OMI Research Newsletter – November 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

News. The Machine Learning and Financial Econometrics Conference will take place on 12 and 13 December 2023 at the Oxford-Man Institute. Academics from various disciplines will address fundamental problems in quantitative finance with a strong focus on machine learning and data-driven models. Click here for a tentative schedule.

Research focus. Noteworthy contributions in this month's edition include works from the OMI, Columbia University, and University of Oxford. Cartea, Duran-Martin and Sánchez-Betancourt (2023) predict toxic trading flow that a broker receives from clients using a Bayesian neural network trained with a novel and efficient online procedure called PULSE. Nutz, Webster and Zhao (2023) study centralised trading desks that must continuously unwind a dynamic trading flow while minimising transaction costs. Finally, Demirel and Holz (2023) propose a data augmentation method for time-series tasks that connect intra-class samples together.

Machine Learning & Statistics

Detecting Toxic Flow

Á. CARTEA, G. DURAN-MARTIN, L. SÁNCHEZ-BETANCOURT

OXFORD-MAN INSTITUTE OF QUANTITATIVE FINANCE, QUEEN MARY UNIVERSITY OF LONDON

This paper develops a framework to predict toxic trades that a broker receives from her clients. Toxic trades are predicted with a novel online Bayesian method which we call the projection-based unification of last-layer and subspace estimation (PULSE). PULSE is a fast and statistically-efficient online procedure to train a Bayesian neural network sequentially. We employ a proprietary dataset of foreign exchange transactions to test our methodology. PULSE outperforms standard machine learning and statistical methods when predicting if a trade will be toxic; the benchmark methods are logistic regression, random forest, and a recursively-updated maximum-likelihood estimator. We devise a strategy for the broker who uses toxicity predictions to internalise or to externalise each trade received from her clients. Our methodology can be implemented in real-time because it takes less than one millisecond to update parameters and make a prediction. Compared with the benchmarks, PULSE attains the highest PnL and the largest avoided loss for each of the deploy horizons we consider.

Finding Order in Chaos: A Novel Data Augmentation Method for Time Series in Contrastive Learning

B. U. DEMIREL, C. HOLZ

ETH ZURICH

The success of contrastive learning is well known to be dependent on data augmentation. Although the degree of data augmentations has been well controlled by utilizing pre-defined techniques in some domains like vision, time-series data augmentation is less explored and remains a challenging problem due to the complexity of the data generation mechanism, such as the intricate mechanism involved in the cardiovascular system. Moreover, there is no widely recognized and general time-series augmentation method that can be applied across different tasks. In this paper, we propose a novel data augmentation method for quasi-periodic time-series tasks that aims to connect intra-class samples together, and thereby find

order in the latent space. Our method builds upon the well-known mixup technique by incorporating a novel approach that accounts for the periodic nature of non-stationary time-series. Also, by controlling the degree of chaos created by data augmentation, our method leads to improved feature representations and performance on downstream tasks. We evaluate our proposed method on three time-series tasks, including heart rate estimation, human activity recognition, and cardiovascular disease detection. Extensive experiments against state-of-the-art methods show that the proposed approach outperforms prior works on optimal data generation and known data augmentation techniques in the three tasks, reflecting the effectiveness of the presented method.

Inference for Rank-Rank Regressions

D. Chetverikov, D. Wilhelm
UCLA, LMU München

Slope coefficients in rank-rank regressions are popular measures of intergenerational mobility, for instance in regressions of a child's income rank on their parent's income rank. In this paper, we first point out that commonly used variance estimators such as the homoskedastic or robust variance estimators do not consistently estimate the asymptotic variance of the OLS estimator in a rank-rank regression. We show that the probability limits of these estimators may be too large or too small depending on the shape of the copula of child and parent incomes. Second, we derive a general asymptotic theory for rank-rank regressions and provide a consistent estimator of the OLS estimator's asymptotic variance. We then extend the asymptotic theory to other regressions involving ranks that have been used in empirical work. Finally, we apply our new inference methods to three empirical studies. We find that the confidence intervals based on estimators of the correct variance may sometimes be substantially shorter and sometimes substantially longer than those based on commonly used variance estimators. The differences in confidence intervals concern economically meaningful values of mobility and thus lead to different conclusions when comparing mobility in U.S. commuting zones with mobility in other countries.

Few-Shot Learning Patterns in Financial Time-Series for Trend-Following Strategies

K. Wood, S. Kessler, S. J. Roberts, S. Zohren
Oxford-Man Institute, University of Oxford,

Forecasting models for systematic trading strategies do not adapt quickly when financial market conditions change, as in the advent of the COVID-19 pandemic in 2020 when market conditions changed dramatically causing many forecasting models to take loss-making positions. To deal with such situations, we propose a novel time-series trend-following forecaster to quickly adapt to new market conditions, referred to as regimes. We leverage recent developments from the deep learning community and use few-shot learning. We propose the Cross Attentive Time-Series Trend Network – X-Trend – which takes positions attending over a context set of financial time-series regimes. X-Trend transfers trends from similar patterns in the context set to make predictions and take positions for a new distinct target regime. X-Trend quickly adapts to new financial regimes with a Sharpe ratio increase of 18.9% over a neural forecaster and 10-fold over a conventional Time-series Momentum strategy during the turbulent market period from 2018 to 2023. Our strategy recovers twice as quickly from the COVID-19 drawdown compared to the neural-forecaster. X-Trend can also take zero-shot positions on novel unseen financial assets obtaining a 5-fold Sharpe ratio increase versus a neural time-series trend forecaster over the same period. X-Trend both forecasts next-day prices and outputs a trading signal. Furthermore, the cross-attention mechanism allows us to interpret the relationship between forecasts and patterns in the context set.

HyperAttention: Long-context Attention in Near-Linear Time

I. Han, R. Jayaram, A. Karbasi, V. Mirrokni, D. P. Woodruff, A. Zandieh
Yale University, Carnegie Mellon University

We present an approximate attention mechanism named HyperAttention to address the computational challenges posed by the growing complexity of long contexts used in Large Language Models (LLMs). Recent work suggests that in the worst-case scenario, quadratic time is necessary unless the entries of the attention matrix are bounded or the matrix has low stable rank. We introduce two parameters which measure: (1) the max column norm in the normalized attention matrix, and (2) the ratio of row norms in the unnormalized attention matrix after detecting and removing large entries. We use these fine-grained parameters to capture the hardness of the problem. Despite previous lower bounds, we are able to achieve a linear time sampling algorithm even when the matrix has unbounded entries or a large stable rank, provided the above parameters are small. HyperAttention features a modular design that easily accommodates integration of other fast low-level implementations, particularly FlashAttention. Empirically, employing Locality Sensitive Hashing (LSH) to identify large entries, HyperAttention outperforms existing methods, giving significant speed improvements compared to state-of-the-art solutions like FlashAttention. We validate the empirical performance of HyperAttention on a variety of different long-context length datasets. For example, HyperAttention makes the inference time of ChatGLM2

50% faster on 32k context length while perplexity increases from 5.6 to 6.3. On larger context length, e.g., 131k, with causal masking, HyperAttention offers five-fold speedup on a single attention layer.

Mathematical Finance

Unwinding Stochastic Order Flow: When to Warehouse Trades

M. Nutz, K. Webster, L. Zhao
Columbia University

We study how to unwind stochastic order flow with minimal transaction costs. Stochastic order flow arises, e.g., in the central risk book (CRB), a centralized trading desk that aggregates order flows within a financial institution. The desk can warehouse in-flow orders, ideally netting them against subsequent opposite orders (internalization), or route them to the market (externalization) and incur costs related to price impact and bid-ask spread. We model and solve this problem for a general class of in-flow processes, enabling us to study in detail how in-flow characteristics affect optimal strategy and core trading metrics. Our model allows for an analytic solution in semi-closed form and is readily implementable numerically. Compared with a standard execution problem where the order size is known upfront, the unwind strategy exhibits an additive adjustment for projected future in-flows. Its sign depends on the autocorrelation of orders; only truth-telling (martingale) flow is unwound myopically. In addition to analytic results, we present extensive simulations for different use cases and regimes, and introduce new metrics of practical interest.

A generalization of the rational rough Heston approximation

J. Gatheral, R. Radoičić
The City University of New York

We extend the rational approximation of the solution of the rough Heston fractional ODE in [Gatheral, Radoičić, 2019] to the case of the Mittag-Leffler kernel. We provide numerical evidence of the convergence of the solution.

Economics

An Economic Test for an Unlawful Agreement to Adopt a Third-Party's Pricing Algorithm

J. E. Harrington Jr
University of Pennsylvania

A growing market is the supply of pricing algorithms by software developers. While there is an efficiency rationale for outsourcing pricing, anticompetitive concerns have been expressed when competitors in a market adopt the same pricing algorithm. These concerns have resulted in private litigation claiming a third-party company (who developed the pricing algorithm) and firms (who adopted it) had an unlawful agreement. This study develops an empirical test for determining whether firms' adoption decisions are coordinated and thus a violation of competition law. If adoption decisions are coordinated then adopters' average price is increasing in the number of firms that adopt, while if adoption decisions are independent then adopters' average price does not depend on the number of firms that adopt. This economic evidence could serve as a plus factor in proving the conduct of a third-party developer and adopting firms is in violation of Section 1 of the Sherman Act.

Less Than Meets the Eye: Simultaneous Experiments as a Source of Algorithmic Seeming Collusion

X. Lambin
ESSEC Business School

This article explains the supra-competitive profits observed in Calvano et al. (2020) and subsequent literature. We reveal an important mistake in the original paper and refute the idea that the phenomenon arises from collusive mechanisms, where each agent exerts the threat of punishment. Instead, our analysis reveals that both the high prices and the apparent punishment schemes result directly from simultaneous experimentation and the learning inertia inherent in most reinforcement learning techniques. Additionally, we demonstrate that such seeming collusion can arise in memoryless environments, even with myopic agents, and much faster than previously thought. Our findings challenge the policy implications of prior research and invite new, simpler approaches to address the issue of algorithmic supra-competitive pricing.

Calibrated Clustering and Analogy-Based Expectation Equilibrium

P. Jehiel, G. Weber
Paris School of Economics

Families of normal-form two-player games are categorized by players into K analogy classes applying the K-means clustering technique to the data generated by the distributions of opponent's behavior. This results in Calibrated Analogy-Based Expectation Equilibria in which strategies are analogy-based expectation equilibria given the analogy partitions and analogy partitions are derived from the strategies by the K-means clustering algorithm. We discuss various concepts formalizing this, and observe that distributions over analogy partitions are sometimes required to guarantee existence. Applications to games

with linear best-responses are discussed highlighting the differences between strategic complements and strategic substitutes.

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 PhD students at the institute: Álvaro Arroyo, Patrick Chang, Gerardo Durán-Martín, and Marcello Monga.

References

Cartea, Álvaro, Gerardo Duran-Martin, and Leandro Sánchez-Betancourt. 2023. "Detecting Toxic Flow." *Available at SSRN*.

Chetverikov, Denis, and Daniel Wilhelm. 2023. "Inference for Rank-Rank Regressions." *arXiv preprint arXiv:2310.15512*.

Demirel, Berken Utku, and Christian Holz. 2023. "Finding Order in Chaos: A Novel Data Augmentation Method for Time Series in Contrastive Learning." *arXiv preprint arXiv:2309.13439*.

Gatheral, Jim, and Radoš Radoičić. 2023. "A generalization of the rational rough Heston approximation." *arXiv preprint arXiv:2310.09181*.

Han, Insu, Rajesh Jarayam, Amin Karbasi, Vahab Mirrokni, David P Woodruff, and Amir Zandieh. 2023. "HyperAttention: Long-context Attention in Near-Linear Time." *arXiv preprint arXiv:2310.05869*.

Harrington Jr, Joseph E. 2023. "An Economic Test for an Unlawful Agreement to Adopt a Third-Party's Pricing Algorithm." *Available at SSRN 4520245*.

Jehiel, Philippe, and Giacomo Weber. 2023. "Calibrated Clustering and Analogy-Based Expectation Equilibrium."

Lambin, Xavier. 2023. "Less than meets the eye: simultaneous experiments as a source of algorithmic seeming collusion." *Available at SSRN 4498926*.

Nutz, Marcel, Kevin Webster, and Long Zhao. 2023. "Unwinding Stochastic Order Flow: When to Warehouse Trades." *arXiv preprint arXiv:2310.14144*.

Wood, Kieran, Samuel Kessler, Stephen J Roberts, and Stefan Zohren. 2023. "Few-Shot Learning Patterns in Financial Time-Series for Trend-Following Strategies." *arXiv preprint arXiv:2310.10500*.