How to Identify Investor’s types in real financial markets by means of agent based simulation

Filippo Neri
Department of Electrical and Computer Engineering, University of Naples, Naples, ProMarket 11 srl, Milan

ABSTRACT
The paper proposes a computational adaptation of the principles underlying principal component analysis with agent based simulation in order to produce a novel modeling methodology for financial time series and financial markets. Goal of the proposed methodology is to find a reduced set of investor’s models (agents) which is able to approximate or explain a target financial time series. As computational testbed for the study, the learning system L-FABS was chosen which combines simulated annealing with agent based simulation for approximating financial time series. Two experimental case studies showing the efficacy of the proposed methodology are reported.

CCS CONCEPTS
• Computing methodologies: • Artificial intelligence: • Learning paradigms; • Applied computing; • Economics;

KEYWORDS
Simulated annealing, agent based simulation of financial time series, machine learning, adaptation of principal component analysis, S&P500, DJIA

ACM Reference Format:
Filippo Neri. 2021. How to Identify Investor’s types in real financial markets by means of agent based simulation. In 2021 6th International Conference on Machine Learning Technologies (ICMLT 2021), April 23–25, 2021, Jeju Island, Republic of Korea. ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3468891.3468913

1 INTRODUCTION
The paper shows how the principles of Principal Component Analysis [1] can be combined with Agent based Simulation and Evolutionary Computation resulting in a novel methodology for identifying investor’s types in financial markets. The discussed methodology would discover a computational function whose independent variables are models of investor’s behaviors and whose dependent variable is the target market financial time series. The computational function is implemented by means of software agents. The learning system L-FABS [2, 3], which combines simulated annealing with an agent based simulation, will be used as experimental testbed. Usually linear regression is used in Econometrics to explain financial time series. However if the model requires nonlinear interactions or recursive combination of time indexed variables or is procedurally defined, it would be meaningless to test for linear correlation or other linear relationships among the variables. And more sophisticated approaches have to be used like that described in this study.

The rest of the paper is organized as follows: Section 2 comments on how agent based modeling can be used to model financial time series; Section 3 and 4 describe our methodology and how the approximation error between two time series can be measured; Section 5 and 6 report the commented experimental analysis; Section 7 shows a comparison of L-FABS to other learning systems, and Section 8 draws our conclusions.

2 AGENT BASED MODELING AS A COMPUTATION TOOL FOR EVALUATING MODELS OF INVESTORS
State-of-the-art literature shows that agent based modeling is a flexible modeling methodology for simulating several types of domains [4–8] including consumer markets [9], economies [10] or societies [11] and financial time series [3, 12, 13]. Moreover examples of how evolutionary computation [14, 15] and agent based modeling can be used to deal with economic tasks can be found in [16, 17]. The cited papers have been selected with the only intent to provide examples of the listed approaches and without any claim to be an exhaustive list of previous work on the topic. In this study, a financial market is modeled by a set of agents reproducing the behavior of the investors. This study will use a specific machine learning system: the Learning Financial Agent Based Simulator L-FABS [2] as experimental testbed. Different research aspects of L-FABS have been studied in [3, 18–22]. For sake of completeness, one can note that alternative approaches to model investors’ or customers’ behavior in a variety markets or trading situations also exists as discussed for instance in [23, 24].

3 APPLYING THE PCA PRINCIPLES TO FIND A REDUCED SET OF AGENT BASED MODELS OF INVESTORS
The approach here described can be used to discover simpler explanatory/predictive models of a target financial time series in terms of a computational combination of a set of investors’ behaviors. As computational tool to define and manage different models of investors, it is used the system L-FABS. L-FBAS is essentially an agent based simulator combined with a machine learning algorithm. The machine learning algorithm is used to discover good simulation’s parameters so that the simulated time series can closely approximate the target one. If this were the case, the learned computation
model, which would include several agent based models, could then be thought as a simplified computational representation of the financial market that has generated the target financial time series.

By applying the PCA principles to the investors/agent models learned by L-FABS then a reduced set of investors’ models may be found, with respect to the input one, while still maintaining an high explanatory power for the target time series. Here is how the PCA principles are adapted and employed in our methodology:

a) start with an hypothetical and possibly redundant set Orig of investor’s behavior models and run Orig in L-FABS. The output agent based model produced by Orig and its approximation error will act as the benchmark model and error.

b) measure the explanatory power of each of the models in Orig
c) add to an initially empty set Reduced the individual model with the highest explanatory power in Orig, that is with the lowest approximation error. Remove the selected model from Orig.
d) repeat step c) until the approximation error of the set Reduced, when run in L-FABS, is better than or close enough to the approximation error of model Orig as built in step a).

The informed reader would have recognized in the above algorithm a classic hill climbing procedure as described in any artificial intelligence textbook. Of course the author is aware of the limitations of getting stuck on local maxima when using a simple hill climbing method but the selection of the best optimization method is not the focus of the research discussed here. Here the intention is to make the point and empirically showing that PCA principles can be successfully combined with agent based simulation to discover sets of investors’ models. It is left to a future work the investigation of if and how using a better optimization function could improve the composition of the discovered reduced set of investors’ models. This future investigation would also open the door to exploring connections with meta-learning studies that have appeared in the machine learning community, just as an example reference [25].

4 MEASURING THE EXPLANATORY POWER OF MODELS

In order to measure the explanatory power of a set of investors’ behaviors, first they will be codified into a set of agents that can be run into L-FABS, then L-FABS will be run with the objective to approximate a target financial time series and, finally, it is measured the Mean Absolute Percentage Error (MAPE) of the predicted time series with respect to the target one [3]. The Mean Absolute Percentage Error or MAPE function has been chosen to measure the approximation error between two time series because it is commonly used in Statistics when two data samplings have to be compared. MAPE is defined as:

\[
MAPE(X, Y) = \frac{1}{N} \sum_{i=1}^{N} \left| \frac{x_i - y_i}{x_i} \right|
\]

Given two time series X and Y, the lower the values for MAPE, the closer the two are. Thus the lower the MAPE value, the highest the explanatory power of the agent based model (set of investor’s behaviors) run in L-FABS.

5 EXPERIMENTAL ANALYSIS

The financial time series selected for our experiments consists of a period of the SP500 index from 3 Jan 2008 to 20 Aug 2010. As usual when working with learning systems, L-FABS is trained on a part of the dataset, the learning set, and then the remaining part of the dataset will be used as test set to assess the performances of the learned model. The original data period is then divided in a learning set: SP500 data from 3 Jan 2008 to 31 Dec 2008 and a test set: SP500 from 2 Jan 2009 to 20 Aug 2010.

Also, the interested reader, may note that L-FABS is configured in the partial knowledge (PK) operating modality. In the PK modality, only the starting point of the time series, \( t=0 \), is given to L-FABS in order to initialize the simulation. In this configuration, the time series model in L-FABS will move from one predicted value for the time series to the next without knowing/using the correct value of the time series at time \( t-1 \) in order to estimate a value for time \( t \).

The PK scenario is selected because the predicted time series will not make use of any other information apart from the value of the target time series at time 0 and the information coded and expressed by the set of models of investors’ behaviors. In the experimental settings two different sets of agent based models of investor’s behavior are explored denoted as Configuration A (Table 1 ) and Configuration B (Table 2 ). To keep things simple, four types of investors (Financial Agents) are chosen to capture the variety of investment decisions and the variety of size of financial transactions that occur in real financial markets. According to each investor’s type, many agents (Financial Agents) are then created in the simulation with similar but not identical behavior. The four types of investors can be thought as: individual investors (and the likes), banks (and the likes), hedge funds (and the likes), and central banks\(^1\) (and the likes). They differ in term of the size of the assets they can invest in financial markets and for their risk/reward appetite. In addition, their numerical presence is also different. As already said two different configurations of investor types (or two different set of agent based models) will be studied and they are identified as Configuration A and Configuration B.

Case Configuration A

Let start by observing the performances of L-FABS when all the investors’ models are used: as it can be seen in fig.1 a, L-FABS is able to output a predicted time series that is very close to the real one. The corresponding MAPE errors for all configurations are reported in Table 3. Values are averaged over 10 runs. To keep the table readable, the study report the standard deviation only for the lowest and closest values of MAPE.

If all the models for the investors are disabled, the output of L-FABS become a constant value equal to the value of the target time series at time 0 as expected and causing a very high MAPE error as reported in fig.1 b.

Let us consider what happens when only one type of investors can act in approximating the financial time series, figg. 1c, 1d, 1e, and 1f. As it can be seen from the graphs, each type of investor has

\(^1\)A little thought objection to our choice to include Central Banks among the factors influencing financial markets has to be easily and strongly rejected considering how in recent decades, Central Banks have acted to 'pump up' financial markets by adding large quantities of liquidity to the related faltering real economies. Thus Central Banks, who have always acted in the background, have finally lost their image of neutral agents with respect to the financial systems.
an explanatory power ranging from high to low when it came to model the target time series. If one also looks at the MAPE values, one can observe that just by using the Institutional Investors/Banks type of investors’ model one can achieve optimal approximation of the target time series. Thus all the other investors’ types are redundant in this case. Applying the adapted principal component methodology, described in Section 3, to model selection in this case is then trivial: in fact it will reduce the original set of investor’s types by selecting the model associated with Institutional Investors/Banks.

**Case Configuration B**

Let us then observe the results obtained when L-FABS is run in the Configuration B case. Again starting with all the investors’ models, in fig.2 a, L-FABS is able to output a predicted time series that is very close to the target one. The corresponding MAPE errors for all configurations are reported in Table 4. When only one type of investors is used, for example see fig.2 b and 2c, each type of investors displays its own explanatory power ranging from high to low when it came to model the target time series. The behavior of using only the models for Retail/Private Investors or Hedge Funds is the same as observed in Case Configuration A and so not reported given their limited explanatory power. Instead we report in fig.2 what will happen if both of them are used to predict the target time series.

The combined approximation error is better than their separate ones but still very far from the benchmark MAPE value of the original set of models. Also in Configuration B, and this is a significant difference with respect to Configuration A, none of the models for the various investors’ types when taken individually is able to approximate very well the target time series. By applying the adapted principal component methodology to this case, then one can start by selecting the model for Govt. type of investors plus the model for Institutional Investors/Banks as first candidates for a reduced set of models of types of investors’ behaviors. One can thereafter run L-FABS with the two models active, and will obtain the graph in fig.2 and corresponding MAPE value shown in...
Table 2: Configuration B of Investor types

| Investor type | Total Assets per i.t.(in millions) | Number of investors |
|---------------|-----------------------------------|--------------------|
| Individual    | 0.1                               | 150                |
| Funds         | 100                               | 100                |
| Banks         | 1000                              | 245                |
| Govt/C.B.     | 100000                            | 5                  |

Figure 2: Actual and predicted time series by L-FABS under the experimental settings Configuration B.

Table 3: MAPE values for Case Configuration A.

| case | MAPE       |
|------|------------|
| a    | 3.32 ± 0.04% |
| b    | 13.14 ± 0.00% |
| c    | 9.48 ± 0.00%   |
| d    | 13.05 ± 0.00% |
| e    | 11.29 ± 0.00% |
| f    | 3.15 ± 0.02%   |

Table 4: MAPE values for Case Configuration B.

| case | MAPE       |
|------|------------|
| a    | 3.31 ± 0.06%   |
| b    | 3.92 ± 0.20%   |
| c    | 7.14 ± 0.00%   |
| d    | 9.70 ± 0.00%   |
| e    | 3.38 ± 0.05%   |

6 COMMENTARY ON THE EXPERIMENTS

Note that reported set of experiments have only explored how the change in the balance of available assets among the types of investors, please see in Table 5, can alter the composition of the reduced set of investors’ models.
Table 5: Types of Investors

| Investor type       | Total Assets in Configuration A (in millions) | Total Assets in Configuration B (in millions) |
|---------------------|----------------------------------------------|----------------------------------------------|
| Individual          | 15                                           | 15                                           |
| Funds               | 10000                                        | 10000                                        |
| Banks               | 245000                                       | 245000                                       |
| Govt/Central Banks  | 50000                                        | 50000                                        |

The Total Assets column shows that in Configuration A case, a type of investors has the majority of available assets to invest, whereas in Configuration B case the balance of available assets among investors have been changed and now there are two types of investors with similar investment capacity.

However, the adapted principal component methodology described here is not limited to the evaluation of only quantitative changes in a model's parameters. In fact, one could use the same methodology of model simplification to explore algorithmic differences in the decision-making process of the investors. Thus, one could explore how different ways to participate in the market could result in more complex or simpler models of the time series. This point will be object of future research.

7 EXPERIMENTAL COMPARISON OF L-FABS TO OTHER SYSTEMS

The comparison of L-FABS to other learning systems is beyond the scope of the paper, though a brief comment about its performances with respect to other learning algorithms is shown. Results are taken from [3] and please refer to the cited paper for more information on the empirical evaluation of L-FABS on several other time series. Table 5 compares L-FABS, a Particle Swarm Optimization algorithm (PSO) [26], and a Multi-Layer Perceptron (MLP) [27] when operating on time series from the SP500 and DJIA respectively. These results have been reported with the aim to locate L-FABS as better than those obtained by MLP and are comparable to those obtained by PSO. The figures for PSO and MLP are as reported by the authors with no confidence intervals given.

Table 6: Experimental results, averaged over 10 runs, for Dataset DJIA and Dataset SP500

| Dataset | Day to predict | PSO MAPE % | MLP MAPE % | L-FABS MAPE % |
|---------|----------------|------------|------------|---------------|
| SP500   | 1              | 0.66       | 1          | 0.714 ± 0.009 |
| SP500   | 7              | 1.47       | 3.11       | 1.424 ± 0.015 |
| DJIA    | 1              | 0.65       | 1.06       | 0.709 ± 0.007 |
| DJIA    | 7              | 1.47       | 5.64       | 1.443 ± 0.011 |

8 CONCLUSION

The study proposes a computational adaptation of the principles underlying Principal Component Analysis and implemented them in the agent-based simulator L-FABS. The paper has also pointed out how easily and agent-based simulation could allow for a parallel implementation and thus scale when markets made by a large number of investors are to be studied. The methodology proposed can be applied to the task of finding a reduced set of models of investor’s behaviors used to approximate a target financial time series and its generative market. The reported methodology can thus be used to evaluate the explanatory power of a set of investor’s models with respect to a given market. Thus, allowing also to artificially reproduce the behavior of the target market, in terms of its generated financial time series, by using L-FABS. The authors believe that the use of simple behavioral models, such as those found by L-FABS, can allow for a better understanding of the underlying and usually hidden mechanism that result in macro behavior like those captured by a market index. Two case studies have been employed to show the efficacy of the proposed methodology in two instances of real financial time series. As a future research, the authors plan to explore how to use the agent-based modeling to find simplest explanatory models in new domains such as when tackling control problems [28] or concept drift adaptation tasks [29].

REFERENCES

[1] I. Jolliffe, Principal Component Analysis. Springer Verlag, 1986.
[2] F. Neri, “Agent-based modeling under partial and full knowledge learning settings to simulate financial markets,” AI Communications, vol. 25, no. 4, pp. 295–304, 2012.
[3] F. Neri, “A comparative study of a financial agent based simulator across learning scenarios,” Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 7103, pp. 86–97, 2012.
[4] R. Axelrod, “Agent-based modeling as a bridge between disciplines,” in Handbook of Computational Economics, Vol. 2: Agent-Based Computational Economics, Handbooks in Economics Series (K. L. Judd and L. Tesfatsion, eds.), pp. 1562–1583, Amsterdam, The Nederlands, North-Holland, 2006.
[5] C. Bruun, “The economy as an agent-based whole simulating schumpeterian dynamics,” Industry and Innovation, vol. 10, 2003.
[6] L. Tesfatsion, “Agent-based computational economics: Growing economies from the bottom up,” Artif. Life, vol. 8, no. 1, pp. 55–82, 2002.
[7] S. Chen, “Agent-based artificial markets in the ai-econ research center: A retrospect from 1995 to the present,” Systems, Control and Information, vol. 46, no. 9, pp. 23–30, 2002.
[8] I. Garcia-Margarith, I. Plaza, and F. Neri, “Abs-mindburnout: An agent-based simulator of the effects of mindfulness-based interventions on job burnout,” Journal of Computational Science, vol. 36, 2019.
[9] S. Schulenburg, P. Ross, and S. Bridge, “An adaptive agent-based economic model,” in Learning Classifier Systems: From Foundations to Applications, volume 1813 of Lecture Notes in Artificial Intelligence, pp. 265–284, Springer-Verlag, 2000.
[10] M. A. H. Dempster, T. W. Payne, Y. Romahi, and G. W. P. Thompson, “Computational learning techniques for intraday fx trading using popular technical indicators,” IEEE Transactions on Neural Networks, vol. 12, no. 4, pp. 744–754, 2001.
[11] J. M. Epstein and R. Axtell, Growing artificial societies: social science from the bottom up. Washington, DC, USA: The Brookings Institution, 1996.
[12] W. B. Arthur, J. H. Holland, B. LeBaron, R. Palmer, and P. Taylor, “Asset pricing under endogenous expectation in an artificial stock market,” in The Economy as an Evolving Complex System II, pp. 15–44, Santa Fe Institute Studies in the Sciences of Complexity Lecture Notes, 1997.
[13] A. O. I. Hoffmann, S. A. Delce, J. H. von Eije, and W. Jager, “Artificial multi-agent stock markets: Simple strategies, complex outcomes,” in Advances in Artificial Economics, volume 584 of Lecture Notes in Economics and Mathematical Systems, pp. 167–176, Springer-Verlag, 2006.
[14] M. Camilleri and F. Neri, “Parameter optimization in decision tree learning by using simple genetic algorithms,” WSEAS Transactions on Computers, vol. 13, pp. 582–591, 2014.
[15] M. Camilleri, and F. Neri, and M. Papoutsidakis, “An algorithmic approach to parameter selection in machine learning using meta-optimization techniques,” WSEAS Transactions on Systems, vol. 13, no. 1, pp. 203–212, 2014.
[16] B. Alexandrova-Kabadjova, A. Krause, and E. Tsang, “An agent-based model of interactions in the payment card market,” in Intelligent Data Engineering and Automated Learning-IDEAL 2007, pp. 1063–1072, Springer, 2007.
[17] S. G. Rodríguez, D. Quintana, I. M. Galván, and P. Isasi, “Portfolio optimization using spea2 with resampling,” in in Intelligent data engineering and automated learning - IDEAL 2011: 12th International Conference, LNCS 6936, pp. 127–134, Springer, 2011.
[18] F. Neri and I. Margaríto, “Simulating and modeling the dax index and the uso etf financial time series by using a simple agent-based learning architecture,” Expert Systems, vol. 37, no. 4, 2020.
[19] F. Neri, "Combining machine learning and agent based modeling for gold price prediction," Artificial life and Evolutionary Computation, vol. Commu- nications in Computer and Information Science 900, pp. 91–100, 2019.
[20] F. Neri, "Case study on modeling the silver and nasdaq financial time series with simulated annealing," Advances in Intelligent Systems and Computing, vol. 746, pp. 755–763, 2018.
[21] F. Neri, "Learning predictive models for financial time series by using agent based simulations," Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 7198, pp. 202–221, 2012.
[22] F. Neri, "Learning and predicting financial time series by combining natural computation and agent simulation," Lecture Notes in Computer Science, vol. 6625, pp. 111–119, 2011.
[23] B. Lariviere and D. Van den Poel, "Banking behaviour after the lifecycle event of moving in together: An exploratory study of the role of marketing investments," European Journal of Operational Research, vol. 183, no. 1, pp. 345–369, 2007.
[24] F. Neri, "Open research issues on computational techniques for financial applications," WSEAS Transactions on Systems, vol. 13, no. 1, pp. 390–391, 2014.
[25] M. Reif, F. Shafait, and A. Dengel, "Meta-learning for evolutionary parameter optimization of classifiers," Machine Learning, vol. 87, no. 3, pp. 357—380, 2012.
[26] J. Kennedy and R. Eberhard, "Particle swarm optimization,” in Int. Conf. on Neural Networks, pp. 1942–1948, IEEE press, 1995.
[27] Zirilli, Financial prediction using Neural Networks. International Thompson Computer Press, 1997.
[28] A. Marino and F. Neri, "Pid tuning with neural networks," Intelligent Information and Database Systems, vol. LNCS 11431, pp. 476–487, 2019.
[29] F. Neri. "Domain Specific Concept Drift Detectors for Predicting Financial Time Series”. Submitted. Draft copy at https://arxiv.org/abs/2103.14079, 2021.