VOL. 2, NO. 8

OMI Research Newsletter – May 2023

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

Noteworthy contributions in this month's edition include works from the OMI, University of Oxford, New York University, University of British Columbia, and University College London. In this work, the authors use Generalised Regression Neural Networks for online multivariate time series forecasting. Here, the authors extend the Neural Network based PDE solvers to prevent ill-conditioning from complex PDEs. Finally, the authors of this work define an equilibrium-inspired distance metric for the space of all normal-form games and uncover a distance-preserving equilibrium-invariant embedding.

Machine Learning & Statistics

A Stable and Scalable Method For Solving Initial Value PDEs With Neural Networks

M. FINZI, A. POTAPCZYNSKI, M. CHOPTUIK, AND A. G. WILSON

NEW YORK UNIVERSITY, UNIVERSITY OF BRITISH COLUMBIA

Unlike conventional grid and mesh based methods for solving partial differential equations (PDEs), neural networks have the potential to break the curse of dimensionality, providing approximate solutions to problems where using classical solvers is difficult or impossible. While global minimization of the PDE residual over the network parameters works well for boundary value problems, catastrophic forgetting impairs the applicability of this approach to initial value problems (IVPs). In an alternative local-in-time approach, the optimization problem can be converted into an ODE on the network parameters and the solution propagated forward in time; however, we demonstrate that current methods based on this approach suffer from two key issues. First, following the ODE produces an uncontrolled growth in the conditioning of the problem, ultimately leading to unacceptably large numerical errors. Second, as the ODE methods scale cubically with the number of model parameters, they are restricted to small neural networks, significantly limiting their ability to represent intricate PDE initial conditions and solutions. Building on these insights, we develop Neural IVP, an ODE based IVP solver which prevents the network from getting ill-conditioned and runs in time linear in the number of parameters, enabling us to evolve the dynamics of challenging PDEs with neural networks.

OFTER: an Online Pipeline for Time Series Forecasting

N. MICHAEL, M. CUCURINGU, AND S. HOWISON

OXFORD-MAN INSTITUTE, UNIVERSITY OF OXFORD

We introduce OFTER, a time series forecasting pipeline tailored for mid-sized multivariate time series. OFTER utilizes the non-parametric models of k-nearest neighbors and Generalized Regression Neural Networks, integrated with a dimensionality reduction component. To circumvent the curse of dimensionality, we employ a weighted norm based on a modified version of the maximal correlation coefficient. The pipeline we introduce is specifically designed for online tasks, has an interpretable output, and is able to outperform several state-of-the art baselines. The computational efficacy of the algorithm, its online nature, and its ability to operate in low signal-to-noise regimes, render OFTER an ideal approach for financial multivariate time series problems, such as daily equity forecasting. Our work demonstrates that while deep learning models hold significant promise for time series forecasting, traditional methods carefully integrating mainstream tools remain very competitive alternatives with the added benefits of scalability and interpretability.

Variational Bayes Made Easy

M. E. Khan
AIP

Variational Bayes is a popular method for approximate inference but its derivation can be cumbersome. To simplify the process, we give a 3-step recipe to identify the posterior form by explicitly looking for linearity with respect to expectations of well-known distributions. We can then directly write the update by simply "reading-off" the terms in front of those expectations. The recipe makes the derivation easier, faster, shorter, and more general.

Diffusion Models for Constrained Domains

N. Fishman, L. Klarner, V. De Bortoli, E. Mathieu, and M. Hutchinson
University of Oxford, ENS Ulm, University of Cambridge

Denoising diffusion models are a recent class of generative models which achieve state-of-the-art results in many domains such as unconditional image generation and text-to-speech tasks. They consist of a noising process destroying the data and a backward stage defined as the time-reversal of the noising diffusion. Building on their success, diffusion models have recently been extended to the Riemannian manifold setting. Yet, these Riemannian diffusion models require geodesics to be defined for all times. While this setting encompasses many important applications, it does not include manifolds defined via a set of inequality constraints, which are ubiquitous in many scientific domains such as robotics and protein design. In this work, we introduce two methods to bridge this gap. First, we design a noising process based on the logarithmic barrier metric induced by the inequality constraints. Second, we introduce a noising process based on the reflected Brownian motion. As existing diffusion model techniques cannot be applied in this setting, we derive new tools to define such models in our framework. We empirically demonstrate the applicability of our methods to a number of synthetic and real-world tasks, including the constrained conformational modelling of protein backbones and robotic arms.

Generative Modeling For Time Series via Schrödinger Bridge

M. Hamdouche, P-H. Labordere, and H. Pham
Université Paris Cité

We propose a novel generative model for time series based on Schrödinger bridge (SB) approach. This consists in the entropic interpolation via optimal transport between a reference probability measure on path space and a target measure consistent with the joint data distribution of the time series. The solution is characterized by a stochastic differential equation on finite horizon with a path-dependent drift function, hence respecting the temporal dynamics of the time series distribution. We can estimate the drift function from data samples either by kernel regression methods or with LSTM neural networks, and the simulation of the SB diffusion yields new synthetic data samples of the time series. The performance of our generative model is evaluated through a series of numerical experiments. First, we test with a toy autoregressive model, a GARCH Model, and the example of fractional Brownian motion, and measure the accuracy of our algorithm with marginal and temporal dependencies metrics. Next, we use our SB generated synthetic samples for the application to deep hedging on real-data sets. Finally, we illustrate the SB approach for generating sequence of images.

Optimal Trading in Automatic Market Makers with Deep Learning

S. Jaimungal, Y. Saporito, M. de Souza, and Y. Thamsten
University of Toronto, Universidade Federal Fluminense

This article explores the optimisation of trading strategies in Constant Function Market Makers (CFMMs) and centralised exchanges. We develop a model that accounts for the interaction between these two markets, estimating the conditional dependence between variables using the concept of conditional elicitability. Furthermore, we pose an optimal execution problem where the agent hides their orders by controlling the rate at which they trade. We do so without approximating the market dynamics. The resulting dynamic programming equation is not analytically tractable, therefore, we employ the deep Galerkin method to solve it. Finally, we conduct numerical experiments and illustrate that the optimal strategy is not prone to price slippage and outperforms naïve strategies.

Resurrecting Recurrent Neural Networks for Long Sequences

A. Orvieto, S. L. Smith, A. Gu, A. Fernando, C. Gulcehre, R. Pascanu, and S. De
ETH Zurich, DeepMind

Recurrent Neural Networks (RNNs) offer fast inference on long sequences but are hard to optimize and slow to train. Deep state-space models (SSMs) have recently been shown to perform remarkably well on long sequence modeling tasks, and have the added benefits of fast parallelizable training and RNN-like fast inference. However, while SSMs are superficially similar to RNNs, there are important differences that make it unclear where their performance boost over RNNs comes from. In this paper, we show that careful design of

deep RNNs using standard signal propagation arguments can recover the impressive performance of deep SSMs on long-range reasoning tasks, while also matching their training speed. To achieve this, we analyze and ablate a series of changes to standard RNNs including linearizing and diagonalizing the recurrence, using better parameterizations and initializations, and ensuring proper normalization of the forward pass. Our results provide new insights on the origins of the impressive performance of deep SSMs, while also introducing an RNN block called the Linear Recurrent Unit that matches both their performance on the Long Range Arena benchmark and their computational efficiency.

Structured State Space Models for In-Context Reinforcement Learning

C. Lu, Y. Schroecker, A. Gu, E. Parisotto, J. Foerster, S. Singh, and F. Behbahani

University of Oxford, DeepMind

Structured state space sequence (S4) models have recently achieved state-of-the-art performance on long-range sequence modeling tasks. These models also have fast inference speeds and parallelisable training, making them potentially useful in many reinforcement learning settings. We propose a modification to a variant of S4 that enables us to initialise and reset the hidden state in parallel, allowing us to tackle reinforcement learning tasks. We show that our modified architecture runs asymptotically faster than Transformers and performs better than LSTM models on a simple memory-based task. Then, by leveraging the model's ability to handle long-range sequences, we achieve strong performance on a challenging meta-learning task in which the agent is given a randomly-sampled continuous control environment, combined with a randomly-sampled linear projection of the environment's observations and actions. Furthermore, we show the resulting model can adapt to out-of-distribution held-out tasks. Overall, the results presented in this paper suggest that the S4 models are a strong contender for the default architecture used for in-context reinforcement learning.

Incorporating Unlabelled Data into Bayesian Neural Networks

M. Sharma, T. Rainforth, Y. W. Teh, and V. Fortuin

University of Oxford, University of Cambridge, Helmholtz AI

We develop a contrastive framework for learning better prior distributions for Bayesian Neural Networks using unlabelled data. With this framework, we propose a practical BNN algorithm that offers the label-efficiency of self-supervised learning and the principled uncertainty estimates of Bayesian methods. Finally, we demonstrate the advantages of our approach for data-efficient learning in semi-supervised and low-budget active learning problems.

On Over-Squashing in Message Passing Neural Networks: The Impact of Width, Depth, and Topology

F Di Giovanni, L. Giusti, F. Barbero, G. Luise, P. Liò, and M. Bronstein

University of Oxford, University of Cambridge, Sapienza University

Message Passing Neural Networks (MPNNs) are instances of Graph Neural Networks that leverage the graph to send messages over the edges. This inductive bias leads to a phenomenon known as over-squashing, where a node feature is insensitive to information contained at distant nodes. Despite recent methods introduced to mitigate this issue, an understanding of the causes for oversquashing and of possible solutions are lacking. In this theoretical work, we prove that: (i) Neural network width can mitigate over-squashing, but at the cost of making the whole network more sensitive; (ii) Conversely, depth cannot help mitigate over-squashing: increasing the number of layers leads to over-squashing being dominated by vanishing gradients; (iii) The graph topology plays the greatest role, since over-squashing occurs between nodes at high commute time. Our analysis provides a unified framework to study different recent methods introduced to cope with over-squashing and serves as a justification for a class of methods that fall under 'graph rewiring'.

Mathematical Finance

Rough Volatility, Path-dependent PDEs and Weak Rates of Convergence

O. Bonesini, A. Jacquier, and A. Pannier

Imperial College London, Université Paris Cité

In the setting of stochastic Volterra equations, and in particular rough volatility models, we show that conditional expectations are the unique classical solutions to path-dependent PDEs. The latter arise from the functional Ito formula developed by [Viens & Zhang (2019). A martingale approach for fractional Brownian motions and related path dependent PDEs]. We then leverage these tools to study weak rates of convergence for discretised stochastic integrals of smooth functions of a Riemann-Liouville fractional Brownian motion with Hurst parameter $H \in (0, 1/2)$. These integrals approximate log-stock prices in rough volatility models. We obtain weak error rates of order 1 if the test function is quadratic and of order $H + 1/2$ for smooth test functions.

A Multilevel Stochastic Approximation Algorithm for Value-at-Risk and Expected Shortfall Estimation

S. Crépey, N. Frikha, and A. Louzi
Université Paris Cité, Université Paris 1 Panthéon-Sorbonne

We propose a multilevel stochastic approximation (MLSA) scheme for the computation of the Value-at-Risk (VaR) and the Expected Shortfall (ES) of a financial loss, which can only be computed via simulations conditional on the realization of future risk factors. Thus, the problem of estimating its VaR and ES is nested in nature and can be viewed as an instance of a stochastic approximation problem with biased innovation. In this framework, for a prescribed accuracy ϵ, the optimal complexity of a standard stochastic approximation algorithm is shown to be of order ϵ^3. To estimate the VaR, our MLSA algorithm attains an optimal complexity of order $\epsilon^{2\delta}$, where $\delta < 1$ is some parameter depending on the integrability degree of the loss, while to estimate the ES, it achieves an optimal complexity of order $\epsilon^2 \left|\ln \epsilon\right|^2$. Numerical studies of the joint evolution of the error rate and the execution time demonstrate how our MLSA algorithm regains a significant amount of the lost performance due to the nested nature of the problem.

Towards Systematic Intraday News Screening: a Liquidity-Focused Approach

M. Rosenbaum and J. Zhang
École polytechnique

News can convey bearish or bullish views on financial assets. Institutional investors need to evaluate automatically the implied news sentiment based on textual data. Given the huge amount of news articles published each day, most of which are neutral, we present a systematic news screening method to identify the "true" impactful ones, aiming for more effective development of news sentiment learning methods. Based on several liquidity-driven variables, including volatility, turnover, bid-ask spread, and book size, we associate each 5-min time bin to one of two specific liquidity modes. One represents the "calm" state at which the market stays for most of the time and the other, featured with relatively higher levels of volatility and trading volume, describes the regime driven by some exogenous events. Then we focus on the moments where the liquidity mode switches from the former to the latter and consider the news articles published nearby impactful. We apply naive Bayes on these filtered samples for news sentiment classification as an illustrative example. We show that the screened dataset leads to more effective feature capturing and thus superior performance on short-term asset return prediction compared to the original dataset.

Economics

Equilibrium-Invariant Embedding, Metric Space, and Fundamental Set of 2×2 Normal-Form Games

L. Marris, I. Gemp, and G. Piliouras
University College London, DeepMind

Equilibrium solution concepts of normal-form games, such as Nash equilibria, correlated equilibria, and coarse correlated equilibria, describe the joint strategy profiles from which no player has incentive to unilaterally deviate. They are widely studied in game theory, economics, and multiagent systems. Equilibrium concepts are invariant under certain transforms of the payoffs. We define an equilibrium-inspired distance metric for the space of all normal-form games and uncover a distance-preserving equilibrium-invariant embedding. Furthermore, we propose an additional transform which defines a better-response-invariant distance metric and embedding. To demonstrate these metric spaces we study 2×2 games. The equilibrium-invariant embedding of 2×2 games has an efficient two variable parameterization (a reduction from eight), where each variable geometrically describes an angle on a unit circle. Interesting properties can be spatially inferred from the embedding, including: equilibrium support, cycles, competition, coordination, distances, best-responses, and symmetries. The best-response-invariant embedding of 2×2 games, after considering symmetries, rediscovers a set of 15 games, and their respective equivalence classes. We propose that this set of game classes is fundamental and captures all possible interesting strategic interactions in 2×2 games. We introduce a directed graph representation and name for each class. Finally, we leverage the tools developed for 2×2 games to develop game theoretic visualizations of large normal-form and extensive-form games that aim to fingerprint the strategic interactions that occur within.

Optimism and Overconfidence

L. Sinander
University of Oxford

I revisit the standard moral-hazard model, in which an agent's preference over contracts is rooted in costly effort choice. I characterise the behavioural content of the model in terms of empirically testable axioms, and show that the model's parameters are identified. I propose general behavioural definitions of relative (over)confidence and optimism, and characterise these in terms of the parameters of the moral-hazard model. My formal results are rooted in a simple but powerful insight: that the moral-hazard model is closely related to the well-known 'variational' model of choice under uncertainty.

Non-Diversified Portfolios with Subjective Expected Utility

C. P. Chambers and G. Gerasimou
Glasgow University, Georgetown University

Although portfolio diversification is the typical strategy followed by risk-averse investors, extreme portfolios that allocate all funds to a single asset/state of the world are common too. Such asset-demand behavior is compatible with risk-averse subjective expected utility maximization under beliefs that assign a strictly positive probability to every state. We show that whenever finitely many extreme asset demands are rationalizable in this way under such beliefs, they are simultaneously rationalizable under the same beliefs by: (i) constant absolute risk aversion; decreasing absolute risk aversion/increasing relative risk aversion (DARA/IRRA); risk-neutral; and risk-seeking utility indices at all wealth levels; (ii) a distinct class of DARA/IRRA utility indices at some strictly positive fixed initial wealth; and (iii) decreasing relative risk aversion utility indices under bounded wealth. We also show that, in such situations, the observable data allow for sharp bounds to be given for the relevant parameters in each of the above classes of risk-averse preferences.

The Effect of Incentives on Beliefs and Choices in Games: An Experiment

T. Esteban-Casanelles and D. Gonçalves
King's College London, University College London

How and why do incentive levels affect strategic behavior? This paper examines an experiment designed to identify the causal effect of scaling up incentives on choices and beliefs in strategic settings by holding fixed opponents' actions. In dominance-solvable games, higher incentives increase action sophistication and best-response rates and decrease mistake propensity. Beliefs tend to become more accurate with higher own incentives in simple games. However, opponents with higher incentive levels are harder to predict: while beliefs track opponents' behavior when they have higher incentive levels, beliefs about opponents also become more biased. We provide evidence that incentives affect cognitive effort and that greater effort increases performance and predicts choice and belief sophistication. Overall, the data lends support to combining both payoff-dependent mistakes and costly reasoning.

The Focal Quantal Response Equilibrium

M. Kovach and G. Tserenjigmid
Virginia Tech, UC Santa Cruz

We propose a generalization of Quantal Response Equilibrium (QRE) built on a simple premise: some actions are more focal than others. In our model, which we call the Focal Quantal Response Equilibrium (Focal QRE), each player plays a stochastic version of Nash equilibrium as in the QRE, but some strategies are focal and thus are chosen relatively more frequently than other strategies after accounting for expected utilities. The Focal QRE is able to systemically account for various forms of bounded rationality of players, especially regret-aversion, salience, or limited consideration. The Focal QRE is also useful to explain observed heterogeneity of bounded rationality of players across different games. We show that regret-based focal sets perform relatively well at predicting strategies that are chosen more frequently relative to their expected utilities.

Nash Equilibria for Relative Investors with (non)Linear Price Impact

N. Bäuerle and T. Göll
Karlsruhe Institute of Technology

We consider the strategic interaction of n investors who are able to influence a stock price process and at the same time measure their utilities relative to the other investors. Our main aim is to find Nash equilibrium investment strategies in this setting in a financial market driven by a Brownian motion and investigate the influence the price impact has on the equilibrium. We consider both CRRA and CARA utility functions. Our findings show that the problem is well-posed as long as the price impact is at most linear. Moreover, numerical results reveal that the investors behave very aggressively when the price impact is beyond a critical parameter.

Acknowledgement

The OMI Research Newsletter was curated by Álvaro Cartea, Director of the Oxford-Man Institute, and Fayçal Drissi, D.Phil. Candidate. Paper suggestions are from fellow students at the institute: Álvaro Arroyo, Patrick Chang, Gerardo Durán-Martín, and Marcello Monga.

References

Bäuerle, Nicole, and Tamara Göll. 2023. "Nash equilibria for relative investors with (non) linear price impact." *arXiv preprint arXiv:2303.18161*.

Bonesini, Ofelia, Antoine Jacquier, and Alexandre Pannier. 2023. "Rough volatility, path-dependent PDEs and weak rates of convergence." *arXiv preprint arXiv:2304.03042*.

Chambers, Christopher P, and Georgios Gerasimou. 2023. "Non-diversified portfolios with subjective expected utility." *arXiv preprint arXiv:2304.08059*.

Crépey, Stéphane, Noufel Frikha, and Azar Louzi. 2023. "A Multilevel Stochastic Approximation Algorithm for Value-at-Risk and Expected Shortfall Estimation." *arXiv preprint arXiv:2304.01207*.

Di Giovanni, Francesco, Lorenzo Giusti, Federico Barbero, Giulia Luise, Pietro Lio, and Michael Bronstein. 2023. "On Over-Squashing in Message Passing Neural Networks: The Impact of Width, Depth, and Topology." *arXiv preprint arXiv:2302.02941*.

Esteban-Casanelles, Teresa, and Duarte Gonçalves. 2023. "The Effects of Incentives on Choices and Beliefs in Games: An Experiment." *arXiv preprint arXiv:2304.00412*.

Finzi, Marc, Andres Potapczynski, Matthew Choptuik, and Andrew Gordon Wilson. 2023. "A Stable and Scalable Method for Solving Initial Value PDEs with Neural Networks." *arXiv preprint arXiv:2304.14994*.

Fishman, Nic, Leo Klarner, Valentin De Bortoli, Emile Mathieu, and Michael Hutchinson. 2023. "Diffusion Models for Constrained Domains." *arXiv preprint arXiv:2304.05364*.

Hamdouche, Mohamed, Pierre Henry-Labordere, and Huyên Pham. 2023. "Generative modeling for time series via Schr {\" o} dinger bridge." *arXiv preprint arXiv:2304.05093*.

Jaimungal, Sebastian, Yuri F Saporito, Max O Souza, and Yuri Thamsten. 2023. "Optimal Trading in Automatic Market Makers with Deep Learning." *arXiv preprint arXiv:2304.02180*.

Khan, Mohammad Emtiyaz. 2023. "Variational Bayes Made Easy." *arXiv preprint arXiv:2304.14251*.

Kovach, Matthew, and Gerelt Tserenjigmid. 2023. "The Focal Quantal Response Equilibrium." *arXiv preprint arXiv:2304.00438*.

Lu, Chris, Yannick Schroecker, Albert Gu, Emilio Parisotto, Jakob Foerster, Satinder Singh, and Feryal Behbahani. 2023. "Structured State Space Models for In-Context Reinforcement Learning." *arXiv preprint arXiv:2303.03982*.

Marris, Luke, Ian Gemp, and Georgios Piliouras. 2023. "Equilibrium-Invariant Embedding, Metric Space, and Fundamental Set of 2x2 Normal-Form Games." *arXiv preprint arXiv:2304.09978*.

Michael, Nikolas, Mihai Cucuringu, and Sam Howison. 2023. "OFTER: An Online Pipeline for Time Series Forecasting." *arXiv preprint arXiv:2304.03877*.

Orvieto, Antonio, Samuel L Smith, Albert Gu, Anushan Fernando, Caglar Gulcehre, Razvan Pascanu, and Soham De. 2023. "Resurrecting recurrent neural networks for long sequences." *arXiv preprint arXiv:2303.06349*.

Sharma, Mrinank, Tom Rainforth, Yee Whye Teh, and Vincent Fortuin. 2023. "Incorporating Unlabelled Data into Bayesian Neural Networks." *arXiv preprint arXiv:2304.01762*.

Sinander, Ludvig. 2023. "Optimism and overconfidence." *arXiv preprint arXiv:2304.08343*.

Zhang, Jianfei, and Mathieu Rosenbaum. 2023. "Towards systematic intraday news screening: a liquidity-focused approach." *arXiv preprint arXiv:2304.05115*.