# Quant Strats Europe 2025: The Next Frontier in Quantitative Finance: From Alpha Models to Autonomous Market Systems

Dr. Paul Alexander Bilokon

Thalesians Ltd

Imperial College London

15 October, 2025

# Overview

- Advances in AI and High-Performance Computing (HPC) are transforming quantitative finance.
- The evolution: from static alpha models $\rightarrow$ adaptive, learning systems $\rightarrow$ autonomous market ecosystems.
- Today's discussion:
  1. Advances in classical and deep econometrics
  2. Deep filtering and Bayesian inference at scale
  3. Reinforcement learning as portfolio optimization
  4. Sequence models: Transformers and LSTMs
  5. Hardware acceleration: LLMs on FPGAs
  6. The xdb computational engine

# Advances in Classical Econometrics

**Paper:** *Regularized Jump Models for Regime Identification and Feature Selection*
**Authors:** Edward Selig and Paul Bilokon (SSRN #4950423)

**Key Concepts:**

- Extends the jump modeling framework to support regime identification and feature selection.
- Through a simulation experiment, we find evidence that these new models outperform the standard and sparse jump models, both in terms of regime identification and feature selection.

**Summary:**

- Jump modeling coupled with modern machine learning yields superior results.

# Deep Econometrics

**Paper:** *Deep Econometrics: Representation Learning in Economic Systems*
**Authors:** Paul Bilokon, et al. (SSRN #5286898)

**Key Concepts:**

- Merges neural architectures with econometric estimation.

**Summary:**

- Econometric systems become differentiable and data-adaptive.
- Enables end-to-end learning of causal macro-financial representations.
- Provides interpretability bridges between ML and economic theory.

# Deep Filtering

**Paper:** *Deep Filtering: Data Assimilation and Sequential Inference in Finance*
**Authors:** Robert Stok, Paul Bilokon, Joseph Simonian, *Journal of Financial Data Science*, Vol. 6, No. 2

**Key Concepts:**

- Combines Kalman and particle filters with deep neural networks.
- Learns non-linear observation and transition functions.
- Enables real-time state estimation under model uncertainty.

**Summary:**

- Provides a universal framework for filtering financial time series.
- Applied to volatility estimation, hidden liquidity, and signal extraction.
- Foundation for next-generation autonomous trading agents.

# Reinforcement Learning and Portfolio Theory

**Paper:** *Optimal Allocation with Continuous Sharpe Ratio Covariance Bandits*
**Authors:** Valeria Varlashova and Paul Alexander Bilokon, *Journal of Financial Data Science*, Vol. 7, No. 3, 2025.

**Key Concepts:**

- RL generalizes classical portfolio optimization (Markowitz, Merton).
- State-action-reward framework maps to dynamic allocation.
- Policy gradients and actor–critic methods yield adaptive strategies.

**Summary:**

- Market interaction becomes an iterative learning process.
- Risk management adapts to changing reward landscapes.
- Enables autonomous allocation agents trained via simulation.

# Transformers and LSTMs in Trading

**Paper:** *Transformers Versus LSTMs for Electronic Trading*
**Authors:** Paul Bilokon and Yitao Qiu (SSRN #4577922)

**Paper:** *Exploring the Advantages of Transformers for High-Frequency Trading*
**Authors:** Fazl Barez, Paul Bilokon, Arthur Gervais, and Nikita Lisitsyn (SSRN #4364833)

**Key Concepts:**

- Sequence-to-sequence architectures capture long-range dependencies.
- LSTMs for temporal memory; Transformers for global attention.
- Enables joint modeling of prices, order flow, and sentiment.

**Summary:**

- Outperforms classical ARIMA/GARCH in multi-horizon forecasting.
- Transformer-based market models reveal hierarchical temporal structure.
- Supports the vision of self-organizing, interpretable alpha generation.

# LLMs on FPGAs

**Paper:** *Optimizing transformer neural network for real-time outlier detection on FPGAs*
**Authors:** Ilia Sobakinskikh and Paul Bilokon (SSRN #4880184)

**Key Concepts:**

- Hardware acceleration for inference and fine-tuning of LLMs.
- Exploits FPGA parallelism and low latency for trading environments.
- Integrates with event-driven architectures for autonomous decisioning.

**Summary:**

- Achieves 10–50$\times$ efficiency gains over CPU inference.
- Reduces energy footprint while increasing execution determinism.
- Foundation for real-time market interpretation by AI agents.

# xdb: A Heterogeneous Computational Engine

**Project:** *xdb – A CPU/GPU/FPGA Computational Engine Based on APL*
**Author:** Paul Bilokon

**Key Concepts:**

- A data-parallel array language integrating CPU, GPU, and FPGA computation.
- Based on APL semantics: tacit, vectorized, symbolic.
- Supports JIT compilation, memory locality optimization, and functional concurrency.

**Summary:**

- Designed for low-latency quantitative research and live trading.
- Enables unified development across heterogeneous hardware.
- Paves the way for autonomous, self-optimizing market systems.

# Conclusion

- Quant finance is evolving from static models to self-learning ecosystems.
- Integration of AI, econometrics, and HPC enables fully autonomous systems.
- The next decade will see convergence of research, hardware, and capital allocation.

**Learn more:**

$\texttt{https://papers.ssrn.com/sol3/cf}_{d}ev/AbsByAuth.cfm?per_{i}d = 1462568$