QUANTITATIVE RESEARCH

Momentum - Not an Anomaly

Combinatorial Foundations of Momentum and Factor Structure

Why Trend-Following is a Mathematical Necessity, Not a Statistical Fluke

Ashish Ranjan
LevUp
hello@levup.in

Abstract

The momentum premium is not a market anomaly. It is a mathematical theorem.

Any sequence of more than n^2 daily returns *must* contain a monotone subsequence of length at least $n+1$, a momentum or mean-reversion pattern—by the Erdős–Szekeres theorem of 1935. No distributional assumptions. No behavioural premises. No notion of market efficiency. Just combinatorics.

The literature on momentum investing rests on an implicit premise: that momentum is a regularity requiring explanation by risk premia, behavioural biases, or institutional frictions. We challenge this premise at its root.

We show that momentum signatures *must* exist in any equity return series of sufficient length, as a direct consequence of the Erdős–Szekeres theorem (1935), a result in combinatorics with no connection to financial markets. The theorem requires no distributional assumption, no model of investor behaviour, and no notion of market efficiency. It holds for any sequence of distinct real numbers.

We extend this in three directions. First, Ramsey's theorem (1930) establishes that factor structure clusters of mutually correlated stocks is likewise combinatorially inevitable in any sufficiently large equity universe. Second, Dilworth's theorem (1950) reveals momentum and mean reversion as mathematical duals: the depth of momentum structure exactly determines the minimum number of mean-reversion segments needed to tile the series. Third, the classical result $\mathbb{E}[L_T] \sim 2\sqrt{T}$ for the longest increasing subsequence of a random permutation of length T, together with the Baik–Deift–Johansson theorem on Tracy–Widom fluctuations, yields a principled duration baseline and a formal significance test for observed momentum runs.

Since momentum cannot *not* exist, the research agenda shifts from *whether* it exists to *at what time scales it is identifiable and exploitable*.

Introduction - The anomaly that was never an anomaly

The momentum effect, namely the tendency of stocks that have performed well over the past three to twelve months to continue outperforming over the subsequent three to twelve months, is among the most replicated findings in empirical asset pricing. Jegadeesh and Titman [1993] documented the effect in US equities. Asness et al. [2013] showed it persists across asset classes and geographies. Moskowitz et al. [2012] established it in time-series form. The empirical record is, by now, uncontroversial.

Yet the literature has treated momentum, almost universally, as a *puzzle*. The standard framing is: momentum exists, it is inconsistent with simple efficient market benchmarks, and it therefore requires explanation. A substantial industry of explanations has followed: risk-based [Fama and French, 1996], behavioural [Barberis et al., 1998, Hong and Stein, 1999], and institutional [Jegadeesh and Titman, 2001].

This paper argues that the puzzle framing is wrong from the start.

Central Claim

Momentum is not an anomaly. It is a theorem.

In any equity return series of length $T > n^2$, a monotone subsequence of daily returns of length at least $n + 1$ must exist with certainty, under no distributional or market-structure assumptions. This is the Erdős–Szekeres theorem of 1935, applied to a return series.

A monotone increasing subsequence is a momentum pattern. A monotone decreasing subsequence is a mean-reversion pattern. Both must coexist in any sufficiently long return series. Their coexistence is not a feature of markets; it is a feature of sequences.

The distinction matters. If momentum is an anomaly, it could in principle disappear; the market could become "more efficient" and the premium could be arbitraged away. If momentum is a theorem, it cannot disappear. It will be present in any return data generated by any process. What can change is whether it is *identifiable in real time* and whether it is *profitably exploitable after costs*, but those are separate questions that belong to a separate layer of analysis.

Three Questions, Three Layers

The paper's organising framework is a decomposition into three strictly distinct questions, which we call layers.

The Three-Layer Framework

Layer 1: Existence. Must the pattern exist? **Yes, always, by the Erdős–Szekeres theorem.** This layer is the subject of the present paper. Its conclusions are unconditional.

Layer 2: Identification. Given that the pattern exists, can it be detected in real time from the data available up to the current date? This is a statistical inference problem. Its answer is noisy and probabilistic. Tools: hidden Markov models, Kalman filters, Markov-switching regressions, breakpoint detection.

Layer 3: Profitability. Given that the pattern is identified, can it be traded profitably net of transaction costs, market impact, and competitive arbitrage? This is an economic question. Its answer depends on capacity, implementation, and the degree of crowding.

The *anomaly framing* confuses layers. It treats a Layer 1 certainty as if it were a Layer 3 observation in need of a risk or behavioural explanation. The correct interpretation: the existence of momentum is guaranteed by mathematics; its identifiability is a statistical challenge; its profitability is an economic one. The three are not the same question.

Results

R1. Momentum must exist (Erdős–Szekeres, 1935). Any sequence of $T > n^2$ distinct daily returns contains a monotone subsequence of length $n + 1$. Setting $n = 15$ and $T = 252$ (one trading year), the theorem guarantees a momentum or mean-reversion run of at least 16 days. No model of markets is involved.

R2. Factor structure must exist (Ramsey, 1930). In any sufficiently large equity universe, the pairwise return-correlation graph must contain a monochromatic clique, a set of stocks that all move together. This is a factor. Factor structure is not discovered by PCA; it is guaranteed before PCA is run.

R3. Momentum and mean reversion are mathematical duals (Dilworth, 1950). The longest momentum run in a return series exactly equals the minimum number of mean-reversion segments required to partition the entire series. The two strategies are not competing explanations for the same data. They are two descriptions of one poset structure.

R4. Momentum duration scales as $\sqrt{T}$ (Logan–Shepp / Vershik–Kerov, 1977; Baik–Deift–Johansson, 1999). Under the null of a structureless return series, the expected longest momentum run is $\mathbb{E}[L_T] \sim 2\sqrt{T}$, with fluctuations of order $T^{1/6}$ following the

Tracy–Widom distribution. This provides a mathematically principled significance test for observed momentum runs and implies a square-root law for lookback window selection.

The anomaly framing has had real costs. It directed research toward explaining momentum rather than exploiting it. It created a false sense of fragility: if momentum is an anomaly, perhaps it will eventually disappear. It led practitioners to treat the *existence* of momentum as the uncertain variable, when in fact it is the certain one. The uncertain variables are identification and profitability.

Knowing the correct layer structure leads to sharper questions. Instead of asking *why does momentum exist?*, the right question is: *at what lookback horizon is the momentum subsequence most identifiable in real time?* Instead of asking *will momentum persist?*, the right question is: *will the momentum premium remain large enough to survive transaction costs and crowding?*

Structure of the Paper

Section establishes notation and mathematical preliminaries. Section states and proves the Erdős–Szekeres theorem and derives its financial consequences. Section covers Ramsey theory and the inevitability of factor structure. Section covers Dilworth's theorem and the momentum–mean-reversion duality. Section derives the $2\sqrt{T}$ baseline and the Tracy–Widom significance test. Section synthesises the three-layer framework. Section 1 states honest limitations. Section 1.0.1 concludes.

Preliminaries

Let $\mathcal{S}$ denote a traded equity and let $\{P_t\}_{t=0}^{T}$ denote its closing price sequence over T trading days. Define the daily log-return series

$$r_t \;=\; \log\!\left(\frac{P_t}{P_{t-1}}\right), \quad t = 1, \dots, T.$$

We work throughout with the sequence $\mathbf{r} = (r_1, r_2, \dots, r_T) \in \mathbb{R}^T$. Since returns are continuous random variables, distinct elements are guaranteed almost surely ($\mathbb{P}[r_i = r_j] = 0$ for $i \neq j$). No further distributional assumptions are imposed unless explicitly stated.

Definition (Subsequence and Length). A *subsequence* of $\mathbf{r}$ is a sequence obtained by selecting indices $1 \leq t_1 < t_2 < \cdots < t_k \leq T$ and retaining $(r_{t_1}, \dots, r_{t_k})$. The integer k is the *length* of the subsequence. The indices $t_1, \dots, t_k$ need not be consecutive.

Definition (Monotone Subsequences). A subsequence $(r_{t_1}, \dots, r_{t_k})$ is

(i) *strictly increasing* if $r_{t_1} < r_{t_2} < \cdots < r_{t_k}$;

(ii) *strictly decreasing* if $r_{t_1} > r_{t_2} > \cdots > r_{t_k}$.

We write $\mathrm{LIS}(\mathbf{r})$ and $\mathrm{LDS}(\mathbf{r})$ for the lengths of the *longest increasing subsequence* and *longest decreasing subsequence* of $\mathbf{r}$, respectively.

Financial Meaning of Monotone Subsequences

An **increasing** subsequence at dates $t_1 < \cdots < t_k$ with $r_{t_1} < \cdots < r_{t_k}$ indicates that the selected daily returns improve monotonically through time. In market terms, this ordered structure can arise in several economically distinct settings:

- *Trend reinforcement.* Returns are positive and increasing: $+0.2\%$, $+0.5\%$, $+1.1\%, \dots$ This is the canonical momentum profile.
- *Downside attenuation.* Returns remain negative but move steadily toward zero: -2.0%, -1.1%, $-0.3\%, \dots$ Formally, this is still an increasing subsequence and is often associated with bottom formation.
- *Regime inflection.* Returns cross from negative to positive: -0.6%, $+0.1\%$, $+0.9\%, \dots$ This is the discrete signature of a change in return regime.

A **decreasing** subsequence captures the mirror-image cases: trend exhaustion on the upside, intensifying downside pressure, and negative regime inflection. The unifying point is that these are economically different manifestations of the same ordered structure. The theorem therefore identifies the combinatorial backbone underlying a broad class of momentum and mean-reversion phenomena.

Definition (Partial Order and Poset). A *partial order* on a set X is a binary relation $\preceq$ satisfying reflexivity, antisymmetry, and transitivity. The pair $(X, \preceq)$ is a *partially ordered set* (poset). Elements $x, y \in X$ are *comparable* if $x \preceq y$ or $y \preceq x$; otherwise they are *incomparable*.

Definition (Chains and Antichains). In a poset $(X, \preceq)$:

(i) A *chain* is a subset $C \subseteq X$ in which every pair of elements is comparable.

(ii) An *antichain* is a subset $A \subseteq X$ in which no two elements are comparable.

Definition (Complete Graph and Colouring). The *complete graph* K_N has vertex set $\{1, \dots, N\}$ and contains every pair $\{i, j\}$ as an edge. A 2-*colouring* of K_N assigns each edge one of two colours (hereafter red or blue). A *monochromatic clique* of size k is a set of k vertices all of whose $\binom{k}{2}$ pairwise edges receive the same colour.

Momentum Must Exist - The Erdős–Szekeres Theorem

The Erdős–Szekeres theorem rests on the Pigeonhole Principle. Before stating the theorem, we give two examples that build intuition for the financial application.

Theorem (Pigeonhole Principle). *If $n + 1$ or more objects are distributed among n categories, at least one category contains at least two objects.*

The analytical skill, which is also the skill of a good quant researcher, lies not in the principle itself but in the *choice of objects and categories*. The following example foreshadows the financial application directly.

Example (Partial Sums and Price Levels). *In any sequence of* 11 *integers, some block of consecutive terms has sum divisible by* 11.

Proof. Let $a_1, \dots, a_{11}$ be the sequence. Form partial sums $S_0 = 0$ and $S_k = a_1 + \cdots + a_k$ for $k = 1, \dots, 11$. Each S_k mod 11 belongs to $\{0, 1, \dots, 10\}$, giving eleven residues. With twelve partial sums and eleven residues, pigeonhole gives $S_i \equiv S_j \pmod{11}$ for some $i < j$. Then $a_{i+1} + \cdots + a_j = S_j - S_i \equiv 0 \pmod{11}$. □

Remark (The Skeleton of Mean Reversion). Replace "divisible by 11" with "return to a reference price level." Partial sums of log-returns are log-prices. With finitely many meaningful price buckets, the log-price process must revisit buckets purely combinatorially. This is the mathematical skeleton of mean reversion, derived without any model of markets.

The Erdős–Szekeres Theorem

Theorem (Erdős–Szekeres, 1935). *Every sequence of more than n^2 distinct real numbers contains either a strictly increasing subsequence of length $n + 1$ or a strictly decreasing subsequence of length $n + 1$ (or both).*

Proof. Let $\mathbf{a} = (a_1, \dots, a_{n^2+1})$ be a sequence of $n^2 + 1$ distinct real numbers. For each index

$i \in \{1, \ldots, n^2 + 1\}$, define

$$x_i := \text{length of the longest strictly increasing subsequence of } \mathbf{a} \text{ ending at } a_i,$$
$$y_i := \text{length of the longest strictly decreasing subsequence of } \mathbf{a} \text{ ending at } a_i.$$

Assign to each element the ordered pair (label) $(x_i, y_i) \in \mathbb{N} \times \mathbb{N}$.

Let $i < j$. Since all elements are distinct, either $a_i < a_j$ or $a_i > a_j$.

- *Case* $a_i < a_j$. Any increasing subsequence of **a** ending at a_i can be extended by appending a_j, yielding an increasing subsequence ending at a_j of strictly greater length. Hence $x_j \geq x_i + 1$, so $(x_i, y_i) \neq (x_j, y_j)$.
- *Case* $a_i > a_j$. Any decreasing subsequence ending at a_i can be extended by a_j, giving $y_j \geq y_i + 1$ and again $(x_i, y_i) \neq (x_j, y_j)$.

Thus all $n^2 + 1$ labels are pairwise distinct.
Now suppose for contradiction that $x_i \leq n$ and $y_i \leq n$ for every i. Then every label belongs to the grid $G = \{1, \ldots, n\}^2$, which has exactly n^2 elements. But we have $n^2 + 1$ distinct labels drawn from G, which is an immediate contradiction by the Pigeonhole Principle.

The contradiction establishes that some element a_i satisfies $x_i \geq n + 1$ (a strictly increasing subsequence of length $n + 1$ exists) or $y_i \geq n + 1$ (a strictly decreasing subsequence of length $n + 1$ exists). □

Proposition (Momentum and Mean-Reversion Are Inevitable). *Let* $\mathbf{r} = (r_1, \ldots, r_T)$ *be any daily return series with* $T > n^2$ *and all returns distinct. Then*

$$\mathrm{LIS}(\mathbf{r}) \geq n + 1 \quad \textit{or} \quad \mathrm{LDS}(\mathbf{r}) \geq n + 1 \qquad \textit{(or both).}$$

This holds for every return-generating process*, under* no distributional assumptions.

Proof. Apply the Erdős–Szekeres theorem to the sequence $(r_1, \ldots, r_T)$ of $T > n^2$ distinct real numbers. □

The final sentence of the momentum inevitability proposition deserves emphasis. It is the core of the reframing. The result holds whether returns are drawn from a normal distribution or a power law, whether markets are efficient or not, whether investors are rational or behavioural. It holds for Reliance Industries and for a hypothetical market on Mars. It is not a statement about markets; it is a statement about sequences.

Instantiation: Indian and Global Equity Markets

The table below instantiates the momentum inevitability proposition for standard lookback windows used in quantitative equity strategies.

Window	T (days)	n s.t. $n^2 < T$	Guaranteed run $\geq$	Approx. months
3 months	63	7	8 days	0.4 mo
6 months	126	11	12 days	0.6 mo
12 months	252	15	16 days	0.8 mo
24 months	504	22	23 days	1.1 mo
36 months	756	27	28 days	1.4 mo

Each row states: given a return series of that length, the Erdős–Szekeres theorem guarantees a monotone subsequence of at least the stated length, with *mathematical certainty*, for every stock in the universe, every day.

Remark (Coexistence and Time Scale). The disjunction in the momentum inevitability proposition is generally non-exclusive: both an increasing and a decreasing subsequence of length $n + 1$ coexist in the same return series, drawn from *different* subsets of dates. This mathematical coexistence is the formal explanation for the empirical observation [Jegadeesh and Titman, 1993, Lo and MacKinlay, 1990] that momentum and mean reversion operate simultaneously at different time horizons. They are not competing models. They are co-present features of any sequence.

Remark (The 58-Year Gap). The Erdős–Szekeres theorem was proved in 1935. Jegadeesh and Titman [1993] documented momentum empirically in 1993. For 58 years, the mathematical guarantee sat in the combinatorics literature while the empirical finance literature searched successfully for what the mathematics had already declared inevitable.

Factor Structure Must Exist - Ramsey Theory

The Erdős–Szekeres theorem operates on a single return series. In practice, a quantitative equity portfolio is built across a universe of hundreds or thousands of stocks. The question becomes:

In a large universe of stocks, what structure in the cross-section is mathematically inevitable?

The answer is provided by Ramsey's theorem: in any large enough graph with coloured edges, a monochromatic clique must exist. Applied to the correlation graph of an equity universe, this gives the inevitability of factor structure.

Ramsey's Theorem

Theorem ($R(3,3) = 6$, Ramsey 1930). *In any 2-colouring of the edges of K_6, there exists a monochromatic triangle.*

Proof. Fix any vertex v. It has 5 edges to the remaining 5 vertices. By pigeonhole, at least $\lceil 5/2 \rceil = 3$ edges share one colour; say vertices u_1, u_2, u_3 are each connected to v by a red edge. Now examine the edges among $\{u_1, u_2, u_3\}$:

- If any edge $\{u_i, u_j\}$ is red, then v, u_i, u_j form a red triangle.
- If no edge among $\{u_1, u_2, u_3\}$ is red, all three are blue, giving a blue triangle.

A monochromatic triangle exists in either case. □

Theorem (Ramsey, General Form). *For all positive integers r and s, there exists a minimal positive integer $R(r,s)$, the* Ramsey number*, such that every 2-colouring of $K_{R(r,s)}$ contains a red clique of size r or a blue clique of size s.*

The Equity Correlation Graph

Definition (Correlation Graph). Let $\mathcal{U} = \{S_1, \ldots, S_N\}$ be an equity universe. For a common estimation window, let $\rho_{ij} = \operatorname{corr}(r^{(i)}, r^{(j)})$ denote the pairwise return correlation. Fix a

Table 1: Known and estimated Ramsey numbers $R(r, s)$. The open problem $R(5, 5)$ has resisted resolution for decades despite enormous effort.

r	s	$R(r, s)$
3	3	6
3	4	9
3	5	14
4	4	18
4	5	25
5	5	$\in$ [43, 48] (open)
6	6	$\in$ [102, 165] (open)

threshold $\tau \in (0, 1)$. The *correlation graph* $G(\mathcal{U}, \tau)$ is K_N with edges coloured:

$$e_{ij} = \begin{cases} \text{red} & \text{if } \rho_{ij} > +\tau, \\ \text{blue} & \text{if } \rho_{ij} < -\tau, \\ \text{grey (absent)} & \text{otherwise.} \end{cases}$$

Proposition (Factor Structure Is Inevitable). *For any $\tau \in (0, 1)$ and integers $r, s \geq 1$, if $\mathcal{U}$ contains at least $R(r, s)$ stocks for which every pairwise correlation satisfies $|\rho_{ij}| > \tau$, then $\mathcal{U}$ contains either*

(i) a set of r stocks all pairwise positively correlated (a factor cluster*), or*

(ii) a set of s stocks all pairwise negatively correlated (a natural hedge cluster*).*

Proof. Apply Ramsey's general theorem to the red–blue subgraph of $G(\mathcal{U}, \tau)$. □

Factor Structure: The Ramsey Interpretation

A **factor cluster**, a red clique in the factor-structure proposition, is a set of stocks that all move together. By definition, they share a common driver: a sector exposure, a macro sensitivity, a style characteristic. That common driver is a *factor* in the sense of Fama and French.

Principal component analysis (PCA) on a return covariance matrix finds these clusters algebraically. The first eigenvector is the market factor; subsequent eigenvectors are sector and style factors. What the factor-structure proposition establishes is that these objects *must exist* before any data is examined. PCA does not discover factors. It computes the algebraic representation of objects whose existence Ramsey's theorem already guaranteed.

A **natural hedge cluster**, a blue clique, is a set of stocks with systematically opposing sensitivities. Long/short equity funds are built on exactly these structures. The mathematical guarantee: such hedging clusters must exist in any large enough universe, regardless of the economic regime.

The Turán Bound

Turán's theorem complements Ramsey by providing a quantitative threshold on correlation density sufficient to force a factor cluster.

Theorem (Turán, 1941). *The maximum number of edges in a graph on N vertices with no clique of size $r+1$ is*

$$\mathrm{ex}(N, K_{r+1}) = \left(1 - \frac{1}{r}\right)\frac{N^2}{2}.$$

Application. In the correlation graph $G(\mathcal{U}, \tau)$, if the fraction of stock pairs with $\rho_{ij} > \tau$ exceeds $1 - 1/r$, a factor cluster of size $r+1$ must exist. This is a data-driven criterion, computable from any estimated correlation matrix, for the guaranteed existence of a factor of prescribed size.

Momentum and Mean Reversion are Duals
Dilworth's Theorem

The Erdős–Szekeres theorem guarantees the existence of momentum and mean-reversion signatures but treats them as alternatives. Dilworth's theorem reveals that they are not alternatives; they are two faces of one mathematical object.

Definition (Return Series Partial Order). Given $\mathbf{r} = (r_1, \ldots, r_T)$, define a partial order $\preceq$ on $\{1, \ldots, T\}$ by

$$i \preceq j \iff i \leq j \text{ and } r_i \leq r_j.$$

Index i precedes j in this order if j is later in time *and* carries a weakly larger return.

Under the return-series partial-order definition:

- A **chain** in $(\{1, \ldots, T\}, \preceq)$ consists of indices $t_1 < \cdots < t_k$ with $r_{t_1} \leq \cdots \leq r_{t_k}$, precisely an **increasing subsequence of returns**, i.e., a **momentum pattern**.

- An **antichain** consists of indices $s_1 < \cdots < s_k$ (w.l.o.g.) with $r_{s_1} > r_{s_2} > \cdots > r_{s_k}$, precisely a **decreasing subsequence of returns**, i.e., a **mean-reversion pattern**.

The poset language thus translates directly: momentum runs are chains, mean-reversion segments are antichains.

Theorem (Dilworth, 1950). *In any finite poset, the maximum size of an antichain equals the minimum number of chains needed to cover the entire poset (partition it into chains).*

Proof sketch (full proof in Appendix A). Let ω denote the maximum antichain size. Any chain intersects any antichain in at most one element; hence any chain cover requires at least ω chains, giving min cover $\geq \omega$. The matching direction, namely that a cover of exactly ω chains exists, is established by applying Hall's marriage theorem to a bipartite graph whose two vertex sets are two copies of the poset, with an edge from x^+ to y^- whenever $x \prec y$ (strict). The maximum matching size equals ω by a Kőnig-type argument, and the transitive closure of the matching yields the desired chain partition. □

Theorem (Mirsky, 1971). *In any finite poset, the maximum size of a chain equals the minimum number of antichains needed to partition the poset.*

Dilworth's and Mirsky's theorems together constitute the *Dilworth-Mirsky min-max duality*.

Financial Duality - One Poset, Two Strategies

Proposition (Momentum–Mean-Reversion Duality). *For any return series* $\mathbf{r} = (r_1, \ldots, r_T)$ *with the return-series partial order:*

$$\mathrm{LIS}(\mathbf{r}) = \min\left\{k : \mathbf{r} \text{ can be partitioned into } k \text{ decreasing subsequences}\right\}, \tag{1}$$

$$\mathrm{LDS}(\mathbf{r}) = \min\left\{k : \mathbf{r} \text{ can be partitioned into } k \text{ increasing subsequences}\right\}. \tag{2}$$

Proof. Apply Dilworth's theorem to (1) and Mirsky's theorem to (2), identifying chains with increasing subsequences and antichains with decreasing subsequences under the return-series partial-order definition. □

The Central Duality Result

Equation (1) states:

The depth of the momentum structure, namely the length of the longest momentum run, exactly determines the minimum number of mean-reversion segments required to fully tile the return series.

Equation (2) states:

The depth of the mean-reversion structure exactly determines the minimum number of momentum runs required to fully tile the return series.

These are not two different strategies competing for the same data. They are two dual descriptions of a single underlying mathematical object: the return series poset. A quant fund running momentum and mean-reversion simultaneously is not expressing a philosophical view about markets. It is exploiting a mathematical duality that guarantees complementary structure in every return series.

How Long Does Momentum Last - The $2\sqrt{T}$ Result

From Existence to Duration

The momentum inevitability proposition guarantees that a momentum run of length $n + 1$ exists in any return series of length $T > n^2$. A natural follow-up question is quantitative:

In a completely structureless return series, one with no genuine momentum, how long would the longest momentum run typically be?

This is the baseline. A quant needs it to distinguish genuine momentum signals from statistical noise.

The Longest Increasing Subsequence Problem

We model a structureless return series as a uniformly random permutation of $\{1, 2, \ldots, T\}$, the combinatorial null hypothesis in which all orderings are equally likely. Denote by L_T the length of the longest increasing subsequence of such a permutation.

Theorem (Logan–Shepp; Vershik–Kerov, 1977).

$$\frac{L_T}{\sqrt{T}} \xrightarrow{\mathbb{P}} 2 \quad \text{as } T \to \infty.$$

Equivalently, $\mathbb{E}[L_T] \sim 2\sqrt{T}$ *as* $T \to \infty$.

Proof Idea. The key is the Robinson–Schensted–Knuth (RSK) correspondence, a bijection between permutations of $\{1, \ldots, T\}$ and pairs of standard Young tableaux of the same shape. Under RSK, L_T equals the length λ_1 of the first row of the corresponding random Young tableau. The limit shape of a uniformly random Young tableau, computed via variational methods applied to the hook-length formula, is a specific curve whose first-row length is exactly $2\sqrt{T}$. Full details are given in Appendix B.

Tracy-Widom Fluctuations

The fluctuations of L_T around its mean are characterised by a celebrated and surprising result.

Theorem (Baik–Deift–Johansson, 1999).

$$\frac{L_T - 2\sqrt{T}}{T^{1/6}} \xrightarrow{d} F_2 \quad \text{as } T \to \infty,$$

where F_2 *is the* Tracy–Widom GUE distribution*, the limiting distribution of the largest eigenvalue of a matrix drawn from the Gaussian Unitary Ensemble, rescaled appropriately.*

Remark (Universality and Random Matrices). F_2 appears in an extraordinary range of contexts: the largest eigenvalue of large random matrices, interface growth models (KPZ universality), directed last-passage percolation, and by the Baik–Deift–Johansson theorem, the longest increasing subsequence. In equity markets, F_2 also governs the fluctuations of the largest eigenvalue of sample return covariance matrices [Marčenko and Pastur, 1967, Plerou et al., 2002]. This connects the momentum duration baseline directly to the random matrix approach to factor model specification: the same distribution that tests whether a momentum run is genuine also tests whether a PCA factor is genuine or a statistical artefact of finite-sample estimation.

The Momentum Duration Baseline

Momentum Duration Baseline: $\mathbb{E}[L_T] \approx 2\sqrt{T}$

Window	T	$2\sqrt{T}$	$T^{1/6}$	95% CI	Signal at $\ell = 45$
1 month	21	9.2	1.6	(6.4, 12.0)	$Z = 22.4$
3 months	63	15.9	2.0	(12.0, 19.8)	$Z = 14.6$
6 months	126	22.4	2.3	(18.0, 26.8)	$Z = 9.8$
12 months	252	31.7	2.6	(26.6, 36.8)	$Z = 5.1$
36 months	756	54.9	3.0	(49.0, 60.8)	$Z = -3.3$

95% CI approximated as $2\sqrt{T} \pm 2T^{1/6}$. The last column (Z-score for an observed run $\ell = 45$) shows that the same run is highly significant over 12 months but is *below* the baseline over 36 months, demonstrating that run significance is window-relative. The F_2 99% quantile is $q_{0.99} \approx 2.02$; any Z-score exceeding this rejects the null of no momentum at the 1% level.

Proposition (Significance Test for Momentum Runs). *Let $\ell = \mathrm{LIS}(\mathbf{r})$ be the observed longest momentum run over a window of T trading days. Under the null hypothesis that $\mathbf{r}$ is a random permutation (no genuine momentum), the statistic*

$$Z = \frac{\ell - 2\sqrt{T}}{T^{1/6}}$$

converges in distribution to F_2 as $T \to \infty$. The null is rejected at level α whenever $Z > q_{1-\alpha}(F_2)$, where selected quantiles of F_2 are tabulated in Appendix C.

Example (Significance of an Observed Momentum Run). A 12-month return series ($T = 252$) exhibits a momentum run of $\ell = 45$ trading days. The test statistic is

$$Z = \frac{45 - 2\sqrt{252}}{252^{1/6}} = \frac{45 - 31.75}{2.57} \approx 5.15.$$

Since $q_{0.999}(F_2) \approx 2.69$, the run is significant well beyond the 0.1% level. The probability of observing $\ell \geq 45$ in a structureless 252-day series is less than one in a thousand. The analyst can conclude, with high confidence, that a genuine momentum regime is present, not noise.

Remark ($\sqrt{T}$ Scaling: A Law for Strategy Design). The Logan–Shepp–Vershik–Kerov theorem implies that the expected momentum signal duration scales as $\sqrt{T}$, not linearly. Doubling the lookback window multiplies the expected signal duration by $\sqrt{2} \approx 1.41$, not 2. The empirical observation that "longer lookbacks yield diminishing returns" is not a practical finding; it is a corollary of the $2\sqrt{T}$ theorem. A lookback window selection algorithm should incorporate this square-root law as a structural prior.

The Three-Layer Framework

Layer 1:

The results of Sections – collectively constitute Layer 1: the set of unconditional, assumption-free mathematical guarantees.

Layer 1 Summary: Unconditional Combinatorial Guarantees

L1.1. Momentum and mean reversion exist (Erdős–Szekeres, 1935): Any return series of length $T > n^2$ contains a monotone subsequence of length $n + 1$. No model, no distribution, no market assumption required.

L1.2. Factor structure exists (Ramsey, 1930): Any sufficiently large equity universe contains a mutually correlated cluster, a factor. Guaranteed before any data is examined.

L1.3. Momentum and mean reversion are duals (Dilworth–Mirsky, 1950/1971): The depth of the momentum structure exactly determines the complexity of the mean-reversion structure, and vice versa. They are two descriptions of one poset.

L1.4. Momentum duration scales as $\sqrt{T}$ (Logan–Shepp / Vershik–Kerov, 1977): The expected longest run in structureless data is $2\sqrt{T}$. Fluctuations follow the Tracy–Widom distribution, enabling formal significance tests (Baik–Deift–Johansson, 1999).

The Gap to Layer 2: Identification

Layer 1 establishes existence with certainty. It says nothing about *where* in the series the pattern sits, *when* it starts, or *how long* any specific instance lasts. Extracting that information from real-time data is the Layer 2 problem, a statistical inference problem with irreducible uncertainty.

Key Layer 2 tools, in order of increasing flexibility:

- **Hidden Markov Models.** Market regime (momentum vs. mean-reversion) modelled as a hidden discrete state; returns are noisy emissions. The Baum–Welch algorithm estimates the transition matrix; the Viterbi decoder recovers the most likely regime path.
- **Kalman Filtering.** A latent momentum signal evolves as a linear state-space system; returns are noisy linear measurements. The Kalman filter delivers the minimum-variance linear estimate of the current signal level.
- **Markov-Switching Regressions [Hamilton, 1989].** Returns follow regime-specific autoregressive processes; the Hamilton filter produces posterior regime probabilities.
- **Structural Break Detection.** Bai–Perron tests and CUSUM statistics identify regime changes retrospectively; useful for strategy evaluation and regime labelling.

None of these inherits the certainty of Layer 1. They produce probability estimates of a hidden truth, not guarantees.

The Gap to Layer 3: Profitability

A perfectly identified momentum signal need not generate profit. The barriers at Layer 3:

- **Transaction costs.** Bid-ask spreads, market impact, and clearing fees impose a gross return threshold. A momentum signal generating 20bps gross is unviable if implementation costs 25bps.
- **Capacity.** Market impact scales super-linearly with order size. A strategy viable at $50m AUM may destroy its own signal at $5bn.
- **Competitive arbitrage and alpha decay.** Once a pattern is identifiable, capital flows toward it. Competition drives the net expected return toward zero. The pattern continues to exist (Layer 1 guarantees it); the profit from trading it does not.
- **Crowding.** Correlated liquidations across funds sharing the same signals amplify losses in stress periods. The August 2007 quant deleveraging is the canonical example: mathematically sound strategies collectively created a destructive feedback loop.

The key conclusion: the anomaly framing confuses Layer 1 with Layer 3. The existence of momentum (Layer 1) is guaranteed; the profitability of momentum strategies (Layer 3) is not.

Researchers who ask “will momentum eventually be arbitraged away?” are asking a Layer 3 question but phrasing it as if it threatens the Layer 1 guarantee. It does not.

Diagram of the Three Layers

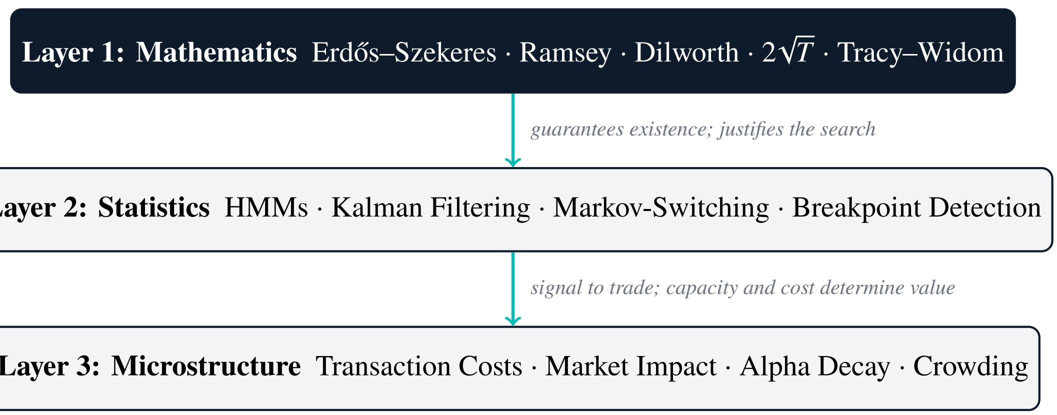

Figure 1: The three-layer decomposition. Layer 1 is unconditional. Layers 2 and 3 involve irreducible uncertainty. The anomaly framing incorrectly treats Layer 1 results as if they were Layer 3 observations.

1. Limitations and Scope Conditions

No theoretical framework is complete without a precise statement of scope. The results in this paper are strongest at the level of mathematical existence and materially weaker at the level of statistical identification and portfolio implementation. We therefore state explicitly where the combinatorial results remain decisive and where economic interpretation requires additional judgment.

Layer 1 Limitations

1.0.1. Subsequences Are Non-Contiguous

The monotone subsequences guaranteed by the Erdős–Szekeres theorem need not be contiguous. The 11 dates forming a guaranteed momentum pattern in a 101-day window may be days $\{3, 11, 24, 38, 52, \ldots\}$, dispersed across the sample and separated by intervening reversals. A portfolio manager, by contrast, bears exposure across the full holding interval and cannot isolate only the observations that appear in the subsequence.

Translating the guarantee to a contiguous trending block requires different mathematics: results on *runs* in random sequences, which are probabilistic and weaker. The Erdős–Szekeres theorem should therefore be interpreted as a structural statement about ordered subsequences, not as a direct guarantee of contiguous, tradeable trends.

Ramsey Numbers Grow Rapidly

For large clique sizes r, Ramsey numbers satisfy $R(r,r) \geq 2^{r/2}$. Guaranteeing a factor cluster of size 20 may require a universe of more than $2^{10} = 1024$ stocks satisfying the correlation threshold. For small clusters ($r \leq 6$), the Ramsey numbers are manageable. For large r, however, the guarantee becomes too coarse to be operational on its own. At that point, economic structure, model specification, and empirical estimation carry more weight than the worst-case combinatorial bound.

Correlation Threshold Sensitivity

The colouring in the correlation-graph definition depends on the threshold τ. Correlations estimated from finite data carry sampling error and are non-stationary (rising sharply in crises). The combinatorial structure is robust to the choice of τ in a formal sense; the financial interpretation of the resulting clusters is not. In practice, the economic meaning of a Ramsey-style cluster depends materially on estimation window, market regime, and the stability of the correlation measure itself.

Layer 1 / $2\sqrt{T}$ Limitations

The Null Model Is Misspecified

The Logan–Shepp–Vershik–Kerov theorem is derived for uniformly random permutations. Actual return series exhibit heavy tails, volatility clustering, serial correlation, and structural breaks. The $2\sqrt{T}$ baseline is best viewed as a benchmark under a maximally unstructured null; it is not calibrated to the full empirical behavior of asset returns. Observed run lengths in live data may therefore fall below the benchmark in highly choppy regimes or exceed it materially in persistent trends.

Non-Stationarity

Layer 1 results hold for fixed sequences. When the data-generating process shifts (regime changes, structural breaks), historical calibration of Layer 2 models becomes unreliable. Layer 1 is immune to non-stationarity; the practical usefulness of $2\sqrt{T}$ as a significance baseline depends on stationarity within the estimation window.

Reflexivity and Market Adaptation

The deepest limitation is reflexivity. Mathematics operates on realised sequences. Financial markets, by contrast, are generated by agents who observe, learn, and adapt to those sequences. Once a pattern becomes widely recognised, capital is allocated in response, and the underlying return-generating process can change.

The Layer 1 guarantee is robust: the pattern continues to exist in the new sequence as well (it is a sequence; the Erdős–Szekeres theorem applies). But the *economic value* of the pattern is not robust: competitive trading drives excess returns toward zero. This is precisely the Layer 3 problem and the core reason the three-layer decomposition is necessary. The existence of momentum is a mathematical statement. The identification of momentum is a statistical problem. The monetisation of momentum is an economic question contingent on crowding, turnover, financing conditions, mandates, and implementation costs.

Conclusion

The argument of this paper can be stated succinctly. A return series is, at the most basic level, a sequence of distinct real numbers. Once the problem is cast in that language, the existence of momentum-type structure no longer depends on a particular asset-pricing model, behavioural mechanism, or market microstructure assumption. By the Erdős–Szekeres theorem (1935), any sufficiently long sequence must contain a monotone subsequence of nontrivial length. In financial language, momentum and mean-reversion signatures are therefore not exceptional features of return data; they are mathematically unavoidable features of ordered sequences.

The same logic extends beyond a single time series. Ramsey's theorem (1930) implies that sufficiently large equity universes must contain internally coherent correlation clusters, providing a combinatorial foundation for factor structure. Dilworth's theorem (1950) then shows that momentum and mean reversion are not competing narratives imposed on the same data, but dual descriptions of the same underlying partial order. In parallel, the Logan–Shepp / Vershik–Kerov result establishes a natural baseline of $2\sqrt{T}$ for the longest increasing subsequence in an unstructured series, while the Baik–Deift–Johansson theorem supplies Tracy–Widom fluctuations around that baseline, making formal significance assessment possible.

Taken together, these results materially change the appropriate framing of the momentum literature. The primary question is not whether momentum exists; at the level of combinatorial structure, it must. The economically relevant questions are instead whether the embedded pattern can be identified sufficiently early, whether it survives implementation frictions, and whether it remains monetisable at institutional scale. Those are Layer 2 and Layer 3 questions. They are empirical and operational in character, and they properly define the frontier of serious quantitative research.

Implications for the Research Agenda

Conventional momentum research is typically framed around two questions: why momentum exists and whether it will persist. The combinatorial perspective developed here suggests a more disciplined starting point. Existence is not the open issue. Once a return series is sufficiently long, monotone structure is guaranteed by theorem. The economically relevant questions concern identification, implementation, and extraction.

For an institutional quantitative research programme, Layer 1 shifts the agenda in three concrete respects:

- **R1. From existence to detection.** The relevant question is not whether momentum is present, but which horizons, filters, and state variables allow the underlying monotone structure to be identified reliably in real time. This is the Layer 2 problem.
- **R2. From persistence to monetisation.** The relevant question is not whether momentum can be "arbitraged away" in the abstract, but under what turnover, capacity, funding, and transaction-cost constraints the pattern remains economically meaningful after implementation frictions. This is the Layer 3 problem.
- **R3. From taxonomy to magnitude.** The debate over whether momentum is "risk-based" or "behavioural" does not govern its existence. Rather, it speaks to the size, timing, and cross-sectional expression of the premium once the combinatorial baseline is acknowledged.

The practical implication is straightforward: the frontier of serious momentum research is not to establish that the phenomenon exists, but to determine the conditions under which it can be detected early, sized correctly, and harvested efficiently.

Directions for Further Research

OP1. Contiguous trend guarantees. Erdős–Szekeres guarantees monotone subsequences, not contiguous runs. A corresponding result for subarrays would materially narrow the gap between the theorem's existence claim and the holding-period constraints faced in live portfolio construction.

OP2. Adaptive Ramsey bounds. Can Ramsey-theoretic guarantees be sharpened using the empirical structure of financial correlation matrices, including block structure and low effective rank? Sharper bounds would translate into more useful conditions on universe size required to support factors of a prescribed breadth.

OP3. Tracy-Widom calibration for heavy-tailed returns. The BDJ theorem applies to random permutations. An analogue for heavy-tailed or autocorrelated sequences would make the significance-test proposition more directly applicable to realised return data without relying on the random-permutation approximation.

Erdős and Szekeres proved their theorem while thinking about points in the plane. They were not thinking about stock markets. The applicability of their result to equity return series is not a coincidence; it is a consequence of the fact that structure in sequences is universal. Markets produce sequences. The mathematics of sequences applies. The 58-year lag between the theorem (1935) and its empirical rediscovery (1993) is a measure of the distance between the mathematical and financial communities. This paper is an attempt to reduce that distance.

A. Proof of Dilworth's Theorem via Hall's Theorem

Theorem (Hall's Marriage Theorem). *Let $H = (A \cup B,\ E)$ be a bipartite graph. A matching that saturates every vertex of A exists if and only if $|N(S)| \geq |S|$ for every $S \subseteq A$, where $N(S)$ denotes the neighbourhood of S in B.*

Proof of Dilworth's theorem. Let $(X, \preceq)$ be a finite poset and ω the maximum antichain size. As argued in the proof sketch, any chain cover has size $\geq \omega$. We construct a chain cover of size exactly ω.

Bipartite graph. Construct $H = (X^+ \cup X^-,\ E)$ where X^+ and X^- are two copies of X, with an edge (x^+, y^-) whenever $x \prec y$ strictly in the poset.

From matching to chain partition. A matching M in H specifies a set of comparable pairs (x, y) with $x \prec y$. Taking the transitive closure of M (viewed as a DAG on X) yields a partition of X into chains: each chain corresponds to one connected component. The number of chains equals $|X| - |M|$.

Maximising the matching. Let $\nu = \nu(H)$ denote the maximum matching size. By Kőnig's minimax theorem (a consequence of Hall's theorem), the maximum matching in a bipartite graph equals the minimum vertex cover. One shows by a standard antichain–vertex-cover duality argument that the minimum vertex cover of H has size $|X| - \omega$, giving $\nu = \omega$. Hence the minimum chain cover has size $|X| - \nu = |X| - \omega$; but since the ω antichains also cover X and the minimum is ω, the number of chains in the optimal partition is ω. □

B. The Robinson-Schensted-Knuth Correspondence

The *RSK correspondence* is a bijection

$$\sigma \in S_T \ \longleftrightarrow\ \big(P(\sigma),\, Q(\sigma)\big),$$

where $P(\sigma)$ and $Q(\sigma)$ are standard Young tableaux (SYT) of the same shape $\lambda \vdash T$ (a partition of T). The construction proceeds by iterative row-insertion: inserting $\sigma(1), \ldots, \sigma(T)$ into P, with Q recording when each cell of the shape was created.

The key property for the Logan–Shepp–Vershik–Kerov theorem is:

$$L_T(\sigma) \ = \ \lambda_1,$$

the length of the *first row* of the common shape λ. Thus the problem of computing $\mathbb{E}[L_T]$ reduces to computing $\mathbb{E}[\lambda_1]$ under the uniform distribution on S_T, which (via the hook-length formula on the number of SYT of each shape and variational methods applied to the resulting optimisation problem) yields $\mathbb{E}[\lambda_1] \sim 2\sqrt{T}$.

C. Tracy-Widom Distribution: Selected Quantiles

Table 2: Quantiles of the Tracy–Widom GUE distribution F_2. The statistic $Z = (\ell - 2\sqrt{T})/T^{1/6}$ is compared against these thresholds.

Level α	$q_\alpha = F_2^{-1}(\alpha)$
0.50	−1.269
0.75	−0.595
0.90	0.450
0.95	0.979
0.99	2.023
0.999	2.689

An observed run ℓ is significant at the 1% level if

$$\frac{\ell - 2\sqrt{T}}{T^{1/6}} > 2.023.$$

References

Asness, C. S., Moskowitz, T. J., and Pedersen, L. H. (2013). Value and momentum everywhere. *Journal of Finance*, 68(3):929–985.

Baik, J., Deift, P., and Johansson, K. (1999). On the distribution of the length of the longest increasing subsequence of random permutations. *Journal of the American Mathematical Society*, 12(4):1119–1178.

Barberis, N., Shleifer, A., and Vishny, R. (1998). A model of investor sentiment. *Journal of Financial Economics*, 49(3):307–343.

Dilworth, R. P. (1950). A decomposition theorem for partially ordered sets. *Annals of Mathematics*, 51(1):161–166.

Erdős, P. and Szekeres, G. (1935). A combinatorial problem in geometry. *Compositio Mathematica*, 2:463–470.

Fama, E. F. and French, K. R. (1996). Multifactor explanations of asset pricing anomalies. *Journal of Finance*, 51(1):55–84.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57(2):357–384.

Hong, H. and Stein, J. C. (1999). A unified theory of underreaction, momentum trading, and overreaction in asset markets. *Journal of Finance*, 54(6):2143–2184.

Jegadeesh, N. and Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance*, 48(1):65–91.

Jegadeesh, N. and Titman, S. (2001). Profitability of momentum strategies: An evaluation of alternative explanations. *Journal of Finance*, 56(2):699–720.

Lo, A. W. and MacKinlay, A. C. (1990). When are contrarian profits due to stock market overreaction? *Review of Financial Studies*, 3(2):175–205.

Logan, B. F. and Shepp, L. A. (1977). A variational problem for random Young tableaux. *Advances in Mathematics*, 26(2):206–222.

Marčenko, V. A. and Pastur, L. A. (1967). Distribution of eigenvalues for some sets of random matrices. *Mathematics of the USSR-Sbornik*, 1(4):457–483.

Mirsky, L. (1971). A dual of Dilworth's decomposition theorem. *American Mathematical Monthly*, 78(8):876–877.

Moskowitz, T. J., Ooi, Y. H., and Pedersen, L. H. (2012). Time series momentum. *Journal of Financial Economics*, 104(2):228–250.

Plerou, V., Gopikrishnan, P., Rosenow, B., Amaral, L. A. N., Guhr, T., and Stanley, H. E. (2002). Random matrix approach to cross correlations in financial data. *Physical Review E*, 65(6):066126.

Ramsey, F. P. (1930). On a problem of formal logic. *Proceedings of the London Mathematical Society*, 30(1):264–286.

Tracy, C. A. and Widom, H. (1994). Level-spacing distributions and the Airy kernel. *Communications in Mathematical Physics*, 159(1):151–174.

Turán, P. (1941). Eine Extremalaufgabe aus der Graphentheorie. *Mat. Fiz. Lapok*, 48:436–452.

Vershik, A. M. and Kerov, S. V. (1977). Asymptotic behaviour of the Plancherel measure of the symmetric group and the limit form of Young tableaux. *Soviet Mathematics Doklady*, 18:527–531.