Stochastic Modelling of Big Data in Finance, Insurance and Energy Markets

Anatoliy Swishchuk (University of Calgary, AB, Canada),
Rudi Zagst (TUM, Munich, Germany),
Wenjun Jiang (University of Calgary,AB, Canada)

May 19, 2023–May 21, 2023

1 Overview of the Field

Since the 2008 global financial crisis, insurance and energy markets have become closely linked with financial markets, with energy prices exhibiting more financial characteristics, e.g., similar volatile behaviour, creating a need for cooperation in such area as data science, and to communicate with highly trained professionals who work with financial, insurance and energy market data, and who can model those data. Big data and data analytics have recently become an essential part of the financial, insurance and energy markets services industry. The 2-Day workshop "Stochastic Modelling of Big Data In Finance, Insurance and Energy markets" has brought together academics, industry people and graduate students to share their ideas, to share their research experience, and to learn more about big data and to know how to model those data. This is the first data science workshop that was devoted to stochastic modelling of big data in finance, insurance, and energy markets. The topics of the workshop included stochastic and statistical modelling of big data in insurance, finance, and energy markets.

2 Recent Developments and Open Problems

Systemic risk, or instabilities, occur in many complex systems: In ecology (diversity of species), in climate change, in material behavior (phase transitions), in finance (financial crisis), etc. In finance, e.g., two types of trading in equities are widely practiced today: High-frequency (limit-order and market) trading and statistical arbitrage or market neutral (generalized) pairs trading. These types of trading account for well over two thirds the volume traded today. It is not yet clear how to quantify the systemic risk, or the market instabilities generated by these types of trading. Main problem: how to deal with big data arising in electronic markets for algorithmic and high-frequency (milliseconds) trading that contain two types of orders, limit orders and market orders. More than half of the markets in today's highly competitive and relentlessly fast-paced financial world now use a limit order book (LOB) mechanism to facilitate trade. Same problems arise in insurance and energy markets associated with big data. The workshop was intended to bring together professionals from finance, insurance and energy markets areas to discuss these open problems and to find the ways for solving them.

3 Presentation Highlights

From the **insurance** side the talks were presented by Wenjun Jiang, Ye Wang and Rudi Zagst. Wenjun Jiang's was devoted to revisiting the optimal reinsurance design under adverse selection and Value-at-Risk. Thus, the design of optimal reinsurance contracts from a monopolistic reinsurer's perspective in the presence of information asymmetry was revisited. The reinsurer aims to maximize its expected profit and the insurers select the contracts based on their Value-at-Risk (VaR) preferences. It was shown that the optimal contract menu is also the first-best solution. Some numerical examples are presented to show more implications of our results [3,4]. Ye Wang's talk considered optimal reinsurance with Vajda condition and Range-Value-at-Risk It was studied an optimal reinsurance problem where the insurer's risk-adjusted liability gets minimized. To consider both robustness and tail risk, the insurer is assumed to apply Range-Value-at-Risk (RVaR) to evaluate its risk. The closed-form solution to the problem, which includes the results in Chi and Weng (2013) as special cases, was derived. Some comparative studies and sensitivity analysis were also carried out through numerical examples [6]. Rudi Zagst's talk was devoted to Optimal investment strategies for pension funds with regulation-conform dynamic pension payment management in the absence of guarantees. It was considered the post-retirement phase optimization problem for a specific pension product in Germany that comes without guarantees. The continuous-time optimization problem is defined consisting of two specialties: first, we have a product-specific pension adjustment mechanism based on a certain capital coverage ratio which stipulates compulsory pension adjustments if the pension fund is underfunded or significantly overfunded; second, due to the retiree's fear of and aversion against pension reductions, it was introduced a total wealth distribution to an investment portfolio and a buffer portfolio to lower the probability of future potential pension shortenings. A numerical case study on optimization and simulation completed the work with highlighting the benefits of the proposed model [1,2].

The talks from the **finance** side were presented by Alex Badescu, Dominik de Witte, Luca Lalor, Jinniao Qiu, Ana Karen Contreras, Matthias Scherer and Yang Yang. Alex Badescu's talk studied the impact of long memory on modelling asset returns and pricing options in discrete time. A general pricing framework based on affine multi-component volatility models that admit $ARCH(+\infty)$ representations, which not only nests a plethora of option pricing models from the literature, but also allows for the introduction of novel fractionally integrated processes for valuation purposes was proposed. An extensive empirical analysis which includes single and joint calibrations of a variety of short and long memory models to historical returns and $S\&P500$ options was carried out [11,12]. Dominik de Witte's talk had shown how to practically obtain the optimal portfolio withrespect to expected power utility maximization if the price processes are drivenby a geometric Levy process. Based on a result by Nutz (2012), who showedthat the optimal portfolio weights are constant in this setting even if we includebudget constraints, we make use of the Stochastic Gradient Descent Algorithmto arrive at the optimal solution even for a high-dimensional asset universe. The main subject of Luca Lalor's talk was to introduce an algorithmic and High-Frequency Trading model where the price process is of the jump-diffusion type. Previous research modelled the jumps through a diffusion approximation, while here the jumps are modelled directly. Preliminary results, using an Implicit-Explicit Finite Difference Scheme, for an Optimal Acquisition algorithmic trading problem have been presented. Here the jump part of the Jump-Diffusion price process was a function of a Poisson process. The future models will account for the non-Markovian property seen in LOB data [7]. Jinniao Qiu talked about the rough volatility. Rough volatility is a new paradigm in finance. He discussed about the option pricing problems for rough volatility models. As the framework is non-Markovian, the value function for a European option is not deterministic; rather, it is random and satisfies a backward stochastic partial differential equation (BSPDE) or so-called stochastic Black-Scholes equation. The well-posedness of such kinds of BSPDEs will be discussed. These BSPDEs are also used to approximate American option prices. Moreover, a deep learning-based method was investigated for the numerical approximations to such BSPDEs and associated non-Markovian pricing problems. Examples were presented for both European and American options. This talk was based on joint work with Christian Bayer and Yao Yao [13]. In Ana Karen Contreras' talk, review solutions for optimal control problems such as optimal acquisition, optimal liquidation, and market making, which are considered primary purposes in trading activity, were presented. Diffusion processes model the price. The optimal solution using the Semi-Markov process versus the general compound Hawkes process in the limit order market was computed. Comparison the optimal solutions expressed in terms of parameters describing the arrival rates and the mid-price and find a more general solution that more authentically describes the observed

price process in HFT markets. In addition, the comparison was explicit with real-world data from LOBster data [8]. Matthias Scherer's talk considered a vector of bankruptcy times with Marshall-Olkin multivariate exponential distribution implies a simple, yet reasonable, continuous-time dynamic model for dependent credit-risky assets with an appealing trade-off between tractability and realism. Within this framework the maximization of expected power utility of terminal wealth requires the maximization of a concave function on a polygon, a numerical problem whose complexity grows exponentially in the number of considered assets. It was demonstrated how to solve this seemingly impractical numerical problem reliably and efficiently in order to prepare the model for practical use cases. Joint work with J-F. Mai and A. Blagoeva [1,2,21]. Yang Yang's talk investigated some stochastic path-dependent volatility (SPDV) models as well as their performance in stock and energy markets. First the SPDV models was shown to capture the rich spot-volatility dynamics in various financial markets. Then it was moved on to a specific kind of SPDV model and discuss its applications in the natural gas market. Both calibration and pricing problems were discussed. The talk is based on joint work with Jinniao Qiu and Antony Ware [9].

Talks from **energy markets** were presented by Tylar Jia, Ruediger Kiesel, Deniz Sezer and Zuming Sun. Tylar Jia's talk presented a methodology to incorporate large-scale atmospheric information into short-term wind speed forecast in Alberta using two publicly accessible datasets. The first dataset was used for atmospheric clustering by applying the k-means algorithm and the hidden Markov model on atmospheric variables related to wind speed. The second dataset was used to test the proposed time series regime-switching models and mixture models that integrate the clustering results to predict 6-hour ahead wind speed at 23 weather stations in Alberta, Canada [5]. Ruediger Kiesel's talk was devoted to managing climate risk. By now it is widely agreed that climate change poses a substantial risk to financial markets and institutions. He discussed risk management strategies in this context and advocate the use of a pre-commitment approach. Utilizing our general framework, the speaker turned to several specific examples relating to the measurements of risks for insurance and financial companies. The presenter also developed a method to assess the credibility of net-zero commitments, which may be applied to control the carbon budget in loan and investment portfolios. The talk relied on joint work with Gerhard Stahl (HDI), Andrej Bajic (Deloitte FSI-Audit-Garage), Alexander Blasberg and Kateryna Chekriy (both University Duisburg-Essen [16, 17]). Deniz Sezer's talk, was devoted to an overview of various methodologies used by my research team for meso- and sub-meso scale problems related to wind power. Meso-scale research focuses on probabilistic modeling of wind power produced at wind farms distributed across a large geographical region such as a province to make inferences about the aggregate wind power and applications to power system planning and management. Sub-meso-scale modeling involves resolving finer features of the wind flow that can affect wind production in nearby farms. Alberta specific wind data was considered in this talk [5]. Zuming Sun's talk was devoted to the Polynomial Process and Polynomial Regression Model for French Electricity Prices. Empirical experience reveals that electricity prices can be affected by the dynamics of residual demand, power generation capacity of each commodity and spot prices of each generation technology. A model involving a polynomial map of polynomial processes, a so called PMPP model, for electricity prices was presented. The work was focused on the French electricity market. Joint talk with Tony Ware and Thomas Deschatre (EDF Lab, Paris-Saclay, France [10]). Statistical properties of big data were discussed and presented in Florian Brueck, Aleksey Min and Ralf Werner's talks. Florian Brueck investigated the influence of parameter estimation on the asymptotic distribution of the two-sample test based on the Maximum-Mean-Discrepancy proposed by Gretton et al and show that its asymptotic distribution is influenced by parameter estimation. To circumvent the problem of determining covariances of an infinite sum of Chi-squared distributed random variables and the parameter estimators, we propose a new two-sample test based on Maximum-Mean-Discrepancy. Moreover, it was deduced a new test for model comparison based on Maximum-Mean-Discrepancy, which is also robust under parameter estimation [21]. Aleksey Min's talk was devoted to Stationary vine copula models for multivariate time series. A time series consists of multiple observations indexed by time. Classical time series models allow for only linear dependence between variables and time points. More recently, several nonlinear time series models based on copulas were proposed. Vine copulas are graphical models for the dependence and can conveniently capture cross-sectional and temporal dependence of multivariate time series. It was also proposed computationally efficient methods for estimation, simulation, prediction, and uncertainty quantification. The theoretical results allow for misspecified models. The talk is based on joint work with Daniel Krüger and Thomas Nagler [18]. Ralf Werner's talk was devoted to asymptotic non-parametric confidence intervals for the uniform estimator in nested simulations. In the European Union, life insurers face severe

challenges when computing the Solvency Capital Requirement (SCR) based on internal models. Mathematically speaking, one has to compute the 99.5% quantile (VaR) of the 1-year loss distribution of basic own funds (BoFs). Therefore, it was suggested a novel non-parametric asymptotic confidence interval for the quantile estimator which significantly improves upon the only existing one (which, albeit, is not specialized to the quantile estimation case). Besides a few illustrating numerical examples and comparison to the existing method, it was also briefly discussed the main ideas for the corresponding proofs if time allows. The last part of the talk was devoted to a discussion of the implications concerning the application of modern machine learning methods. This was a joint work with Maximilian Klein [19].

There exists a need for highly trained professionals such as data specialists who can work with financial and energy markets data and who can model those data. In addition, big data and data analytics have become an essential part of the financial services industry. Thus, the workshop has brought together academics, industry people and graduate students to share their ideas, to share their research experience and to learn more about big data and their modelling. This workshop has also aim to train young talents and included mentorship component in the program. There were 4 people from industry, who shared their industry experience with 9 graduate students, presented at the workshop: Scott Dalton (Ovintiv), Clifford Kitchen (NBC), Kevin Malenfant (Auspicecapital) and Nima Safaian (Cenovus). It was performed during the **industry panel** of the workshop.

The social program included two walks to Tunnel Mountain and along the Bow River.

4 Scientific Progress Made

The workshop have brought together academics, industry people and graduate students to share their ideas, to share their research experience and to learn more about big data and their modelling.

This workshop also accomplished its aim to train young talents and included mentorship component and industry panel in the program.

5 Outcome of the Meeting

Several research projects and working papers were created based on the workshop's topics. They were associated with stochastic modelling of insurance, finance and energy markets real data. Below are just two out of many testimonies from the workshop's participants:

"Dear professor, I wanted to thank you for giving me the opportunity to join the BIRS workshop. It was insightful and inspiring. The workshop was very well organized and it was a very good learning experience for me. Thank you for giving me the opportunity to participate."

"Thank you so much for your invitation to this wonderful workshop! The workshop gave me some new ideas for my research and a lot of motivation!"

References

[1] Glau, K.; Linders, D.; Min, A.; Scherer, M.; Schneider, L.; Zagst, R. Innovations in Insurance, Risk- and Asset Management, World Scientific, 450 pages

[2] Glau, K.; Grbac, Z. Scherer, M.; Zagst, R. (Eds.) Innovations in Derivatives Markets Springer, 449 pages

[3] Jiang, Wenjun. "Pareto-optimal insurance under heterogeneous beliefs and incentive compatibility". Scandinavian Actuarial Journal 2022.9 (2022): 775-793. Print.

[4] Liang, Zhihang, Zou, Jushen and Jiang, Wenjun. "Revisiting the optimal insurance design under adverse selection: distortion risk measures and tail-risk overestimation". Insurance: Mathematics and Economics 104. (2022): 200-221. Print.

[5] Tylar Jia: Short-term wind speed forecasting with regime-switching and mixture models at multiple weather stations. Lunch at the Lab finance seminar talk, Oct 4, 2022.

[6] Ye Wang: Optimal Reinsurance Under Vajda Condition and Range-Value-At-Risk. Lunch at the Lab finance seminar talk, Mar 22, 2022.

[7] Luca Lalor: A Numerical Solutions to an Algorithmic and HFT Problems with a Jump-Diffusion Price Processes. Lunch at the Lab finance seminar talk, Apr 4, 2022.

[8] Ana Karen Roldan Contreras: Stochastic Optimal Control Problems in Limit Order Books for Semi-Markov Process vs General Compound Hawkes Process. Lunch at the Lab finance seminar talk, Jan 24, 2023.

[9] Yang Yang: Stochastic Path-Dependent Models and Analysis. Lunch at the Lab finance seminar talk, Oct 26, 2021.

[10] Zuming S: Polynomial Process and Polynomial Regression Model for French Electricity Prices. Lunch at the Lab finance seminar talk, Mar 1, 2022.

[11] Augustyniak, Maciej, Badescu, Alexandru and Bégin, Jean-François. "A Discrete-Time Hedging Framework with Multiple Factors and Fat Tails: On What Matters". Journal of Econometrics 232.2 (2023): 416-444. Print.

[12] Augustyniak, Maciej, Badescu, Alexandru and Boudreault, Mathieu. "On the measurement of hedging effectiveness for long term investment guarantees". Journal of Risk and Financial Management (to appear), Print.

[13] Qiu, Jinniao and Zhang, Jing. "Stochastic Differential Games with Random Coefficients and Stochastic Hamilton-Jacobi-Bellman-Isaacs Equations". Annals of Applied Probability 33.2 (2023): 689–730. Print.

[14] Scherer, M. A Multivariate Claim Count Model for Applications in Insurance. Springer, 2018.

[15] Mai, J.-F.; Scherer, M. Simulating Copulas. Stochastic Models, Sampling Algorithms, and Applications. World Scientific, 2017.

[16] Blasberg, A.; Kiesel, R.: Climate Risk in Structural Credit Models, 2023.

[17] Kremer, M.; Kiesel, R.; Paraschiv, F.: An econometric model for intraday electricity trading. In: Philosophical Transactions of the Royal Society A, Jg. 379 (2021) Nr. 2022. doi:10.1098/rsta.2019.0624

[18] Czado, C., Bax, K., Sahin, Ö., Nagler, T., Min, A. and Paterlini, S.: Vine copula based dependence modeling in sustainable finance. The Journal of Finance and Data Science, 2022.

[19] P Groetzner, R Werner:Multiobjective optimization under uncertainty: A multiobjective robust (relative) regret approach, European Journal of Operational Research 296 (1), 101-115

[20] Ariel Neufeld & Marcel Nutz, 2015. "Robust Utility Maximization with Lévy Processes," Papers 1502.05920, arXiv.org, revised Mar 2016.

[21] Brück F., Mai J. and Scherer M.: Exchangeable min-id sequences: Characterization, exponent measures and non-decreasing id-processes. Extremes, 2022.