About the Symposium

The Summer Research Symposium in Mathematics showcases undergraduate research in mathematical modeling, data science, and pure mathematics.

This brochure features extended abstracts from student projects completed as part of two research programs: **CMUSR**, focused on pure and applied mathematics, and **CSUREMM**, focused on mathematical modeling and simulation.

CMUSR (Columbia Math Undergraduate Summer Research) is a ten-week intensive research program for rising juniors and seniors in mathematics and related fields. Students work closely with faculty and graduate mentors on advanced research topics in small teams.

Program Lead: Evan Sorensen
PhD Mentor: Xinyi (Sara) Zhang

Supported by: Department of Mathematics and a research grant from Prof. Ivan Corwin.

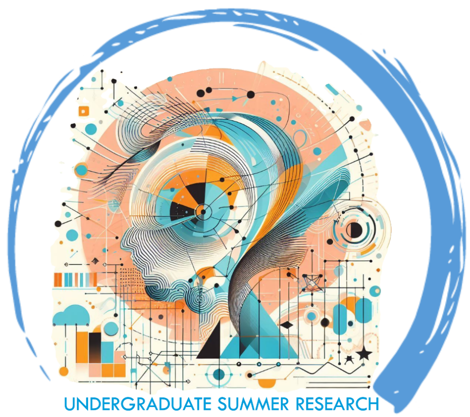
CSUREMM (Columbia Summer Undergraduate Research Experiences in Mathematical Modeling) is a ten-week program that provides hands-on experience in advanced mathematical modeling and interdisciplinary research. Students work in collaborative teams under mentorship to develop and explore original research projects.

Program Lead: George Dragomir
Faculty Adviser: Dobrin Marchev (Statistics)
Mentor: Vihan Pandey

Supported by: Department of Mathematics, Department of Statistics, Irving Institute for Cancer Dynamics, and the Barnard SRI Program.

2025

Undergraduate Summer Research Symposium



Department of Mathematics
Columbia University

Invariance of Polymer Partition Functions Under the Discrete Periodic Pitman Transform

Caio Hermano Maia de Oliveira, Ryan Xu, Jasper Kra-Caskey

This project studies a discrete integrable model of polymers on a vertex-weighted, vertically periodic lattice, where weights repeat with a fixed column period. The model is related to the KPZ universality class and admits exact solvability. We focus on the discrete periodic Pitman transform, a bijective transformation acting on the lattice weights. Using matrix-encoding techniques, we prove that this transformation preserves both point-to-point and multi-path partition functions. As a corollary, we show that the transform satisfies a braid relation analogous to that satisfied by the full-line Pitman transform, highlighting underlying representation-theoretic symmetries. Our results extend recent invariance theorems to the periodic setting and reveal structural invariance in the geometry of polymer paths.

Distributions for the Zero-Temperature Discrete Periodic Pitman Transform

Eva Engel, Ivan Wong, Oleksandr Lazorenko

This project explores the zero-temperature discrete periodic Pitman transform, a piecewise-linear transformation arising in queueing theory and last-passage percolation. Motivated by analogies to Burke's theorem, which asserts the invariance of Poisson input and output processes in certain queues, we prove a corresponding Burke-type property for the periodic transform. Interpreting the output as a last-passage path maximum, we give a combinatorial description of its distribution in terms of the weight configuration. We further establish sufficient conditions for distributional identities to hold, especially in the cases of geometric and exponential weight distributions. These results contribute to a broader understanding of symmetries and limiting behavior in integrable probabilistic systems.

From Headlines to Holdings: Deep Learning for Smarter Portfolio Decisions

Yun Lin, Jiawei Lou, Jinghe Zhang

Recent advances in deep learning have fueled interest in its application to stock portfolio optimization. Traditional approaches predict individual asset returns in isolation and apply mean-variance optimization (MVO), often overlooking inter-asset dependencies and compounding errors. This project proposes an end-to-end deep learning framework that directly learns portfolio weights. The model combines Long Short-Term Memory (LSTM) networks to capture temporal patterns and Graph Attention Networks (GAT) to represent dynamic inter-stock relationships. It also integrates sentiment analysis of financial news to reflect market psychology. Using data from nine major stocks across six sectors, our model outperforms equal-weighted and CAPM-based MVO portfolios in cumulative return and Sharpe ratio. The study highlights the potential of combining deep learning and natural language processing to improve portfolio management.

When Crypto Sneezes: Forecasting U.S. Macroeconomic Indicators Using Cryptocurrency Volatility

Tianyi Shen, Tianqi Wang, Xintong Ye

While prior research shows that crypto markets respond to macroeconomic shocks, the reverse relationship remains underexplored. This project investigates whether volatility in major cryptocurrencies, both stablecoins and non-stablecoins, can help forecast U.S. macroeconomic indicators. We compare SARIMA baseline models to SARIMAX models enhanced with lagged crypto volatility under varying macroeconomic conditions. Model performance is evaluated through out-of-sample forecast accuracy using MAPE and p-values. Results indicate that incorporating lagged signals from non-stablecoins improves forecasting accuracy for several macroeconomic indicators, suggesting that cryptocurrency volatility may offer predictive value for economic analysis.

Structure Over Signal: A Globalized Approach to Multi-relational GNNs

Aruzhan Abil, Amber Li, Juno Marques Oda

Stock price prediction remains challenging due to the non-linear and volatile nature of financial data. While Graph Neural Networks (GNNs) can model inter-stock dependencies, they often overlook structural supply chain shifts caused by macroeconomic shocks. We propose OmniGNN, an attention-based multi-relational GNN that incorporates heterogeneous node and edge types, along with a sector node acting as a global intermediary for rapid shock propagation. The model leverages Graph Attention Networks for adaptive message passing and Transformers to capture temporal dynamics across multiplex relations. Experiments show that OmniGNN outperforms existing models, especially during periods of market disruption such as the COVID-19 pandemic.

Improving S&P 500 Volatility Forecasting Through Regime-Switching Methods

Ava Blake, Nivika Gandhi, Anurag Jakkula

We propose regime-switching models to improve forecasts of S&P 500 volatility by capturing structural changes across time. Using eleven years of SPX data, we compute daily realized volatility from 5-minute intraday returns, adjusted for trading irregularities. Feature engineering incorporates VIX-based lags, short-term reversal factors, realized kurtosis, and jump variation. Our methods include soft Markov switching with EM-weighted regressions, distributional clustering via Mood tests and spectral methods, and a coefficient-based soft clustering algorithm that extracts HAR coefficients and clusters them with Bayesian GMM. Evaluated across pre-, during-, and post-COVID periods, the coefficient-based approach consistently outperforms baselines, highlighting the value of soft clustering and regime-aware forecasting during periods of uncertainty.

Demographic Predictors of Flu Vaccine Uptake Pre-, During, and Post-COVID-19

Jacob Hahn, Olivia Huang, Athena Ke

This project investigates how demographic factors influence influenza vaccine uptake across U.S. counties during three periods: pre-, during, and post-COVID-19. Building on prior research in COVID-19 vaccination behavior, we apply supervised learning to predict flu vaccination rates using political, socioeconomic, and healthcare-related features. Pre-COVID, political affiliation, race, and education were most predictive; during COVID, occupation-related features gained prominence; and post-COVID, socioeconomic indicators and political affiliation were most influential. Urban-rural comparisons reveal shifting predictors in urban areas, while rural counties remained more stable. These findings highlight the evolving influence of occupation and local politics on vaccine behavior and support more targeted health policy strategies.

Supercolonies, Social Structure, and Temporal Population Dynamics in Communal Nesting Birds

Lucas Azenha, Luiz Gabiatti Backes, Pranav Suresh

We study large-scale cooperation in a population of cooperatively breeding birds using over 6 million RFID detections spanning five years. Persistent dyadic interactions form the basis of dynamic social networks, revealing 108 emergent *supercolonies* – robust social structures defined by long-term inter-colony cooperation. Male-female dyads drive long-range connectivity, while female-female ties support within-colony cohesion. These structures exhibit replicable forms such as hubs and chains, offering insight into spatially distributed social organization. Our approach provides a generalizable framework for analyzing decentralized cooperation in dynamic ecological systems.