

Motif Discovery & Repeated Substring Mining in Financial Time Series

A comparative motif analysis of S&P 500 valuation states.

Assigned by:

Adam Kayal 211584271

Nofar Bazak 212148340

Dataset and Key Design Choice

1

Dataset:

S&P 500 historical data

2

Time span:

(1871-2023)

3

Signal analyzed:

Real Price

4

Additional valuation signal:

PE10

sp500

PE10	Real Earnings	Real Dividend	Real Price	Long Interest Rate	Consumer Price Index	Earnings	Dividend	SP500	Date
33.25	0	0	4515.77	4.09	306.13	0	0	4515.77	2023-09-01
32.91	0	0	4459.48	4.17	305.98	0	0	4457.358695652170	2023-08-01
33.38	0	0	4514.51	3.9	305.69	0	0	4508.075500000000	2023-07-01
32.41	181.77	68.94	4359.88	3.75	305.11	181.17	68.71	4345.372857142860	2023-06-01
31.14	180.35	68.99	4173.45	3.57	304.13	179.17	68.54	4146.1731818181800	2023-05-01
31.15	178.78	69	4159.03	3.46	303.36	177.17	68.38	4121.467368421050	2023-04-01
30.27	177.76	69.22	4027.13	3.66	301.67	175.17	68.21	3968.5591304347800	2023-03-01
31.32	177.43	68.97	4151.39	3.75	300.84	174.3633333333300	67.78	4079.684736842110	2023-02-01
30.69	177.59	68.92	4052.77	3.53	299.17	173.55666666666700	67.35	3960.6565	2023-01-01



Why Look for Motifs in Financial Time Series?



Financial time series are **complex**, **noisy**, and **non-stationary**



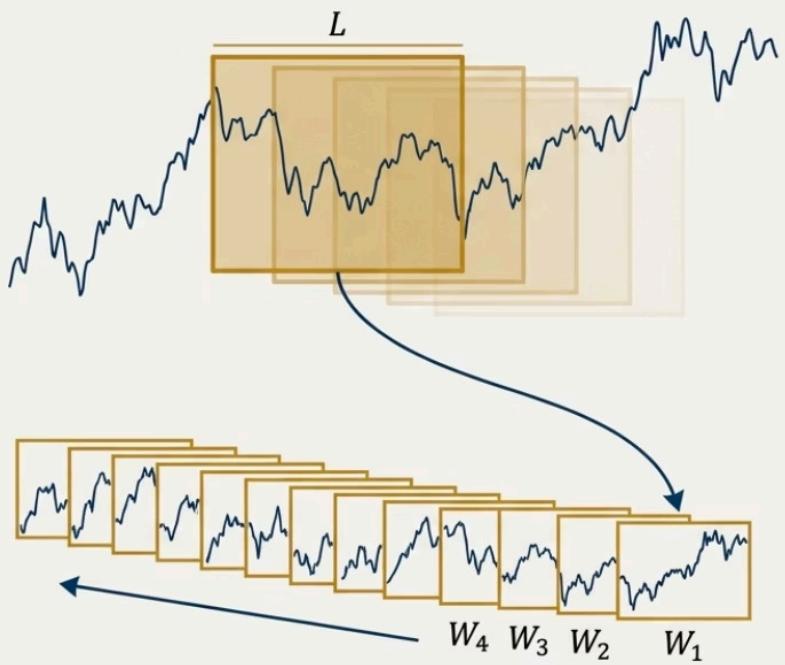
Traditional statistics miss **structural patterns**



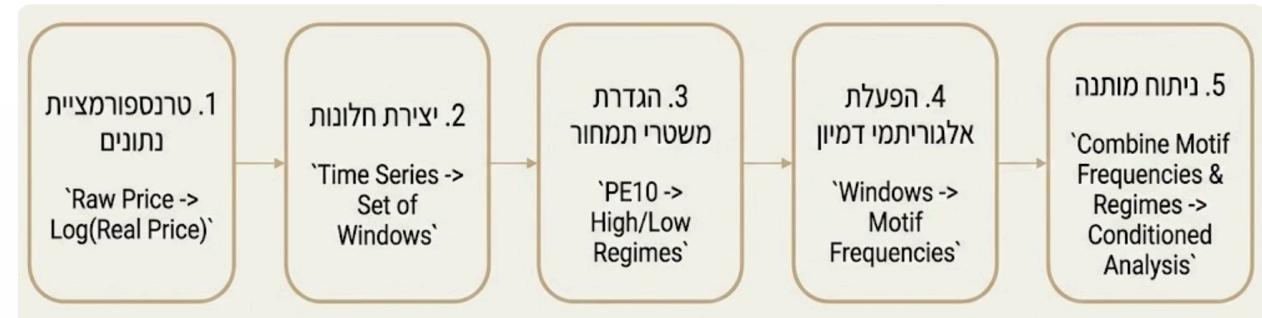
Motif discovery reveals **recurring behaviors**, not exact values



Goal: understand ***structure***, not predict prices



how does it work?



Regime Split: Understanding Market Dynamics



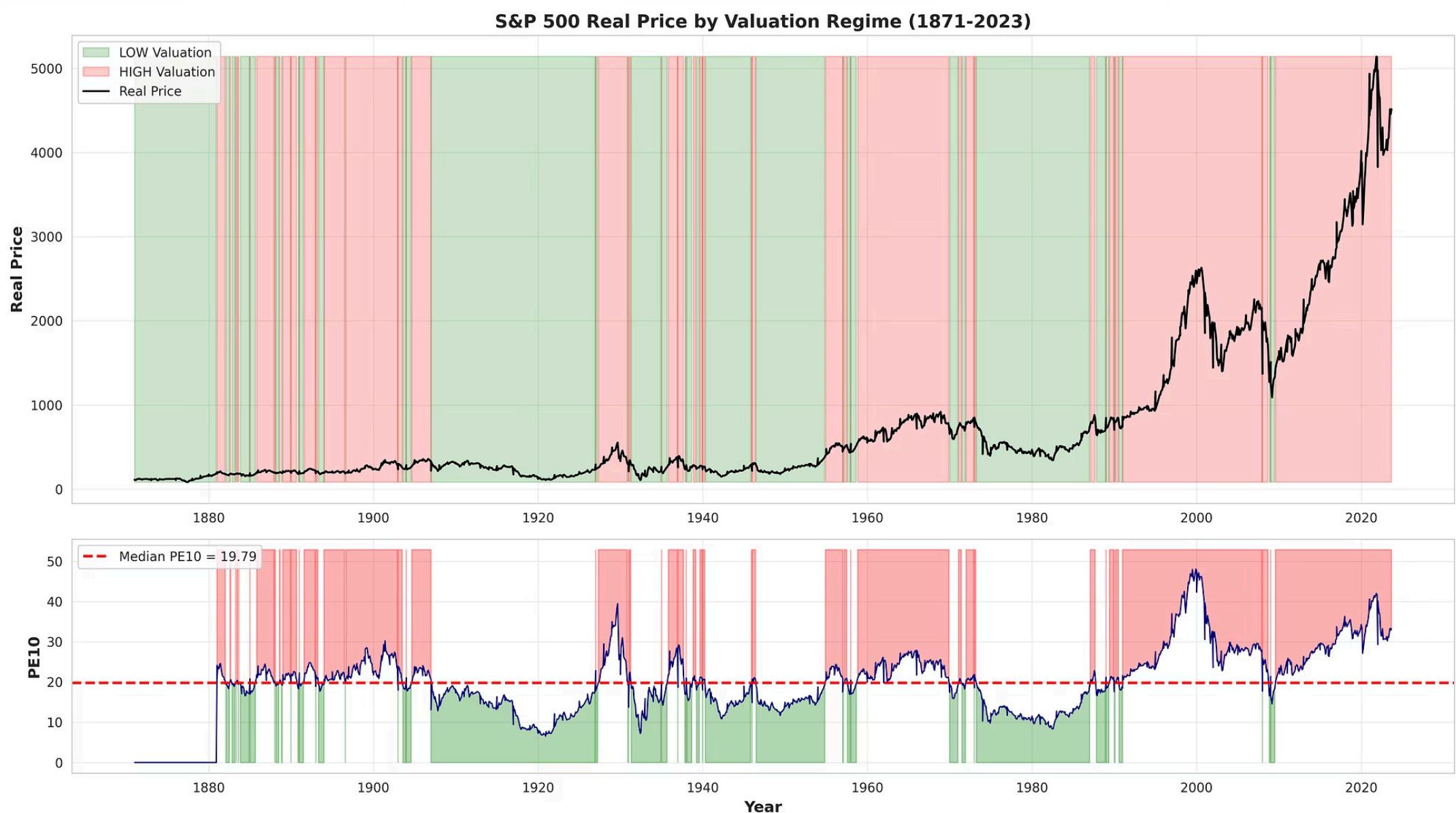
Low Valuation Periods

Characterized by **long and stable** market periods, often preceding periods of growth. These offer consistent, predictable patterns for motif discovery.



High Valuation Periods

Marked by **clustered and volatile** market movements, typically reflecting rapid changes and uncertainty. Motif discovery here tends to be more dynamic and complex.



 [adamkiy.github.io](https://github.com/adamkiy)



OPM & CTM Pattern Discovery – S&P 500

Finds patterns with the same relative movements. Scale-invariant: [10,30,20,40] matches [100,300,200,400]. Perfect for finding "shapes" across different price levels.

Key Research Questions

1

Do LOW and HIGH valuation regimes exhibit different pattern structures?

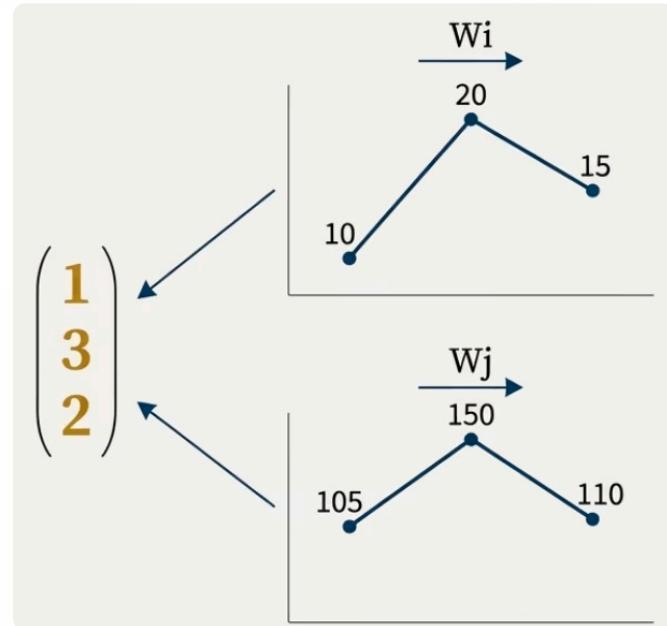
2

Can we identify early warning signals for regime transitions?

OPM vs. CTM: Adapting to Market Complexity

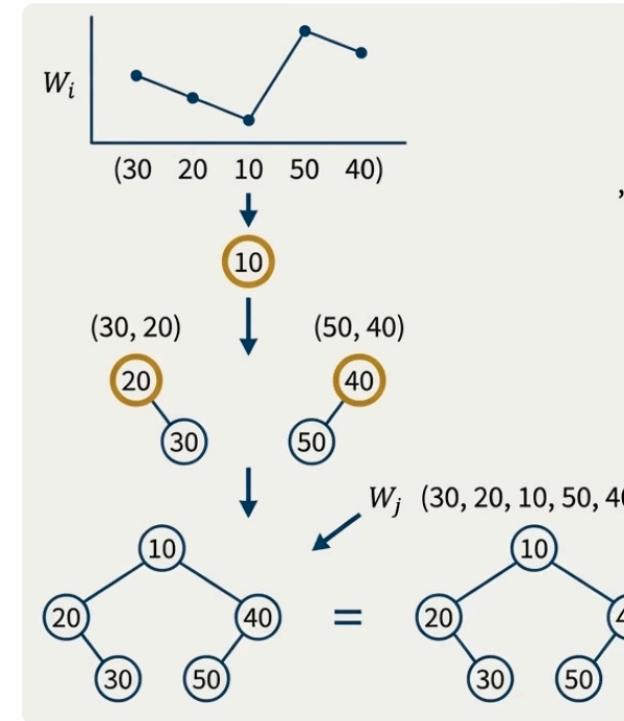
OPM (Directional Motifs)

- Sensitive to **relative price changes**
- Capture **short-term, volatile patterns**
- Frequent and diverse motifs

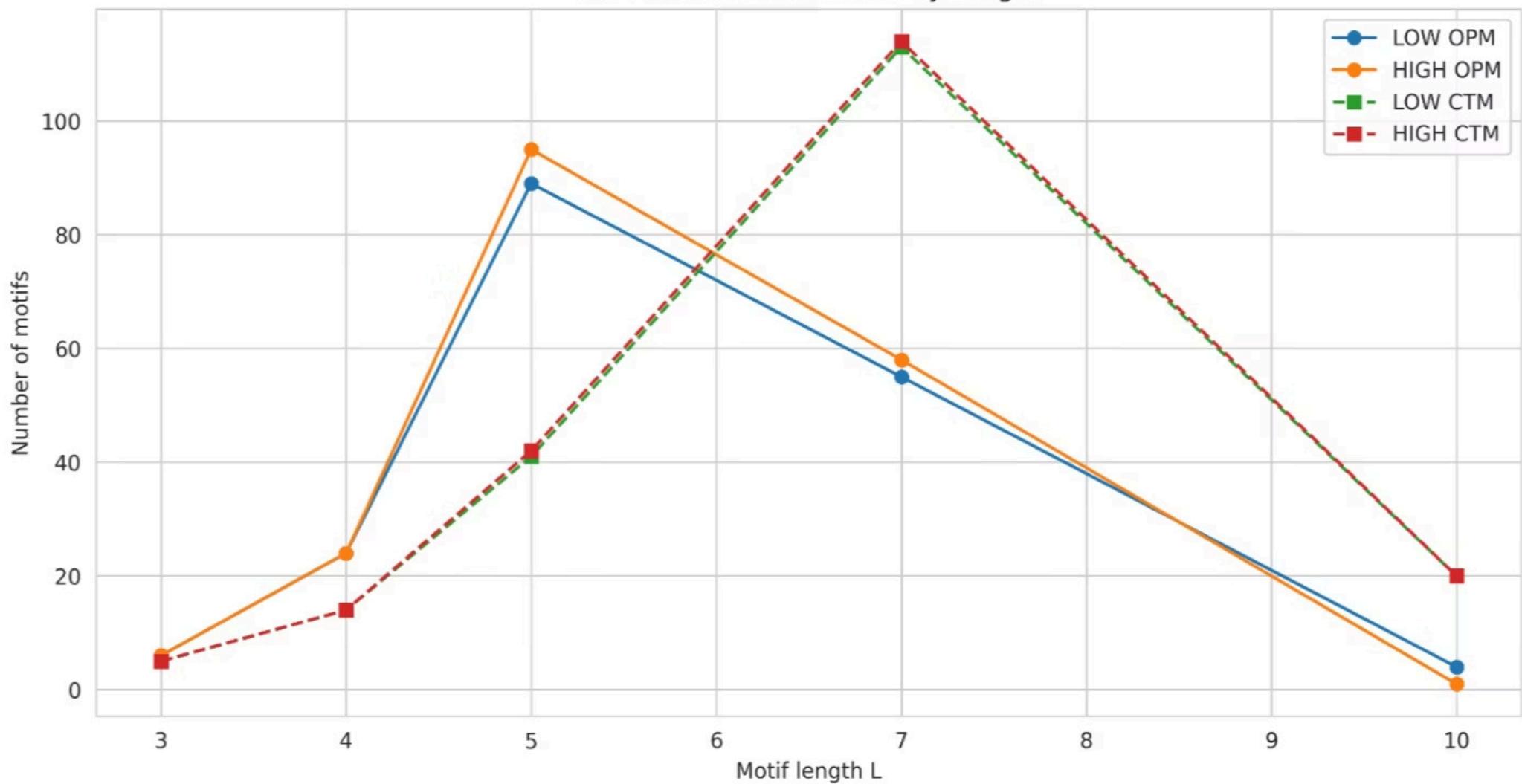


CTM (Structural)

- Focus on **structural similarity**
- Capture **long-term, stable patterns**
- Filter out minor fluctuations



OPM vs CTM Motif Counts by Length



The 2008 Crisis: Regime Collapse in Action

- Motif analysis highlights early warning signs of regime shifts
- Useful for risk management and predicting market instability

1

Pre-Crisis (2003-2007)

- Bull run, high valuations
- Easy credit & housing boom
- Low volatility, consistent upward trends
- Stable market motifs (OPM & CTM)

2

Crisis (2008)

- Subprime mortgage collapse
- Extreme volatility, erratic behavior
- Motif fragmentation: patterns disintegrate



OPM Across Regimes



Low Valuation Regimes

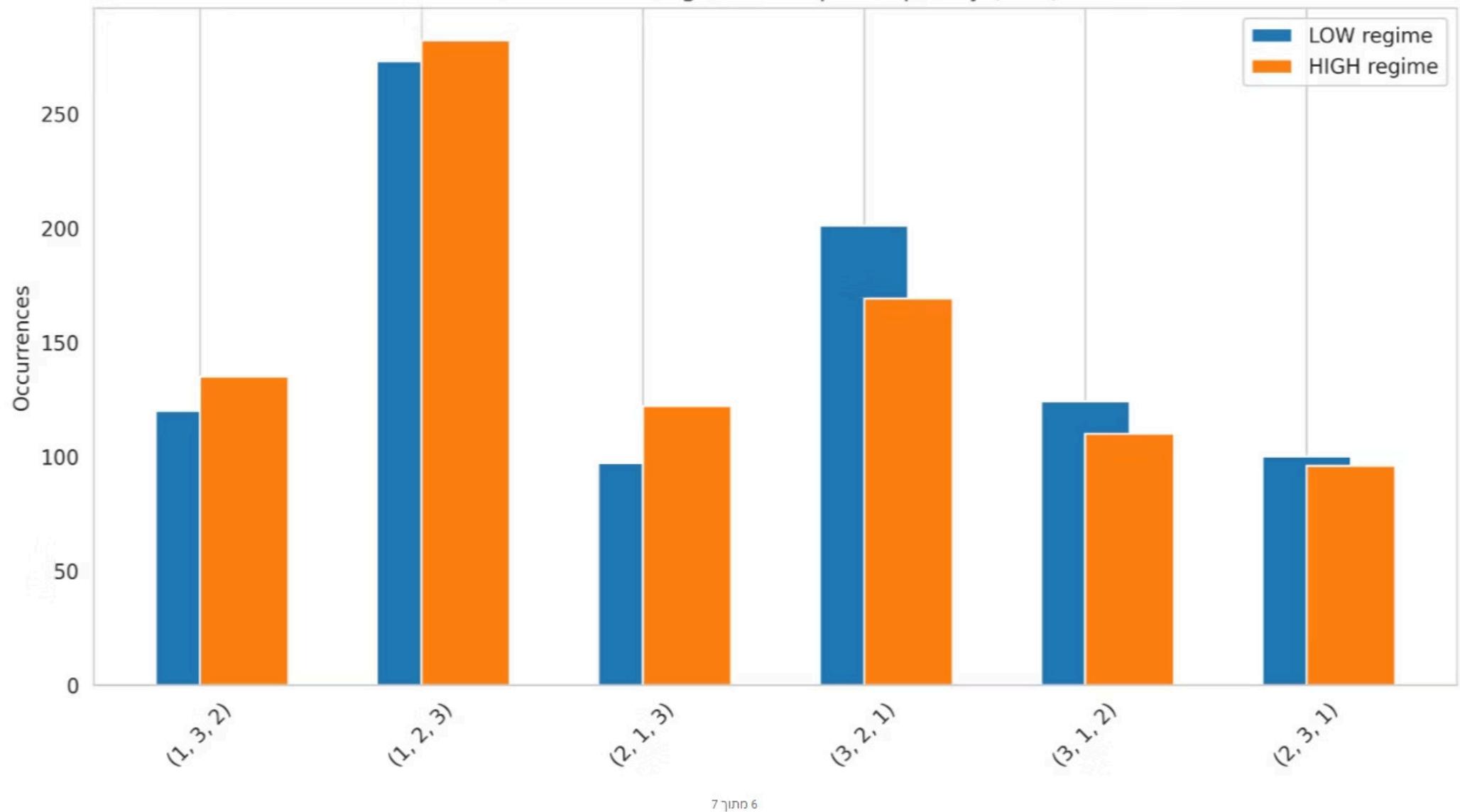
- Greater variety of order patterns
- Rapid pattern turnover
- Complex rank structures

High Valuation Regimes

- Pattern concentration
- Extended pattern duration
- Simpler rank structures

These differences reveal how investor sentiment directly shapes market pattern stability and predictability.

Order-Preserving Motif Shape Frequency (L=3)



CTM Across Regimes



Low Valuation Regimes

- Complex parent-distance signatures
- Frequent structural breaks
- High volatility fingerprints



High Valuation Regimes

- Simple parent-distance signatures
- Extended structural stability
- Predictable hierarchies

These differences reveal how investor sentiment fundamentally shapes market pattern stability and predictability.

Key Insight: The Paradox

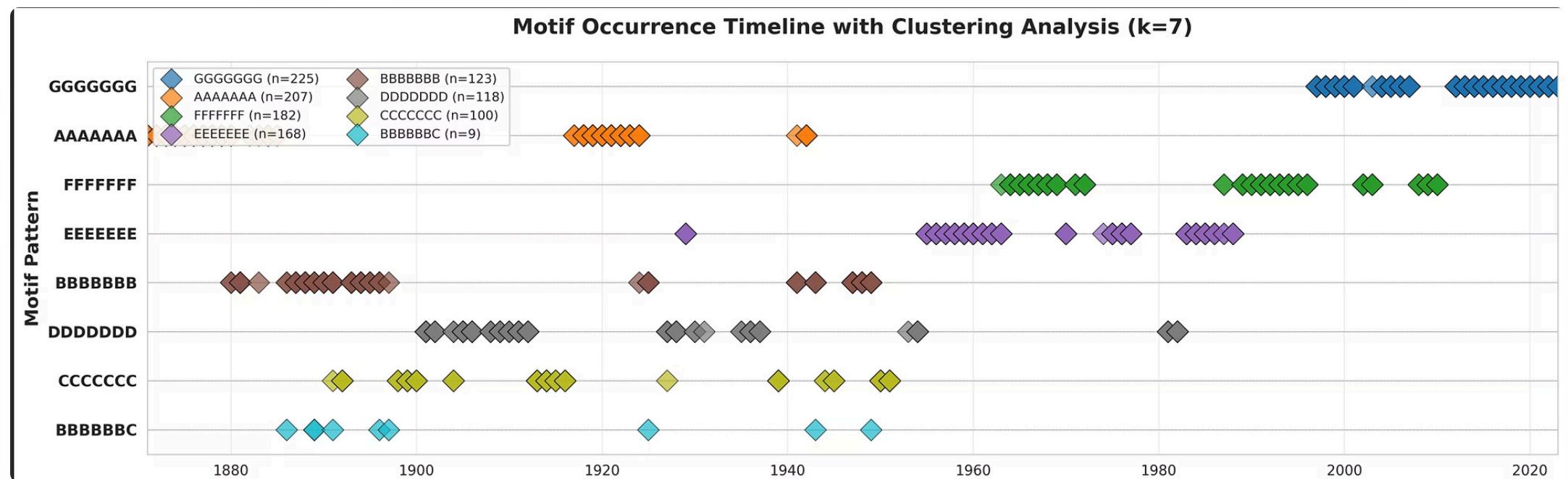
There's a fascinating paradox here:

For OPM (directional patterns):

- LOW regime = MORE diversity, LESS persistence
- HIGH regime = LESS diversity, MORE persistence

For CTM (structural patterns):

- LOW regime = MORE complexity, LESS stability
- HIGH regime = LESS complexity, MORE stability



Our Three Questions

1. Do LOW and HIGH valuation regimes exhibit different pattern structures?

YES, dramatically different

LOW (Cheap Markets): chaotic & unique, short patterns (5-7), high volatility

HIGH (Expensive Markets): uniform, long patterns (10+), smooth structures

3. Can we identify early warning signals for regime transitions?

- **HIGH → LOW (Crash):** pattern fragmentation, repeated motifs, CTM changes, OPM diversity spike
- **LOW → HIGH (Recovery):** pattern consolidation, reduced motif diversity, OPM convergence, extended motif lengths

Bottom Line

Pattern structures shift significantly between cheap and expensive markets – these changes can serve as early warning signals for regime transitions.

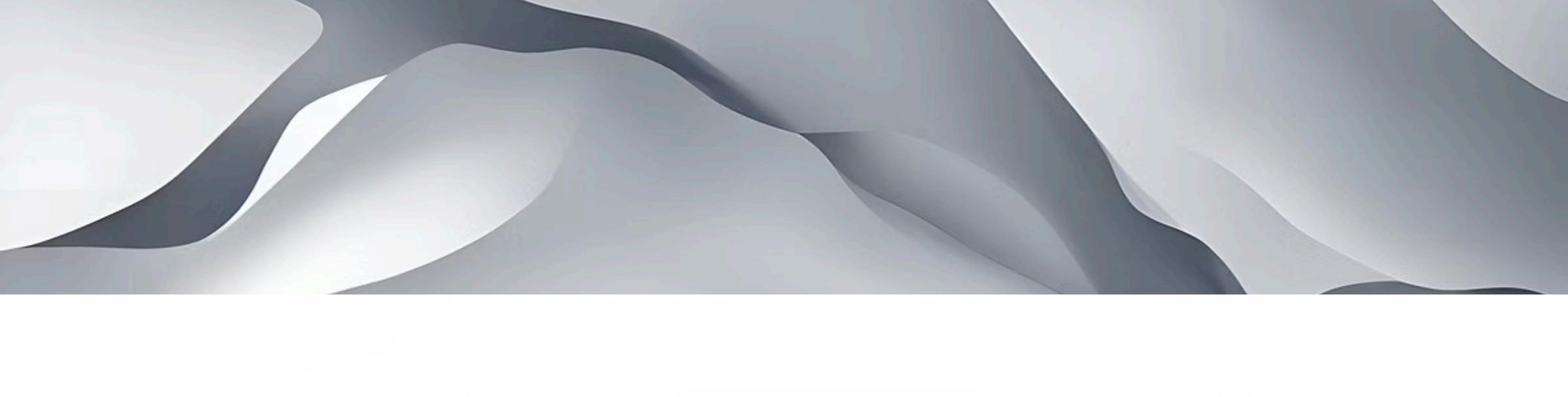
Classic chart structures



Head and Shoulders

Double Top

Cup and Handle



Thank you for your
attention

Adam & Nofar