\mathcal{A}$ represents the stochastic differentiation operator, and subject to the terminal condition $V(x,T) = D(x)$[66].

F. Path of Interaction

Our first few simulations, despite not fulfilling the burden of proof, opened our eyes up to issues associated to optimality, need for more scientific rigour and perhaps an alternative way to fulfill this burden of proof. The concept of Path of Interaction that we introduce next is an attempt at addressing this alternative methodology.

1) HFTE Game: One way to control our simulation issues, is to perhaps take a step back in complexity in order to gain momentum in constructing a theory with more rigor. With this in mind we have chosen to inspire ourself from the scientific method used by Axelrod [6], [7] extended by Nowak's [104], [126], and to introduce a mathematical object, similar in spirit to the PD matrix used as a battle ground (Figure 16) by the name of Path of Interaction. In order to do this rigorously. Let's first go through few definitions.

Definition (Dynamic Mini Order-Book): We will call a Dynamic Mini Order-Book o, the sequence of length l of static snapshots of the order-book ${}^{a_2,a_1}\mathtt{M}_{b_1,b_2}$ of asked and bid volumes a_i/b_i where i corresponds to the depth of the order book and $\mathtt{M}$ its mid price.

Remark In the context of our study we will take $l = 4$.

Definition (Ranking Rule): A Ranking Rule are the set of directives that decides the Birth, Death and Survival processes of any Strategy Ecosystem.

[66]the randomness has disappeared.

Definition (Environment): We will call an Environment e of size i a set of evolving strategy, $S = s^a, s^b, \ldots, s^i$ of HFFF spanning the one from Figure 7 with potential to interact with each other one after the other via an order-book, ${}^{a_2,a_1}\mathtt{M}_{b_1,b_2}$.

Remark Note that the Ranking Rules we assume going forward are the one described by Figure 18. The environment can then evolves according to a set of Ranking Rules.

Definition (HFTE Game): We will call an HFTE Game the sequence of Environments composed of 2 strategies, $S = s^a, s^b, \ldots, s^i$ of HFFF spanning the one from Figure 7 with a dynamic mini order-book and P&L.

Definition (Full Order-Book (FOB)): An OB will be called full if and only if it has a volume of 1 on all the depth of the OB.

Definition (Path of Interaction Table): We will call an Path of Interaction Table an HFTE Game decomposed in its most infinitesimal steps.

The top row of the table points to the strategy involved in the relevant column. The row below (2nd row from the top) provides the stage of the HFTE Game. The 3rd row corresponds to the trading signal. The game starts in a states of in which none of the two strategies has a position (Signal = "N/A") on the order book. Because each strategy needs some form of information on the order book, we take as assumption that there is a random seed on the order book. There is four possibilities of random seeds corresponding to whether the price has been going up or down last and whether the order book has increased its OI or decreased it. These four situations are symbolized by the following set of symbols: ⇈, ⇅, ⇵ and ⇊. We have chosen the case of ⇈ to illustrate our examples arbitrarily. The 4th row corresponds to the order book state. The latter can be either scarse or full. We will see that this latter point matters but for now let us illustrate this point with an example. In Table III, we start $\mathtt{P}^{1,1,1,1,1,1}$ meaning that at the current price P, we have one order to sell at the first 6 depths of the order book. The 5th row corresponds to the current price (last completed order) or the mid price if no order was completed in the current iteration. The 6th row corresponds to the Open Interest. If the buy side of the order book has one of its orders matched then the OI decreases by 1 (-1 if the opposite occurs). The 7th row corresponds to the price change. If no order is matched, then the price is approximated by the mid price. The last row corresponds to the profit and loss. In order to illustrate the Path of Interaction we propose to go through the details of a TF strategy interacting with another TF. Algorithm 3 represents our simplified TF strat and Table III represent the Path of Interaction of two strategies following the systematic rules of Algorithm 3. In this table, for convenience sake[67], we have chosen to represent only one side of the order book: $\mathtt{P}^{1,1,1,1,1,1}$ (for display purposes seeing that the price only takes one direction in the simulations). Since both strategies

Algorithm 3: Simplified TF Strategy

Input: $s, \Delta O, \Delta P$
Output: o

Our simple TF Strategy copies last update's trend while disregarding OI
if $\Delta P > 0$ **then**
  order← 1
else if $\Delta P < 0$ **then**
  order← -1
else
  order← 0
end
return order

follow the trend, and that the order book is full, the price keeps increasing, their respective P&L keeps increasing and the OI imbalance keeps decreasing. Table III can therefore be seen as a rigorous proof that the TF strategies interacting with each other is "self fulfilling", a terminology we introduce more rigorously next. We introduce before the concept of Invasion Flow Chart which is mirror concept of evolutionary dynamics applied to quantitative strategies through the mean of the HFTE Game instead of the PD Matrix. We go first through few formal definitions.

Definition (Invasion): We will call a strategy, s invasive with respect to an environment, e when the P&L of s increases through the HFTE Games taking place in the environment e.

For instance, if we assume that, the more complex a network is, the more likely it is to invade, up to a point where overfitting makes the network obsolete in it performance then we would expect to see an invasion flowchart like the one in Figure 20. Indeed if we assume a TF brings some sort of information innovation from a random strategy and if we assume that the MLR sees more information than the TF and so on then Figure 20 represents a flow chart that exhibits the idea that TF strategies would invade an environment composed of random strategies, that TF would in turn be invaded by MLR, which would be invaded by XORs etc ... This chart also assume that beyond XOR strategies, the complexity would be such that it would equate to a random strategy or would alternatively take a complex path which would lead to a farmer like strategy. We will illustrate later on in this Section that hypothesis illustrated by Figure 20 is not necessarily verified.

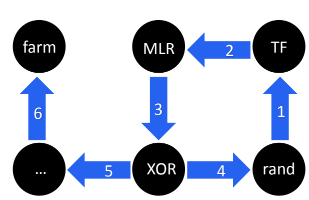


Fig. 20: Illustration for a hypothetical Strategy Invasion Map

[67]The price dynamics goes in only one direction in this case.

Strategy	seed $\upuparrows$	TF1	TF2	TF1	TF2	TF1	TF2
Iteration	0		1		2		3
Signal	N/A	+1	+1	+1	+1	+1	+1
OB	$P^{1,1,1,1,1,1}$	${}_{0}P^{1,1,1,1,1}$	${}_{0,0}P^{1,1,1,1}$	${}_{0,0,0}P^{1,1,1}$	${}_{0,0,0,0}P^{1,1}$	${}_{0,0,0,0,0}P^{1}$	${}_{0,0,0,0,0,0}P$
Mid	100	101	+102	103	104	105	106
ΔOI	$+1$	-1	-2	-3	-4	-5	-6
ΔPrice	$+1$	$+1$	$+1$	$+1$	$+1$	$+1$	$+1$
P&L	$[0, 0]$		$[1, 0]$		$[2, 1]$		$[3, 2]$

TABLE III: Path of Interaction for 2 TF Strategies with $\upuparrows$ Seeds and Full OB

Definition (Self-Fulfilling): We will call a strategy, s Self-Fulfilling when it is Invasive with respect to an environment composed of strategies like itself.

2) Strategy Tournament: Before we discuss our Strategy Tournament, in order to avoid the classification issues mentioned earlier in the paper, we take their most simple forms. First introduce the simplified MLR strategy formalized in Algorithm 4. The idea of this simplified version is that Price and OB imbalance both contribute in defining the trading signal. The last simplified strategy will be the simplified XOR in Algorithm 5.

Algorithm 4: Simplified MLR Strategies

Input: $s, \Delta O, \Delta P$
Output: o

```
Simplified MLR Strategy follows the trend until basic
  OB imbalance
if ΔO + 2 × ΔP > 0 then
  | order← 1
else if ΔO + 2 × ΔP < 0 then
  | order← −1
else
  | order← 0
end
return order
```

Algorithm 5: Simplified XOR Strategies

Input: $s, \Delta O, \Delta P$
Output: o

```
Defining simplified XOR Strategy
if (ΔO > 0) & (ΔP > 0) then
  | order← 1
else if (ΔO > 0) & (ΔP < 0) then
  | order← −1
else if (ΔO < 0) & (ΔP > 0) then
  | order← −1
else if (ΔO < 0) & (ΔP < 0) then
  | order← 1
else
  | order← 0
end
return order
```

A Path of Interaction tournament was implemented in the context of 15 possible games on 7 different timescales: $0, 2, 3, 5, 11, 23, 47$. The choice of these timescales may be a little odd at first glance but the idea was to increase the timescale on average by a factor of two while at the same time picking prime numbers. The idea of the latter is related to an intuition that we had over potential cycles occurring in these games. Though formalizing these possible cycles is premature we thought avoiding chances of getting cycles would make analyzing these interactions easier.

Remark In order to use some conventions around strategy sequences for HFTE games we have chosen the following notation $\xhookrightarrow{s_1} s_2$ and $s_2 \xhookrightarrow{s_1} s_3$ to mean, for the first case, that strategy s_1 changes first the OB, then s_2 (and the sequence continues until the end of the timescale) and, for the second case s_3 impacts the OB after s_2 (before, again going back to s_1). For example, $TF \hookrightarrow TF$ means that the environment is composed of two TF strategies and $MLR \xhookrightarrow{TF} XOR$ refers to an HFTE game composed of a TF, MLR and XOR strategy which OB impact sequence is one which mimics the intuitive order laid down by the $\hookrightarrow$ symbol (TF, first, MLR, second and XOR, third). These symbols are expended into their full form in Tables IV and V but we thought it would be useful to have a text friendlier version for the analysis.

Table IV represents the results of these games for two strategies interacting and Table V represents the same for 3 strategies. We can make several interesting observations.

Proposition The TF strategy is self-fulfilling on a OB that is full.

Proof: We have illustrated this point with Table III. Though only on 4 iterations, the proof can be expanded on longer timescales. ■

Remark The intuition we had [86] around the TF acting like a prey increasing exponentially in frequency in the absence of predator is confirmed. The first connections to the Lotka-Volterra 3-species predator/prey model is established. It is worthy to note however that there is a benefit in starting first

Scenario	$TF \hookrightarrow TF$	$TF \hookrightarrow MLR$	$TF \hookrightarrow XOR$	$MLR \hookrightarrow TF$	$MLR \hookrightarrow MLR$	$MLR \hookrightarrow XOR$	$XOR \hookrightarrow TF$	$XOR \hookrightarrow MLR$	$XOR \hookrightarrow XOR$
Code	s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8	s_9
P&L$_0$	$\begin{bmatrix}0\\0\end{bmatrix}$	$\begin{bmatrix}0\\0\end{bmatrix}$	$\begin{bmatrix}0\\0\end{bmatrix}$	$\begin{bmatrix}0\\0\end{bmatrix}$	$\begin{bmatrix}0\\0\end{bmatrix}$	$\begin{bmatrix}0\\0\end{bmatrix}$	$\begin{bmatrix}0\\0\end{bmatrix}$	$\begin{bmatrix}0\\0\end{bmatrix}$	$\begin{bmatrix}0\\0\end{bmatrix}$
ΔP_0	0	0	0	0	0	0	0	0	0
P&L$_2$	$\begin{bmatrix}0\\0\end{bmatrix}$	$\begin{bmatrix}0\\0\end{bmatrix}$	$\begin{bmatrix}0\\0\end{bmatrix}$	$\begin{bmatrix}0\\0\end{bmatrix}$	$\begin{bmatrix}0\\0\end{bmatrix}$	$\begin{bmatrix}0\\0\end{bmatrix}$	$\begin{bmatrix}0\\0\end{bmatrix}$	$\begin{bmatrix}0\\0\end{bmatrix}$	$\begin{bmatrix}0\\0\end{bmatrix}$
ΔP_2	0	0	0	0	0	0	0	0	0
P&L$_3$	$\begin{bmatrix}3\\2\end{bmatrix}$	$\begin{bmatrix}3\\2\end{bmatrix}$	$\begin{bmatrix}-1\\2\end{bmatrix}$	$\begin{bmatrix}3\\2\end{bmatrix}$	$\begin{bmatrix}3\\2\end{bmatrix}$	$\begin{bmatrix}-1\\2\end{bmatrix}$	$\begin{bmatrix}3\\2\end{bmatrix}$	$\begin{bmatrix}3\\2\end{bmatrix}$	$\begin{bmatrix}-1\\2\end{bmatrix}$
ΔP_3	3	3	-1	3	3	-1	3	3	-1
P&L$_5$	$\begin{bmatrix}9\\7\end{bmatrix}$	$\begin{bmatrix}-9\\8\end{bmatrix}$	$\begin{bmatrix}-1\\4\end{bmatrix}$	$\begin{bmatrix}-1\\10\end{bmatrix}$	$\begin{bmatrix}9\\5\end{bmatrix}$	$\begin{bmatrix}4\\5\end{bmatrix}$	$\begin{bmatrix}3\\5\end{bmatrix}$	$\begin{bmatrix}3\\9\end{bmatrix}$	$\begin{bmatrix}7\\-2\end{bmatrix}$
ΔP_5	6	-3	-3	5	-5	-4	5	-4	3
P&L$_{11}$	$\begin{bmatrix}45\\40\end{bmatrix}$	$\begin{bmatrix}-39\\21\end{bmatrix}$	$\begin{bmatrix}-1\\4\end{bmatrix}$	$\begin{bmatrix}-50\\78\end{bmatrix}$	$\begin{bmatrix}15\\12\end{bmatrix}$	$\begin{bmatrix}14\\7\end{bmatrix}$	$\begin{bmatrix}-9\\26\end{bmatrix}$	$\begin{bmatrix}-16\\21\end{bmatrix}$	$\begin{bmatrix}7\\-2\end{bmatrix}$
ΔP_{11}	15	11	-3	13	15	-6	11	11	3
P&L$_{23}$	$\begin{bmatrix}198\\187\end{bmatrix}$	$\begin{bmatrix}-216\\48\end{bmatrix}$	$\begin{bmatrix}-1\\4\end{bmatrix}$	$\begin{bmatrix}-326\\387\end{bmatrix}$	$\begin{bmatrix}39\\38\end{bmatrix}$	$\begin{bmatrix}74\\11\end{bmatrix}$	$\begin{bmatrix}-87\\122\end{bmatrix}$	$\begin{bmatrix}-21\\54\end{bmatrix}$	$\begin{bmatrix}7\\-2\end{bmatrix}$
ΔP_{23}	33	26	-3	28	36	-10	23	27	3
P&L$_{47}$	$\begin{bmatrix}828\\805\end{bmatrix}$	$\begin{bmatrix}-703\\354\end{bmatrix}$	$\begin{bmatrix}-1\\4\end{bmatrix}$	$\begin{bmatrix}-1580\\1707\end{bmatrix}$	$\begin{bmatrix}96\\99\end{bmatrix}$	$\begin{bmatrix}290\\19\end{bmatrix}$	$\begin{bmatrix}-459\\530\end{bmatrix}$	$\begin{bmatrix}-27\\54\end{bmatrix}$	$\begin{bmatrix}7\\-2\end{bmatrix}$
ΔP_{47}	69	-75	-3	58	78	-18	47	27	3

TABLE IV: P&L in Path of Interaction for 2 Strategies with $\upuparrows$ Seeds and Full OB

Scenario	$MLR \xhookrightarrow{TF} XOR$	$XOR \xhookrightarrow{TF} MLR$	$TF \xhookrightarrow{MLR} XOR$	$XOR \xhookrightarrow{MLR} TF$	$TF \xhookrightarrow{XOR} MLR$	$MLR \xhookrightarrow{XOR} TF$
Code	s_{10}	s_{11}	s_{12}	s_{13}	s_{14}	s_{15}
P&L$_0$	$\begin{bmatrix}0\\0\\0\end{bmatrix}$	$\begin{bmatrix}0\\0\\0\end{bmatrix}$	$\begin{bmatrix}0\\0\\0\end{bmatrix}$	$\begin{bmatrix}0\\0\\0\end{bmatrix}$	$\begin{bmatrix}0\\0\\0\end{bmatrix}$	$\begin{bmatrix}0\\0\\0\end{bmatrix}$
ΔP_0	0	0	0	0	0	0
P&L$_2$	$\begin{bmatrix}0\\0\\0\end{bmatrix}$	$\begin{bmatrix}0\\0\\0\end{bmatrix}$	$\begin{bmatrix}0\\0\\0\end{bmatrix}$	$\begin{bmatrix}0\\0\\0\end{bmatrix}$	$\begin{bmatrix}0\\0\\0\end{bmatrix}$	$\begin{bmatrix}0\\0\\0\end{bmatrix}$
ΔP_2	0	0	0	0	0	0
P&L$_3$	$\begin{bmatrix}3\\2\\0\end{bmatrix}$	$\begin{bmatrix}-2\\3\\3\end{bmatrix}$	$\begin{bmatrix}3\\2\\0\end{bmatrix}$	$\begin{bmatrix}-1\\2\\0\end{bmatrix}$	$\begin{bmatrix}-1\\-2\\2\end{bmatrix}$	$\begin{bmatrix}3\\2\\0\end{bmatrix}$
ΔP_3	3	-2	3	-1	-1	3
P&L$_5$	$\begin{bmatrix}4\\0\\4\end{bmatrix}$	$\begin{bmatrix}-2\\-8\\0\end{bmatrix}$	$\begin{bmatrix}8\\-5\\7\end{bmatrix}$	$\begin{bmatrix}5\\6\\4\end{bmatrix}$	$\begin{bmatrix}-11\\-3\\20\end{bmatrix}$	$\begin{bmatrix}-3\\7\\-4\end{bmatrix}$
ΔP_5	4	3	-4	-5	-6	-3
P&L$_{11}$	$\begin{bmatrix}-26\\60\\-21\end{bmatrix}$	$\begin{bmatrix}-108\\61\\-36\end{bmatrix}$	$\begin{bmatrix}17\\-7\\-34\end{bmatrix}$	$\begin{bmatrix}14\\-18\\-37\end{bmatrix}$	$\begin{bmatrix}10\\-26\\25\end{bmatrix}$	$\begin{bmatrix}-48\\57\\-47\end{bmatrix}$
ΔP_1	17	-14	7	6	18	16
P&L$_{23}$	$\begin{bmatrix}-65\\-13\\119\end{bmatrix}$	$\begin{bmatrix}-164\\131\\-39\end{bmatrix}$	$\begin{bmatrix}57\\-35\\-198\end{bmatrix}$	$\begin{bmatrix}54\\-62\\-201\end{bmatrix}$	$\begin{bmatrix}128\\-93\\-74\end{bmatrix}$	$\begin{bmatrix}137\\145\\-96\end{bmatrix}$
ΔP_{23}	31	-32	15	14	45	-46
P&L$_{47}$	$\begin{bmatrix}-3127\\2500\\-231\end{bmatrix}$	$\begin{bmatrix}-720\\250\\-289\end{bmatrix}$	$\begin{bmatrix}233\\-187\\-910\end{bmatrix}$	$\begin{bmatrix}230\\-246\\-913\end{bmatrix}$	$\begin{bmatrix}187\\-230\\-588\end{bmatrix}$	$\begin{bmatrix}553\\621\\13\end{bmatrix}$
ΔP_{47}	-54	-54	31	30	97	-104

TABLE V: P&L in Path of Interaction for 3 Strategies with $\upuparrows$ Seeds and Full OB as the TF1 does better at the end in this HFTE game.

Proposition A strategy A can invade a strategy B but the latter can invade strategy B if the seed or and the sequence in which these strategies are started changes.

Proof: The MLR strategy invades the TF strategy on the longer times scales (column s_2 of Table IV) but when the MLR starts the HFTE game (column s_4 of Table IV) then TF invades the MLR strategy. The same remark can be made when XOR take the MLR spot in the same HFTE set up (column s_3 and s_7 of Table IV). ■

Proposition The Dominance relation is not transitive.

Proof: This comes to exposing that if a strategy A dominates a Strategy B and Strategy B dominates Strategy C, this does not mean that Strategy A will dominate Strategy C. An example of counterexample is s_2, s_6 and s_3 of Table IV. ■

Proposition Having a more complex strategy does not mean a higher P&L.

Proof: We can observe in column s_7 of Table IV, that the TF strategy invades the XOR strategy over the first 47 iterations even-though the XOR strategy involves a hidden layer, on the contrary to the TF strategy that consist of only 1 input. ■

Proposition All strategies in an Ecosystem can make money even if the market goes down.

Proof: See s_6 example in Table IV. ■

Proposition In a situation of twin strategies interacting within an HFTE game, starting first is not necessarily advantageous.

Proof: See s_5 in Table IV for the example. ■
Finally we notice, in Figure 21 that the 3 strategies ecosystems exhibited more fluctuations than the 2 strategies ecosystems which tend to support the conjecture than more diversity in an ecosystem of strategies induces more instability to the market.

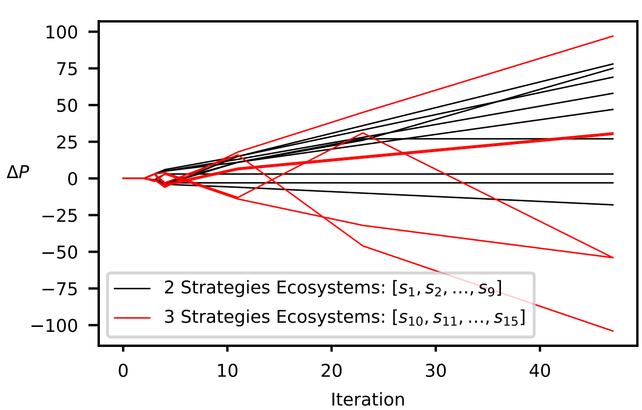


Fig. 21: Market instability and additional strategy.

Conjecture 2: All strategies in an ecosystem can make money at the same time but all cannot lose money at the same time.

Remark We noticed this interesting fact with our sample of HFTE games but have not been able to found a counter example yet nor been able to rigorously prove it.

We have built in this Section the foundations associated to the Bottom-Up approach to algorithmic trading. We first achieved that tackling the problem using a simple genetic algorithm methodology which we have abandoned to a series of problem associated, but not limited to classification. We took this opportunity to shown possible connections to the world of evolutionary dynamics, predator prey models, as well as how the mathematical tools associated to these fields that can be brought in the world of Quantitative Finance. In order to perhaps take a step in ordet to gain momentum in the scientific approach we formalized the HFTE game as well as the Path of Interaction concept. We have also given few examples of such games and also presented few interesting results. Market participants are however quite secretive when it comes to their financial strategy. The only observable data on the market is essentially the price dynamics. We explore in the next Section how inference can be constructed in the Bottom-Up approach when the price dynamic alone is available.

IV. Stability of Financial Systems and Multi-Target Tracking

In this section we take a look at another example in which ML can revolutionize classic Mathematical Finance as it lays down the foundations for controlling systemic risk in a challenging electronic trading environment where speed and secrecy are of utmost importance. More specifically we first offer a literature review of the Multi-Target Tracking starting with the linear and then moving to non linear methods. We then expand the study by connecting some of the concepts in the previous two Sections with particle filtering applied to scenario modelling.

A. Classic Methods in Multi-Target Tracking

Multi-Target Tracking (MTT), which deals with state space estimation of moving targets, has applications in different fields [9], [75], [129], the most intuitive ones being perhaps radar and sonar.

1) Linear Methods: The Kalman Filter (KF) is a mathematical tool which provides the best estimation (in a MSE sense) of some dynamical process,(x_k), perturbed by noise and influenced by a controlled process. The estimation is based upon observations which are functions of these dynamics (y_k). A review can be found in [116]. The observations of the KF are usually referred to in the literature as x_k and the dynamics are given by Equation (20).

$$x_k = F_k x_{k-1} + B_k u_k + w_k \tag{20}$$

where 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 (again we will assume later as null); w_k is the process noise which is assumed to be drawn from a zero mean multivariate normal distribution with covariance 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 (21).

$$y_k = H_k x_k + v_k \tag{21}$$

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

Algorithm 6: Kalman Filter

Input: array of weights w_1^N
Output: array of weights w_1^M resampled

Predicted state:
$\hat{x}_{k|k-1} \leftarrow F_k \hat{x}_{k-1|k-1} + B_{k-1} u_{k-1}$
$P_{k|k-1} \leftarrow F_k P_{k-1|k-1} F_k^{\mathrm{T}} + Q_{k-1}$

Update state:
Innovation (or residual)
$\tilde{y}_k \leftarrow y_k - H_k \hat{x}_{k|k-1}$
Covariance
$S_k \leftarrow H_k P_{k|k-1} H_k^T + R_k$
Optimal Kalman gain
$K_k \leftarrow P_{k|k-1} H_k^{\mathrm{T}} S_k^{-1}$
Updated state estimate
$\hat{x}_{k|k} \leftarrow \hat{x}_{k|k-1} + K_k \tilde{y}_k$
Updated estimate covariance
$P_{k|k} \leftarrow (I - K_k H_k) P_{k|k-1}$

Return state:
Return
w_1^M

vectors ($\{w_1, \ldots, w_k\}$, $\{v_1 \ldots v_k\}$) at each step are 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 above is only valid for the optimal Kalman gain. Usage of other gain values require a more complex formula. The KF methodology has been summarized by Algorithm 6. *Proof:* Please see original papers [63], [64]. ∎

Although the KF presents obviously lots of benefits in tracking, its linear constraints makes it not a first choice in the exercise we have at hand. 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 (22).

$$\begin{cases} x_k &= f(x_{k-1}, u_k) + w_k \\ y_k &= h(x_k) + v_k \end{cases} \tag{22}$$

The algorithm is very similar to the one described in Algorithm (6) but with couple of modifications highlighted below Algorithm (7)[68].

Proof: The proof for algorithm 7 is very similar to the proof of algorithm 6 with couple of exceptions, first F_k and H_k approximations at the first order of F_k and H_k, we obtain a truncation error which technically 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 requires us to continue our literature review. However note, as we will see in Section IV-C.2 that this methodology might be a good candidate when we use the dual mathematical biology problem is used instead of the rough direct method.

2) Non-Linear Methods: Importance sampling (IS) was first introduced in [93] and was further discussed in several books including in [51]. 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

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

Algorithm 7: Extended Kalman Filter

Input: array of weights w_1^N
Output: array of weights w_1^M resampled

Predicted state:
$\hat{x}_{k|k-1} \leftarrow f(\hat{x}_{k-1|k-1}, u_k)$
$P_{k|k-1} \leftarrow F_k P_{k-1|k-1} F_k^{\mathrm{T}} + Q_{k-1}$

Update state:
Innovation (or residual)
$\tilde{y}_k \leftarrow y_k - h(\hat{x}_{k|k-1})$
Covariance
$S_k \leftarrow H_k P_{k|k-1} H_k^{\mathrm{T}} + R_k$
Optimal Kalman gain
$K_k \leftarrow P_{k|k-1} H_k^{\mathrm{T}} S_k^{-1}$
Updated state estimate
$\hat{x}_{k|k} \leftarrow \hat{x}_{k|k-1} + K_k \tilde{y}_k$
Updated estimate covariance
$P_{k|k} \leftarrow (I - K_k H_k) P_{k|k-1}$

of $p(x)$.

$$\int f(x)p(x)dx = \int f(x)\frac{p(x)}{q(x)}q(x)dx \tag{23a}$$

$$\hat{f} = \frac{1}{N_p}\sum_{i=1}^{N_p} W(x^{(i)})f(x^{(i)}) \tag{23b}$$

In Equation (23a) we write the integration problem in the more appropriate form with Equation (23b) the numerical approximation where N_p, usually describes the number of independent samples drawn from $q(x)$ to obtain a weighted sum to approximate $\hat{f}$,

$$W(x^{(i)}) = \frac{p(x^{(i)})}{q(x^{(i)})} \tag{24a}$$

$$W(x^{(i)}) \propto p(x^{(i)})q(x^{(i)}) \tag{24b}$$

and where $W(x^{(i)})$ in Equation (24a) is the Radon-Nikodym derivative of $p(x)$ with respect to $q(x)$ or called in engineering the importance weights. Equation (24b) suggests that if the normalizing factor for $p(x)$ is not known, the importance weights can only be evaluated up to a normalizing constant. To ensure that $\sum_{i=1}^{N_p} W(x^{(i)}) = 1$, we normalize the importance weights to obtain Equation (25).

$$\hat{f} = \frac{\frac{1}{N_p}\sum_{i=1}^{N_p} W(x^{(i)})f(x^{(i)})}{\frac{1}{N_p}\sum_{i=1}^{N_p} W(x^{(i)})} = \frac{1}{N_p}\sum_{i=1}^{N_p} \tilde{W}(x^{(i)})f(x^{(i)}) \tag{25}$$

where $\tilde{W}(x^{(i)}) = \frac{W(x^{(i)})}{\sum_{i=1}^{N_p} W(x^{(i)})}$ are called the normalized importance weights. The variance of importance sampler estimate [17] in Equation (25) is given by $Var_q[\hat{f}] = \frac{1}{N_p}Var_q[f(x)W(x)] = \frac{1}{N_p}Var_q[f(x)p(x)/q(x)] = \frac{1}{N_p}\times \int[\frac{f(x)p(x)}{q(x)} - \mathbb{E}_p[f(x)]]^2 q(x)dx = \frac{1}{N_p}\times \int[(\frac{(f(x)p(x))^2}{q(x)}) - 2p(x)f(x)\mathbb{E}_p[f(x)]]dx + \frac{(\mathbb{E}_p[f(x)])^2}{N_p} = \frac{1}{N_p}\times \int[(\frac{(f(x)p(x))^2}{q(x)})]dx - \frac{(\mathbb{E}_p[f(x)])^2}{N_p}$.

The variance can be reduced when an appropriate $q(x)$ is chosen to either match the shape of $p(x)$ so as to approximate the true variance; or to match the shape of $|f(x)|p(x)$ so as to further reduce the true variance.

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

3) 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 [68] is optimal in terms of variance. Figure 22 gives an illustration of the Stratified Sampling and the corresponding algorithm is described in algorithm 8. We see at the top of the Figure 22 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. Sequential Monte Carlo methods (SMC), also known as Particle Filters (PF) 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

Algorithm 8: Resample

Input: array of weights w_1^M
Output: array of weights w_1^M resampled

Sample:
$u^0 \sim \mathcal{U}[0, 1/M]$

Resample:
for $m = 1$ *to* N **do**
  $i^{(m)} \leftarrow \left\lfloor (w_n^{(m)} - u^{(m-1)}m) \right\rfloor + 1$
  $u^{(m)} \leftarrow u^{(m)} + \frac{i^{(m)}}{M} - w_n^{(m)}$
end

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 (26a) and (26b) gives an example of such system.

$$x_k = f(x_{k-1}) + w_k \tag{26a}$$

$$y_k = h(x_k) + v_k \tag{26b}$$

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 (26a) and (26b) are our state space equations. If

Algorithm 9: Sequential Monte Carlo

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

1 **Sample**:
2 **for** $L = 1$ *to* N_P **do**
3 $\quad x_k^{(L)} \sim \pi(x_k|x_{1:k-1}^{(L)}, y_{1:k})$
4 **end**
5 **for** $L = 1$ *to* N_P **do**
6 $\quad \hat{w}_k^{(L)} \leftarrow w_{k-1}^{(L)} \frac{p(y_k|x_k^{(L)})p(x_k^{(L)}|x_{k-1}^{(L)})}{\pi(x_k^{(L)}|x_{1:k-1}^{(L)}, y_{1:k})}$
7 **end**
8 **for** $L = 1$ *to* N_P **do**
9 $\quad w_k^{(L)} \leftarrow \frac{\hat{w}_k^{(L)}}{\sum_{J=1}^{P} \hat{w}_k^{(J)}}$
10 **end**
11 $\hat{N}_{eff} \leftarrow \frac{1}{\sum_{L=1}^{P}\left(w_k^{(L)}\right)^2}$
12 **Reample**:
13 draw N_P particles from the current particle set with probabilities proportional to their weights. Replace the current particle set with this new one.
14 **if** $\hat{N}_{eff} < N_{thr}$ **then**
15 $\quad$ **for** $L = 1$ *to* N_P **do**
16 $\quad\quad w_k^{(L)} \leftarrow 1/N_P$.
17 $\quad$ **end**
18 **end**

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 (27).

$$\int f(x_k)p(x_k|y_1, \ldots, y_k)dx_k \approx \frac{1}{N_P}\sum_{L=1}^{N_P} f(x_k^{(L)}) \tag{27}$$

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 $(w_k^{(L)}, x_k^{(L)})\ :\ L = 1, \ldots, N_P$. The importance weights $w_k^{(L)}$ are approximations to the relative posterior probability measure of the particles such that $\sum_{L=1}^{P} w_k^{(L)} = 1$. SIR is a essentially a recursive version of importance sampling. Like in IS, the expectation of a function $f(\cdot)$ can be approximated like described in Equation (28).

$$\int f(x_k)p(x_k|y_1, \ldots, y_k)dx_k \approx \sum_{L=1}^{N_P} w^{(L)} f(x_k^{(L)}) \tag{28}$$

The algorithm performance is dependent on the choice of the proposal $\pi(x_k|x_{1:k-1}, y_{1:k})$ distribution with the optimal proposal distribution being $\pi(x_k|x_{0:k-1}, y_{0:k})$ in Equation (29).

$$\pi(x_k|x_{1:k-1}, y_{1:k}) = p(x_k|x_{k-1}, y_k) \tag{29}$$

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 22 gives an illustration of the algorithm just described.

Note that on line 5 of algorithm 9, $\hat{w}_k^{(L)}$, simplifies to $w_{k-1}^{(L)} p(y_k|x_k^{(L)})$, when $\pi(x_k^{(L)}|x_{1:k-1}^{(L)}, y_{1:k}) = p(x_k^{(L)}|x_{k-1}^{(L)})$. Algorithm (9) summarizes the SMC methodology.

B. *Scenario Tracking Algorithm*

1) Introduction: Recently, SMC methods [25], [26], [77], especially when it comes to the data association issue, have been developed. Particle Filters (PF) [44], [67], have recently become a popular framework for 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 [52]. 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 [122], [124], [91], [136], 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 [102]. The M-best feasible solutions is also a new useful finding in SMC [102], [70], [11], [73], [12]. Articles [123], [125] were proposed to cope with both the multitarget detection and tracking scenario but according to [101] 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 [101] 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

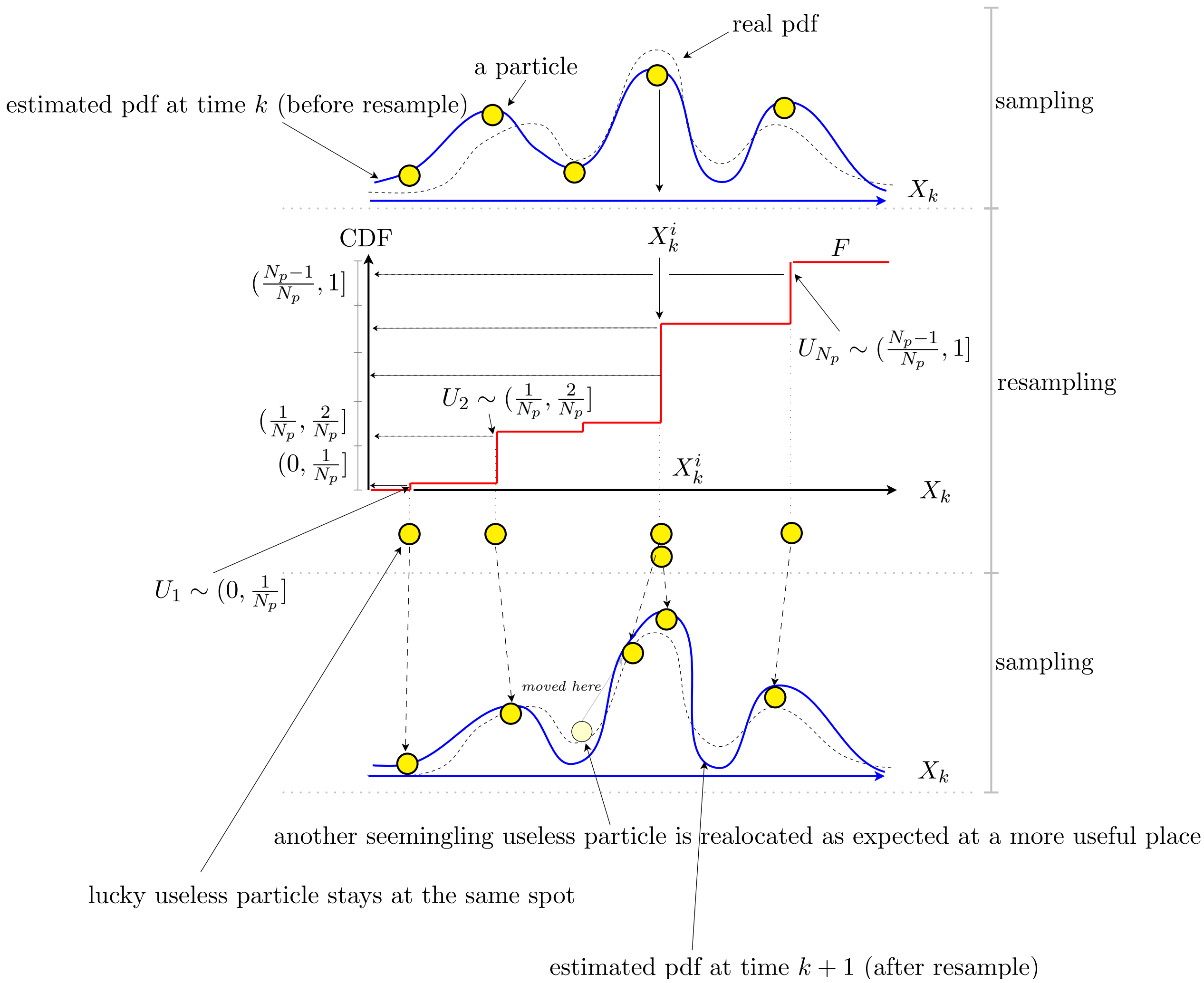


Fig. 22: Stratified Sampling illustration

[103] 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 [74]. That method provided a solution to deal with both time-varying number of targets, and measurement-to-target association issues. 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 [92], [123], [125], [136], 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 [101]. 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) [9], [37]. Its major drawback though is that its tracks tend to coalesce when targets are closely spaced [36] 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 [13], [53]. Unfortunately the track swap problems still remain. Also performance of the EKF [52] is known to be limited by the linearity of the data model on the contrary to SMC based tracking algorithms developed by [57], [47], [46], [54]. This issue of data association can also be sampled via Gibbs sampling [54]. Also because target detection and initialization were not covered by this framework algorithms developed in [134] were suggested in order to improve detection and tracking performance. The algorithm suggested in [134] 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 [117] 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 [117], 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 [134]. In [134], 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 [3] 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 labelling 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 modelling 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 [3] 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) [62] 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) [2] . 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 in [66]. 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 [134]. For non-linear and non-Gaussian models, generally speaking SMC's [26], [78] known alternatively as we have seen as PF [44], [67] but also seldom CONDENSATION [59] 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 described, 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 [66]. It has been defined in [134] that there exists numerous strategies available to solve the data association problem but they could be categorised as either single frame assignment methods, or multi-frame assignment methods. The multi-frame assignment problem can be solved using Lagrangian relaxation [111]. Another algorithm the Multiple Hypotheses Tracker (MHT) [112] 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. The Nearest Neighbour Standard Filter (NNSF) [9] 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) [9], [37] 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 [107], [118] 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) [41], [130] 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 [23]. Unfortunately the PMHT is not suitable for sequential applications because considered a batch strategy. Moreover [138] 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 [121], [120], [38], [65] 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 [5], [45] but the simulations dealt only with a single target. In the article [55] the distribution and the hypotheses of the association is computed using a Gibbs sampler, [42] at each iterations. This method, similar to the one described in [22], uses MCMC [43] 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 on-line applications. Doucet [46] presents a method where the associations are sampled from a well chosen importance distribution. Although intuitively appealing it is, however, reserved to Jump Markov Linear Systems (JMLS) [27]. The follow up of this strategy, based on the UKF and the Auxiliary Particle Filter (APF) [109], so that applicable to Jump Markov Systems (JMS) is presented in [28]. Similar in [57], 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 [14], [119]. 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 [106], and where a selection strategy is addressed to solve this problem. In [134] 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 [121], [120] 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 [134] 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 [57].

C. HFTE SMC Tracking Methodology

Now that we have done a comprehensive review of tracking methodologies we would like to apply our findings to the HFTE formulated problem from Subsection **??**. We will describe first the brute force method in Subsection IV-C.1, then remind the connection the HFTE model has with the predator prey model in Subsection III-C in order to introduce the dual methodology in Subsection IV-C.2.

1) Direct Approach: The direct approach would consist of tracking not only the number of alive strategies and number of births but on top of that the HFFF of each of the live strategies, namely, if each particle is associated to our state space θ then we can summarize our state space by Equation (30)).

$$\theta \triangleq \left\{ \; N_t^s, N_t^b, N_t^d, \cup_{i=1}^{N_t^a} S_i, \cup_{i=1}^{N_t^a} \mathcal{H}_i, \cup_{i=1}^{N_t^a} \mathcal{P}_i, \mathcal{O} \; \right\} \quad (30)$$

with N_t^s, the number of survived strategies, N_t^b, the number of born strategies, N_t^d, the number of dead strategies and N_t^a, the number of alive strategies[69]. As we can see from Equation (30)), not only we need to keep track of the alive strategies through time but also of their HFFF which itself has seemingly insurmountable dimenssionality challenges, notably an exponential likelihood for clustering. If you add to this specific point that you need to keep track of all possible sets of orderbooks $\mathcal{O}$ and P&L $\mathcal{P}$, with a little experience in signal processing, one will realize that this path is a lost battle before it really began. The question is whether there is still something that can be done, this is what we will attempt at answering in Section IV-C.2 but before let's recall the connections the HFTE model has with predator prey models in the following Subsection III-C.

2) Dual Approach: One way to address this indirect approach comes from realizing couple of points. First, a particle filtering exercise based on Equation (30)) is impossible with the technology, and the research status quo. Second, market participants will never divulge their strategies to allow regulators to come up with original risk management systems that we have discussed [86]. However, what can be done is to assume, the HFTE model [86] behaves like the N_t^a Species Lotka-Volterra and therefore the fluctuations of the financial systems can be studied indirectly by studying the mirror ecosystem model of Equation (31).

$$\begin{cases} \frac{dx}{dt} & = ax - bxy_1 \\ \frac{dy_1}{dt} & = a_1 x y_1 - b_1 y_1 y_2 \\ \frac{dy_2}{dt} & = a_2 y_1 y_2 - b_2 y_2 y_3 \\ \vdots & = \quad \vdots \quad - \quad \vdots \\ \frac{dy_{N_t^a-1}}{dt} & = a_{N_t^a-1} y_{N_t^a-2} y_{N_t^a-1} - b_{N_t^a-1} y_{N_t^a-1} z \\ \frac{dy_{N_t^a}}{dt} & = a_{N_t^a} y_{N_t^a-1} y_{N_t^a} - b_{N_t^a} y_{N_t^a} z \\ \frac{dz}{dt} & = -a_{N_t^a} z + b_{N_t^a} y_{N_t^a} z \end{cases} \quad (31)$$

However, Equation (31)) can be further simplified by assuming that the P&L lost by a strategy is linearly gained by another which leads us to assume that $a = a_1 = \ldots = a_{N_t^a} = 1$ and $b = b_1 = \ldots = b_{N_t^a} = 1$. Finally the regulators may assume that the maximum number of strategies M can be fixed to the number of market participants. This is obviously arguable on the basis that a market participant may have multiple strategies but we can assume that this latter multiple strategy is itself a strategy. These simplifications

[69] $N_t^a = N_t^s + N_t^b$ or $N_t^a = N_{t-1}^a + N_t^b - N_t^d$.

give Equation (32).

$$\begin{cases} \frac{dx}{dt} &= x - xy_1 \\ \frac{dy_1}{dt} &= xy_1 - y_1y_2 \\ \frac{dy_2}{dt} &= y_1y_2 - y_2y_3 \\ \vdots &= \quad \vdots \quad\quad \vdots \\ \frac{dy_{N_t^a-1}}{dt} &= y_{N_t^a-2}y_{N_t^a-1} - y_{N_t^a-1}z \\ \frac{dy_{N_t^a}}{dt} &= y_{N_t^a-1}y_{N_t^a} - y_{N_t^a}z \\ \frac{dz}{dt} &= -z + y_{N_t^a}z \end{cases} \tag{32}$$

We can observe in Figure 23 a simulation of this simplified approach which makes the task of assigning a traditional SMC to the scenarios more achievable than the brute force direct approach.

Definition (HFTE Dual Tracking Methodology): We will call the High Frequency Trading Ecosystem Dual (HFTED) tracking methodology, the Lotka-Volterra mirror problem we have described in this section.

It should become straightforward to see that using the dual methodology reduces the dimensionality issue enough to allow most of the SMC methods to work in an otherwise open problem in signal processing. We will leave the different SMC methodologies implementation with the dual specified problem as an exercise for the motivated reader or PhD student looking for a paper idea.

3) Simplified Simulation: We present here an application of the results from Section III-F to our tracking methodology presented in this Section. For this we assume the state space is limited to a set of 15 scenarios spanned by up to 3 different types of strategies[70] acting on the OB in different sequences. In order to manage complexity we have also assumed that there is no birth or death processes involved in our scenarios. Algorithm 10 describes our simplified study in pseudo code. Note that the traditional resampling algorithm as developed by Doucet [24] has been substituted by the term $W_{t-1}^s + \lambda_r$ in the line $w_t^s \to \lambda_e \times \frac{L_s}{W} + (1 - \lambda_e - \lambda_r) \times W_{t-1}^s + \lambda_r \times 1/15$. We also added a small noise function to the market observed prices in order to make observations more realistic. The results from the series of simulations are presented in Figure 24. What we can observe is that every scenario had already clearly emerged by iteration 23, the second row from the bottom on all 15 scenarios. By iteration 47, the density is very clear, so much so that the only reason it is not a Dirac function is due to the resampling methodology introduced in that effect.

Algorithm 10: Particle Filter Simplified HFTE Strategies

Input: ΔP, I, w_t, λ_e, λ_r
Output: w_t
$w_{t-1} \leftarrow w_t$
$W \leftarrow 0$
for $0 \leqslant s \leqslant 15$ **do**
  H_I^s is is our Scenario Hash Table for scenario s and iteration I
  $L_s \leftarrow \exp(-\Delta P - H_I^s)$
  $W = W + L_s$
end
for $0 \leqslant s \leqslant 15$ **do**
  $w_t^s \leftarrow \lambda_e \times \frac{L_s}{W} + (1 - \lambda_e - \lambda_r) \times W_{t-1}^s + \lambda_r \times 1/15$
end
return w_t

Algorithm 11: Scenarios Hash Table H_I^s

Input: I
Output: array position
if $I == 0$ **then**
  return 0
else if $I == 2$ **then**
  return 1
else if $I == 3$ **then**
  return 2
else if $I == 5$ **then**
  return 3
else if $I == 11$ **then**
  return 4
else if $I == 23$ **then**
  return 5
else if $I == 47$ **then**
  return 6
else
  return 'issue with iteration recognition'
end

D. Another Practical Application

1) Problem Formulation: The context is the following: we are working in a bank and have been given instructions to build a realistic market simulators on which one can test strategies. You are given a set of strategies $\Omega = \{S_1, S_2, \ldots, S_n\}$ that can be replicated through the HFFF that we have seen in Figure 7. You also assume that you have a history oh P&L distribution for each element of Ω. This simulated market ought to be composed of an ecosystem of all possible theoretical strategies which frequency is unknown and which should react in such a way that the P&L distribution of all strategies for which we have historical data ought to perform in a similar manner.

2) Proposed Solution: We need to defined a particle filter on the scenarios described in the problem formulation. In doing so we need to both define a slightly different likelihood function as well as a very different resampling solution. We need to create a likelihood function for the particles associated to the scenario being investigated. This likelihood function should be itself function of the relative entropy between the expected P&L distribution and the one realized by the simulated market. The Kullback-Leibler divergence [69], of equation (**??**) can be a simple enough measure for the individual strategies being simulated. More specifically for

[70] Exact formalization has been given by Algorithms 3, 4 and 5.

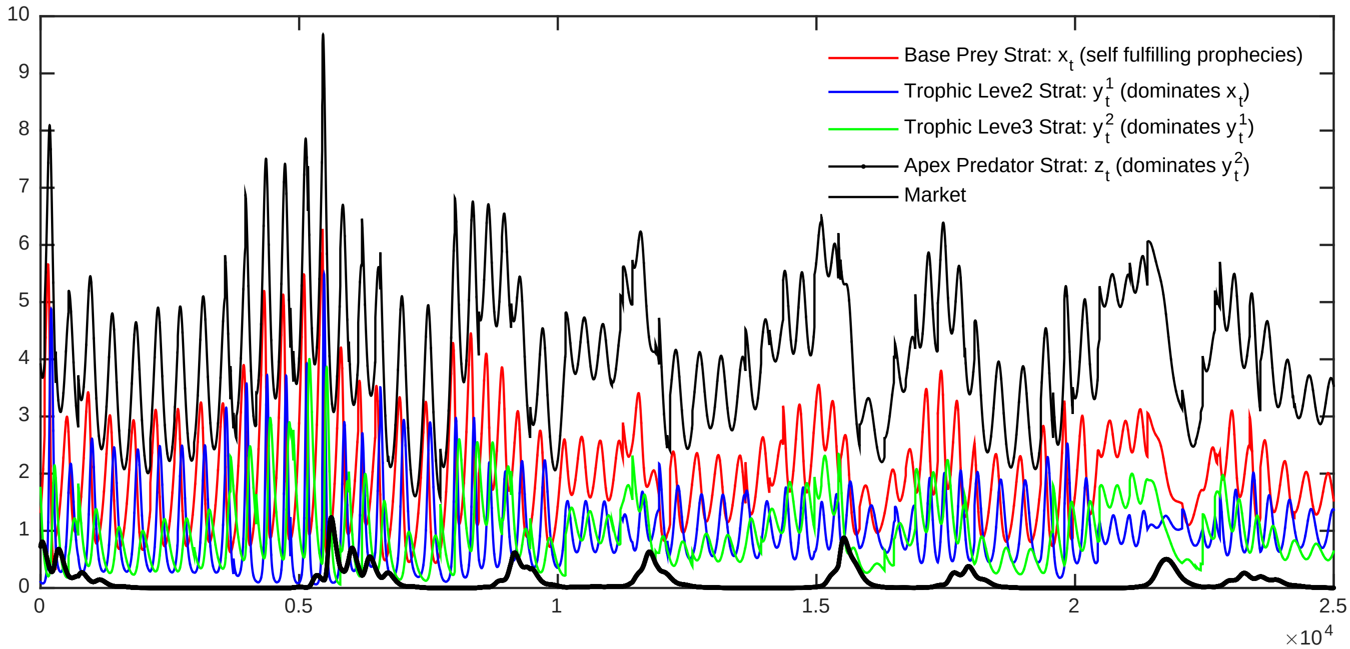
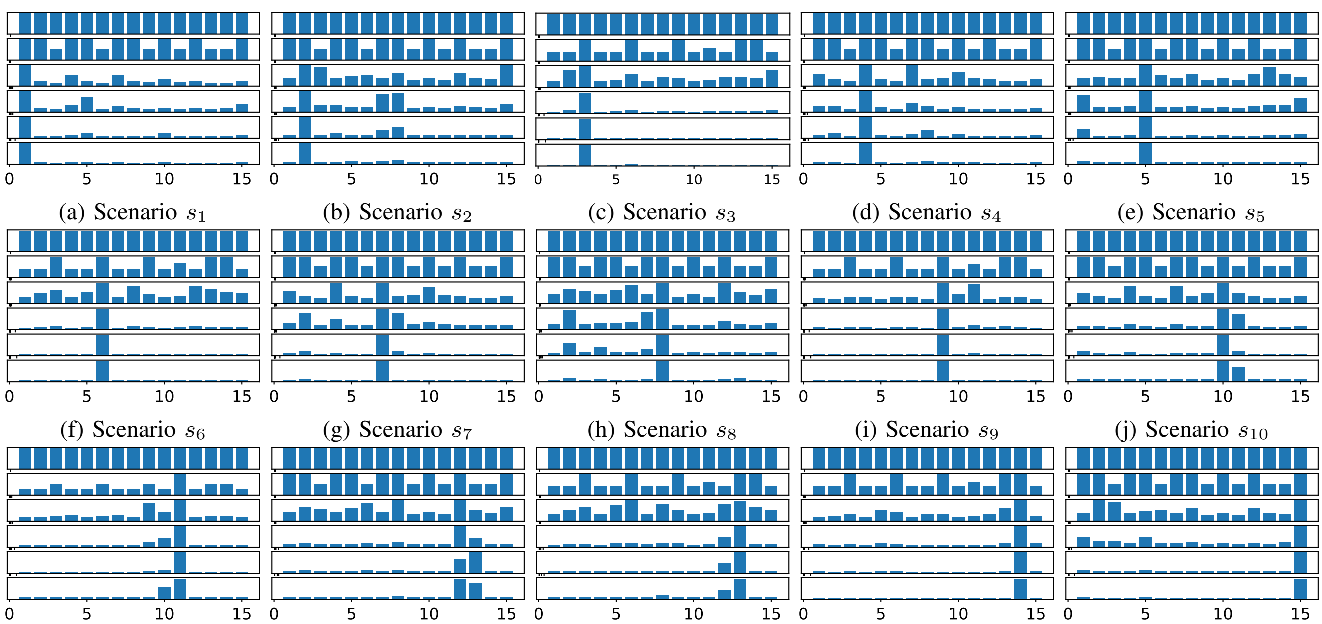


Fig. 23: Simplified Stochastic 4-species Lotka-Volterra of Equation (31)



(a) Scenario s_1 (b) Scenario s_2 (c) Scenario s_3 (d) Scenario s_4 (e) Scenario s_5

(f) Scenario s_6 (g) Scenario s_7 (h) Scenario s_8 (i) Scenario s_9 (j) Scenario s_{10}

(k) Scenario s_{11} (l) Scenario s_{12} (m) Scenario s_{13} (n) Scenario s_{14} (o) Scenario s_{15}

Fig. 24: Particle Filter on market scenarios on $[2, 3, 5, 11, 23, 47]$ milestones

discrete probability distributions where H, is the historical distribution and S, the simulated distribution, the Kullback-Leibler divergence from S to H is given by Equation (33).

$$D_{\mathrm{KL}}(H\|S) = -\sum_i H(i)\,\log\frac{S(i)}{H(i)}, \tag{33}$$

We can then define the unnormalized likelihood of a market scenario i by Equation (34a) where N_k represents the number of strategies alive as initially defined by equation (14).

$$L^{(i)} = \sum_{i=1}^{N_k} D_{\mathrm{KL}}(H\|S) \tag{34a}$$

$$\tilde{L}^{(i)} = \frac{L^{(i)}}{\sum_{i=1}^{N_k} L^{(i)}} \tag{34b}$$

Note here the resampling is at first glance done very differently. Indeed we are no longer looking to directly find the most profitable strategies[71] but the right frequency of each strategy in the ecosystem so as to make the latter more realistic (as compared measure of the historical P&L distribution). A good argument could be made that selecting for the best strategies using the genetic algorithm resampling methodology of our original paper [87] summarized by figure 18 would converge to a mature market that will hopefully converge to the ecosystem that would best accommodate the simulation of the historical P&Ls of each strategy. This is not necessarily a bad approach but a little restrictive. An additional way to adjust the resampling here would perhaps be to inspire ourselves from what would be a more rigorous invasion chart of the one from figure 20. The resampling methodology would then consist of increasing/decreasing the neighboring strategies that impact the P&L of the strategies which Kullback-Leibler divergence are the closest to 1 (most off their historical P&L).

3) Example: For instance, let us assume in figure 20 that we are sure of the invasion relation depicted by arrow 2 and 3 and that our Ω only consists of MLR, XOR and TF strategies. We have also sampled few markets with a random frequency of each of these strategies and are now examining how to adjust the frequencies of each strategy in each scenario so as to converge with a set of senarios in which the P&L of each strategy is similar to the historical one. In each scenario we calculate the Kullback-Leibler divergence of each strategy and have found that the MLRs are the one which are the closest to 1. This means that in this precise scenario we need to increase the set of strategies which would benefit in making the MLR closer to their historical P&L. In this case, let's say that the P&L resulting in the simulation for the MLR is significantly below it's historical one. We therefore need to increase the frequency of TF strategies so as to encourage the MLR invasion and therefore their increase in P&L and therefore an anticipated decrease in their simulated Kullback-Leibler divergence score. Had the P&L been overestimated, we would have instead increased the frequency of XOR strategies.

V. Conclusion

A. Summary

We have started this paper by pointing to a puzzling observation from the newly born high frequency commodities market which because of its extreme youth and therefore immaturity makes it a great case study for a high frequency market at inception and therefore for our purpose. More specifically as we have seen with Figure 1, on 06/08/2011, fascinating patterned oscillations occurred in the commodities market. These oscillations cannot be explained by the Top-Down assumption in Quantitative Finance (e.g. the Brownian motion). We have proposed in this paper to study these oscillations with the bottom-up approach instead. The latter theory was developed in 3 Sections. We first expressed, in Section II, classic Financial Strategies in HFFF format and shown the incentive for going from from a simple perceptron, to shallow and finally deep learning. We also established connections to fields that are traditionally associated to mathematical biology, in Section III, namely predator-prey models and evolutionary dynamics. This was done in order to express the bottom-up approach at the infinitesimal level. More specifically we developed the concept of Path of Interaction in an HFTE Game. Finally we looked at how the financial market composition could be tracked through time with MTT in Section IV.

B. Current & Future Research

Our first few simulations opened our eyes up to issues associated to optimality and need for more scientific rigor. We have classified these points of improvement in half a dozen issues listed below.

1) Classification Simplification: As mentioned before the direct simulation approach [86] is too challenging and the results perhaps too convoluted to filter out the essence of the paper. For this reason we proposed to study the problem using fixed HFFFs, of Figures 8, 10 and 13. Though this simplifies the problem it also means there is human intervention in the strategy pool chosen. This latter intervention, though convenient raises the question of whether what seems to be equivalent strategies are equivalent after all. Less human interventions should take place going forward.

2) The State Space can be improved: Choosing three types of strategies greatly limits our state space which makes our tracking methodology easier but not as realistic as we wish ultimately. Additional strategies must be incorporated and more HFTE games must be included in our database of scenarios. This could be the work of many years and could be addressed in the form of creating an online database in which interested scientists could deposit their findings in object oriented format for simulation purposes.

3) Order-book Dynamics: Many of the markets are driven by different rules for the OB. We need to incorporate these different rules in our HFTE games as the latter rules obviously impact the outcome of the games.

4) Increased HFFF complexity does not equate to Invasion: It has been speculated that the need for a bigger brain in men is partly due to the need for humans to elaborate deceitful strategies with their rivals and cooperative strategies with their allies. It is therefore not entirely ridiculous to associate increased neural network branching (to be roughly understood as increase in cranial size) with increased strategy complexity. However, increased intelligence does not necessarily equate to survival as we can see in the shark population, considered like an apex predator in the sea but with a relatively small brain that has not evolved for millennial. We are very much at the early stages in defining NN complexity and dominance. A clear picture did not necessarily emerge from the first simulations.

5) Birth & Death Processes: We need to incorporate a Birth and Death Process to our MTT to make more realistic scenarios. In order to do that we need to incorporate the OB in the likelihood function instead of using only the price dynamics. This will undoubtedly make the programming

[71] Or rather, not just yet.

exercise more challenging but will at the same time bring more value to the research in the long run.

6) Complex Food Webs: We have seen in Section III-F.1 that we have taken $l = 4$ in our Path of Interaction sequence. Would the Path of Interaction results change if we increase the sequence's length? In the context of the Path of Interaction study, is there a more rigorous way to connect some of the Lotka-Volterra predator prey models to these interactions? It seems intuitively more likely that the strategy ecosystem should rather be a complex food web. Can we enhance the idea of the simple Lotka-Volterra predator prey model to more complex food webs? More specifically what are the strategies that would create a stable and unstable food web? The concept of Path of Interaction is meant to be a bridge connecting the gap between strategy formalization to evolutionary dynamics but this bridge in not yet clearly specified.

7) Diversity & Stability: One other legitimate question that we can ask ourselves is whether the HFFF a complex enough network to model all financial strategies? And if not all, does it encompass enough strategies to convey something interesting and meaningful when you make the strategies interact with each other. In this context our first chapter and paper [86] ended with the proposed "Diversity & the Financial Markets" conjecture below which is currently an open problem that is interesting to mention in the context of future research:

Conjecture 3 (Diversity & the Financial Markets): Diversity in financial strategies in the market leads to its instability.

Remark Note this conjecture has be studied partially with simulations and can be perhaps indirectly studied or at least intuitively using some of the finding in mathematical biology. More specifically the one associated with diversity in ecosystem and stability[72]. However we are still a long way before being able to answer this question.

References

[1] Omar Al-Jarrah, Paul Yoo, Sami Muhaidat, George Karagiannidis, and Kamal Taha. Efficient machine learning for big data: A review. 2:8793, 09 2015.

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

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

[4] Mark Robert Anderson. Twenty years on from deep blue vs kasparov: how a chess match started the big data revolution. 2017.

[5] 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.

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

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

[8] Louis Bachelier. Theorie de la speculation. *Annales Scientifiques de lcole Normale Suprieure*, 3 (17):21–86, 1900.

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

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

[72] Though no consensus is reached there either.

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

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

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

[14] 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.

[15] Jean-Philippe Bouchaud. Economics needs a scientific revolution. *Nature*, 455:1181, 2008.

[16] Mark Buchanan. Meltdown modelling: Could agent-based computer models prevent another financial crisis? *Nature*, 460:680–682, 2009.

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

[18] Matt Burgess. Google's deepmind wins historic go contest 4–1. 2016.

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

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

[21] Kenneth Cukier and Viktor Mayer-Schoenberger. The rise of big data: How it's changing the way we think about the world. *Foreign Affairs*, 92(3):28–40, 2013.

[22] 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.

[23] 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.

[24] A. Doucet. On sequential simulation-based methods for bayesian filtering. Technical report, 1998.

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

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

[27] 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.

[28] 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.

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

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

[31] 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.

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

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

[34] Shaokun Fan, Raymond Y.K. Lau, and J. Leon Zhao. Demystifying big data analytics for business intelligence through the lens of marketing mix. *Big Data Res.*, 2(1):28–32, March 2015.

[35] J. Doyne Farmer and Duncan Foley. The economy needs agent-based modelling. *Nature*, 460:685–686, 2009.

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

[37] 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.

[38] 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.

[39] Martin Gardner. Mathematical games. *Scientific American*, 1970.

[40] Jim Gatheral. The volatility surface. page 37, 2006.

[41] 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.

[42] 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.

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

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

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

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

[47] 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.

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

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

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

[51] 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.

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

[53] 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.

[54] 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.

[55] 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.

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

[57] 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.

[58] 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.

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

[60] Xiaolong Jin, Benjamin W. Wah, Xueqi Cheng, and Yuanzhuo Wang. Significance and challenges of big data research. *Big Data Res.*, 2(2):59–64, June 2015.

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

[62] 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.

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

[64] 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.

[65] 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.

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

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

[68] 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.

[69] S. Kullback and R. A. Leibler. On information and sufficiency. *Ann. Math. Statist.*, 22(1):79–86, 03 1951.

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

[71] Jae-Gil Lee and Kang Minseo. Geospatial big data: Challenges and opportunities. 2, 02 2015.

[72] Clara Leonard. Apres la crise, lenseignement de la finance repense. Sep 2013.

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

[74] 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*.

[75] 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.

[76] Alexandr Mikhailovich Liapunov. The general problem of the stability of motion. *Doctoral dissertation, University of Kharkov*, 1892.

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

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

[79] 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.

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

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

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

[83] Babak Mahdavi-Damghani. UTOPE-ia. *Wilmott Magazine*, 60:28–37, 2012.

[84] Babak Mahdavi-Damghani. The non-misleading value of inferred correlation: An introduction to the cointelation model. *Wilmott*, 67:50–61, 2013.

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

[86] Babak Mahdavi-Damghani. Introducing the HFTE model: a multi-species predator prey ecosystem for high frequency quantitative financial strategies. *Wilmott Magazine*, 89:52–69, 2017.

[87] Babak Mahdavi-Damghani. Introducing the HFTE model: a multi-species predator prey ecosystem for high frequency quantitative financial strategies. *Wilmott*, 89:52–69, 2017.

[88] Babak Mahdavi-Damghani and Andrew Kos. De-arbitraging with a weak smile. *Wilmott Magazine*, 2013.

[89] Babak Mahdavi-Damghani and Stephen Roberts. Deciphering price formation in the high frequency domain: Systems & evolutionary dynamics as keys for construction of the high frequency trading ecosystem. *University of Oxford - Oxford-Man Institute of Quantitative Finance*, 2017.

[90] Babak Mahdavi-Damghani, D. Welch, C. O'Malley, and S. Knights. The misleading value of measured correlation. *Wilmott*, 62:64–73, 2012.

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

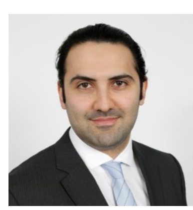
[92] R.P.S. Mahler. Multitarget bayes filtering via first-order multitarget moments. *IEEE Transactions on Aerospace and Electronic Systems*, 39(4), 2003.

[93] 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.

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

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

[96] 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.
[97] David D. L. Minh and Do Le (Paul) Minh. Understanding the hastings algorithm. *Communications in Statistics - Simulation and Computation*, 44(2):332–349, 2015.
[98] Marvin Minsky and Seymour Papert. *Perceptrons: An Introduction to Computational Geometry*. 2nd edition with corrections, first edition 1969 edition, 1972.
[99] Nanex. Strange days june 8'th, 2011 - natgas algo, 2011.
[100] Radford M. Neal. Suppressing random walks in markov chain monte carlo using ordered overrelaxation. *University of Toronto, Department of Statistics*, 1995.
[101] 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.
[102] 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.
[103] 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.
[104] Martin Nowak. *Evolutionary dynamics : exploring the equations of life*. 2006.
[105] Basel Committee on Banking Supervision. Basel 3: A global regulatory framework for more resilient banks and banking systems. page 11, 2011.
[106] 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.
[107] L.Y. Pao. Multisensor multitarget mixture reduction algorithms for target tracking. *AIAA Journal of Guidance, Control and Dynamics*, 17, 1994.
[108] S. L. Pimm and J. H. Lawton. On feeding on more than one trophic level. *Nature*, 275:542–544, 1978.
[109] M. K. Pitt and N. Shephard. Filtering via simulation: Auxiliary particle filter. *Journal of the American Statistical Association*, 94:590–599, 1999.
[110] Warren McCulloch; Walter Pitts. A logical calculus of ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5:115–133, 1943.
[111] 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.
[112] D. Read. An algorithm for tracking multiple targets. *IEEE Transactions on Automation and Control*, 24(6):84–90, 1979.
[113] 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.
[114] Stephen Roberts and Martin Tegner. Sequential sampling of gaussian latent variable models. *University of Oxford*, 2018.
[115] Frank Rosenblatt. *Principles of neurodynamics: perceptrons and the theory of brain mechanisms*. 1962.
[116] S. Roweis and Z. Ghahramani. A unifying review of linear gaussian models. *Neural Comput*, 11(2):305–345, 1999.
[117] 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.
[118] 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.
[119] D.J. Salmond and H. Birch. A particle filter for track-before-detect. *in Proceedings of the American Control Conference*, pages 3755–3760, 2001.
[120] 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.
[121] 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.
[122] H. Sidenbladh. Multi-target particle filtering for the probability hypothesis density. *Proceedings 6th International Conference on Information Fusion*, pages 800–806, 2003.
[123] H. Sidenbladh. Multi-target particle filtering for the probability hypothesis density. *Proceedings 6th International Conference on Information Fusion*, pages 800–806, 2003.
[124] H. Sidenbladh and S.L. Wirkander. Particle filtering for finite random sets. *IEEE Transactions on Aerospace and Electronic Systems*, 2003.
[125] H. Sidenbladh and S.L. Wirkander. Particle filtering for finite random sets. *IEEE Transactions on Aerospace and Electronic Systems*, 2003.
[126] Marton Nowak; Karl Sigmund. A strategy of win-stay, lose-shift that outperforms tit-for-tat in the prisoner's dilemma game. *Nature*, 364:56–58, 1993.
[127] Justin Sirignano and Rama Cont. Universal features of price formation in financial markets: perspectives from deep learning. *Quantitative Finance*, 2018.
[128] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, and Ruslan Salakhutdinov Ilya Sutskever and. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15:1929–1958, 2014.
[129] L.D. Stone, C.A. Barlow, and T.L. Corwin. Bayesian multiple target tracking.
[130] 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.
[131] Nassim Nicholas Taleb. *The Black Swan*. Allen Lane, 2007.
[132] 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.
[133] M. R. Vargas, B. S. L. P. de Lima, and A. G. Evsukoff. Deep learning for stock market prediction from financial news articles. In *2017 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA)*, pages 60–65, June 2017.
[134] 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.
[135] Cedric Villani. Cedric villani: la simulation numrique est un enjeu majeur pour la societe, 2012.
[136] B.N. VO, S.S Singh, and A. Doucet. Sequential monte carlo implementation of the phd filter for multi-target tracking. *in Proceedings of the International Conference on Information Fusion*, pages 792–799, 2003.
[137] Vito Volterra. Variazioni e fluttuazioni del numero dindividui in specie animali conviventi. *Mem. Acad. Lincei Roma*, pages 31–113, 1926.
[138] P. Willett, Y. Ruan, and R. Streit. Pmht: Problems and some solutions. *IEEE Transactions on Aerospace and Electronic Systems*, 38(3):738–754, 2002.
[139] P. Yodzis. The stability of real ecosystems. *Nature*, 289:674–676, 1981.



Babak Mahdavi Damghani has been working in the financial industry within a broad range of functions (Exotics & High Frequency Trading, Structuring, FO & Risk Quantitative Analytics and Clearing), through all major asset classes (Equities, Commodities, FX, Rates and Hybrid) in both the buy and sell sides across different geographical locations. He did his undergraduate education at the University of Pennsylvania in what would now correspond to Financial Engineering. His post-graduate studies were accomplished in several areas of the applied mathematical & computational sciences at the University of Cambridge & University of Oxford. He is the founder of EQRC and currently doing research in Machine Learning at the University of Oxford. He is also the author of numerous publications, including cover stories of Wilmott magazine and mathematical models (eg: Cointelation, IVP & HFTE model) that are now taught in the CQF.