Scalable Nearest-Neighbor Gaussian Graphical Models for Demographic Heterogeneity in GLP-1 Outcomes

Gaoqianxue Liu, Drexel University

Treatment response to GLP-1 receptor agonists varies widely across demographic groups, yet the joint relationships among clinical biomarkers are rarely modeled across strata. Existing joint Gaussian graphical models couple groups through a global prior and become computationally costly as the number of strata grows. We introduce the Nearest-Neighbor Gaussian Graphical Model (NNGGM), which adapts the Vecchia approximation from spatial statistics, linking each demographic group mainly to its most similar neighbors. This substantially reduces computational cost and stabilizes estimation in small and unbalanced subgroups. Applied to a TriNetX cohort of chronic kidney disease patients initiating GLP-1 therapy, the method reveals substantial heterogeneity in biomarker networks and clear local structure across demographic strata. NNGGM offers a scalable and robust framework for studying how clinical dependency structure differs across populations. Although motivated by clinical biomarkers, the method applies broadly in settings where dependency networks are compared across many related groups, such as genomics, neuroimaging, and finance.

Paired Portfolio Trading: A Statistical Approach

Sara Mezuri, Graduate Reasearch Assistant

Traditional pairs trading focuses on selecting a pair of individual stocks. However, in practice, it is difficult to identify pairs that may exhibit similar behavior over an extended period of time. In this work, we instead take a more drastic approach of paired portfolio trading, where, using the statistical techniques, instead of looking for pairs, we carefully construct the paired portfolios. We do so by using the multivariate techniques of canonical correlations and cointegration.

We apply and illustrate our approach on the daily data for ten large-cap U.S. equities from January 2021 to December 2023. We construct a host of paired portfolios from these ten equities, which have a high possibility of cointegrating and hence allow us the potential for profitable paired portfolio trading. Our strategy is almost market neutral in that the net investment of money in the market is minimal. We develop several entry and exit rules and refine our approach by using a variety of modelling techniques, such as orthogonal regression.

We also evaluate the strategy for the future out-of-sample data from January 2024 through December 2025. Our analysis shows that the strategy is profitable for all cointegrated pairs in-sample, achieving a 100% win rate, while also maintaining profitability out-of-sample.

Methodological opportunities in genomic data analysis to advance health equity

Leandra Braeuninger, University College London

The causes and consequences of inequities in genomic research and medicine are complex and widespread. However, it is widely acknowledged that underrepresentation of diverse populations in human genetics research risks exacerbating existing health disparities. Efforts to improve diversity are ongoing, but an often-overlooked source of inequity is the choice of analytical methods used to process, analyse and interpret genomic data. This choice can influence all areas of genomic research, from genome-wide association studies and polygenic score development to variant prioritization and functional genomics. New statistical and machine learning techniques to understand, quantify and correct for the impact of biases in genomic data are emerging within the wider genomic research and genomic medicine ecosystems. At this crucial time point, it is important to clarify where improvements in methods and practices can, or cannot, have a role in improving equity in genomics. Here, we review existing approaches to promote equity and fairness in statistical analysis for genomics, and propose future methodological developments that are likely to yield the most impact for equity.