Discovering Algebraic Group Structures such as $\mathbb{Z}_4 \times \mathbb{Z}_2$ in Stock Market Time Series Data via Information-Theoretic Methods

Modulo Four, NBE

January 18, 2025

Abstract

This white paper details a procedure for discovering specific information-theoretical algebraic groups in financial time series data, with a focus on uncovering a structure analogous to the group $\mathbb{Z}_4 \times \mathbb{Z}_2$. We present the methodology in general terms, including threshold-based screening of mutual information (MI) values.

1 Introduction

Financial time series often exhibit intricate dependencies. One way to capture potential relationships among different instruments is via *mutual information* (MI). By computing MI on two fronts—*within* the same time step ($\mathrm{MI}_{\mathrm{within}}$) and across a one-step lag ($\mathrm{MI}_{\mathrm{lagged}}$)—we can observe both synchronous and directional interactions between instruments.

In particular, we seek to isolate small subsets of instruments whose pairwise interactions align with an *algebraic group* structure, specifically a $\mathbb{Z}_4 \times \mathbb{Z}_2$ pattern:

- **Two directed 4-cycles in** $\mathrm{MI}_{\mathrm{lagged}}$, for instance:

$$S_1 \to S_2 \to S_3 \to S_4 \to S_1 \quad \text{and} \quad S_5 \to S_6 \to S_7 \to S_8 \to S_5.$$

- **Four 2-cycles in** $\mathrm{MI}_{\mathrm{within}}$ that *connect the two 4-cycles in order*. Specifically, each stock in the first 4-cycle has a corresponding stock in the second 4-cycle, linked by a 2-cycle:

$$S_1 \leftrightarrow S_5, \quad S_2 \leftrightarrow S_6, \quad S_3 \leftrightarrow S_7, \quad S_4 \leftrightarrow S_8.$$

The presence of such an 8-instrument arrangement hints at a deeper organizational feature akin to a group of symmetries, but in an *information-theoretic* sense.

2 Method Overview

2.1 Data Preprocessing

We begin with a dataset of stock market time series, each column representing one instrument's daily (or periodic) price. The procedure is:

1. Sort all rows by ascending date.

2. Convert each instrument's price vector into a discrete movement vector in $\{0, 1\}$, with "1" for upward movement from one day to the next, and "0" otherwise.

This yields binary sequences for each instrument over a common time horizon.

2.2 Computing Mutual Information

We then compute two matrices of pairwise MI:

- $\mathrm{MI}_{\text{within}}(X, Y)$: Both sequences X and Y are aligned to the same time step indices.
- $\mathrm{MI}_{\text{lagged}}(X, Y)$: Sequence X is shifted by one time step relative to Y, thus capturing potential directional relationships.

Given two binary sequences X and Y of length n,

$$\mathrm{MI}(X, Y) \;=\; \sum_{x=0,1} \sum_{y=0,1} p(x, y) \, \log\left[\tfrac{p(x,y)}{p(x)\,p(y)}\right],$$

where $p(\cdot)$ and $p(\cdot, \cdot)$ are empirical frequencies over the observed data.

2.3 Information-Theoretical $\mathbb{Z}_4 \times \mathbb{Z}_2$ Groups

The crux of our pipeline is a systematic search for 8-instrument combinations satisfying a structure analogous to $\mathbb{Z}_4 \times \mathbb{Z}_2$. Specifically, each valid combination:

1. *Lagged 4-cycles*:

$$S_1 \to S_2 \to S_3 \to S_4 \to S_1, \quad S_5 \to S_6 \to S_7 \to S_8 \to S_5,$$

   wherein every directed edge has $\mathrm{MI}_{\text{lagged}}$ above a certain **lagged threshold**.

2. *Within 2-cycles*:

$$(S_1 \leftrightarrow S_5),\; (S_2 \leftrightarrow S_6),\; (S_3 \leftrightarrow S_7),\; (S_4 \leftrightarrow S_8),$$

   with each direction's $\mathrm{MI}_{\text{within}}$ exceeding a chosen **within threshold**.

We interpret these cycles collectively as an "algebraic group" in the sense that the 4-cycles correspond to a cyclic factor (akin to $\mathbb{Z}_4$), and the 2-cycles pair up instruments in a manner reminiscent of a $\mathbb{Z}_2$ factor.

2.4 Thresholds and Screening

To decide whether a given edge in $\mathrm{MI}_{\text{within}}$ or $\mathrm{MI}_{\text{lagged}}$ is "strong enough," we apply *thresholds*:

- $\text{threshold}_{\text{lagged}}$ for all directed edges in the 4-cycles.
- $\text{threshold}_{\text{within}}$ for both directions in each 2-cycle pair.

Only if an edge's mutual information *exceeds* these thresholds do we consider it present in the group structure. By adjusting the thresholds, we can tune the stringency of group detection. If the thresholds are too high, fewer (or no) groups appear; if too low, the structure might become trivial or abundant.

2.5 Implementation Outline

1. **Parse Data**: Load CSV data (dates vs. instruments), sorting chronologically.
2. **Preprocess Movements**: Convert each price series into a 0/1 sequence.
3. **Compute MI**: Create two symmetric matrices:
   - $\text{MI}_{\text{within}}$ for same-step pairs.
   - $\text{MI}_{\text{lagged}}$ for pairs offset by one time step.
4. **Generate Combinations**: Enumerate all 8-instrument subsets (or a practical sample if N is large).
5. **Check $\mathbb{Z}_4 \times \mathbb{Z}_2$ Pattern**: For each subset, verify:
   - Two distinct 4-cycles in $\text{MI}_{\text{lagged}}$.
   - Four cross 2-cycles in $\text{MI}_{\text{within}}$.
   - All edges above respective thresholds.
6. **Output**: Save the discovered groups, listing each edge's MI value. If none satisfy the pattern, the result is empty.

Notes on Simplifications

While this paper focuses on a binary (0/1) encoding of price movements ("up" vs. "otherwise") and on identifying a $\mathbb{Z}_4 \times \mathbb{Z}_2$ structure, the framework can be readily extended in two ways:

1. Multi-State Movements. Instead of reducing each daily price change to a 0/1 label, one could use a 3-state (or larger) discrete variable to capture more nuanced movement categories (e.g., "up," "down," "flat"). The mutual information calculation then proceeds analogously over the resulting discrete probability distributions.

2. Different Algebraic Groups. Although we have illustrated the detection of an 8-instrument group with two 4-cycles (lagged MI) and four cross 2-cycles (within MI), one could target alternative group structures by introducing additional or different "generators." For example, one could look for a $\mathbb{Z}_m \times \mathbb{Z}_n$ pattern by defining more or differently lagged MIs. The core procedure of threshold-based screening across various MI matrices remains the same, but would allow for richer symmetry and group-theoretic patterns.

3. Application to Other Time-Series Domains. The same information-theoretic algebraic approach applies to any time-series data beyond finance. Whether analyzing operational metrics within a business, or other categories of sequential data (e.g. biological signals, sensor readings, etc.), one can compute within-step and lagged mutual information, define suitable thresholds or additional "generators" of interaction, and detect analogous group structures following the same procedure.

3 Conclusion

In this white paper, we have demonstrated how *information-theoretical algebraic groups* of type $\mathbb{Z}_4 \times \mathbb{Z}_2$ can be unveiled in stock market time series data through a threshold-based analysis of both within-time-step and lagged mutual information (MI).

More broadly, our approach involves:

1. **Discretizing time-series data** to capture relevant features (such as binary movement indicators or higher-order states).

2. **Computing mutual information** in multiple configurations (e.g. same-step vs. lagged) to quantify both synchronous and directional relationships.

3. **Searching for group-like patterns**, such as the $\mathbb{Z}_4 \times \mathbb{Z}_2$ structure, by imposing algebraic constraints on how time series must interconnect via MI.

4. **Applying thresholds** to ensure that only sufficiently strong interactions qualify as edges within these prospective group structures.

These techniques need not be limited to financial data; they can be employed wherever discrete time-series observations are available and one wishes to identify structured interaction patterns—whether in biology, operational metrics, or other domains. By leveraging the powerful framework of mutual information and linking it to algebraic group constructs, researchers can gain a novel lens into complex, multi-way dependencies within evolving systems.