VOL. 3, NO. 9

OMI Research Newsletter – July 2024

The Oxford-Man Institute of Quantitative Finance (OMI) is a world-leading centre for interdisciplinary research in financial markets. The OMI addresses fundamental problems in quantitative finance with a strong focus on machine learning and data driven models.

*The **OMI Research Newsletter** is a curated summary that covers the most recent research, insights, and tools related to financial decision making.*

Highlights

Research focus. Key contributions in this month's edition include works from the OMI, École Polytechnique, University of Tokyo, and Imperial College London. Gonon, Jacquier and Wiedemann (2024) propose a method to smooth implied volatility surfaces with neural operators. Elomari-Kessab et al. (2024) study the joint micro-dynamics of asset prices and order flow in limit order books. Waldon et al. (2024) uses physics-informed neural networks to approximate the optimal policy of an agent in unknown and non-stationary environments. Finally, Kitadai et al. (2024) study the suitability of large language models as a generative model for agents in economic experiments.

Machine Learning & Statistics

Microstructure Modes – Disentangling the Joint Dynamics of Prices & Order Flow

S. ELOMARI-KESSAB, G. MAITRIER, J. BONART, J-P. BOUCHAUD

ÉCOLE POLYTECHNIQUE

Understanding the micro-dynamics of asset prices in modern electronic order books is crucial for investors and regulators. In this paper, we use an order by order Eurostoxx database spanning over 3 years to analyze the joint dynamics of prices and order flow. In order to alleviate various problems caused by high-frequency noise, we propose a double coarse-graining procedure that allows us to extract meaningful information at the minute time scale. We use Principal Component Analysis to construct "microstructure modes" that describe the most common flow/return patterns and allow one to separate them into bid-ask symmetric and bid-ask anti-symmetric. We define and calibrate a Vector Auto-Regressive (VAR) model that encodes the dynamical evolution of these modes. The parameters of the VAR model are found to be extremely stable in time, and lead to relatively high R^2 prediction scores, especially for symmetric liquidity modes. The VAR model becomes marginally unstable as more lags are included, reflecting the long-memory nature of flows and giving some further credence to the possibility of "endogenous liquidity crises". Although very satisfactory on several counts, we show that our VAR framework does not account for the well known square-root law of price impact.

Your Transformer is Secretly Linear

A. RAZZHIGAEV, M. MIKHALCHUK, E. GONCHAROVA, N. GERASIMENKO, I. OSELEDETS, D. DIMITROV, A. KUZNETSOV

HSE UNIVERSITY, LOMONOSOV MOSCOW STATE UNIVERSITY

This paper reveals a novel linear characteristic exclusive to transformer decoders, including models such as GPT, LLaMA, OPT, BLOOM and others. We analyze embedding transformations between sequential layers, uncovering a near-perfect linear relationship (Procrustes similarity score of 0.99). However, linearity decreases when the residual component is removed due to a consistently low output norm of the transformer layer. Our experiments show that removing or linearly approximating some of the most linear blocks of transformers does not affect significantly the loss or model performance. Moreover, in our pretraining experiments on smaller models we introduce a cosine-similarity-based regularization, aimed at reducing layer linearity. This regularization improves performance metrics on benchmarks like Tiny Stories

and SuperGLUE and as well successfully decreases the linearity of the models. This study challenges the existing understanding of transformer architectures, suggesting that their operation may be more linear than previously assumed.

DARE: The Deep Adaptive Regulator for Control of Uncertain Continuous-Time Systems

H. Waldon, F. Drissi, Y. Limmer, U. Berdica, J. N. Foerster, Á. Cartea

Oxford-Man Institute, University of Oxford

A fundamental challenge in continuous-time optimal control (OC) is the efficient computation of adaptive policies when agents act in unknown, uncertain environments. Traditional OC methods, such as dynamic programming, face challenges in scalability and adaptability due to the curse-of-dimensionality and the reliance on fixed models of the environment. One approach to address these issues is Model Predictive Control (MPC), which iteratively computes open-loop controls over a receding horizon. However, classical MPC algorithms typically also assume a fixed environment. Another approach is Reinforcement Learning (RL) which scales well to high-dimensional setups but is often sample inefficient. Certain RL methods can also be unreliable in highly stochastic continuous-time setups and may be unable to generalize to unseen environments. This paper presents the Deep Adaptive Regulator (DARE) which uses physics-informed neural network based approximations to the agent's value function and policy which are trained online to adapt to unknown environments. To manage uncertainty of the environment, DARE optimizes an augmented reward objective which dynamically trades off exploration with exploitation. We show that our method effectively adapts to unseen environments in settings where "classical" RL fails and is suited for online adaptive decision-making in environments that change in real time.

Universal randomised signatures for generative time series modelling

F. Biagini, L. Gonon, N. Walter

University Munich, Imperial College London

Randomised signature has been proposed as a flexible and easily implementable alternative to the well-established path signature. In this article, we employ randomised signature to introduce a generative model for financial time series data in the spirit of reservoir computing. Specifically, we propose a novel Wasserstein-type distance based on discrete-time randomised signatures. This metric on the space of probability measures captures the distance between (conditional) distributions. Its use is justified by our novel universal approximation results for randomised signatures on the space of continuous functions taking the underlying path as an input. We then use our metric as the loss function in a non-adversarial generator model for synthetic time series data based on a reservoir neural stochastic differential equation. We compare the results of our model to benchmarks from the existing literature.

Randomization Inference: Theory and Applications

D. M. Ritzwoller, J. P. Romano, A. M. Shaikh

Stanford University, University of Chicago

We review approaches to statistical inference based on randomization. Permutation tests are treated as an important special case. Under a certain group invariance property, referred to as the "randomization hypothesis," randomization tests achieve exact control of the Type I error rate in finite samples. Although this unequivocal precision is very appealing, the range of problems that satisfy the randomization hypothesis is somewhat limited. We show that randomization tests are often asymptotically, or approximately, valid and efficient in settings that deviate from the conditions required for finite-sample error control. When randomization tests fail to offer even asymptotic Type 1 error control, their asymptotic validity may be restored by constructing an asymptotically pivotal test statistic. Randomization tests can then provide exact error control for tests of highly structured hypotheses with good performance in a wider class of problems. We give a detailed overview of several prominent applications of randomization tests, including two-sample permutation tests, regression, and conformal inference.

Learning conditional distributions on continuous spaces

C. Bénézet, Z. Cheng, S. Jaimungal

Universite d'Evry-Val-d'Essonne, University of Toronto

We investigate sample-based learning of conditional distributions on multi-dimensional unit boxes, allowing for different dimensions of the feature and target spaces. Our approach involves clustering data near varying query points in the feature space to create empirical measures in the target space. We employ two distinct clustering schemes: one based on a fixed-radius ball and the other on nearest neighbors. We establish upper bounds for the convergence rates of both methods and, from these bounds, deduce optimal configurations for the radius and the number of neighbors. We propose to incorporate the nearest neighbors method into neural network training, as our empirical analysis indicates it has better performance in practice. For efficiency, our training process utilizes approximate nearest neighbors search with random binary space partitioning. Additionally, we employ the Sinkhorn algorithm and a sparsity-enforced

transport plan. Our empirical findings demonstrate that, with a suitably designed structure, the neural network has the ability to adapt to a suitable level of Lipschitz continuity locally. For reproducibility, our code is available at https://github.com/zcheng-a/LCD_kNN.

Operator Deep Smoothing for Implied Volatility

L. Gonon, A. Jacquier, R. Wiedemann
Imperial College London

We devise a novel method for implied volatility smoothing based on neural operators. The goal of implied volatility smoothing is to construct a smooth surface that links the collection of prices observed at a specific instant on a given option market. Such price data arises highly dynamically in ever-changing spatial configurations, which poses a major limitation to foundational machine learning approaches using classical neural networks. While large models in language and image processing deliver breakthrough results on vast corpora of raw data, in financial engineering the generalization from big historical datasets has been hindered by the need for considerable data pre-processing. In particular, implied volatility smoothing has remained an instance-by-instance, hands-on process both for neural network-based and traditional parametric strategies. Our general operator deep smoothing approach, instead, directly maps observed data to smoothed surfaces. We adapt the graph neural operator architecture to do so with high accuracy on ten years of raw intraday S&P 500 options data, using a single set of weights. The trained operator adheres to critical no-arbitrage constraints and is robust with respect to subsampling of inputs (occurring in practice in the context of outlier removal). We provide extensive historical benchmarks and showcase the generalization capability of our approach in a comparison with SVI, an industry standard parametrization for implied volatility. The operator deep smoothing approach thus opens up the use of neural networks on large historical datasets in financial engineering.

How Neural Networks Learn the Support is an Implicit Regularization Effect of SGD

P. Beneventano, A. Pinto, T. Poggio
Princeton University, Massachusetts Institute of Technology

We investigate the ability of deep neural networks to identify the support of the target function. Our findings reveal that mini-batch SGD effectively learns the support in the first layer of the network by shrinking to zero the weights associated with irrelevant components of input. In contrast, we demonstrate that while vanilla GD also approximates the target function, it requires an explicit regularization term to learn the support in the first layer. We prove that this property of mini-batch SGD is due to a second-order implicit regularization effect which is proportional to η/b (step size / batch size). Our results are not only another proof that implicit regularization has a significant impact on training optimization dynamics but they also shed light on the structure of the features that are learned by the network. Additionally, they suggest that smaller batches enhance feature interpretability and reduce dependency on initialization.

Finance and Economics

Common Identification and Common Learning

M. W. Cripps
University College London

Cripps, Ely, Mailath and Samuelson (2008) showed that if there are finitely many states, and the signals are i.i.d and finite, then individual learning is sufficient for common learning. In this note we describe what is commonly learned when this sufficient condition does not hold.

Can AI with High Reasoning Ability Replicate Human-like Decision Making in Economic Experiments?

A. Kitadai, S. D. R. Lugo, Y. Tsurusaki, Y. Fukasawa, N. Nishino
The University of Tokyo

Economic experiments offer a controlled setting for researchers to observe human decision-making and test diverse theories and hypotheses; however, substantial costs and efforts are incurred to gather many individuals as experimental participants. To address this, with the development of large language models (LLMs), some researchers have recently attempted to develop simulated economic experiments using LLMs-driven agents, called generative agents. If generative agents can replicate human-like decision-making in economic experiments, the cost problem of economic experiments can be alleviated. However, such a simulation framework has not been yet established. Considering the previous research and the current evolutionary stage of LLMs, this study focuses on the reasoning ability of generative agents as a key factor toward establishing a framework for such a new methodology. A multi-agent simulation, designed to improve the reasoning ability of generative agents through prompting methods, was developed to reproduce the result of an actual economic experiment on the ultimatum game. The results demonstrated that the higher the reasoning ability of the agents, the closer the results were to the theoretical solution than to the real experimental result. The results also suggest that setting the personas of the generative agents may be important

REFERENCES

for reproducing the results of real economic experiments. These findings are valuable for the future definition of a framework for replacing human participants with generative agents in economic experiments when LLMs are further developed.

Mathematical Finance

Stochastic Liquidity as a Proxy for Concave Multi-Asset Propagator Models

J. Muhle-Karbe, C. Tracy

Imperial College London

We study multi-asset propagator models that describe how the self and cross impact of trades on prices decay over time. As for a single asset, the model fit can be improved significantly using concave rather than linear impact functions. Such models are not analytically tractable, but can be approximated effectively by linear models in which stochasticity of the impact level acts as a reduced-form proxy for concavity. Risk-neutral optimal trading problems can be solved in closed form for the reduced-form model, leading to sharp characterizations for the absence of price manipulation and emphasizing that trading uncorrelated assets is not only beneficial from a risk but also from a transaction cost perspective. The relevance of cross impact and the effectiveness of the approximation is illustrated using LOBSTER price and trade data.

Macroscopic Market Making Games

I. Guo, S. Jin, K. Nam

Monash University

In continuation of the macroscopic market making à la Avellaneda-Stoikov as a control problem, this paper explores its stochastic game. Concerning the price competition, each agent is compared with the best quote from the others. We start with the linear case. While constructing the solution directly, the ordering property and the dimension reduction in the equilibrium are revealed. For the non-linear case, extending the decoupling approach, we introduce a multidimensional characteristic equation to study the well-posedness of forward-backward stochastic differential equations. Properties of coefficients in the characteristic equation are obtained via non-smooth analysis. In addition to novel well-posedness results, the linear price impact arises and the impact function can be further decomposed into two parts in some examples.

Acknowledgement

The OMI Research Newsletter is curated by Álvaro Cartea, Director of the Oxford-Man Institute, and Fayçal Drissi, postdoctoral researcher at the institute. Paper suggestions are from fellow members at the institute: Álvaro Arroyo, Patrick Chang, Gerardo Durán-Martín, Marcello Monga, and Harrison Waldon.

References

Beneventano, Pierfrancesco, Andrea Pinto, and Tomaso Poggio. 2024. "How Neural Networks Learn the Support is an Implicit Regularization Effect of SGD." *arXiv preprint arXiv:2406.11110*.

Bénézet, Cyril, Ziteng Cheng, and Sebastian Jaimungal. 2024. "Learning conditional distributions on continuous spaces." *arXiv preprint arXiv:2406.09375*.

Biagini, Francesca, Lukas Gonon, and Niklas Walter. 2024. "Universal randomised signatures for generative time series modelling." *arXiv preprint arXiv:2406.10214*.

Cripps, Martin W. 2024. "Common Identification and Common Learning." *arXiv preprint arXiv:2406.20029*.

Elomari-Kessab, Salma, Guillaume Maitrier, Julius Bonart, and Jean-Philippe Bouchaud. 2024. ""Microstructure Modes"–Disentangling the Joint Dynamics of Prices & Order Flow." *arXiv preprint arXiv:2405.10654*.

Gonon, Lukas, Antoine Jacquier, and Ruben Wiedemann. 2024. "Operator Deep Smoothing for Implied Volatility." *arXiv preprint arXiv:2406.11520*.

Guo, Ivan, Shijia Jin, and Kihun Nam. 2024. "Macroscopic Market Making Games." *arXiv preprint arXiv:2406.05662*.

Kitadai, Ayato, Sinndy Dayana Rico Lugo, Yudai Tsurusaki, Yusuke Fukasawa, and Nariaki Nishino. 2024. "Can AI with High Reasoning Ability Replicate Human-like Decision Making in Economic Experiments?" *arXiv preprint arXiv:2406.11426*.

Muhle-Karbe, Johannes, and Connor Tracy. 2024. "Stochastic Liquidity as a Proxy for Concave Multi-Asset Propagator Models." *Available at SSRN 4885733*.

Razzhigaev, Anton, Matvey Mikhalchuk, Elizaveta Goncharova, Nikolai Gerasimenko, Ivan Oseledets, Denis Dimitrov, and Andrey Kuznetsov. 2024. "Your Transformer is Secretly Linear." *arXiv preprint arXiv:2405.12250*.

Ritzwoller, David M, Joseph P Romano, and Azeem M Shaikh. 2024. "Randomization Inference: Theory and Applications." *arXiv preprint arXiv:2406.09521*.

Waldon, Harrison, Fayçal Drissi, Yannick Limmer, Uljad Berdica, Jakob Nicolaus Foerster, and Alvaro Cartea. 2024. "DARE: The Deep Adaptive Regulator for Control of Uncertain Continuous-Time Systems." *ICML 2024 Workshop: Foundations of Reinforcement Learning and Control–Connections and Perspectives*.