Causal Inference For Portfolio Optimization And Trading

Candidate: Georgii Nigmatulin
georgii.nigm@gmail.com

ABSTRACT

In the fast-paced financial world, traditional tools, while useful for asset allocation and risk control, often fall short in determining cause-and-effect relationships, especially when influenced by macroeconomic confounding factors, spurious correlations, and non-linearity. Causal inference has emerged as a powerful and promising solution for this challenge. This research proposal presents a shift from conventional correlation analyses to causal inference techniques for analyzing financial time-series data. The approach employs causality methods to elucidate intricate market relationships, emphasizing the potential to predict price shifts, optimize portfolios, and simulate responses to economic changes. The proposal aims to: uncover directional relationships between financial instruments and macro-economic variables; identify lead-lag interdependencies for control and pairs trading; transition from covariance matrices in portfolio optimization methods to causal matrices; enable counterfactual analysis of potential market scenarios; enhance robustness and insightful of asset allocation decisions; and integrate causal insights with cutting-edge machine learning paradigms, notably Causal Reinforcement Learning. This study holds the potential to redefine investment strategies and risk management by integrating causal inference and artificial intelligence in financial analytics.

***K*eywords** Causal inference · Correlation · Machine learning · Financial engineering · Macro-economic indicators Directed Acyclic Graphs · Lead-lag relationships · Portfolio optimization · Fed hikes rates · Counterfactual analysis

About author

Georgii Nigmatulin is a Machine Learning Research Engineer at Samsung R&D Institute, focusing on AI features for smartwatches using sensor time-series signals. He earned MSs in Applied Mathematics and Physics from the Moscow Institute of Physics and Technology in 2022. Georgii's interests include applying machine (deep) learning, causal and bayesian methods, reinforcement learning to problems in the field of quantitative finance, particularly to portfolio optimization, market dependency structure, factor investing and predictive analytics. Research experience, which he currently engaged in at Samsung, concerns proceeding signals by ML algorithms, dependency structure, end-to-end development of new features. In academic research, he has worked in the fields of Anomaly detection, Uncertainty quantification, Deep Learning and Active learning in the Skoltech-Sberbank joint laboratory. Previously, at a leading regional bank, he was engaged in the predictive modelling of withdrawals from ATM networks and created a discrete optimization system that controls cash management. As a result of the work, a total of 8 articles were published and 2 patents were registered. Now, undertaking a PhD is an important step for Georgii's professional evolution. With access to esteemed supervisors, academic resources, datasets, and a hub of innovative discussions, the proposed research has the potential to significantly advance the field of machine learning and financial engineering science.

Personal website: https://george-nigm.github.io//
Repository website: https://github.com/george-nigm/causal-portfolio-and-trading

1 Introduction

In the finance domain, understanding the market relationships is vital for trading and portfolio management. While traditional tools, such as partial correlations and covariance matrices offer insights into asset connections, they only highlight associations without clarifying causality [1]. External economic influences can distort these relationships [2]. Moreover, these conventional methods mainly detect linear relationships in stationary data, missing non-linear dynamics, leading to sub-optimal investment strategies [3]. Empirical correlation matrices are limited because they rely solely on observed data, lacking an economic theory. To enhance covariance matrix estimations, researchers use denoising, detoning, shrinkage and hierarchical asset relationship methods. Causal inference surpasses correlation analysis, revealing directional relationships despite confounders and economic shifts [4]. Unlike traditional linear models, it identifies non-linear dependencies, offering profound implications for portfolio optimization. By replacing Modern Portfolio Theory's covariance matrices with causal matrices, portfolios can become more efficient and robust against market volatility [5]. This approach paves the way for innovative trading techniques, utilizing the lead-lag relationships between assets to forecast, control and clustering [6]. Furthermore, causal estimates can provide a platform for counterfactual analysis, allowing the simulation of the effects of various scenarios on stocks and portfolios, such as federal rate adjustments [5]. Moreover, the endeavour of intertwining causal insights with cutting-edge machine learning methods, like Causal Reinforcement Learning, unveils the vast potential in synergizing causal inference and artificial intelligence within the financial domain [7].

The objective of this study is to explore and apply causal inference techniques within the context of financial time-series data. Specifically, research aims to:

1. Determine directional links between financial instruments and macro-economic indicators.
2. Enhance portfolio optimization by transitioning from traditional covariance matrices to causal matrices.
3. Identify and exploit lead-lag relationships among equities for advanced trading algorithms.
4. Model counterfactual responses of equities to economic shifts, like federal rate adjustments.
5. Strengthen portfolio optimization process in the face of economic shifts.
6. Merge the causal framework with cutting-edge machine learning, especially Causal Reinforcement Learning, expanding AI's role in financial analytics.

The results of this work are expected to enhance portfolio optimization processes, identify lead-lag relationships among stocks, facilitate counterfactual analysis, and lay the foundation for advanced machine learning models that incorporate causal insights.

2 Related papers

The causal inference in financial data has emerged as a pivotal research domain. Wang employed linear non-Gaussian structural equation modelling (SEM) to assess risk-return relationships [8]. By integrating Vector Autoregression (VAR) with SEM, Gao identified the impact of the federal funds rate on US macroeconomic indicators [9]. Yang applied a Bayesian network-based method to study interrelationships among 56 US financial factors [10]. Zaremba and Aste evaluated measures such as linear, regularized, kernel generalizations of Granger causality, and transfer entropy within a financial time series, emphasizing the hurdles in grasping nonlinear causality [11]. Ghosh introduced a machine and deep learning framework tailored for causal predicting Asian stock indices [12]. Nauta applied Attention-Based Convolutional Neural Networks to data simulated by the Fama-French Three-Factor Model [13]. Papana explored time-varying properties in financial networks by using the conditional Granger causality index (CGCI) and partial mutual information on mixed embedding (PMIME) [1]. Haigh employed Directed Acyclic Graphs (DAGs) in conjunction with Error Correction Models (ECMs) to examine how transportation rates influence commodity markets [14]. Shi utilized evolving window algorithms to detect change points in causal relationships, particularly examining the money-income link in the US [15]. Zema fused the Directed Acyclic Graph based-Information Shares (DAG-IS) technique with structural vector error correction models (SVECM) for an in-depth look at IBM intraday data [3]. Scaramozzino harnessed transfer entropy to ascertain links between tech sentiments and stock valuations [16]. Finally, Kim proposed a counterfactual mean-variance optimization method tailored for portfolio construction, introducing a counterfactual analysis responsive to changes in the federal rate [5].

3 Proposed approach

Step-by-step overview. Figure 1 illustrates the proposed method. Using this approach, Causal Inference examines the links between stocks and macroeconomic indicators. Causal Discovery detects potential relationships, highlights lead-lag dynamics, and identifies confounders via statistical tests, visualized in Directed Acyclic Graphs (DAGs). At the stage of causal evaluation, the influences between the nodes are quantified, which are represented by coefficients for the Structural Equation Model (SEM). These coefficients form adjacency matrices at different time lags when intersecting with each other, offering an alternative to covariance matrices for many portfolio optimization models, such as Mean-Variance Portfolio Theory, Hierarchical Risk Parity and Black-Litterman. This approach enables Counterfactual Analysis to simulate scenarios and estimate causal impacts, enhancing pairs trading and portfolio strategies. Portfolio performance is evaluated using metrics like Sharpe, Sortino, Maximum Drawdown, and Alpha from backtesting. This compares strategies driven by cause with covariance-based ones using historical data.

Methods. Various methods have been used to study causal relationships in multivariate time-series data. Granger Causality, foundational in time-series, uses statistical tests to determine predictive relationships between time series, evolving from bivariate to advanced techniques like Vector Autoregressive (VAR) models, regularization, kernels and deep learning. Constraint-based methods, such as the Sprites-Glymour-Scheines (SGS), Peter-Clark (PC) algorithm, and Peter-Clark Momentary Conditional Independence (PCMCI), rely on tests of conditional independence. Score-based methods, e.g., on-combinatorial Optimization via Trace Exponential and Augmented Lagrangian for Structure Learning (NOTEARS) and Greedy Equivalence Search (GES), evaluate models using Bayesian or information-theoretic metrics. Functional Causal Model-Based (FCM) strategies like Linear Non-Gaussian Acyclic Model (LiNGAM), VAR-LiNGAM, and those based on Additive Noise Models (ANM) use Independent Component Analysis (ICA) to deduce causal ties. Each offers specific benefits tailored to data types and research queries, equipping researchers to dissect data's causal relationships. [17]

Counterfactual analysis and limitations. Counterfactual analysis uses tools like Potential Outcomes Framework, do-calculus, Propensity Score Matching, Difference-in-Differences, and Instrumental Variables to assess hypothetical scenarios for pairs trading and portfolio optimization. However, it has limitations such as confounding factors, limited validity, and method assumptions. Financial analytics faces non-stationarity, non-linearity, high dimensionality, noise, and market disturbances, which can be addressed with meticulous methods and advanced statistics. [18].

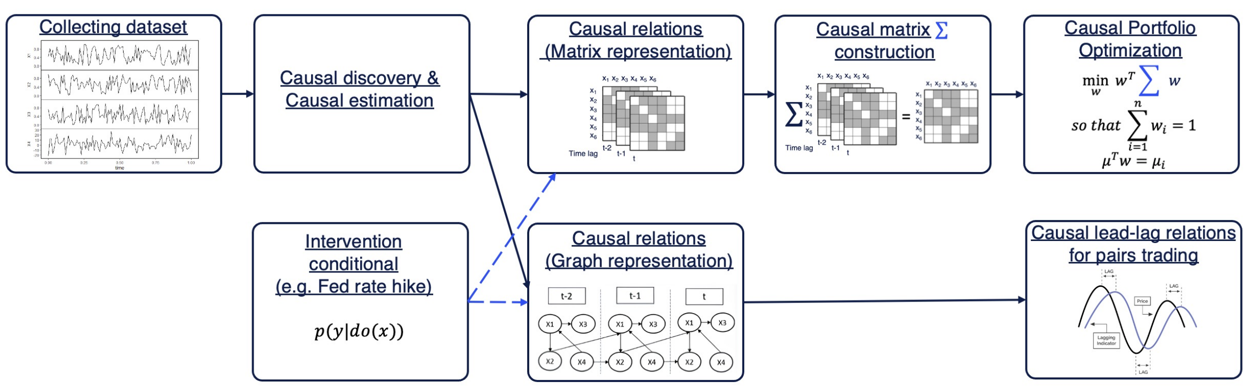

Figure 1: Causality usage diagram for portfolio optimization and trading.

4 Conclusion

Financial markets, known for their complexity and dynamism, require methodologies that surpass traditional correlation analyses. This proposal emphasizes the need to transition from associations to causations in financial analysis, suggesting causal inference as a pivotal tool. Combining causal inference with time-series analytics can provide deeper market insights, optimizing investment strategies such as pairs trading through lead-lag relationships. Causality instead of covariance can revolutionize financial analysis, portfolio optimization, and risk management. Counterfactual analysis can forecast equity or portfolio responses to economic changes, such as federal rate changes. Combining causal inference with AI, like Causal Reinforcement Learning, expands possibilities. This synergy may create remarkable innovations, strengthen financial systems and deepen market understanding. Backed by the substantial literature, the proposal outlines strategies to achieve these objectives, arming stakeholders with tools for improved vision and efficiency.

Words count: 1497 words

References

[1] Angeliki Papana, Catherine Kyrtsou, Dimitris Kugiumtzis, and Cees Diks. Financial networks based on granger causality: A case study. *Physica A: Statistical Mechanics and its Applications*, 482:65–73, 2017.

[2] Jakob Runge, Andreas Gerhardus, Gherardo Varando, Veronika Eyring, and Gustau Camps-Valls. Causal inference for time series. *Nature Reviews Earth & Environment*, pages 1–19, 2023.

[3] Sebastiano Michele Zema. Directed acyclic graph based information shares for price discovery. *Journal of Economic Dynamics and Control*, 139:104434, 2022.

[4] Samantha Kleinberg. *Causality, probability, and time*. Cambridge University Press, 2013.

[5] Kwangho Kim, Alan Mishler, and José R Zubizarreta. Counterfactual mean-variance optimization. *arXiv preprint arXiv:2209.09538*, 2022.

[6] Yichi Zhang, Mihai Cucuringu, Alexander Y Shestopaloff, and Stefan Zohren. Robust detection of lead-lag relationships in lagged multi-factor models. *arXiv preprint arXiv:2305.06704*, 2023.

[7] Maximilian Seitzer, Bernhard Schölkopf, and Georg Martius. Causal influence detection for improving efficiency in reinforcement learning. *Advances in Neural Information Processing Systems*, 34:22905–22918, 2021.

[8] Zitian Wang and Shaohua Tan. Automatic linear causal relationship identification for financial factor modeling. *Expert Systems with Applications*, 36(10):12441–12445, 2009.

[9] Zhe Gao, Zitian Wang, Lili Wang, and Shaohua Tan. Linear non-gaussian causal discovery from a composite set of major us macroeconomic factors. *Expert Systems with Applications*, 39(12):10867–10872, 2012.

[10] Jianjun Yang, Yunhai Tong, Xinhai Liu, and Shaohua Tan. Causal inference from financial factors: Continuous variable based local structure learning algorithm. In *2014 IEEE Conference on Computational Intelligence for Financial Engineering Economics (CIFEr)*, pages 278–285, 2014.

[11] Anna Zaremba and Tomaso Aste. Measures of causality in complex datasets with application to financial data. *Entropy*, 16(4):2309–2349, 2014.

[12] Indranil Ghosh, Rabin K Jana, and Manas K Sanyal. Analysis of temporal pattern, causal interaction and predictive modeling of financial markets using nonlinear dynamics, econometric models and machine learning algorithms. *Applied Soft Computing*, 82:105553, 2019.

[13] Meike Nauta, Doina Bucur, and Christin Seifert. Causal discovery with attention-based convolutional neural networks. *Machine Learning and Knowledge Extraction*, 1(1):19, 2019.

[14] Michael S Haigh and David A Bessler. Causality and price discovery: An application of directed acyclic graphs. *The Journal of Business*, 77(4):1099–1121, 2004.

[15] Shuping Shi, Stan Hurn, and Peter CB Phillips. Causal change detection in possibly integrated systems: Revisiting the money–income relationship. *Journal of Financial Econometrics*, 18(1):158–180, 2020.

[16] Roberta Scaramozzino, Paola Cerchiello, and Tomaso Aste. Information theoretic causality detection between financial and sentiment data. *Entropy*, 23(5):621, 2021.

[17] Chang Gong, Di Yao, Chuzhe Zhang, Wenbin Li, and Jingping Bi. Causal discovery from temporal data: An overview and new perspectives. *arXiv preprint arXiv:2303.10112*, 2023.

[18] Ricardo Masini and Marcelo C Medeiros. Counterfactual analysis with artificial controls: Inference, high dimensions, and nonstationarity. *Journal of the American Statistical Association*, 116(536):1773–1788, 2021.