OMI Research Newsletter – March 2025

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.*

Announcement

The Workshop on Decentralised Finance and Market Microstructure, co-organised by the University of Toronto, Columbia University and OMI, will be held on 8–9 May 2025 at the Fields Institute for Research in Mathematical Sciences, University of Toronto. Please find the Call for Contributed Papers and Posters.

The OMI Machine Learning in Quantitative Finance Conference 2025 will be held on 5–6 June 2025 at the Mathematical Institute, University of Oxford.

Highlights

Key contributions in this month's edition: Calvo-Ordoñez et al. (2025) examine how regularization and a shifted network approach enable wide neural networks to account for observation noise and arbitrary prior mean. Marton and Schneider (2025) introduce a recurrent decision tree model that learns temporal dependencies with memory and gradient descent. Chong, Delerue and Mies (2025) propose a rate-optimal estimator for the Hurst parameter of mixed martingales. Cartea et al. (2025) find that intraday volume shocks positively predict overnight stock returns but not intraday returns, challenging conventional explanations in the literature. Corhay et al. (2025) investigate the impact of data technology on firms' market power and asset prices. Harvey, Mazzoleni and Melone (2025) show that institutional portfolio rebalancing creates predictable price patterns, leading to significant costs for investors and opportunities for front-running by informed traders. Li et al. (2025) introduce an innovative semiparametric framework of duration-based volatility estimation. Fukasawa, Maire and Wunsch (2025) develop a mathematical model for liquidity provision in decentralised exchanges, analysing arbitrage, impermanent loss, and optimal market-making under transaction fees.

Machine Learning & Statistics

Observation Noise and Initialization in Wide Neural Networks

S. CALVO-ORDOÑEZ, J. PLENK, R. BERGNA, Á. CARTEA, J. M. HERNÁNDEZ-LOBATO, K. PALLA, K. CIOSEK

UNIVERSITY OF OXFORD, UNIVERSITY OF CAMBRIDGE, SPOTIFY

Performing gradient descent in a wide neural network is equivalent to computing the posterior mean of a Gaussian Process with the Neural Tangent Kernel (NTK-GP), for a specific choice of prior mean and with zero observation noise. However, existing formulations of this result have two limitations: i) the resultant NTK-GP assumes no noise in the observed target variables, which can result in suboptimal predictions with noisy data; ii) it is unclear how to extend the equivalence to an arbitrary prior mean, a crucial aspect of formulating a well-specified model. To address the first limitation, we introduce a regularizer into the neural network's training objective, formally showing its correspondence to incorporating observation noise into the NTK-GP model. To address the second, we introduce a *shifted network* that enables arbitrary prior mean functions. This approach allows us to perform gradient descent on a single neural network, without expensive ensembling or kernel matrix inversion. Our theoretical insights are validated empirically, with experiments exploring different values of observation noise and network architectures.

Decision Trees That Remember: Gradient-Based

Learning of Recurrent Decision Trees with Memory

S. Marton, M. Schneider
University of Mannheim, Boehringer Ingelheim

Neural architectures such as Recurrent Neural Networks (RNNs), Transformers, and State-Space Models have shown great success in handling sequential data by learning temporal dependencies. Decision Trees (DTs), on the other hand, remain a widely used class of models for structured tabular data but are typically not designed to capture sequential patterns directly. Instead, DT-based approaches for time-series data often rely on feature engineering, such as manually incorporating lag features, which can be suboptimal for capturing complex temporal dependencies. To address this limitation, we introduce ReMeDe Trees, a novel recurrent DT architecture that integrates an internal memory mechanism, similar to RNNs, to learn long-term dependencies in sequential data. Our model learns hard, axis-aligned decision rules for both output generation and state updates, optimizing them efficiently via gradient descent. We provide a proof-of-concept study on synthetic benchmarks to demonstrate the effectiveness of our approach.

Dimension-free Regret for Learning Asymmetric Linear Dynamical Systems

A. Marsden, E. Hazan
Google DeepMind

Previously, methods for learning marginally stable linear dynamical systems either required the transition matrix to be symmetric or incurred regret bounds that scale polynomially with the system's hidden dimension. In this work, we introduce a novel method that overcomes this trade-off, achieving dimension-free regret despite the presence of asymmetric matrices and marginal stability. Our method combines spectral filtering with linear predictors and employs Chebyshev polynomials in the complex plane to construct a novel spectral filtering basis. This construction guarantees sublinear regret in an online learning framework, without relying on any statistical or generative assumptions. Specifically, we prove that as long as the transition matrix has eigenvalues with complex component bounded by $1/\text{polylog}(T)$, then our method achieves regret $\tilde{O}(T^{9/10})$ when compared to the best linear dynamical predictor in hindsight.

Robust and Conjugate Spatio-Temporal Gaussian Processes

W. Laplante, M. Altamirano, A. Duncan, J. Knoblauch, F.-X. Briol
University College London, Imperial College London, Alan Turing Institute,

State-space formulations allow for Gaussian process (GP) regression with linear-in-time computational cost in spatio-temporal settings, but performance typically suffers in the presence of outliers. In this paper, we adapt and specialise the robust and conjugate GP (RCGP) framework of Altamirano et al. (2024) to the spatio-temporal setting. In doing so, we obtain an outlier-robust spatio-temporal GP with a computational cost comparable to classical spatio-temporal GPs. We also overcome the three main drawbacks of RCGPs: their unreliable performance when the prior mean is chosen poorly, their lack of reliable uncertainty quantification, and the need to carefully select a hyperparameter by hand. We study our method extensively in finance and weather forecasting applications, demonstrating that it provides a reliable approach to spatio-temporal modelling in the presence of outliers.

Least Squares Variational Inference

Y. Le Fay, N. Chopin, S. Barthelmé
CREST-ENSAE, GIPSA-Lab@CNRS

Variational inference consists in finding the best approximation of a target distribution within a certain family, where 'best' means (typically) smallest Kullback-Leiber divergence. We show that, when the approximation family is exponential, the best approximation is the solution of a fixed-point equation. We introduce LSVI (Least-Squares Variational Inference), a Monte Carlo variant of the corresponding fixed-point recursion, where each iteration boils down to ordinary least squares regression and does not require computing gradients. We show that LSVI is equivalent to stochastic mirror descent; we use this insight to derive convergence guarantees. We introduce various ideas to improve LSVI further when the approximation family is Gaussian, leading to a $O(d^3)$ complexity in the dimension d of the target in the full-covariance case, and a $O(d)$ complexity in the mean-field case. We show that LSVI outperforms state-of-the-art methods in a range of examples, while remaining gradient-free, that is, it does not require computing gradients.

Rate-Optimal Estimation of Mixed Semimartingales

C. H. Chong, T. Delerue, F. Mies
Hong Kong University of Science and Technology, Helmholtz Munich, Delft University of Technology

Consider the sum $Y = B + B(H)$ of a Brownian motion B and an independent fractional Brownian motion $B(H)$ with Hurst parameter $H \in (0, 1)$. Even though $B(H)$ is not a semimartingale, it was shown by Cheridito (*Bernoulli*, 7 (2001), 913–934) that Y is a semimartingale if $H > 3/4$. Moreover, Y is locally equivalent to B in this case, so H cannot be consistently estimated from local observations of Y. This paper pivots on another unexpected feature

in this model: if B and $B(H)$ become correlated, then Y will never be a semimartingale, and H can be identified, regardless of its value. This and other results will follow from a detailed statistical analysis of a more general class of processes called mixed semimartingales, which are semiparametric extensions of Y with stochastic volatility in both the martingale and the fractional component. In particular, we derive consistent estimators and feasible central limit theorems for all parameters and processes that can be identified from high-frequency observations. We further show that our estimators achieve optimal rates in a minimax sense.

Finance and Economics

Volume Shocks and Overnight Returns

Á. Cartea, M. Cucuringu, Q. Jin, M. Wilson
University of Oxford, University of California, Los Angeles

We study the effect of intraday volume shocks on stock returns during overnight and intraday periods. We discover a significant positive relationship between volume shocks and subsequent overnight returns, while no such effect exists during the next intraday session. This pattern is consistent regardless of the market capitalization of stocks. Well-known asset pricing risk factors and common explanations that associate abnormal trading volume with investor attention and cost of capital cannot account for the distinct intraday and overnight patterns we observe. We employ linear and machine learning models to forecast volume shocks and construct portfolios that monetize the positive correlation between volume shocks and overnight stock returns. Our approach addresses the issue that volume shock is only known after the close auction when one trades stocks; we show that this issue of non-tradability does not explain the observed relationship between volume shock and overnight stock returns.

Data, Markups, and Asset Prices

A. Corhay, K. Hu, J. E. Li, J. Tong, C.-Y. Tsou
University of Toronto, University of Warwick, University of Oxford, University of Manchester

This paper investigates how data technology affects firms' market power and asset prices. Using a novel dataset tracking firms' employment of data scientists, we document three key empirical findings: firms with higher proportions of data scientists exhibit larger markups, have higher information quality proxied by lower sales forecast errors, and earn higher stock returns. Specifically, a long-short portfolio strategy based on firms' data scientist ratios generates significant annual excess returns of approximately 4%. To quantitatively rationalize these empirical findings, we develop a heterogeneous firm model in which firms optimally hire data scientists to learn about unobserved consumer tastes. The model demonstrates how data enables firms to improve demand forecasting accuracy and extract higher markups. Importantly, supply-constrained firms have stronger incentives to hire data scientists, leading to countercyclical data scientist hiring that amplifies their exposures to aggregate risk through an operating leverage channel. We provide empirical evidence supporting our model mechanism.

The Aggregated Equity Risk Premium

V. Azevedo, C. Riedersberger, M. Velikov
RPTU Kaiserslautern-Landau, Pennsylvania State University

We propose a new approach for predicting the equity risk premium (ERP) that first estimates expected returns on individual stock before aggregating them to the market level. Our deep learning combination forecast aggregates firm-level return predictions from neural networks of varying complexity, trained on a comprehensive two-dimensional feature set of post-publication firm-level characteristics and aggregate macroeconomic variables. Using this aggregation method, we achieve an out-of-sample R^2 of 2.74% in a sample from 2000 to 2021. The forecasts demonstrate strong economic significance in trading strategies even with transaction costs. While the market generated a return of 376% over this period, a simple market-timing strategy based on our model's forecast signs yields a net cumulative return of approximately 768%. Our results show that aggregating firm-level predictions can lead to profitable market timing signals, challenging the conventional wisdom that the ERP is unpredictable out-of-sample and suggesting that valuable market-wide information can be extracted from the cross-section of individual stocks.

The Unintended Consequences of Rebalancing

C. R. Harvey, M. G. Mazzoleni, Alessandro Melone
Duke University, Capital Group, Ohio State University

Institutional investors engage in trillions of dollars of regular portfolio rebalancing, often based on calendar schedules or deviations from allocation targets. We document that such rebalancing has a market impact and generates predictable price patterns. When stocks are overweight, funds sell stocks and buy bonds, leading to a decrease in equity returns of 17 basis points over the next day. Our results are robust to controls for momentum, reversals, and macroeconomic information. Importantly, we estimate that current rebalancing practices cost investors about $16 billion annually-or $200 per U.S. household. Moreover, the predictability of these trades

enables certain market participants to profit by front-running the orders of large institutional funds. While rebalancing remains a fundamental tool for investors, our findings highlight the costs associated with prevailing strategies and emphasize the need for innovative approaches to mitigate these costs.

Can AI Solve the Peer Review Crisis? A Large Scale Experiment on LLM's Performance and Biases in Evaluating Economics Papers

P. Pataranutaporn, N. Powdthavee, P. Maes
Massachusetts Institute of Technology, Nanyang Technological University

We investigate whether artificial intelligence can address the peer review crisis in economics by analyzing 27,090 evaluations of 9,030 unique submissions using a large language model (LLM). The experiment systematically varies author characteristics (e.g., affiliation, reputation, gender) and publication quality (e.g., top-tier, mid-tier, low-tier, AI generated papers). The results indicate that LLMs effectively distinguish paper quality but exhibit biases favoring prominent institutions, male authors, and renowned economists. Additionally, LLMs struggle to differentiate high-quality AI-generated papers from genuine top-tier submissions. While LLMs offer efficiency gains, their susceptibility to bias necessitates cautious integration and hybrid peer review models to balance equity and accuracy.

Collective Intelligence in Dynamic Networks

F. Mudekereza
Massachusetts Institute of Technology

We revisit DeGroot learning to examine the robustness of social learning outcomes in dynamic networks – networks that evolve randomly over time. Randomness stems from multiple sources such as random matching and strategic network formation. Our main contribution is that random dynamics have double-edged effects depending on social structure: while they can foster consensus and boost collective intelligence, they can have adverse effects such as slowing down the speed of learning and causing long-term disagreement. Collective intelligence in dynamic networks requires balancing people's average influence with their average trust as society grows. We also find that the initial social structure of a dynamic network plays a central role in shaping long-term beliefs.

Decoupling Interday and Intraday Volatility Dynamics with Price Durations

Y. Li, I. Nolte, S. Nolte, S. Yu
University of Manchester, Lancaster University, University of Oxford

This paper introduces a novel framework for duration-based volatility estimation. Instead of a constant threshold commonly adopted in the literature, we employ a daily predicted threshold that adapts to changes in daily volatility. This innovation allows us to disentangle daily and intraday volatility dynamics from price durations, which greatly simplifies the parametric modelling of price durations and hence leads to more accurate volatility estimators. Simulation results demonstrate superior finite-sample performance of our duration-based estimators for both spot and integrated volatility compared to some established methods. An empirical application based on intraday data for the SPDR S&P 500 ETF highlights the improved forecasting accuracy of our integrated volatility estimator within a standard daily volatility forecasting framework. Furthermore, an intraday analysis of spot volatility estimation shows that our method can capture the immediate and substantial impact of FOMC news announcements on market volatility.

Mathematical and Computational Finance

Liquidity Provision of Utility Indifference Type in Decentralized Exchanges

M. Fukasawa, B. Maire, M. Wunsch
University of Osaka, Quantena, ZHAW School of Management and Law

We present a mathematical formulation of liquidity provision in decentralized exchanges. We focus on constant function market makers of utility indifference type, which include constant product market makers with concentrated liquidity as a special case. First, we examine no-arbitrage conditions for a liquidity pool and compute an optimal arbitrage strategy when there is an external liquid market. Second, we show that liquidity provision suffers from impermanent loss unless a transaction fee is levied under the general framework with concentrated liquidity. Third, we establish the well-definedness of arbitrage-free reserve processes of a liquidity pool in continuous-time and show that there is no loss-versus-rebalancing under a nonzero fee if the external market price is continuous. We then argue that liquidity provision by multiple liquidity providers can be understood as liquidity provision by a representative liquidity provider, meaning that the analysis boils down to that for a single liquidity provider. Last, but not least, we give an answer to the fundamental question in which sense the very construction of constant function market makers with concentrated liquidity in the popular platform Uniswap v3 is optimal.

A Sea of Coins: The Proliferation of Cryptocurrencies in UniswapV2

M. Naviglio, F. Tarantelli, F. Lillo

Scuola Normale Superiore, INFN Sezione di Pisa, Universit'a di Bologna

Blockchain technology has revolutionized financial markets by enabling decentralized exchanges (DEXs) that operate without intermediaries. Uniswap V2, a leading DEX, facilitates the rapid creation and trading of new tokens, offering high return potential but exposing investors to significant risks. In this work, we analyze the financial impact of newly created tokens, assessing their market dynamics, profitability and liquidity manipulations. Our findings reveal that a significant portion of market liquidity is trapped in honeypots, reducing market efficiency and misleading investors. Applying a simple buy-and-hold strategy, we are able to uncover some major risks associated with investing in newly created tokens, including the widespread presence of rug pulls and sandwich attacks. We extract the optimal sandwich amount, revealing that their proliferation in new tokens stems from higher profitability in low-liquidity pools. Furthermore, we analyze the fundamental differences between token price evolution in swap time and physical time. Using clustering techniques, we highlight these differences and identify typical patterns of honeypot and sellable tokens. Our study provides insights into the risks and financial dynamics of decentralized markets and their challenges for investors.

Selfish Mining under General Stochastic Rewards

M. Bahrani, M. Neuder, S. M. Weinberg

Ritual, Ethereum Foundation, Princeton University

Selfish mining, a strategy where Proof-of-Work consensus participants selectively withhold blocks, allows miners to earn disproportionately high revenue. The vast majority of the selfish mining literature focuses exclusively on block rewards. Carlsten et al. (2016) is a notable exception, which observes that similar strategic behavior may be profitable in a zero-block-reward regime if miners are compensated with transaction fees alone. As of February 2025, neither model fully captures miner incentives. The block reward remains 3.125 BTC, yet some blocks yield significantly higher revenue. For example, congestion during the launch of the Babylon protocol in August 2024 caused transaction fees to spike from 0.14 BTC to 9.52 BTC, a 68× increase in fee rewards within two blocks. We present a framework for considering strategic behavior under more general miner reward functions that could be stochastic, variable in time, and/or ephemeral. This model can capture many existing reward sources (sometimes called Miner/Maximal Extractable Value or MEV) in blockchains today. We use our framework to examine the profitability of cutoff selfish mining strategies for any reward function identically distributed across forks. Our analysis requires a novel reward calculation technique to capture non-linearity in general rewards. We instantiate these results in a combined reward function that much more accurately represents miner incentives as they exist in Bitcoin today. This reward function includes block rewards and linear-in-time transaction fees, which have been studied in isolation. It also introduces a third random reward motivated by the aforementioned transaction fee spike. This instantiation enables us to (i) make qualitative observations, (ii) make quantitative claims, and (iii) confirm the theoretical analysis using Monte Carlo simulations.

Acknowledgement

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

References

Azevedo, Vitor, Christoph Riedersberger, and Mihail Velikov. 2025. "The Aggregated Equity Risk Premium." *SSRN Working Paper*.

Bahrani, Maryam, Michael Neuder, and S Matthew Weinberg. 2025. "Selfish Mining under General Stochastic Rewards." *arXiv preprint arXiv:2502.20360*.

Calvo-Ordoñez, Sergio, Jonathan Plenk, Richard Bergna, Álvaro Cartea, José Miguel Hernández-Lobato, Konstantina Palla, and Kamil Ciosek. 2025. "Observation Noise and Initialization in Wide Neural Networks." *arXiv preprint arXiv:2502.01556*.

Cartea, Álvaro, Mihai Cucuringu, Qi Jin, and Mungo Wilson. 2025. "Volume Shocks and Overnight Returns." *SSRN Working Paper*.

Chong, Carsten H, Thomas Delerue, and Fabian Mies. 2025. "Rate-Optimal Estimation of Mixed Semimartingales." *The Annals of Statistics*, 53(1): 219–244.

Corhay, Alexandre, Kejia Hu, Jun E Li, Jincheng Tong, and Chi-Yang Tsou. 2025. "Data, Markups, and Asset Prices." *SSRN Working Paper*.

Fay, Yvann Le, Nicolas Chopin, and Simon Barthelmé. 2025. "Least squares variational inference." *arXiv preprint arXiv:2502.18475*.

Fukasawa, Masaaki, Basile Maire, and Marcus Wunsch. 2025. "Liquidity Provision of Utility Indifference Type in Decentralized Exchanges." *arXiv preprint arXiv:2502.01931*.

Harvey, Campbell R, Michele G Mazzoleni, and Alessandro Melone. 2025. "The Unintended Consequences of Rebalancing." *SSRN Working Paper*.

Laplante, William, Matias Altamirano, Andrew Duncan, Jeremias Knoblauch, and François-Xavier Briol. 2025. "Robust and Conjugate Spatio-Temporal Gaussian Processes." *arXiv preprint arXiv:2502.02450*.

Li, Yifan, Ingmar Nolte, Sandra Nolte, and Shifan Yu. 2025. "Decoupling Interday and Intraday Volatility Dynamics with Price Durations." *SSRN Working Paper*.

Marsden, Annie, and Elad Hazan. 2025. "Dimension-free Regret for Learning Asymmetric Linear Dynamical Systems." *arXiv preprint arXiv:2502.06545*.

Marton, Sascha, and Moritz Schneider. 2025. "Decision Trees That Remember: Gradient-Based Learning of Recurrent Decision Trees with Memory." *arXiv preprint arXiv:2502.04052*.

Mudekereza, Florian. 2025. "Collective Intelligence in Dynamic Networks." *arXiv preprint arXiv:2502.12660*.

Naviglio, Manuel, Francesco Tarantelli, and Fabrizio Lillo. 2025. "A Sea of Coins: The Proliferation of Cryptocurrencies in UniswapV2." *arXiv preprint arXiv:2502.10512*.

Pataranutaporn, Pat, Nattavudh Powdthavee, and Pattie Maes. 2025. "Can AI Solve the Peer Review Crisis? A Large Scale Experiment on LLM's Performance and Biases in Evaluating Economics Papers." *arXiv preprint arXiv:2502.00070*.