Introducing the HFTE Model: a Multi-Species Predator Prey Ecosystem for High Frequency Quantitative Financial Strategies

Babak Mahdavi-Damghani*

EQRC & Oxford-Man Institute of Quantitative Finance

Abstract— **In this paper we take a new approach to studying electronic trading & systemic risk by introducing the HFTE model. We specify an approach in which agents interact through a topological structure designed to address the complexity demands of most common high frequency strategies but designed randomly at inception. The primitive strategy ecosystem is then studied through a simplified genetic algorithm. The results open up intriguing social and regulatory implications with the helping doors of Mathematical Biology & Game Theory which specific mirror points have been summarized for the sake of illustrating the puzzling findings.**

Keywords: HFTE Model, High Frequency Financial Funnel, HFFF, Multi-Target Tracking, Stability of Financial Systems, Data Analysis and Patterns in Data, Electronic Trading, Systemic Risk, High Frequency Trading, Game Theory, Machine Learning, Predator Prey Models.

I. Introduction

A. Historical Context

After the subprime crisis of 2008 and the resulting social uproar, governments strongly pushed the regulators to develop more efficient risk monitoring systems[1]. Given that the biological ramifications of the unfortunate cost of pattern recognition [15] is inherent to humans and the fact that the historical crises were not directly connected to each other, it became implicitly clear that the next financial crisis could not be in real estate again (at least not immediately ...) but rather elsewhere. The candidate sector under coercion became very quickly the one of algorithmic systematic trading which most famous incident was the flash crash of May 6, 2010, in which the Dow Jones Industrial Average lost almost 10% of its value in matter of minutes. However, the current state of the art risk models are the ones inspired by the last subprime crisis and are essentially models of networks in which each node can be impacted by the connected nodes through contagion [10] perhaps better suited for lower frequency models. Indeed, on 06/08/2011 a seemingly relatively unnoticed event occurred on the natural gas commodities market. We mention here 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 with

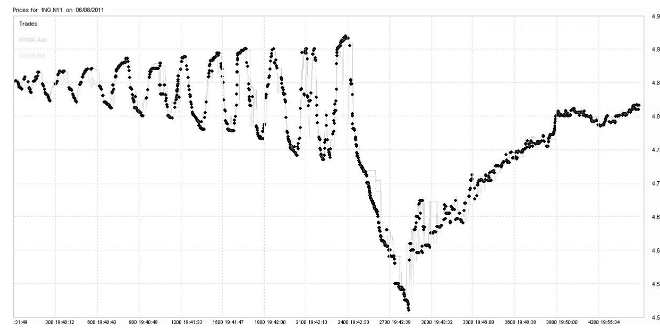

Fig. 1. Natural Gas flash crash of 06/08/2011 [20]

the strong intuition that only interacting agents falling into some sort of quagmire could yield such fascinating series of increasing oscillations followed by a mini crash. Until the arrival of algorithm, never in more than a 100 years of data in countless products at different geographical location anything as clear in terms of oscillation was observed and this observation came into an immature market for electronic trading. I would like to spend few lines developing this point as this may not be clear to the reader. Indeed, commodities has historically been seen as a physical market, this in turn meaning that the prices are driven by supply and demand of commodities which can be consumed, stored and/or produced. This particular point is a unique feature compared to the other markets (Equities, FX, or Rate). Also this figure 1 suggests that the common, though perhaps a bit lazy view, that crashes occur through totally unpredictably [30] events may not be true for algorithmic trading.

B. Scientific method & parallel to Conway Game of Life

In this paper we will take an approach, similar in methodology, to Conway's Game of Life [8], a 4 rules cellular automaton exercise which figure 2 reminds of the rules and figure 3 provides a 3 pictures snapshot of one random simulation. We will apply Conway's methodology to the world of High Frequency Trading (HFT) while adjusting some of the idiosyncratic parts of the exercise. As a reminder, Conway's Game of Life assumes that complexity in an ecosystem[2] arises from simple rules. For instance these rules can lead to a family of three different types of automatons

* bmd@eqrc.co.uk or babak.mahdavidamghani@oriel.ox.ac.uk

[1] In this context risk is viewed as a mixture or Market and Reputational risk.

[2] 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.

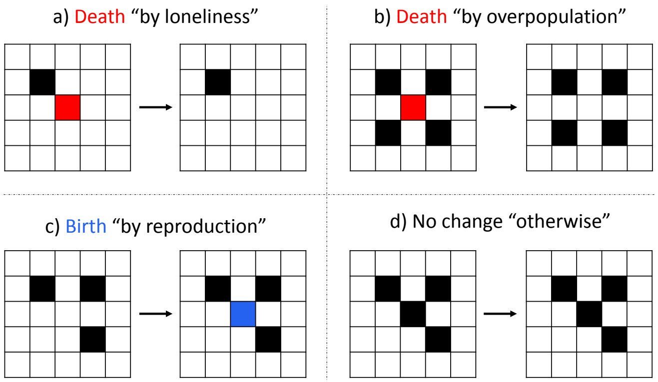

Fig. 2. Conway's Game of Life rules illustrated

(when iterations are increased and the seed is random). As reminder you may get:

- **Stable forms** (for example the "Block", the "Beehive", the "Loaf", the "Boat"). Intuitively the reader may guess that the concept of financial stability may be raised through a similar methodology.
- **Oscillating forms**: for example the "Blinker[3]", the "Toad[4]", the "Beacon[5]", the "Pulsar[6]", the "Pentadecathlon[7]". Intuitively the reader may guess that the concept of financial cycles or HF oscillations like of figure 1 may be induced through a similar scientific methodology.
- **Moving forms**: for example the "Glider" and the "LWSS (Lightweight spaceship)" which may have different sizes and speeds[8].

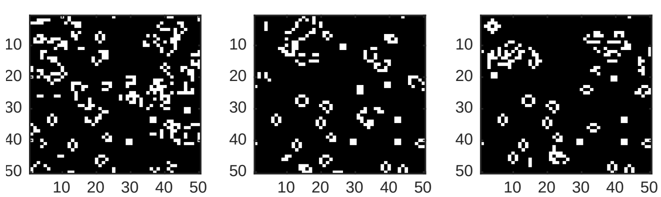

Fig. 3. 3 snapshots of a simulation of Conway's Game of Life.

The parallel to the world of quantitative financial strategies would be the following few points:

- interacting agents lead to the market price fluctuations and more specifically their sole interaction determines the stability or instability of the market depending on what the market is made of in terms of the strategies involved as well as the evolving order-book.
- the market will follow the rules of a zero-player game[9] with, however random seeds.
- agents (eg: strategies) will follow a simple rule for their births and deaths.

C. Market & Orderbooks

1) Caveat: This paper assumes a simplification of the market: that is one product into one single possible market with few market participants who are unable to cheat the system through technology. In reality there exists a plethora of products in many markets in multiple geographical locations and the SEC and the FCA expose new stories of cheats on daily basis. This approach may seem overly simplistic, but, we will see that this simplistic rule abiding approach may open up a new perceptive towards how people see and may want to take actions on the market.

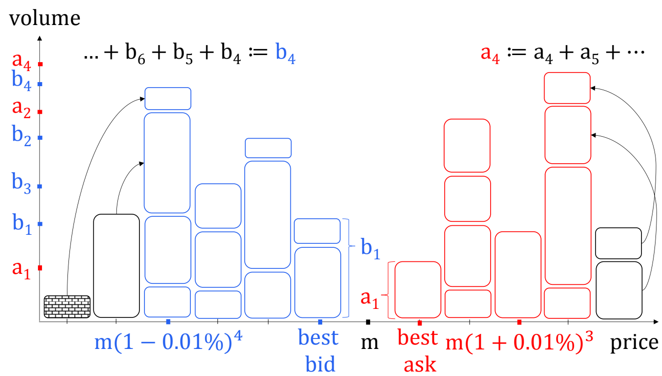

Fig. 4. Order-book visual representation

2) 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 was recorded by few market participants and in which a referee would decide which would win the trade.

Definition 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 apart as shown by equation (1).

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

Remark Figure 4 represent an order book which the previous definition aims at describing.

[3] 2 period iteration
[4] 2 period iteration
[5] 2 period iteration
[6] 3 period iteration
[7] 15 period iteration
[8] idiosyncratic properties from the Game of Life, which parallel to our problem is not necessarily transferable.
[9] Meaning that its evolution is determined by its initial state, requiring no further input.

3) Variable Definition:

Definition 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[10]. Similarly we will define $\{x_{j,1}, x_{j,2}, \ldots x_{j,p}\}_{j=i+1}^{n}$ 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 listed below:

- The price of the underlier itself
- The accumulated volume at different market depth of the order books (4 of 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).

D. Problem Formulation & Agenda

1) Problem Formulation: The connection between machine learning and high frequency trading (HFT) has long been implicitly established via the numerous systematic trading position available in most job searching tools (eFinancialCareers, linkedIn, etc ...). It is however unclear which of the numerous machine learning techniques is most relevant to what high frequency traders wish to accomplish. The field of machine learning itself is quite rich, genetic algorithm, algorithmic game theory, state space models, kalman filter, sequential monte carlo methods, support vector machine, neural networks or even a simple multi-linear regression are some of the key words mentioned in job descriptions which title would not suggests much difference in the tools used by the quants supposed to perform the tasks associated to these jobs. However, what most of these methodologies have in common is that they assume a pattern inherent to the market itself as opposed to taking the market as a consequence of the strategies composing this market[11].

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 see altruism at a different enhanced angle. By analogy we are trying to communicate the idea that the market centered view of the financial system is the wrong way around understanding the fluctuation of the market and that the strategy centered view of the financial system provides the opportunity to look at the market at a different enhanced angle.

Recently the concept of ecosystem of strategies [12] was introduced. Though the idea had great potential the paper assumes a set of static strategies which does offer to some extend an interesting current snapshot of the market but does not offer:

- a history for this snapshot,
- a inspiring future for the field,
- a topology for these strategies (in the form of a DNA) on which one could study the complex problem into another mathematical domain[12] easier to solve,
- a sense of how to study the stability of the markets as suggested by the term ecosystem and its biological meaning,
- social insight about how this should impact the regulatory horizon,
- a connection to other fields[13] which mirror concepts and properties could be used to increase our mathematical weaponry in the context of analyzing critical concepts such as stability or cycles.

Definition We call **HFTE** the **H**igh **F**requency **T**rading **E**cosystem model which attempts at answering the 6 bullet points just raised, the subject of this paper.

Remark The naming of the model proved a bit challenging. Combination of the following phrases were assessed:

- ”High Frequency Trading”,
- ”Quantitative Strategies”,
- ”Multi-Species Predator Prey Ecosystem”,
- ”Financial Automata”.

but ultimately HFTE prevailed due to its connection to HFT which almost anyone in the Financial industry knows of the acronym and ”E” (for ecosystem) which really is the key idea from the paper.

2) Agenda: We will first introduce in section II a generalized Network Topology which we speculate as having enough architectural DNA to have the potential to formalize most Classic Financial Strategies for which we will give few examples. In section III we will specify our Genetic Algorithm (GA) as a mean to study the high frequency (HF) market, which essentially, with section II, is the core mathematical engine (the HFTE model) to help us keep track of the various strategy families' performance in our environment through time. We will analyze in section IV the results and will provide in that occasion a parallel to the world of mathematical biology, more specifically in section IV-C its connection to predator prey models and in section IV-B its link to some of the interesting results in Game Theory namely Evolutionary Dynamics. We will further expand on our findings by providing couple of applications in section V-A, more specifically in high frequency trading and in high level regulatory and government policies. Finally in section VI we will conclude our paper by suggesting potential continuation for research in the HFTE model.

Remark This paper and its second part are[14] at the crossroad

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

[11] Top-Down vs Bottom-Up approach

[12] Geometry

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

[14] the particle filter part is currently being ironed out

of few different fields; Figure 5 is attempting at guiding the reader in these fields for preparatory sake. However, the paper has been written in such a way to be accessible to the biggest possible audience including practitioners at the risk of lacking a bit of rigor from time to time.

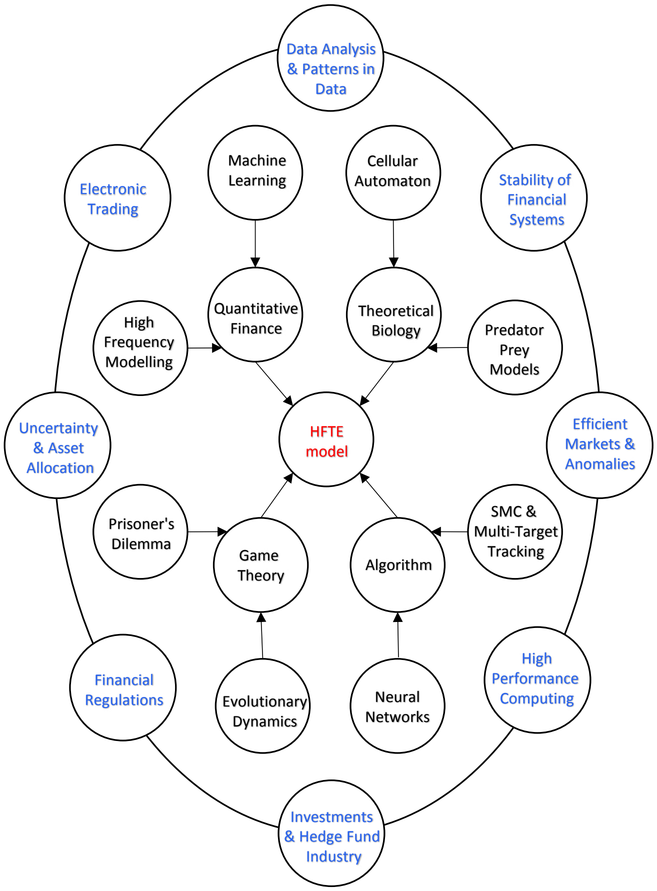

Fig. 5. Academic schemes & fields involved in this paper's and its follow up

II. Network Topology & Classic Financial Strategies

A. Network Topology & learning potential

Two important milestones in Machine Learning are worth reminding as they shed light on why the core building blocks of our HFTE model is a certain way. First, Warren McCulloch and Walter Pitts [23] introduced their threshold logic model in 1943 which is agreed to have guided the research in network topology as it relates to artificial intelligence for more or less a decade. Second, Rosenblatt [25], 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 6 could act as separators[15] and therefore a direct equivalence could be made to the multi-linear regression which we will elaborate on more in details is section II-B.2. One observed limitation of the perceptron as described by

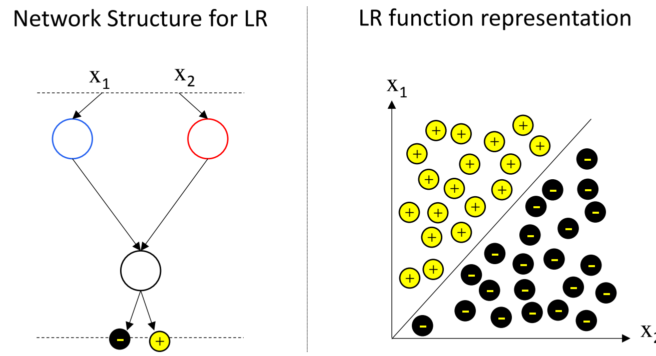

Fig. 6. Simple Neural Network modeling a linear regression

Rosenblatt, in 1969, was that a simple yet critical well known functions such as the XOR function could not be modeled [19]. 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 7 for the illustration). Added, to this we have the well known overfitting [29] problems when it comes to supervised learning which has been there since inception though regular progress is being made in that domain without real breakthrough.

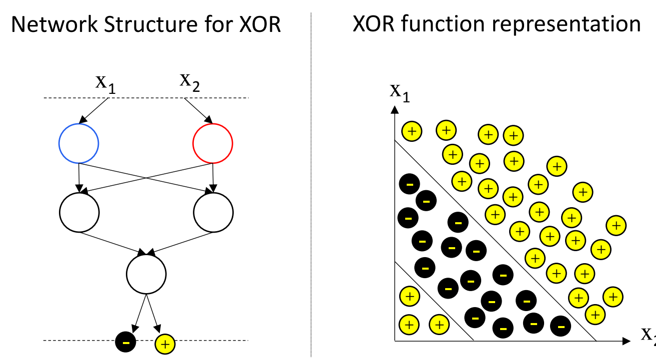

Fig. 7. Feedfoward ANN with 1 hidden layer & the XOR function

B. The Funnel

These few historical rationals are the main drivers which have led us to propose the Funnel, introduced by Martin Nowak [21], as the simplest possible network to model (therefore which minimizes overfitting) the key functions for our application. The area of evolutionary graph theory is quite rich. Many graphs provide interesting properties. We can formalize the learning process from all of our strategies using the topology of figure 8 by providing a set $\mathcal{T}$, as described by equation (2) of weights corresponding to all

[15] 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.

the possible weights of this particular figure.

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

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

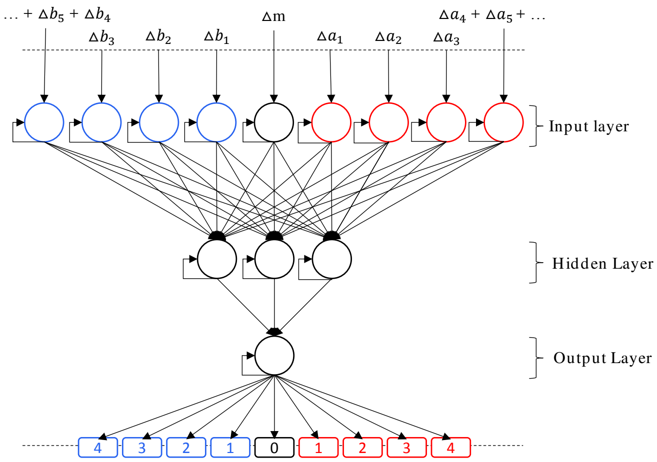

Fig. 8. The High Frequency Financial Funnel

Remark Note that in the context of this paper we have chosen to work with Martin Nowak's [21] funnel, which modification is described in figure 8. This topological structure offers the advantage of linking some interesting bridges between the worlds of:

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

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

Definition Let's call the **H**igh **F**requency **F**inancial **F**unnel (HFFF) to be a topological structure 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).

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

1) The Trend Following Topology: 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 (eg: up or down) in the recent past, then it is more likely to follow the same trend in the immediate future.

Definition The mathematical formulation of 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).

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

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

Remark 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 $w^{h}_{s,3} = 0$. ■

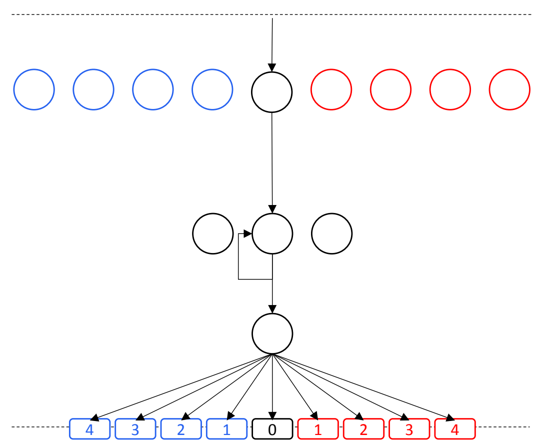

Fig. 9. The EWMA strategy translated in terms of network topology (the weights equal to 0 have not been represented)

Remark The proof is visually illustrated by figure 9 (the weight equal to 0 have not been represented[16]). On a side note the HFFF can also model differences in EWMA's: simply slightly change figure 9 into figure 10. There are 3 different ways to come up to the exact results when handling figure 10. 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 topological DNA which would look like the one from figure 9.

[16]Note that there is different ways to achieve the same numerical results though with a different topology

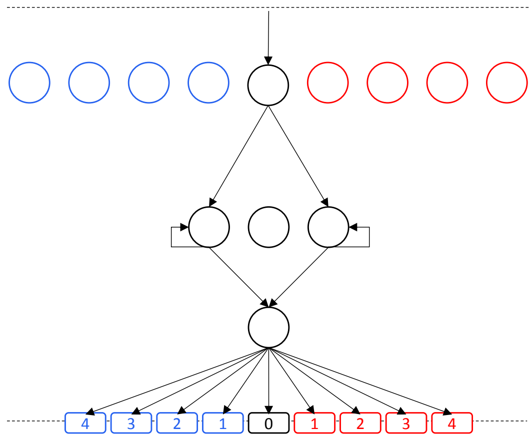

Fig. 10. The difference of two EWMAs strategies translated in terms of network topology (the weights equal to 0 have not been represented)

2) *Multi Linear Regression Topology:* The Multi Linear Regression (MLR) is another well known 101 type strategy traders have been using in the industry.

Definition Given a data set $\{y_i, x_{i-1,1}, \ldots, x_{i-1,9}\}_{i=1}^n$ where n is the sample size, and y_i then our MLR is formalized by the equation below :

$$\begin{aligned} y_i &= \beta_1 x_{i-1,1} + \cdots + \beta_9 x_{i-1,9} + \varepsilon_i \\ &= \mathbf{x}_{i-1}^{\mathrm{T}} \beta + \varepsilon_i, \;\; i = 1, \ldots, n \end{aligned} \tag{5}$$

where T denotes the transpose, so that $\mathbf{x}_{i-1}^{\mathrm{T}}\beta$ is the inner product between vectors x_i and β.

Proposition The HFFF can model multi linear regression like strategies.

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

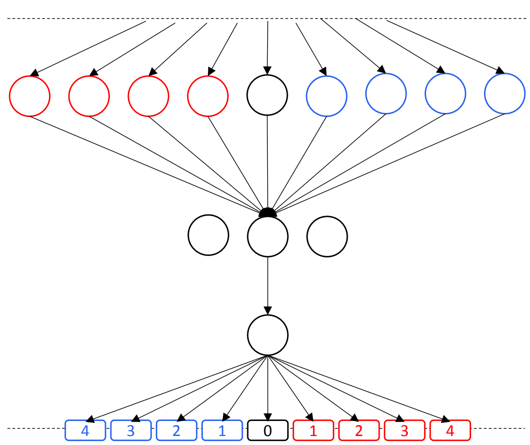

Fig. 11. The MLR strategy translated in terms of network topology

Remark We will make 4 remarks:

- Topologically a MLR can be illustrated by figure 11 (the weights equal to 0 have not been represented).
- As we have explained before, different topologies may lead more or less to the same strategy. Figure 12 is another example of a MLR.
- 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 topological DNA which would look like the one from figure 11.
- Logistic or weighted MLR can be modeled through the same topology of figure 11 by simply changing respectively the activation function (from linear to logistic) and the weights.

3) *XOR Topology:* We recall here the truth table associated by the XOR function in table II-B.3. How is this relevant to HFT? Let's look at the following known HF rational.

I_1	I_2	O
1	1	0
1	0	1
0	1	1
0	0	0

TABLE I
THE TRUTH TABLE OF THE XOR FUNCTION

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

- the price and the OI are rising then the market is bullish,
- the Price is rising but the Open Interest Falling then the market is bearish,
- the Price is falling but the Open Interest rising then the market is bearish,
- the Price is falling and the Open Interest falling then the market is bullish.

Remark These 4 market situations can be summarized by table II-B.3.

Price	Open Interest	Signal
Rising	Rising	Buy
Rising	Falling	Sell
Falling	Rising	Sell
Falling	Falling	Buy

TABLE II
THE RELATIONSHIP BETWEEN OPEN INTEREST, PRICE & SIGNAL

Proposition The HFFF can model XOR like strategies.

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

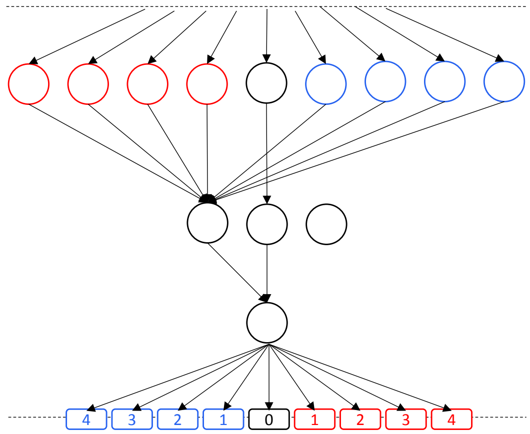

Fig. 12. The XOR strategy translated in terms of network topology

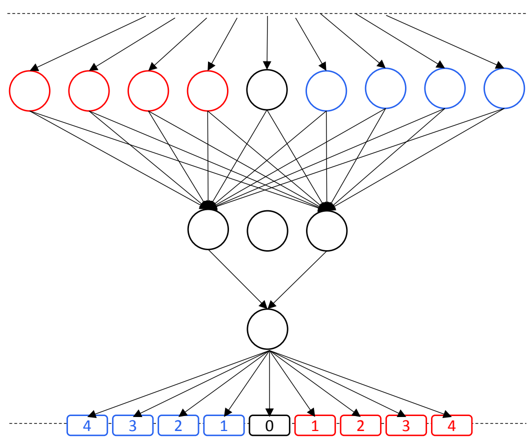

Fig. 13. Another XOR strategy translated in terms of network topology

Remark We will make the following 2 observations:

- The preceding proof is visually illustrated by figure 13 (the weights equal to 0 have not been represented).
- The XOR topology can be designed in various ways[17].

4) Execution strategy: To make the problem more realistic, one needs to formalize an execution strategy which rule would apply to all strategies but still be rule based and function of its topology. In this first paper we will take this simple approach in which all strategies will follow algorithm 1. The idea of this algorithm will be that:

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

In this context algorithm PERCENTILE$(\mathcal{D}^r_{1:t-1}, \alpha_1, \alpha_2)$ returns a value between 0 and 9, the 9 pillar points for our order

[17]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 topological DNA which would look like the one from figure 13.

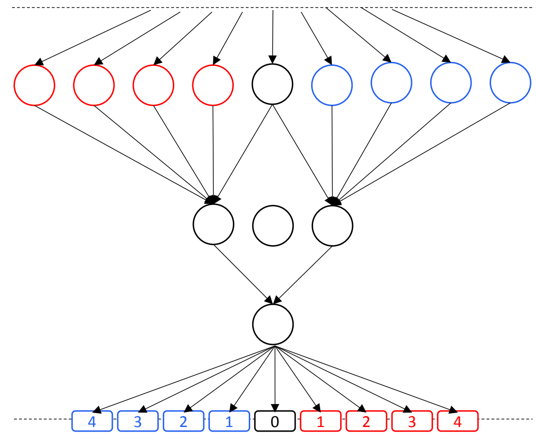

Fig. 14. The XOR strategy translated in terms of network topology

book. The tested input is compared against the α_1 and α_2 percentiles. Given that no history exists in the first iteration and that the first few iterations are not significant, we will randomize the first R_n iterations (though not mentioned in the algorithm 1).

Algorithm 1 EXECUSSION STRATEGY($\mathcal{T}$, $\mathcal{I}_{1:9}$, $\mathcal{D}_{1:t}$)

Require: topology $\mathcal{T}$, array of current inputs $\mathcal{I}_{1:9}$ (8 accumulated volumes and 1 last price) from order book, array of rough previous execution decisions $\mathcal{D}^r_{1:t-1}$

Ensure: strategy of topology $\mathcal{T}$ modifies the order book by putting in a order $\mathcal{O}$ at any of the 9 positions of the orderbook.

{Side comment: A_i for asked price at $i$bps from mid}
{Side comment: B_j for bid price at $j$bps from mid}
$\mathcal{D}^r_t \leftarrow$ CALCULATE$(\mathcal{T}, \mathcal{I}_{1:9})$
if $\mathcal{D}^r_t \in$ PERCENTILE$(\mathcal{D}^r_{1:t-1}, 0, \frac{1}{9})$ **then**
  $\mathcal{O} \leftarrow B_4$
else if $\mathcal{D}^r_t \in$ PERCENTILE$(\mathcal{D}^r_{1:t-1}, \frac{1}{9}, \frac{2}{9})$ **then**
  $\mathcal{O} \leftarrow B_3$
else if $\mathcal{D}^r_t \in$ PERCENTILE$(\mathcal{D}^r_{1:t-1}, \frac{2}{9}, \frac{3}{9})$ **then**
  $\mathcal{O} \leftarrow B_2$
else if $\mathcal{D}^r_t \in$ PERCENTILE$(\mathcal{D}^r_{1:t-1}, \frac{3}{9}, \frac{4}{9})$ **then**
  $\mathcal{O} \leftarrow B_1$
else if $\mathcal{D}^r_t \in$ PERCENTILE$(\mathcal{D}^r_{1:t-1}, \frac{5}{9}, \frac{6}{9})$ **then**
  $\mathcal{O} \leftarrow A_1$
else if $\mathcal{D}^r_t \in$ PERCENTILE$(\mathcal{D}^r_{1:t-1}, \frac{6}{9}, \frac{7}{9})$ **then**
  $\mathcal{O} \leftarrow A_2$
else if $\mathcal{D}^r_t \in$ PERCENTILE$(\mathcal{D}^r_{1:t-1}, \frac{7}{9}, \frac{8}{9})$ **then**
  $\mathcal{O} \leftarrow A_3$
else if $\mathcal{D}^r_t \in$ PERCENTILE$(\mathcal{D}^r_{1:t-1}, \frac{8}{9}, 1)$ **then**
  $\mathcal{O} \leftarrow A_4$
else
  $\mathcal{O} \leftarrow m$ {m for this is a comment}
end if
return $\mathcal{O}$

III. Genetic Algorithm as a mean to study the market through time

In this section we will specify the genetic algorithm which we have used to study our problem. Throughout this subsection we will refer to Micro and Macro increments.

Definition We will define two types of iterations:

- the first type being **Micro** corresponding to an infinitesimal increment in our environment namely, an increment in which a strategy S analyses and in turn changes the order book by placing a order itself.
- the second type being **Macro** corresponding to a generational increment in our environment namely, *a certain equal number* of Micro increment per strategy leading to a calculation of P&L and a survival process[18] based on this P&L.

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 regretion like strategies, N_k^r the number of xo**r** like strategies and N_k^o the number of ***o**ther unclassified* strategies[19]. The relationship between these entities can be summarized by equation (6).

$$N_k = N_k^e + N_k^m + N_k^r + N_k^o \tag{6}$$

A strategy will consist of a topology $\mathcal{T}$, a rolling P&L $\mathcal{P}$ and a common orderbook $\mathcal{O}$ as shown by equation (7).

$$\mathcal{S} \triangleq \{\mathcal{P}, \mathcal{T}, \mathcal{O}\} \tag{7}$$

Remark One may ask why have we 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 IV-C:

- N_k^e behaves in mathematical biology like the number of pr**e**ys in a Lotka Volterra (LV) 3 species equations [3]
- N_k^m behaves in mathematical biology like the number of **m**ixed (both prey and predator) in a LV 3 species equations.
- N_k^r behaves in mathematical biology like the number of super predato**r**s in a LV 3 species 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 The Survivor set[20] is the set of strategies with

[18] explained next

[19] This label will be the same in section IV-C.

[20] or alternatively alive process

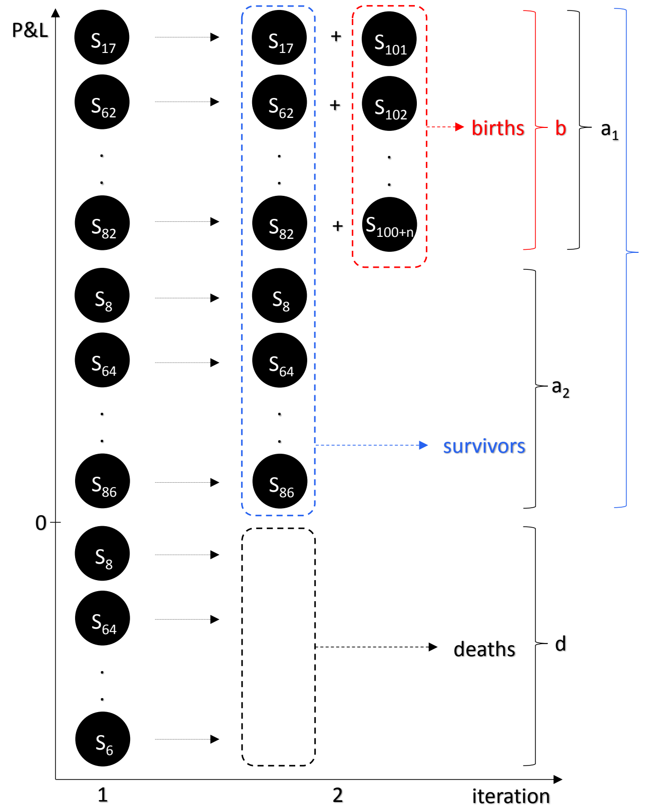

Fig. 15. Illustration for the Death and Birth processes in our GA

a positive P&L. Namely if $\mathcal{S}_a = S_{(1)}, S_{(2)}, \ldots, S_{(s)}$ with $S_{(s)} \geq 0$ and $S_{(s+1)} < 0$. We will subdivide this set by distinguishing:

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

Definition We will call the Birth process, the first half of survived strategies. Namely, if $a_1 = b = \lfloor \frac{s}{2} \rfloor$ the strategies $\mathcal{S}_1 \ldots \mathcal{S}_{a_1}$ will both survive and reproduce and create a set of equal size but with a slightly different topology and with carnality $b = a_1$.

Definition 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 $s = a_1 + a_2$, $a_1 \geq a_2$, $a_1 = b$. Figure 15 illustrates these few definitions.

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[21] which produced

[21] so no crossover in this model

it but be at the same time be a bit different to allow the ecosystem to evolve. We have seen in section II that the DNA of our strategies is essentially their topology $\mathcal{T}$ (which is itself a combination of weights). We will therefore concentrate on performing the re-sampling on the weights of the offspring. Recall that the pdf of the beta distribution, is given by

$$\mathcal{B}eta(x;\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\, x^{\alpha-1}(1-x)^{\beta-1} \tag{8}$$

with $\Gamma(n) = (n-1)!$, $0 \leq x \leq 1$, 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 topology decision making of figure 8. 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. We can see how the shape parameters can achieve these targeted resamplings in figure 16. This way we can prevent too large deviations and rather select small incremental changes and intuitively follow the principles of selection. We can see that the $\mathcal{B}eta(x,1,7)$ or $\mathcal{B}eta(1-x,1,7)$ both concentrate a great deal of the distribution towards 0 and 1 respectively. Likewise $\mathcal{B}eta(x,3,7)$ and $\mathcal{B}eta(x,5,7)$ provide a more Gaussian like distribution towards in between zones which is what we want.

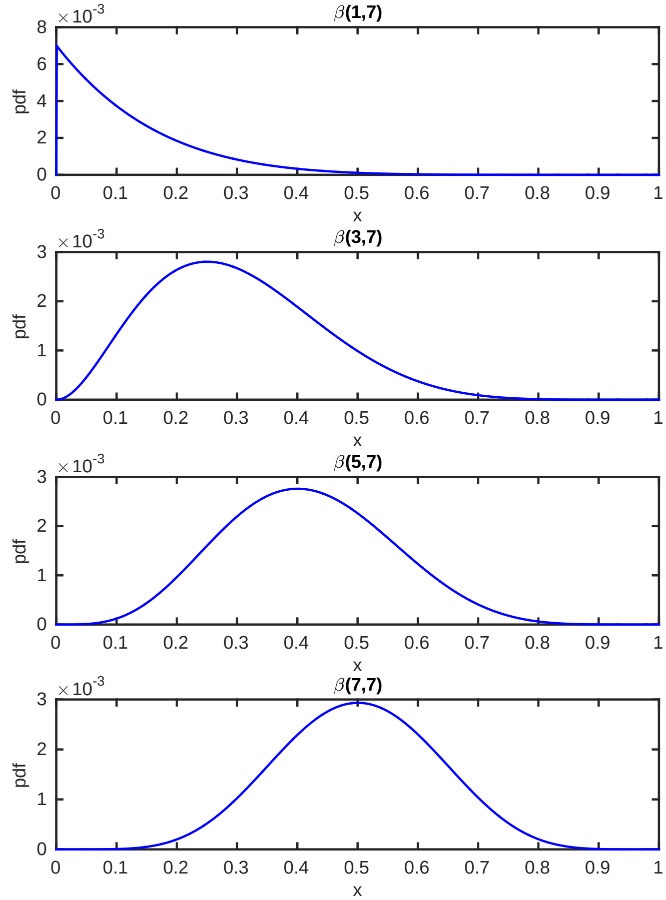

Fig. 16. PDF of the beta distribution for different combinations of (α,β)

$$\begin{aligned}
\mathcal{D}(\tilde{x}) &= 1_{\tilde{x}\leq -\frac{1}{5}}\mathcal{B}eta\Big(\frac{\tilde{x}+1}{2};\alpha(\tilde{x}),\beta\Big) \\
&\quad +1_{\tilde{x}>\frac{1}{5}}\mathcal{B}eta\Big(1-\frac{\tilde{x}+1}{2};\alpha(\tilde{x}),\beta\Big) \\
&\quad +1_{|\tilde{x}|\leq\frac{1}{5}} f(\tilde{x})
\end{aligned} \tag{9}$$

$$\alpha(\tilde{x}) = \begin{cases} 1, & \text{if } 1 > |\tilde{x}| \geq \frac{3}{4} \\ 3, & \text{if } \frac{3}{4} > |\tilde{x}| \geq \frac{1}{2} \\ 5, & \text{if } \frac{1}{2} > |\tilde{x}| \geq \frac{1}{5} \end{cases} \tag{10}$$

$$F(k) = \begin{cases} \frac{1}{10}, & \text{if } k \leq -\frac{1}{5} \\ \frac{8}{10}, & \text{if } |k| < \frac{1}{5} \\ 1, & \text{if } k \geq \frac{1}{5} \end{cases} \tag{11}$$

with $\tilde{x} \in]-1,1[$, $\beta = 7$ and the function $\alpha(\tilde{x})$ modeling the interval of condition, 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.

IV. Results and Parallel to Predator Prey Models & Evolutionary Dynamics

A. *Observations & Interpretation*

1) Results: Many improvements can be made with the coding exercise since it proved to be a challenging task, however, with some of the idiosyncratic simplifications used on the fly we get the results from figure 17 which represents one simulation of the HFTE model from inception using the methodology from this paper.

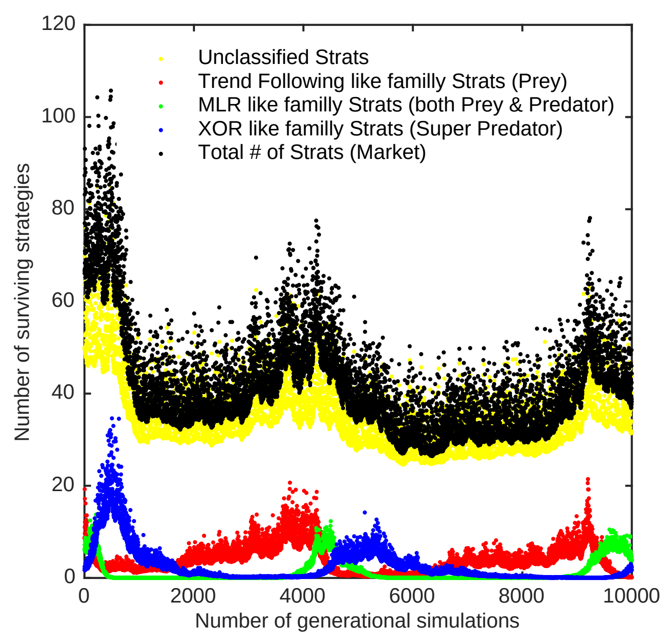

Fig. 17. A HFTE Simulation

2) Observations: Figure 17 is a plot through time for the concentration of each of the strategies explained in details in section II. The yellow dots correspond to unclassified strategies and the black dots correspond to the addition of the 3 different types of strategies on top of the yellow dots, corresponding overall to the total number of alive strategies indirectly corresponding to the general heath of the market (strong when the black dot is high and vice-versa when low). As we have mentioned before, the proper detailed definition of what the 3 strategies are, will be addressed in a subsequent paper but of the sake of keeping intuition we will call topology 1 *Trend Following* (TF), topology 2 *Multi-Linear Regression* (MLR) and topology 3 *XOR*. We will point to two zones:

- zone A: from around simulation 0 to 500
- zone B: from around simulation 3500 to 4500
- zone C: from around simulation 7500 to 9000

We notice the following similarities between zone A and B:

- 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,
- The XOR strategies frequency increases suddenly in the middles of the zones and decreases slowly

3) Interpretation: We propose the following interpretation for the observations from figure 17:

- TF strategies are what people commonly call self fulfilling like prophecies strategies 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 IV-C would be an ultimate prey which given an environment without any predator would never die and actually grow exponentially.
- The XOR strategy is a super predator strategy (similar to the z parameter in section IV-C) and feeds on the MLR strategies.
- MLR are both predator and prey strategies. It feeds onto the TF strategies but are used as preys 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 preys (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.

B. Comparison to Game Theory

In this section we will present few results from the world of Game Theory in order to make the parallel to our results for a better understanding of how these strategies interact with each other.

1) Prisoner's dilemma: The prisoner's dilemma (PD) is a well known Game Theory introductory concept. The way it is usually explained is that couple of criminal associates are taken into separate rooms and being independently interrogated. The prosecutor wants to close the case and send someone in prisons so he offers a deal to both captives. If the criminals both cooperate (C), nobody goes to prison but they each get a heavy fine. If one denounces/deceits (D) the 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 matrix[22] of 2 by 2 with CC, CD, DC and DD with respective payoffs (2,2), (0,3), (3,0) and (1,1). The reason why this game theory concept is within the family of dilemmas is because although the prisoners clearly should cooperate here, given that they do not know what the other is going to do, by expectation (with equal probability for a C and a D) any user should deceit given that the expectation of the payoff for a deceit is 2 as opposed to a 1 for a cooperation.

2) Axelrod's computer tournament: However this dilemma presented in the previous subsection proved to shuffle the rules of payoff strategy optimality when the game became iterative, Robert Axelrod main contribution to the field. Indeed Axelrod [1], [2] designed a computer tournament which aim was to take a look at what strategy would prevail in an iterative format. In that occasion he invited few Mathematicians, Computer Scientists, Economists and Political Scientists to code a strategy they believed could win such tournament with the constraints of a PD rules in which it is not known when the tournament will stop[23]. Many strategies were thrown into this ecosystem in form of a tournament ranging from being simplistic like "Always Deceit" (AD) strategy[24] to many other more complicated strategies which generic representation can be looked at in figure 18b). Surprisingly the Tit For Tat (TFT) strategy came at the top of this tournament. The TFT is considered in the literature to be a nice strategy, meaning that it is never the first to deceit (its first move is by design to be a C), but it is also a strategy that is able to retaliate in situation in which it was previously deceited. Finally, it is a strategy that is able to forgive meaning that if he sees that

[22]figure 18a)

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

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

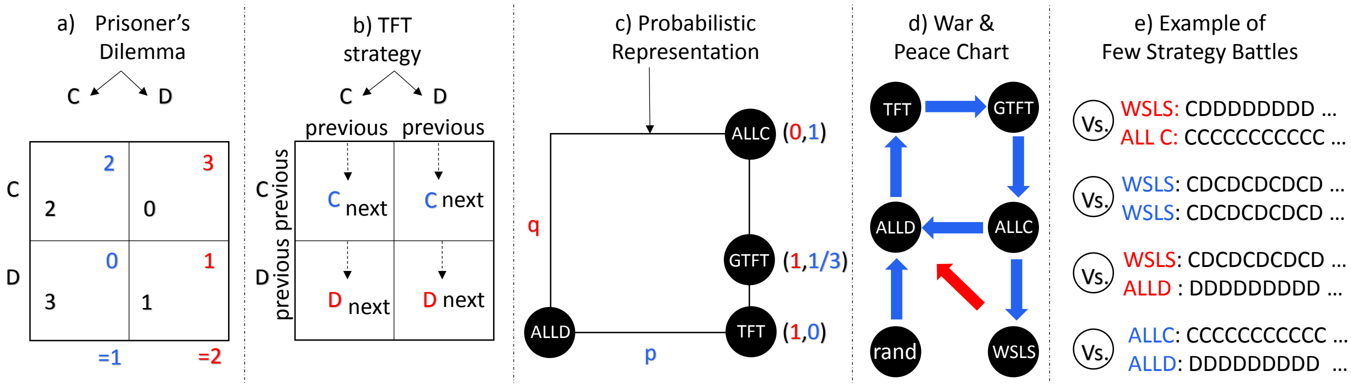

Fig. 18. Some classic Game theory representations

the adversary has decided to cooperate after a deceit, then he switches back to a C.

3) Evolutionary Dynamics: Martin Nowak [21] recently enhanced some of Axelrod's work by introducing new strategies and further developing the concepts of invasion/dominance[25] within a competitive strategic ecosystem. For instance as we can see from figure 18d) 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[26]. Indeed an ecosystem composed of a set of unbiased random strategies (that would randomly C or D) would invite the invasion of an ALLD (always defect) kind. In term the frequency of ALLD would take the ecosystem which would invite the TFT strategy which would benefit from the mutual cooperation when within the same proximity etc ... Figure 18d) 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:

- TFT (*Tit of Tat*) developed in the previous section
- GTFT (*Generous Tit of Tat*) which makes it slightly less grudge prone compared to the TFT as it only deceits for 2 successive D's from the opponent.
- WSLS (*Win-Stay, Lose-Shift*) that outperforms tit-for-tat in the Prisoner's Dilemma game [21], [28]
- ALLD (*Always Deceits*) which is self explanatory
- ALLC (*Always Cooperates*) which is also self explanatory
- rand (*Random Strategy*) which outputs a C or a D with equal probability.

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

C. Theoretical Biology & Predator/Prey models

It was discussed in the 1960s [9] that complexity in an ecosystem insures its stability or to keep the same jargon "communities not being sufficiently complex to damp out oscillations" [7], [11] have a higher likelihood of vanishing. It is widely accepted, in the context of ecosystem simulation, that complexity should always arise from simplicity [17], [4].

1) literature review: The diversity-stability debate in the context of ecosystem modeling has been ongoing since the 1950s [18] with no consensus being ever reached. It was initially believed [18], [14], [6] that given that nature was infinitely complex a more diverse ecosystem should insure more stability. This assertion was however ultimately challenged through rigorous mathematical specification [17], [32], [22] in the 1970s and 1980s by using Lotka Volterra's Predator/Prey model initially published in the 1920's [31], [13] with similar "non-intuitive" results. More recently the work has been extended to small ecosystems of three-species food chain [3]. The intuitive 3 species example we have chosen to discuss is the one containing Sharks (chosen to be the z parameter), Tuna (chosen to be the y parameter) and Small Fishes (chosen to be the x parameter), the idea being that tunas eat small fishes which in turn are eaten by a sharks. Without loss of generality sharks are assumed to die of natural causes and their decomposing bodies go on to feed the small fishes. The set of differential equations has been summarized in equation (12).

$$\begin{cases} \frac{dx}{dt} &= ax - bxy \\ \frac{dy}{dt} &= -cy + dxy - eyz \\ \frac{dz}{dt} &= -fz + gyz \end{cases} \tag{12}$$

where a is the natural growth rate of species x in the absence of predator, d the one of y in the absence of z. We also have b representing the negative predation effect of y on a and e the one of z on y. We also have g which mirrors the efficiency of reproduction of z in the presence of prey y. Note that we assume that x never dies of natural causes (if it's too old then it can't run fast enough to outrun predator y) but this is not the case for z since it is an alpha predator and therefore needs some natural death rate which is symbolized by f. This relatively simple system of three equations has been studied extensively [18] for stability. For example figure 22 represents a particular instance in which the system is unstable. Indeed, we can notice that

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

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

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 [18] 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 19, we have oscillation which are stable through time with the highest peek from the ultimate pray (x) coming first with the highest peek and the the one of the ultimate predator (z) coming last but with the smallest amplitude. This suggest that sophisticated working trading strategies[27] need enough prey like strategies[28] in the same ecosystem to get them to be profitable. One other interesting observation is that the total ecosystem population as depicted in the thick black line from the same figure suggest that it itself oscillates which may not necessarily be intuitive. Indeed one could have speculated that the loss of a species directly benefits the other and that therefore the total population should stay constant. This interesting observation suggest that the oscillations of a financial market may likewise be subject of similar dynamics: a financial ecosystem may go through periods in which it thrives followed by period in which it declines, the economy itself is cyclical with, some may argue oscillations which are more and more important like one depicted by the unstable ecosystem from figure 22. The stunning similarities of the

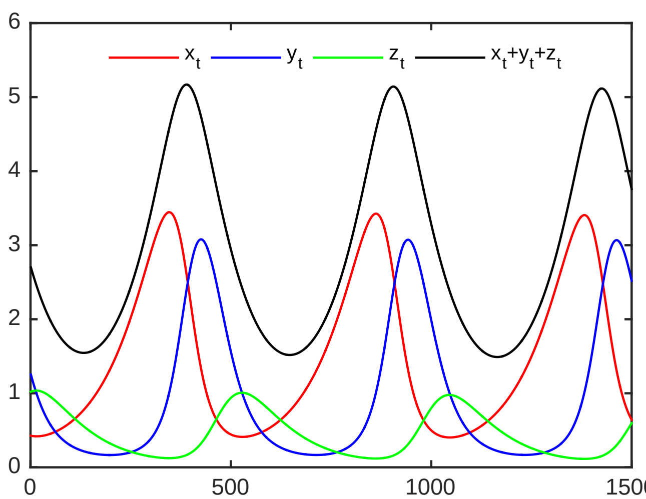

Fig. 19. The Lotka Volterra three-species food chain equation 12 with $x_1 = 0.5$, $y_1 = 1$, $z_1 = 2$ and $a = b = c = d = e = f = g = 1\%$

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 zero sum game like roots.

[27]perhaps from top algorithmic desks in top tier investment banks?
[28]perhaps the retail clients of the world?

Remark Note that these results corroborates some of the connections between utility functions and the Lotka Volterra model recently discussed. [27], [26], [24].

2) The interesting case of the simplified Four-Species Stochastic Lotka Volterra: If the reader is not entirely convinced of the results from figure 17, more specifically how these results can lead to the kind of real observations from figure 1 since the full formalization and boundaries of the 3 strategies are not provided and that therefore the analysis of section IV-A.3 is put in question, let's look at the following simplified stochastic system which is essentially a transform of the 4 species Lotka Volterra. The transform are listed below:

1) Assuming that the P&L lost by a strategy is linearly gained by another leads us to assume that $a = a_1 = \ldots = a_4 = 1$ and $b = b_1 = \ldots = b_4 = 1$.
2) The regulators may assume that the maximum number of strategies M can be fixed to the number of market participants: here we have chosen four[29]. 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.
3) Also we may assume that the notional associated by a certain family type may impact the market. For instance you may assume that a new market participant may enter and impact the market by entering it at a particular level of predation. Meaning that the types of strategies the participant may chose to implement may be in our case:
   - XOR type of strategy (with positive jump in notional upon entry φ_z)
   - MLR type of strategy (with positive jump in notional upon entry φ_{y_2})
   - TF type of strategy (with positive jump in notional upon entry φ_{y_1})
   - Unclassified type of strategy (with positive jump in notional upon entry φ_x)

with $\varphi_i \sim 1_{\epsilon<0.01}\varepsilon$ with $\epsilon, \varepsilon \sim U[0, 1]$. These simplifications give equation (13).

$$\begin{cases} \frac{dx}{dt} &= x - xy_1 + \varphi_x \\ \frac{dy_1}{dt} &= xy_1 - y_1y_2 + \varphi_{y_1} \\ \frac{dy_2}{dt} &= y_1y_2 - y_2y_3 + \varphi_{y_2} \\ \frac{dz}{dt} &= -z + y_3z + \varphi_z \end{cases} \tag{13}$$

Figure 20 is a simulation from this "mirror simplified HFTE model". We can see notice from 5000 until zone 7000 a very similar situation with regular oscillations followed by a crash. If we increase the timescale (figure 21) we actually see that periods of crashes are triggered by an $\varphi_i > 0$ with $i \in \{x, y_a, y_2\}$, followed by an increase in frequency from the ultimate predator strategy z before the correction in the market occurs.

[29]In the n-species simplified stochastic Lotka Volterra, the $n = 2$ and $n = 3$ behave a bit differently but the case $n \geq 4$ is a family on its own when you take a look at the "market" level.

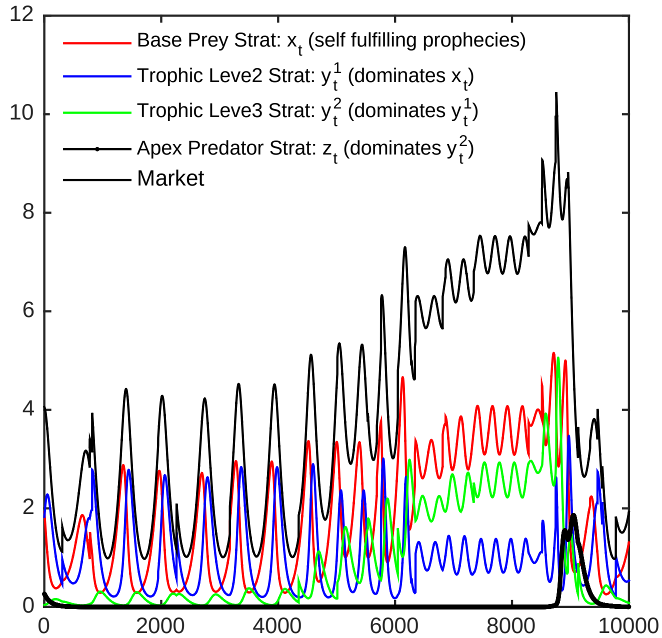

Fig. 20. The simplified stochastic Lotka Volterra 4-species food chain of equation (13).

V. Application & Summary

A. Application

Many applications could be implemented and we may disclose few more in a subsequent paper but for the sake of keeping the paper relatively short we propose couple of applications. The first of these applications will be in trading which methodology will be expended in section V-A.1, and the second application we hope to be a new tentative approach in doing risk management at the regulatory level in section V-A.2.

1) Trading Application: In terms of providing market context, the arrival of High Frequency Trading on the Chinese commodities market has opened up a great deal of opportunities on the systematic trading end because:

- Commodities have historically been a physical trading asset class and the arrival of HFT in commodities is rather new compared to equities and FX therefore opportunities for making money are still quite vast.
- Regulations on the Chinese market makes the life of high frequency trader much easier than in heavily regulated regions like Europe or Americas.

In terms of providing additional information on the data: the data used consists of a week of data split in 12 months for 14 different commodities futures provided by a reputable HFT firm. In order to eliminate the survival bias, 15 random time-series were chosen in the month of December 2015 and 15 other random ones were chosen in the month of November 2015. All the major commodities were represented in these 15 random time-series, however few commodities seem to have limited data. Given that there is a bit of a learning associated in any machine learning exercise of this type, the commodities set that had little data were eliminated. In terms of limitations: the data seems to be generally speaking of good quality however few limitations were observed.

- First the data comes in every 500ms, which means some important information may be lost.
- Some time series have too few data on one day to do proper training (those have been dismissed in the backtesting for now).
- Some files seem to have corrupt data though more investigation is necessary in that domain (for now those red flag files have been dismissed)

If we were to provide the current limits, we would say that the trading signal used is very unsophisticated and can easily improved by:

- better formulating the trend en trend reversal signals
- better modeling the co-movement of the different futures with respect to each other.
- better optimizing the signal with respect to the cost process.

However, despite these limitations, as we can see from the table V-A.1, the in and out of sample statistical performance are highly unlikely to have come about by chance.

Simple Stats	Out of Sample	In Sample
Annualized Sharpe Ratio	19.9	32.6
Average Max Drawdown	0.7%	0.7%

TABLE III
In and Out of Sample Backtesting Trading results

2) Regulatory Implications: The second and last immediate application we will take a look at in the context of this paper is the one of systemic risk. Given that this paper proposes that the fluctuations of the markets are linked to the frequency of the strategies composing the ecosystem of the market, we propose a model which would take advantage of this assumptions to build new approaches in doing 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's develop a bit the argument. Suppose now that we label strategies of figure 9, 11 and 13 by respectively x, y and z and that we use equation (12). If we can somehow guess what the frequency of x, y and z are in the ecosystem, then we can study whether or not the ecosystem is stable [3].

Remark The work around guessing could actually be as simple as asking the market participants to provide the code of their systematic trading strategies under the motivation of national interest and for the sake of the stability of the financial markets. In exchange the regulators would agree on keeping the strategies confidential. This may raise ethics questions as the concept of risk would dangerously flirt with the concept of "strategy destiny" as the regulators would

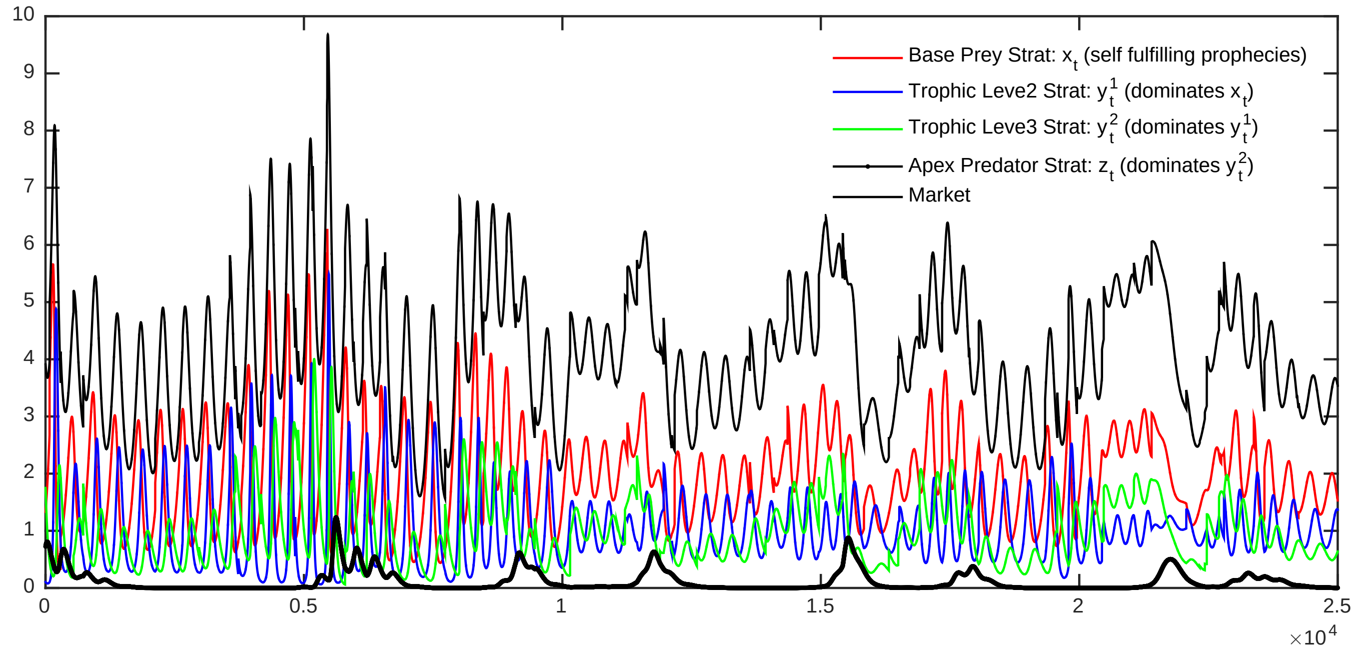

Fig. 21. A longer simplified stochastic Lotka Volterra 4-species food chain of equation (13)

be able to analyze which market participant would end up profiting and losing from the environment before these losses actually occur.

Now going back to the actual mathematical study of the stability of the financial market. Answering if the financial market is stable would come to studying the Jacobian matrix J from equation (14).

$$J(x,y,z) = \begin{bmatrix} a-by & -xb & 0 \\ yd & -c+dx-ez & -ye \\ 0 & -zg & -f+gy \end{bmatrix} \quad (14)$$

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[30]. More specifically if all eigenvalues of $J(x,y,z)$ have negative real parts then our system is asymptotically stable. Figure 22 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 systems of equation (12) is one of them. 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? 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[31]. This is where the section on guidance for future research has been added. In the meantime though in order to encourage the motivated reader to think about the problem of stability in the financial markets we introduce the following conjecture:

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

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

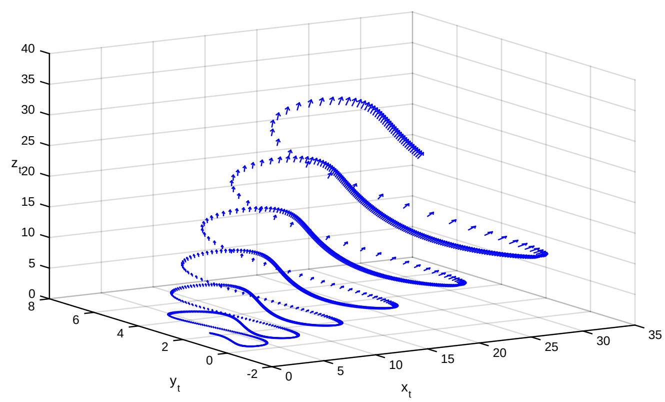

Fig. 22. The Lotka Volterra three-species food Chain equation 12 with $x_1 = 0.5$, $y_1 = 1$, $z_1 = 2$, $a = b = c = d = e = f = 1\%$ and $g = 1.6\%$

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

Remark Note this conjecture can be studied indirectly or at least intuitively using some of the finding in mathematical biology associated with diversity in ecosystem and stability[32].

VI. Summary

A. *Summary*

We have started this paper by pointing to a puzzling observation from the newly born high frequency commodities

[32]though no consensus is reached there either

Ecosystem Size	Main Players	Main Results	Main Problems
3 Strategies	"sort of" the Trend Following, Multi-Linear-Regression & XOR	parallel to Lotka-Volterra (LV) 3-Species Predator Prey model	Maps between network topology & strategy families need to be better specified
4 Strategies	Above + random strategy	Stability of the market can be studied using the parallel mathematical biology problem	same as above
⋮	⋮	⋮	⋮
n Strategies	Above + sets of Ignoratus, Praedor and Servus Dominum	More Complex LV with food web structure + potential strategy with foresight & farming	Diversity in financial strategies in the market lead to its instability?

TABLE IV
ROUGH SUMMARY OF THE DIFFERENT MODEL DISCUSSED

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 that on 06/08/2011 fascinating obviously patterned oscillation occurred on the commodities market. We have proposed in this paper that these oscillation are due to the interactions of the different strategies participating in the market and participating in the fluctuations of the market. We have proposed that these oscillation are of the same nature as the Lotka Volterra model. In order to test our hypothesis we have proposed a topology which we proved is able to model the HFT strategies known to the market participants. More specifically we have illustrated how it can achieve the Trend Following, the MLR and the XOR strategies. This topology is then randomly sampled at inception (the random seed) in this swarm market and a simple genetic algorithm is enforced to allow us to study the market and its participants through time. The results are commensurate with the Lotka Volterra model as well as some of the other Game Theory results, more specifically around strategy invasion that we have also made a comparison to. Finally, we have applied our findings to a simple trading exercise as well as proposed a new way to monitor the markets for regulatory purposes.

B. Guidance for Future Research

1) The regulatory aspect: As we have seen in section V-A.2 the regulatory implications from this research naturally invites us to explore a research project in which we would try to *guess*[33] the frequency of each types of strategies using the LotkaVolterra multi species models as likelihood functions. We propose to use a particle filter on scenarios to achieve this point. We will discuss this particular point in a subsequent paper.

2) The options market: We have recently introduced a new parametrisation of the implied volatility surface [16], [5] and have established that de-arbing is a convoluted mathematical optimization which simplification can be illustrated by figure 23. For the sake of making the notation a bit more

[33]given that we cannot ask the market participants to provide us with their strategies

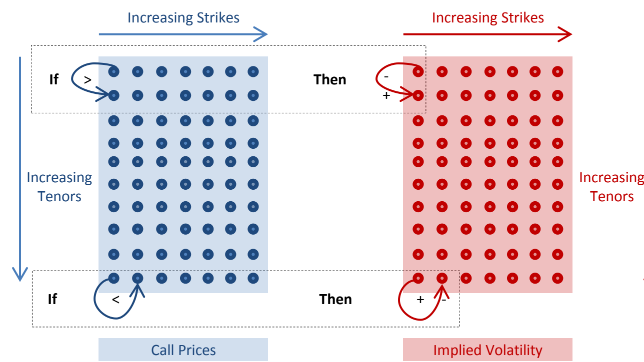

Fig. 23. Visualization for the core simple de-arbing idea

intuitive, we use the notation from table V where equation (15) provides the relevant equivalences.

$$P_t^{i-1,j} := C_t(Ke^{-r2\Delta}, T-\Delta) \quad (15a)$$
$$P_t^{i,j-1} := C_t(K-\Delta)e^{-r\Delta}, T) \quad (15b)$$
$$P_t^{i,j} := C_t(Ke^{-r\Delta}, T) \quad (15c)$$
$$P_t^{i,j+1} := C_t(K+\Delta)e^{-r\Delta}, T) \quad (15d)$$
$$P_t^{i+1,j} := C_t(K, T+\Delta) \quad (15e)$$

where $C_t(.)$ representing the call price under the relevant asset class assumption[34]. We aim at studying the Bayesian

	$P_t^{i-1,j}$	
$P_t^{i,j-1}$	$P_t^{i,j}$	$P_t^{i,j+1}$
	$P_t^{i+1,j}$	

TABLE V
NAMING GRID ASSOCIATED TO FIGURE 23.

probability problem of equation (16).

$$p\Big(\mathbf{P}^{i,j} - l(\mathbf{P}^{i,j})\Big| l(\mathbf{F})\Big) \quad (16)$$

[34]eg: Log-Normal diffusion in Equities, Normal diffusion for rates and Garman Kohlhagen for FX.

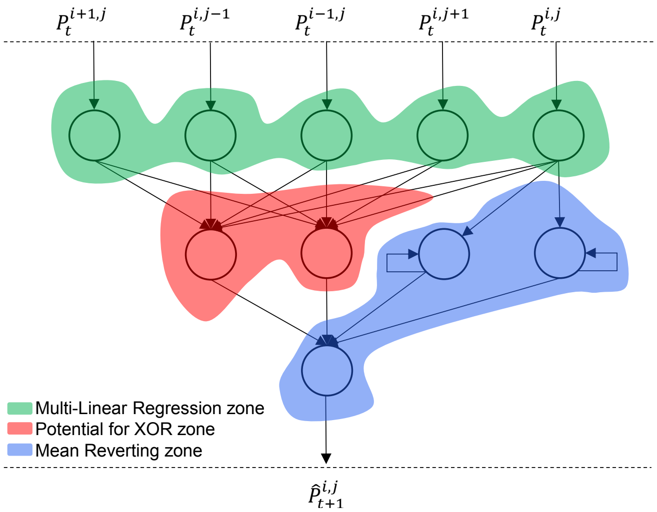

Fig. 24. Illustration of an ANN aiming at predicting the price of an option based on its adjacent points.

where, $\mathbf{F} = \mathbf{P}^{i,j}, \mathbf{P}^{i-1,j}, \mathbf{P}^{i,j-1}, \mathbf{P}^{i,j+1}, \mathbf{P}^{i+1,j}$ in the discrete space, $\mathbf{P} = P_{\tau \in [1,t]}$ and l represents the lag inducing function such that $l(P_{t+1}) = (P_t)$. The implied volatility is very different product than spot because it has a tendency to mean revert, it is very much subject to what the adjacent points are doing and reacts in a lower frequency than spot. Our aim will be to study the HFTE in light of these observation. However it is interesting to notice that, already that these observation could be addressed by a modification of the HFFF (figure 24). Following the rational from section II we need to create an learning architecture that would incorporate the following observations:

- Presumably the price point $P_t^{i,j}$ can be best approximated by the 4 adjacent points, a simple MLR[35] can be used to model this idea (green part of figure 24)
- The second point to notice is that each point of the implied volatility is a mean reverting stochastic process and this can be modeled in terms of network architecture by a spread of EWMA[36] (blue part of figure 24)
- At least one hidden layer to address some of the economical drivers leading to a need for an architecture that could learn XOR like functions like we saw could sometimes be necessary in algorithmic trading from table II-B.3 (red part of figure 24).

Remark Note that the XOR like functions may not be as necessary as the dynamics of spot since vol may be driven by economical factors that are different especially especially if we study the problem at different timescales. This suggests that the red part of figure 24 may at the end be the identity function. For the sake of keeping that door open though, we have left it in our network topology.

We will also see how the parameters introduced in the newly published IVP model [16] can contribute in fine tuning the learning process as well as it execution strategy which requires insight around liquidity.

3) Network Topology For Implied Volatility Point Dynamics Learning: Following the rational from section II we need to create an learning architecture that would incorporate:

- Presumably the price point $P_t^{i,j}$ can be best approximated by the 4 adjacent points, a simple MLR[37] can be used to model this idea (green part of figure 24)
- The second point to notice is that each point of the implied volatility is a mean reverting stochastic process and this can be modeled in terms of network architecture by a spread of EWMA[38] (blue part of figure 24)
- At least one hidden layer to address some of the economical drivers leading to a need for an architecture that could learn XOR like functions like we saw could sometimes be necessary in algorithmic trading from table II-B.3 (red part of figure 24).

C. Possible Strategy Classification

As we have seen in section II, the formalization of these pure prey, mixed prey/predator, and pure predator strategies need to be formalized more rigorously.

1) Guidance on Strategy Naming: The tentative naming inspiration initially came from:

- exploring how ecologists have, by convention, named the different species in latin (eg: *Homo Habilis*, *Homo Neanderthalensis*, *Homo Sapiens* etc...)
- noticing that the naming was descriptive (eg: *Homo Habilis* is supposed to have used tools)
- noticing that being extinct does not necessary mean a species would not have been dominant today in certain conditions (for instance it's not hard to imagine that, had the *Homo Sapiens* never been born, then *Homo Neanderthalensis* would have probably been able to enslave other animals (eg: cattle) almost the same way we do. Therefore ecological niche supersedes any human conception that only the most "superior" species prevail).

2) Core Naming Conventions: It seems that strategy in latin gives *Militarium* and that ancestor seems to be *Antecessoris* and therefore:

- **Ignoratus**, for example the *Militarium Ignoratus* would essentially correspond to a random strategy without much "intelligence" (example a TF in finance) corresponding, in a biological ecosystem, to perhaps a vegetarian (eg: mouse) in an advanced ecological world.
- **Praedor**, for example the *Militarium Praedor* would correspond to a strategy that can dominate the *Militarium Ignoratus* like species. Note that few *Militarium Praedor* could be in competition at one point which could ultimately lead one of the *Militarium Praedor* to go extinct. The analogy to our world could be the one in which Leopard and Lions both compete for the Antelope in Africa but Lions are slightly better at it and

[35] $y_i = \beta_1 x_{i-1,1} + \cdots + \beta_9 x_{i-1,9} + \varepsilon_i$ from equation (5)
[36] $\hat{x}_t = (1-\lambda)x_t + \lambda \hat{x}_{t-1},\ \lambda \in [0,1]$ from equation (4)

[37] $y_i = \beta_1 x_{i-1,1} + \cdots + \beta_9 x_{i-1,9} + \varepsilon_i$ from equation (5)
[38] $\hat{x}_t = (1-\lambda)x_t + \lambda \hat{x}_{t-1},\ \lambda \in [0,1]$ from equation (4)

which may lead ecologist to speculate that the Leopard is now an endangered species. One type of well know extinct ancestor of the big cats is the *saber-toothed cat*. The *Militarium Praedor Antecessoris* in our ecosystem would correspond to that Leopard who would have gone extinct.

- **Servus Dominum**, for example the *Militarium Servus Dominum* would correspond to a strategy that can dominate and have foresight with respect to the *Militarium Ignoratus* or/and the *Militarium Praedor* like species. An example of such biological ecosystem could be one in which Humans[39] who would for example feed a Fox[40] population with Mice[41] in the context of fur farming. Can a strategy have so much foresight and understanding of the market that it can implement this idea on the markets? We will address these questions on subsequent papers.

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 the Oxford Man Institute of Quantitative Finance in the area of Machine Learning. He is also the author of numerous publications, including cover stories of Wilmott magazine and mathematical models (eg: Cointelation & IVP) that are now taught in the CQF.

[39]by analogy the *Militarium Servus Dominum*
[40]by analogy the *Militarium Praedor*
[41]by analogy the *Militarium Ignoratus*

References

[1] Robert Axelrod. The evolution of cooperation. *Basic Books*, 1984.
[2] Robert Axelrod. The complexity of cooperation: Agent-based models of competition and collaboration. *Princeton University Press*, 1997.
[3] Erica Chauvet, Joseph E. Paullet, Joseph P. Previte, and Zac Walls. A lotka-volterra three-species food chain. *Mathematics Magazine*, 75:243–255, 2002.
[4] John Horton Conway. *On Numbers and Games*. 1976.
[5] Babak Mahdavi Damghani and Andrew Kos. De-arbitraging with a weak smile. *Wilmott Magazine*, 2013.
[6] C. S. Elton. Ecology of invasions by animals and plants. *Chapman & Hall, London*, pages 228–233.
[7] C. S. Elton. The ecology of invasions by animals and plants. *Methuen*, 1958.
[8] Martin Gardner. Mathematical games. *Scientific American*, 1970.
[9] Andrew G. Haldane and Robert M. May. Systemic risk in banking ecosystems. *Nature*, 469:351–355, 2011.
[10] Rama Cont Hamed Amini and Andrea Minca. Resilience to contagion in financial networks. *Mathematical Finance*, 26:329–365, 2016.
[11] G. E. Hutchinson. Homage to santa rosalia, or why are there so many kinds of animals? *Nature*, 93:145–159, 1959.
[12] S. Mankad; G. Michailidis; A. Kirilenko. Discovering the ecosystem of an electronic financial market with a dynamic machine-learning method. *Algorithmic Finance*, 2:2:151–165, 2013.
[13] Alfred J. Lotka. Elements of physical biology. *Williams & Wilkins Co.*, 1925.
[14] R. H. MacArthur. Fluctuations of animal populations and a measure of community stability. *Ecology*, 36:533536, 1955.
[15] Babak Mahdavi-Damghani. Utope-ia. *Wilmott Magazine*, 60:28–37, 2012.
[16] Babak Mahdavi-Damghani. Introducing the implied volatility surface parametrisation (ivp):application to the fx market. 2015.
[17] Robert M. May. Simple mathematical models with very complicated dynamics. *Nature*, pages 459–467, 1976.
[18] Kevin Shear McCann. The diversitystability debate. *Nature*, pages 228–233.
[19] Marvin Minsky and Seymour Papert. *Perceptrons: An Introduction to Computational Geometry*. 2nd edition with corrections, first edition 1969 edition, 1972.
[20] Nanex. Strange days june 8'th, 2011 - natgas algo, 2011.
[21] Martin Nowak. *Evolutionary dynamics : exploring the equations of life*. 2006.
[22] S. L. Pimm and J. H. Lawton. On feeding on more than one trophic level. *Nature*, 275:542–544, 1978.
[23] Warren McCulloch; Walter Pitts. A logical calculus of ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5:115–133, 1943.
[24] Angel Manuel Ramos and Tomas Roubicek. Nash equilibria in non cooperative predator-prey games. *Applied. Mathematics and Optimization*, 56:211–241, 2007.
[25] Frank Rosenblatt. *Principles of neurodynamics: perceptrons and the theory of brain mechanisms*. 1962.
[26] Nicola Serra. Possible utility functions for predator-prey game. *Journal of Game Theory*, 3(1):11–18, 2014.
[27] Nicola Serra. Utility functions and lotka-volterra model: A possible connection in predator-prey game. *Journal of Game Theory*, 3(2):31–34, 2014.
[28] 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.
[29] 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.
[30] Nassim Nicholas Taleb. *The Black Swan*. 2007.
[31] Vito Volterra. Variazioni e fluttuazioni del numero dindividui in specie animali conviventi. *Mem. Acad. Lincei Roma*, pages 31–113, 1926.
[32] P. Yodzis. The stability of real ecosystems. *Nature*, 289:674–676, 1981.