Financial Time Series Forecasting using Agent Based Models in Equity and FX Markets

Abstract—We investigate the application of machine learning Agent Based Modelling (ABM) techniques to model and forecast various financial markets including Foreign Exchange and Equities, especially models that could reproduce the time-series properties of the financial variables. We model the economy by considering non-equilibrium economics. We adopt the features that are required for modelling non-equilibrium economics using ABMs and replicate the non-equilibrium nature of the financial markets by considering a set of bounded rational heterogeneous agents, with different strategies that are ranked according to their performance in the market. We consider markets where there are different agents interacting among themselves and forming some sort of patterns. For example, the patterns are equity prices or exchange rates. While the agents have been interacting in the artificial market, the generated patterns (price dynamics) they co-produce would match with the real financial time-series. In order to get the best fit to the real market, we need to search for the best set of artificial heterogeneous agents that represents the underlying market. Evolutionary computing techniques are used in order to search for a suitable set of agent configuration in the market. We verify the forecasting performance of the artificial markets by comparing that with the real financial market by conducting out-of-sample tests.

I. INTRODUCTION

Like fluids flow from a region of high pressure to a trough or low pressure, in financial markets wealth gets transferred from region/individual with excess of it to a region/individual in need of it. Wealth can be stocks (stock market), currencies (FX market) etc and its dynamics are reflected in the form of price changes in the markets. Understanding and studying such markets are of prime importance, as they have a direct influence on an individual’s financial well being, behaviour of business and economy as a whole. Traditionally there are two main approaches that can be used to study the financial markets: (1) consider the time-series of the financial variable as random processes and the dynamics involved is formulated as stochastic differential equations and (2) solving for the financial variable by assuming individual agent’s preferences. We consider the second approach using the ideas of out-of-equilibrium economics. Unlike neoclassical economics, which studies equilibrium models and related patterns that would induce no further reaction from the economic agents, out-of-equilibrium economics tries to model the frustration of the agents who are trying to reflect and react to the patterns in the market. That means, in out-of-equilibrium we study how the agents’ behaviour might not be consistent with the patterns. This approach considers that the agents are heterogeneous and their rationality is bounded. When the agents’ rationality is bounded, it is appropriate to replace their deductive reasoning nature that is assumed in equilibrium economics with inductive reasoning nature. In a way, equilibrium economics is a special case of out-of-equilibrium.

One way of adopting the features that are required for modelling non-equilibrium economics is Agent Based Models. The emergence of agent based modelling can be attributed to various factors. However, the disadvantages of analytical models namely, incapability of generating emergent phenomena, incapability of exercising free choice, Lucas critique for macroeconomic models [1], lead to the widespread use of agent based modelling. The main advantages of ABMs include - the ability to cope with heterogeneous and bounded rational agents, this helps greatly to study decentralized markets [1]. The other advantages of ABMs for the study of financial markets are: (1) ABMs are capable of capturing the emergent phenomena and (2) allowing for unanticipated behaviour to emerge [12]. Moreover, dynamically varying state of the agents can be incorporated in the model over time. This includes tinkering the complexity, which allows us to incorporate various levels of aggregation [13]. In this paper, we consider ABMs and econo-physics methods for modelling the economy that gives better forecasting and throws light on arbitrage opportunities in the given market. Analysing systems with several particles acting homogeneously/heterogeneously is a very common practice in physics. We extend this idea to economics by applying the concepts of the minority game - our ABMs are constructed using different proportions and variants of minority /majority agents (“economic particles”). Related literature can be found in [5], [3], [6]. Apart from the similarities with the existing literature, our work is significantly different in two aspects. First, we introduce better evolutionary computing techniques [19] that could enhance the learning capabilities of the agents in the model that replicate those of the traders in the market and reduce the computing time. Second, we extend the methodology to FX markets and to determine the market composition or mood at any particular period. In other words this provides an comprehensive overview of the market macro-structure.

In order to model the given financial market (foreign exchange, equity etc.), we consider a set of heterogeneous artificial agents that mimic the behaviour of the agents in the real market. Each artificial agent has a set of different trading strategies that are ranked according to their performance in the market. The agents would update the ranking of the strategies from time to time according to their performances in trading. While the artificial agents have been interacting...
and the generated patterns that they co-produce may match with the real financial time-series. If so, one can argue that the artificial stock market can be used to describe the internal organization of the real market and to forecast the near horizon price dynamics. But, there are many possible ways to select the set of artificial agents that could represent the underlying real market and the problem turns out to be identifying the set of artificial agents that could be the best representation of the real market. For searching the extensive space of configurations, we employ a common evolutionary computing technique, Genetic Algorithms (GA). We use Genetic Algorithms (GA) and search for the best possible set of artificial agents by considering usual fitness evaluation, reproduction, crossover and mutation operations in GA. GAs are inspired by the process of natural evolution [9]. In GA, every candidate solution in phenotype space is encoded into genotype space as a binary string. Each individual in genotype space will be evaluated according to the fitness function which is chosen based on the problem. For reproduction, we assign probabilities to the individuals based on their fitness rank in the population in order to perform crossover, mutation, and reproduction operations. The process evaluation and reproduction terminates if the ideal solution is reached or it has run for a sufficient number of generations after which there is no change in the fitness of the population. It would be more interesting to understand the internal organization of the economic agents. After finding the best ABM to a real financial time-series, we study the internal structure (for example, trader’s behaviour and their strategies in the market) at a given time in the financial market. For that we categorize the agents according to the pay-off incentives provided to them like minority agents, majority agents and the variations of them [11]. With the available categories, we study the behaviour of the agents in the financial market at various points in time. In the next section we briefly discuss the related literature. Section 3 describes the methodology for creating artificial financial market and in Section 4, we discuss some of the experimental results and evaluate their performances.

II. BACKGROUND AND LITERATURE

Understanding the underlying pattern and behaviour behind a set of activities, using time-series analysis has been the major area of research over the last few decades. In a quest several models and hypothesis have been proposed for different problem sets. The main aim and the proposed hypothesis not only explain the observed behaviours, but predict the upcoming ones as closely as possible. Analysing past data and predicting the future trends has been extended massively to the financial domain. The main reason being the abundance of high quality time-series data produced in this domain. The data is rich and valuable in terms of the information contained. Mining and extracting such information will give an indication of the upcoming trends. However, the randomness that involves is a major challenge. As the number of possible outcomes increases, so does the complexity. The El-Farol bar problem [7] shows that how ordinary real life problems get complex in nature in terms of the number of possibilities or hypothesis each human being has towards it. These can be modelled using the inductive reasoning models whereby the multiple belief models adapt to the environment they jointly create. The idea can be extended to economic problems such as speculating the stock market using ABMs. The random element here is being the stock prices, which varies over time. Multiple agents with variety of working hypotheses (to explain the price patterns) act upon the most suitable ones and replace them with the new ones if they don’t work. Such an asset pricing model using heterogeneous agents was proposed in [8]. The agents continuously adapt their expectations to the market which are created by these multiple expectations using inductive reasoning.

Agent based systems are not known to produce theorems. They produce times-series of the state variables, which are then analysed with econometric methods to understand the model. They are complex adaptive systems, where multiple agents co-exist. Endowed with resources, agents have three useful characteristics [12].

- Adaptive: An agent may have the ability to learn and adapt its behaviour, based on the accumulated experiences. This requires memory.
- Goal-directed: Agents may have certain measurable goals to achieve, which helps itself to compare its performance and adjust accordingly.
- Heterogeneous: Each agent or a subgroup of agents may be given different properties, for example different goals. However, certain simulations like gas particles or molecules may need homogeneous attributes.

A generic agent structure is illustrated in Figure 1. Unlike traditional game theory, where the rules of the game are held constant, agent based systems do not require the agents to have cognitive abilities [14]. Agent based simulation techniques have been successfully applied to a number of complex problems, ranging over multiple disciplines. Applications range
from traffic simulations [13], predicting spread of epidemics
[15], understanding consumer behaviours [16] and modelling
molecular self assembly [17] etc. [18] points out the need to
use agent based techniques to aid in better decision making for
a healthy economy. LeBaron et al. [2] provides an interesting
overview on the ABMs used in finance. There are several
aspects that need to be addressed while developing such a
model for finance. This includes the type of market to be
modelled, the type of securities to be traded and how to
introduce heterogeneity etc. Evolutionary Computing (EC)
techniques have been a key component of most agent based
models in financial markets. Broadly speaking EC techniques
are meta-heuristic optimization algorithms. These techniques
help searching the enormous space of possibilities and finally
converge to the near optimum solution based on certain given
conditions. Depending on the representation of the problems
and the solutions there are different flavours of evolutionary
computing [10]. One of the most commonly used evolutionary
computing techniques is Genetic Algorithms (GA) introduced
by [9]. The defined problem along with a starting point is
fed to the system. It starts by exploring the space and find
the candidate solutions to form a population. The fittest ones
are selected from the population and breed (crossed-over) to
produce children into the next generation. The size of the pop-
ulation remains the same. Hence the less fit ones are discarded
and a new generation is formed. Based on the predefined
probability some of the bits of the candidate solutions are
mutated, to ensure genetic diversity and exploration of other
solution areas. The process terminates if the ideal solution is
reached or it has run for a sufficient number of generations
after which there is no change in the fitness. The Santa Fe
Artificial Stock Market (SF-ASM) has been a promising step
towards the application of EC techniques and ABMs to stock
markets. The main aim of SF-ASM is to understand the
behaviour of an environment of evolving trading behaviour,
where prediction strategies compete against each other [2].

The basic SF-ASM consisted of number of traders, each with
one unit of risky stock (which pays a random dividend) and
few units of cash. During each period the traders had to decide
how to invest or how much to keep to yield a risk free return.

The effective demand of an agent is the difference of his actual
and expected holdings. Once agents determined their effective
demands, they submit them as well as their partial derivatives
with respect to the price to a specialist, who tries to balance
the effective demands by setting a market clearing price. If the
specialist is not able to find a market clearing price in the first place, an iterative process is started in which new trial prices are announced and agents update their effective
demands and partial derivatives accordingly. The learning and
forecasting is implemented using a modification of Holland’s
condition action classifier. [5] establishes that trading strategies
can generate significant returns based on the media moods.

He uses an agent based model and simple genetic algorithms
techniques. In the artificial market, the interaction of the
participating traders (agents) with the market takes place
according to the four different behaviours of the traders/agents.

For example, the behaviours can be encapsulated in the form of
(1) Minority Game Agents, (2) Majority Game Agents, (3)
Delayed Minority Game Agents and (4) Dollar Game Agents
[11], [5]

III. METHODOLOGY

In equilibrium economics, since the participating agents in
the market are rational and homogeneous, all agents have the
same expectations from the available information $I_t$ at time
$t$. So, by using deductive logic, the homogeneous agents’
expectations of the future price $P_{t+1}$, $E[P_{t+1}|I_t]$ can be
deducted rationally. Where as in the case of heterogeneous
agents, there exists differences in agents’ expectations. In that
case, it is not possible to derive market price expectations
using deductive logic even the agents are perfectly rational.
The only possibility is using subjective expectations and use
inductive reasoning in order to come up with a sensible price
expectations. For that, the agents are supposed to choose
the appropriate predicative models for a given specific data
set. That means, in reality, each agent acts as a statistician
testing his/her set of predictive models for the given data set
and choose the best possible future price expectations. For
example, each predictive model can be encapsulated in the
form of a trading strategy. In this research, using ABMs, we
model the returns time-series of a given financial asset for the
market forecast. We consider that the agents in the model are
using binary predictive models (or binary trading strategies).
The input to a predictive model is a binary string of 0s and 1s
and the output is either 0 or 1. In order to facilitate that, we
extract the underlying binary time-series of a given financial
asset’s price time-series. Let $P_t$ be the price of an asset at time
t, then the binary time-series $P_t^b$ is generated as

$$P_{t+1}^b = \begin{cases} 
1 & \text{if } P_t \geq P_{t-1} \\
0 & \text{if } P_t < P_{t-1} 
\end{cases} \quad (1)$$

We find the ABM (artificial market) that represent the real
stock market in the sense that the binary time-series generated
by the ABM closely matches with the directional changes of
the real market price time-series. We populate the artificial
stock market with a set of $N$ heterogeneous agents. Each agent
($i$) has memory $m$ and holds a fixed number ($s$) of trading
strategies $S_{i1}, S_{i2}, \ldots, S_{is}$ that are ranked according to their
performance in the market. In a way, these trading strategies
would represent as price expectation models of the agents.
For the given size of agent’s memory, we consider the possible
trading strategies that a given agent can hold. For example, for
memory size $m = 3$, there are $2^3$ strategies. One can realize
that the number of strategies increases exponentially with the
size of the agent’s memory. Table I depicts a possible trading
strategy that an agent can hold with memory size 3. In this
case, the agent’s memory is 3 and it can only hold past 3 his-
torical data points of the binary time-series. The trading strategy
gives the decision for a given configuration of historical binary

\begin{table}
\begin{tabular}{|c|c|c|}
\hline
\textbf{Agent Memory Size} & \textbf{Trading Strategies} \\
\hline
3 & \{S_{i1}, S_{i2}, S_{i3}\} \\
\hline
\end{tabular}
\end{table}
data points \((0 = \text{sell}, 1 = \text{buy})\). Statistical independence of the strategies can be verified by using Hamming distance. We can see that the number of strategies increases exponentially with the agent’s memory. Two strategies can be considered as statistically independent if the Hamming distance is half way between 0 and \(2^m\). Challet et al [11] confirm that for the given memory size \(m\), there are at most \(2^m\) statistically independent strategies. Using the methodology of [11], we extract the independent set of trading strategies.

<table>
<thead>
<tr>
<th>Pattern</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>000</td>
<td>0</td>
</tr>
<tr>
<td>001</td>
<td>0</td>
</tr>
<tr>
<td>010</td>
<td>1</td>
</tr>
<tr>
<td>011</td>
<td>0</td>
</tr>
<tr>
<td>100</td>
<td>1</td>
</tr>
<tr>
<td>101</td>
<td>0</td>
</tr>
<tr>
<td>110</td>
<td>1</td>
</tr>
<tr>
<td>111</td>
<td>0</td>
</tr>
</tbody>
</table>

**TABLE I**  
A BINARY TRADING STRATEGY FOR \(m = 3\)

Let \(a_i(t)\) be the action \((-1 = \text{sell} or 1 = \text{buy})\) taken by the agent \(i\) according to its best strategy at time \(t\) and \(A(t)\) represents the aggregate actions made by all the agents in the artificial market.

\[
A(t) = \sum_{i=1}^{N} a_i(t)
\]  

We populate the market with a set of heterogeneous agents grouped under four categories. The description of the agents and the pay-off for their actions are given below.

- **Minority Game Agents** - At time \(t\), a minority agent get rewarded \((c_i(t))\) positively if its trading decision \((a_i(t))\) is opposite to that of the actual market, 
  \[
  c_i(t) = -a_i(t)A(t)
  \]

- **Majority Game Agents** - At time \(t\), a majority agent get rewarded \((c_i(t))\) positively if its trading decision \((a_i(t))\) are same as that of the actual market,
  \[
  c_i(t) = a_i(t)A(t)
  \]

- **$\$-Game Agents** - Also known as one-step delayed majority games. The agents anticipate that the market at \(t\) will be the same as their choice at \(t-1\), and get a positive score \((c_i(t))\) if the belief matches the reality,
  \[
  c_i(t) = a_i(t-1)A(t)
  \]

- **Delayed Minority Game Agents** - Also known as one time step delayed minority game [5]. They base their decision of the belief that the market at time \(t\) will be opposite to that \(t-1\), and get rewarded accordingly,
  \[
  c_i(t) = -a_i(t-1)A(t)
  \]

Depending on the memory sizes, each agent is armed with a set of trading strategies. They rank these strategies based on the rewards it gets by virtue of their performance in the virtual market. As we discussed before, there are many possible ways to select the set of artificial agents that could represent the underlying real market. Each set of agents depicts a particular scenario out of many possible scenarios. In order to get the best fit to the real market, we need to search the entire space of scenarios. One can use techniques from Evolutionary Computing (EC) in order to search for a suitable scenario using Genetic Algorithms (GA) that could approximate the real financial market. While using GA, each scenario will be encapsulated into a binary string (or gene) and use the standard GA steps like fitness evaluation, reproduction, crossover and mutation. Figure 2 depicts the GA process of identifying a suitable ABM that represents the underlying market. We perform the search for the most appropriate gene using the parallel GA technique. Island models provided by ECI (an evolutionary computing tool developed at the George Mason University). Island models implements the idea of parallel evolutionary processes distribute over several clients to speed up the search. Multiple islands (sub-populations) are initialised
with different seeds (starting points). The islands act as clients running their own evolutionary process (GA) in tandem and occasionally exchanging individuals via the server. Provided there is an optimum number of islands, it would help in covering all the possibilities in the enormous search area efficiently and avoids from being stuck in a local optimum as well. A system of $u$ islands each with a population of $v$ individuals is not equivalent to a single island with a population of $uv$. The reason for this is that the island system partitions the search. If one island prematurely converges on a sub-optimal solution it does not affect the evolution happening on the other islands; they are following their own paths. A single large population does not have this in-built resilience.

IV. SIMULATION RESULTS

For modelling the Equity markets, we use the FTSE-100 closing prices. Similarly, for the FX market, we use the daily closing mid prices for the three major pairs-EUR/USD, GBP/USD, USD/JPY. We extract the underlying binary time-series from the given price time-series as mentioned before. The binary time-series is used as input to the predictive models of the agents that are populated for a given ABM. The collective returns ($r_t$) at time $t$ from a given ABM are evaluated as

$$r_t = \frac{A_t}{K},$$

where $A_t$ represents aggregate actions made by the agents from the artificial market and $K$ be a proportional constant [4]. Let $r^{m}_t$ be the log returns at time $t$ from the real market,

$$r^{m}_t = \ln(P_t) - \ln(P_{t-1}).$$

The accumulated absolute differences $|r^{m}_t - r_t|$ over the period of calibration is considered as the proxy for fitness criteria in the GA process. Once we identify the best gene (ABM, represents the underlying market) based on the criteria, the future returns of the market is generated by using the best ABM. Once the returns are obtained from the ABM, the future prices of the market can be estimated as explained below.

A. Price Reconstruction

Using the returns time-series $r_{t+i}, r_{t+i-1}, \ldots, r_t$ and price $P_t$ at time $t$, the price $P_{t+i}$ at time $t+i$ can be derived as shown below.

$$r_{t+1} = \ln\left(\frac{P_{t+1}}{P_t}\right)$$

$$\Rightarrow P_{t+1} = P_t \exp(r_{t+1})$$

Now,

$$\Rightarrow \ln\left(\frac{P_{t+2}}{P_t}\right) = \ln\left(\frac{P_{t+2}}{P_{t+1}}\right) + \ln\left(\frac{P_{t+1}}{P_t}\right)$$

$$\Rightarrow \ln\left(\frac{P_{t+2}}{P_t}\right) = r_{t+2} + r_{t+1}$$

$$\Rightarrow P_{t+2} = P_t \exp(r_{t+2} + r_{t+1})$$

Hence, iteratively it can be derived that

$$P_{t+i} = P_t \exp(r_{t+i} + r_{t+i-1} + \ldots + r_{t+1}).$$

We present some initial experimental results here. Figure 3 depicts the performance of Agent Based Model in FX market for two currencies. Historical EUR/USD, USD/JPY, GBP/USD exchange rates for three months are considered for in-sample training for GA that identifies the best ABM for forecasting the market macro-structure for the coming three months. Table III shows the measure the major stylized facts of the ABMs as compared to the original market. Similarly, ABM’s performance in FTSE is depicted in Figure 4. Although the time-series patterns generated by ABMs are similar(deviations, skewness, kurtosis etc.) to the real financial markets, it can be observed that the time-series generated by ABM is not completely matched in terms of magnitude with that of the financial market. This result is not a surprise as the ABM is selected based on the binary time-series (direction of the market, but not the magnitude). For getting better clarity on the performance of the ABMs, we provide the hit-ratio for both the in-sample and out-of-sample. In order to exploit the arbitrage opportunities, we use the price forecasting from the ABM and constructed a naive trading strategy that buys and sells the asset according to the forecast. The accumulation of wealth from $\$1$ investment is displayed in Figure 5.

It is important to understand the internal organization of the economic agents in an ABM. After finding the optimum ABM representing the real market, looking into the internal organization tells us about the market mood and trends. We consider various proportions of minority and majority agents in ABMs. Table IV depicts that the different proportion of the agents in ABM would capture the facts of the real market. If the ABM that contains more number of majority (minority) agents is closest fit to the real market then one can assume that the trend in the market continues (changes).
TABLE II
HI T RATIO TEST

<table>
<thead>
<tr>
<th>Market</th>
<th>In Sample</th>
<th>Out-of-sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>FTSE</td>
<td>77.14 %</td>
<td>65.91 %</td>
</tr>
<tr>
<td>EUR/USD</td>
<td>74.28 %</td>
<td>68.18 %</td>
</tr>
<tr>
<td>GBP/USD</td>
<td>72.23 %</td>
<td>65.61 %</td>
</tr>
<tr>
<td>USD/JPY</td>
<td>77.19 %</td>
<td>69.86 %</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>EURUSD</th>
<th>GBPUSD</th>
<th>USDJPY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.000183</td>
<td>-4.8E-05</td>
<td>6.9758E-06</td>
</tr>
<tr>
<td>Mean(ABM)</td>
<td>0.000203</td>
<td>6.96E-05</td>
<td>-0.00085</td>
</tr>
<tr>
<td>Std Dev</td>
<td>0.00224</td>
<td>0.002179</td>
<td>0.00274</td>
</tr>
<tr>
<td>Std Dev(ABM)</td>
<td>0.00742</td>
<td>0.015012</td>
<td>0.02118</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>2.4442</td>
<td>2.269279</td>
<td>1.3679</td>
</tr>
<tr>
<td>Kurtosis(ABM)</td>
<td>11.3524</td>
<td>18.37732</td>
<td>9.4544</td>
</tr>
<tr>
<td>Skewness</td>
<td>-0.0193</td>
<td>-0.06039</td>
<td>-0.1912</td>
</tr>
<tr>
<td>Skewness(ABM)</td>
<td>-2.1616</td>
<td>-1.85798</td>
<td>-1.3289</td>
</tr>
</tbody>
</table>

TABLE III
SAMPLE STATISTICS FOR THE LOG RETURNS FOR FX

<table>
<thead>
<tr>
<th></th>
<th>EURUSD</th>
<th>GBPUSD</th>
<th>USDJPY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>5.69E-05</td>
<td>-0.00246</td>
<td>-0.0024</td>
</tr>
<tr>
<td>Std Dev</td>
<td>0.009467</td>
<td>0.01037</td>
<td>0.00945</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>1.64179</td>
<td>19.9872</td>
<td>18.7057</td>
</tr>
<tr>
<td>Skewness</td>
<td>-0.15195</td>
<td>-3.6961</td>
<td>-4.1044</td>
</tr>
</tbody>
</table>

TABLE IV
SAMPLE STATISTICS FOR THE LOG RETURNS FOR FTSE(VARYING PROPORTIONS OF AGENTS)

V. CONCLUSION AND FUTURE WORK

We discussed the application of ABMs in forecasting financial markets using more realistic assumptions according to the non-equilibrium nature of the markets. With the initial simulation results we demonstrated the forecasting capabilities of ABMs in financial markets. The validation of the empirical results are done by comparing the output time-series of the ABMs with the output of real financial time-series. The empirical validation is a very crucial methodological concern in agent-based modelling, and that there are many approaches but no one is dominant. It is very important to spell out very clearly how one can handle this problem. For the future work, in order to verify the significance of the results, we consider Universal Information Criterion (UIC) that estimates the explanatory power of any model that is generating simulated or predicted data.

REFERENCES