

To appear in *Quantitative Finance*, Vol. 00, No. 00, Month 20XX, 1–22

Relative-entropy tests of market efficiency: symbolic benchmarks, finite alphabets, and attainable log-growth

Jan VECER^{*†}

[†] Charles University, Department of Probability and Mathematical Statistics, Sokolovska 83, 18675 Prague 8, Czech Republic

(Received 00 Month 20XX; in final form 00 Month 20XX)

Brouty and Garcin (2023) test a symbolic implication of weak-form efficiency by comparing the next-return sign after each finite history with a conditionally fair sign. This paper reconstructs their entropy argument and identifies the statistic as a conditional Kullback–Leibler divergence whose plug-in estimate is a rescaled likelihood-ratio deviance. The representation extends to finite alphabets and gives a test of whether the conditional law of the next symbol, given a history of length L , equals a specified history-free benchmark. A data-processing theorem shows that deterministic coarsening cannot increase population information. Paired simulations show that information and power need not be ordered alike: binary tests benefit from fewer degrees of freedom for sign-level alternatives, whereas four-state tests detect within-sign dependence erased by coarsening. For overlapping histories, finite-order Markov transition-likelihood theory supplies chi-square inference, local power, and resampling. When benchmark probabilities are offered as normalized state prices, the same relative entropy measures maximal expected log growth relative to cash; the complete-market result accommodates zero physical probabilities. In daily S&P 500 data, evidence around the COVID crash is strongest at the one-day, one-lag specification and weakens for longer memory or disjoint two-day returns. A stylized bounded cash/index allocation yields no statistically significant in-sample log-growth improvement over the best history-free exposure in either crisis window.

Keywords: Market efficiency, conditional independence, likelihood ratio, relative entropy, Kelly criterion, finite alphabets, S&P 500, COVID-19

JEL Classification: G14, C12, D81.

1. Introduction

Weak-form market efficiency concerns whether past prices contain exploitable information about future returns; see Fama (1970). A direct way to make this question operational is to encode returns by a finite symbolic process and examine whether the distribution of the next symbol depends on recent symbolic history. Risso (2008) use the distribution of return signs to study informational efficiency and financial crashes. Brouty and Garcin (2023) turn this symbolic approach into a formal test: they condition the next sign on a history of length L and define market information as the entropy gap between the observed length- $(L + 1)$ distribution and a benchmark with conditionally fair next signs. They also derive finite-sample and asymptotic null laws for the resulting statistic.

^{*}Corresponding author. Email: vecer@karlin.mff.cuni.cz

This paper identifies the Brouty–Garcin statistic as a conditional relative entropy. Kullback–Leibler divergence, also called relative entropy, is the expected logarithm of a likelihood ratio, and its connection with logarithmic growth is classical (Kullback and Leibler 1951, Kelly 1956, Cover and Thomas 2006). Related likelihood-ratio interpretations include transfer entropy in Barnett and Bossomaier (2012) and market-efficiency testing with quoted prediction-market probabilities in Richard and Vecer (2021). The plug-in estimate of the divergence in nats, multiplied by twice the number of transitions, is the transition-likelihood deviance; when the benchmark probabilities are offered as normalized state prices in a complete market, the divergence itself is the maximal expected Kelly log growth relative to cash. The likelihood and resampling results do not depend on this financial interpretation and apply to any finite categorical sequence satisfying the stated finite-order Markov or history-free conditions.

Benchmark choice determines which feature of the symbolic process the statistic measures. A fixed 50–50 sign benchmark records both unconditional sign imbalance and dependence on past signs. This is the appropriate specification when conditionally fair signs are the hypothesis of interest. A history-free Bernoulli or categorical benchmark absorbs the unconditional symbolic distribution and isolates dependence between the selected history and the next symbol. When the history-free benchmark equals the unconditional symbolic law, the divergence reduces to mutual information. The fixed-benchmark and conditional-independence statistics differ exactly by the divergence of the unconditional symbolic law from the fixed benchmark (Proposition 2.2).

A statistical benchmark and a pricing law play distinct economic roles. The statistical benchmark specifies the transition law tested by the likelihood ratio. A normalized vector of Arrow–Debreu prices specifies the budget law for attainable state-contingent payoffs. When the reference probabilities are normalized state prices, relative entropy is the maximal expected Kelly log growth relative to cash in the complete market. In an incomplete market, its attainable component is obtained by projecting the likelihood ratio onto the available payoff family.

The COVID-19 application belongs to a literature that studies market efficiency during the pandemic through several empirical approaches. Baker et al. (2020) document the exceptional US stock-market reaction to the pandemic, while Dima et al. (2021) study contemporaneous efficiency through the VIX. Frezza et al. (2021) infer changing local efficiency from pointwise regularity in a multifractional Brownian-motion model. Choi (2021) and Ammy-Driss and Garcin (2023) use fractal or fractional-stable persistence measures, while Okorie and Lin (2021) apply martingale-difference spectral and conditional-heteroskedasticity tests across horizons. Navratil et al. (2021) use a likelihood-ratio portfolio framework with price and pandemic covariates and examine strategy profitability. The present approach contributes a transition-likelihood measure of symbolic dependence: for a selected alphabet, history length, and benchmark, it tests whether the conditional next-symbol law equals the benchmark and quantifies the departure by relative entropy. Brouty and Garcin (2024) study how entropy-based market information changes with the return horizon and the size of the information set; their analysis motivates the scale and history-length checks reported below.

The paper makes five contributions.

- (i) It reconstructs both entropies in the original Brouty–Garcin definition and derives their exact conditional and joint relative-entropy representations.
- (ii) It proves a coarsening theorem, including equality and strictness conditions, and shows why population information ordering does not imply finite-sample power ordering. A controlled simulation compares binary and four-state tests under common paths, balanced and imbalanced marginals, two memory lengths, and two qualitatively different alternatives.
- (iii) It derives inference from the conditional likelihood of a finite-order Markov process. Consecutive history–outcome pairs may overlap; the score is a martingale difference and the usual chi-square limit follows under fixed dimension. Local power, fixed-alternative variability, bootstrap calibration, and conditional permutation inference are developed in the same transition-likelihood framework.
- (iv) It separates the physical law, a statistical benchmark, and a market-pricing law. The complete-

- market Kelly theorem accommodates zero physical probabilities and empirical zero cells, while the incomplete-market projection characterizes the log growth attainable under a pricing law.
- (v) It revisits daily S&P 500 data around COVID-19. Conditional-permutation calibration and horizon/history checks find the strongest evidence at the one-day, one-lag specification. A separate bounded index/cash calculation evaluates whether this ordering improves log growth over the best history-free exposure.

Section 2 defines the three laws and the conditional likelihood-ratio framework. Section 3 reconstructs the Brouty–Garcin statistic and proves the coarsening result. Section 4 develops transition-likelihood inference. Section 5 reports the finite-sample comparison. Section 6 treats attainable log growth under a pricing law. Section 7 presents the S&P 500 illustration, and Section 8 concludes.

2. Physical, statistical-benchmark, and pricing laws

2.1. *Histories, outcomes, and the statistical null*

Let H be a finite-valued information state or history, with relevant state space $\mathcal{H}$, and let the next symbol X take values in a finite alphabet $\mathcal{A}$. Write

$$\mu(h) := \mathbb{P}^M(H = h), \quad m_h(a) := \mathbb{P}^M(X = a \mid H = h)$$

for the physical history distribution and physical transition kernel. Superscripts identify the governing law; for example, $\mathbb{E}^M$ denotes expectation under $\mathbb{P}^M$. Histories with $\mu(h) = 0$ are excluded from the identifiable model.

A *statistical benchmark* is a specified or parametrically estimated kernel

$$b_h(a) := \mathbb{P}^B(X = a \mid H = h).$$

For the one-step comparison, define the physical and benchmark joint laws by

$$\mathbb{P}_{H,X}^M(h, a) := \mu(h)m_h(a), \quad \mathbb{P}_{H,X}^B(h, a) := \mu(h)b_h(a). \quad (1)$$

Thus $\mathbb{P}^B$ retains, by construction, the physical history marginal and changes only the next-outcome conditional law. The statistical null is

$$H_0 : \quad m_h(a) = b_h(a) \quad \text{for every relevant } (h, a). \quad (2)$$

DEFINITION 1 (Kullback–Leibler divergence and conditional benchmark information) For probability laws $\mathbb{P}_1$ and $\mathbb{P}_0$ on a finite set Ω , their Kullback–Leibler divergence, also called relative entropy, is

$$D_{\text{KL}}(\mathbb{P}_1 \parallel \mathbb{P}_0) := \begin{cases} \sum_{\omega: \mathbb{P}_1(\omega) > 0} \mathbb{P}_1(\omega) \log \frac{\mathbb{P}_1(\omega)}{\mathbb{P}_0(\omega)}, & \mathbb{P}_1 \ll \mathbb{P}_0, \\ +\infty, & \text{otherwise.} \end{cases}$$

The same notation applies to probability vectors. Its base-2 form is

$$D_2(\mathbb{P}_1 \parallel \mathbb{P}_0) := \frac{D_{\text{KL}}(\mathbb{P}_1 \parallel \mathbb{P}_0)}{\log 2}.$$

The history-specific benchmark information is

$$\mathcal{I}_h(\mathbb{P}^M \parallel \mathbb{P}^B) := D_{\text{KL}}(m_h \parallel b_h).$$

When $\mathbb{P}_{H,X}^M \ll \mathbb{P}_{H,X}^B$, its physical-history average is

$$\mathcal{I}(\mathbb{P}^M \parallel \mathbb{P}^B) := \sum_{h \in \mathcal{H}} \mu(h) D_{\text{KL}}(m_h \parallel b_h) = D_{\text{KL}}(\mathbb{P}_{H,X}^M \parallel \mathbb{P}_{H,X}^B). \quad (3)$$

Zero numerator terms contribute zero; a positive numerator over a zero denominator gives $+\infty$. Natural logarithms give nats, while D_2 gives bits.

In the principal empirical benchmark, $b_h(a) = \eta_a$ is history free. Then (2) specifies a time-homogeneous next-symbol law that is independent of histories of the selected length on the selected alphabet.

2.2. A separate market-pricing law and the Kelly rule

Suppose instead that one-period Arrow–Debreu claims are traded. If $\pi_h(a)$ is the undiscounted price of one unit in state a and B_h is the risk-free gross return, the normalized state-price vector is

$$q_h(a) := B_h \pi_h(a), \quad \sum_a q_h(a) = 1.$$

It defines a *pricing law* $\mathbb{P}^Q$. There is no general reason for q_h to equal either m_h or a statistical benchmark b_h .

The finite-outcome state-price formulation is closely related to the analytical multi-outcome Kelly solution in Vecer (2026). The result below states the complete-market rule with support conditions that remain valid when some physical probabilities, including empirical plug-in probabilities, are zero.

For a fixed history, let $W_h(a)$ be terminal gross wealth generated from one unit of initial wealth and define the cash-relative payoff $\ell_h(a) := W_h(a)/B_h$. The undiscounted budget $\sum_a \pi_h(a) W_h(a) = 1$ is then equivalently

$$\sum_{a \in \mathcal{A}} q_h(a) \ell_h(a) = 1. \quad (4)$$

Thus the logarithmic objective below is excess log growth relative to the cash account.

THEOREM 2.1 (Complete-market Kelly rule) *Fix h and suppose $q_h(a) > 0$ for all a , while $m_h(a) \geq 0$. Admit payoff profiles $\ell_h(a) \geq 0$ satisfying (4), with $\ell_h(a) > 0$ whenever $m_h(a) > 0$. Interpret $0 \log 0 = 0$ and $m \log 0 = -\infty$ for $m > 0$. Then*

$$G(h; \ell_h) := \sum_a m_h(a) \log \ell_h(a)$$

is uniquely maximized by

$$\ell_h^*(a) = \frac{m_h(a)}{q_h(a)}. \quad (5)$$

Equivalently, the normalized state-contingent expenditure weights satisfy $q_h(a) \ell_h^*(a) = m_h(a)$: the optimal fraction spent on state a equals its physical probability. The maximal expected excess log

return relative to cash is

$$G^*(h) = D_{\text{KL}}(m_h \| q_h). \quad (6)$$

Proof. For any admissible ℓ_h , the vector $v_h(a) := q_h(a)\ell_h(a)$ is a probability vector. Therefore

$$\sum_a m_h(a) \log \ell_h(a) = \sum_a m_h(a) \log \frac{m_h(a)}{q_h(a)} - \sum_a m_h(a) \log \frac{m_h(a)}{v_h(a)} = D_{\text{KL}}(m_h \| q_h) - D_{\text{KL}}(m_h \| v_h).$$

Gibbs' inequality gives the bound, and equality holds only when $v_h = m_h$. This is equivalent to (5), including its zero components. $\square$

Remark 1 (Strictly positive wealth) If payoff profiles are required to be strictly positive in every state, the same value is attained exactly when m_h has full support. Otherwise it is a supremum, approached for example by replacing m_h with $(1 - \varepsilon)m_h + \varepsilon q_h$ and sending $\varepsilon \downarrow 0$.

Theorem 2.1 applies to the statistical divergence (3) when the benchmark probabilities are offered as normalized state prices, $b_h = q_h$, or when a hypothetical market quoting those odds is considered explicitly. The empirical history-free benchmark of Section 7 is a statistical null, not a pricing law.

2.3. What a history-free benchmark removes

Let $b^0 = (b^0(a))_{a \in \mathcal{A}}$ be a fixed history-free benchmark and let

$$\bar{m}(a) := \sum_h \mu(h) m_h(a)$$

be the unconditional next-symbol law under $\mathbb{P}^M$.

PROPOSITION 2.2 (Drift-dependence decomposition) *If $\bar{m} \ll b^0$, then*

$$\sum_h \mu(h) D_{\text{KL}}(m_h \| b^0) = I_{\mathbb{P}^M}(H; X) + D_{\text{KL}}(\bar{m} \| b^0), \quad (7)$$

where $I_{\mathbb{P}^M}(H; X) = \sum_h \mu(h) D_{\text{KL}}(m_h \| \bar{m})$ is mutual information under the physical law.

Proof. Insert $\bar{m}(a)$ between $m_h(a)$ and $b^0(a)$ inside the logarithm and sum. The term containing $\log\{\bar{m}(a)/b^0(a)\}$ averages to $D_{\text{KL}}(\bar{m} \| b^0)$; the remaining term is mutual information. $\square$

Thus the fixed 50–50 binary statistic combines history dependence with unconditional sign imbalance. Choosing or estimating $b^0 = \bar{m}$ leaves the symbolic conditional-independence target $I_{\mathbb{P}^M}(H; X)$. Since the identity holds for any finite alphabet, the fitted four-state benchmark of Section 7.3 also targets mutual information.

2.4. Dynamic path comparison

Let $X_1, \dots, X_N$ be a symbolic path and let H_t be the information available before X_{t+1} , for example the last L symbols. For the pathwise factorization, assume that H_t is sufficient for the physical one-step law,

$$\mathbb{P}^M(X_{t+1} = a \mid X_{1:t}) = m_t(a \mid H_t) \quad \mathbb{P}^M\text{-a.s.};$$

this holds in particular when H_t is the full past or the physical process is Markov of the stated order. Let $b_t(a | h)$ be the benchmark one-step kernel, and let $\mathbb{P}_N^M$ denote the physical law of $X_{1:N}$. Define the comparison path law

$$\mathbb{P}_N^B(x_{1:N}) := \mathbb{P}^M(x_{1:L}) \prod_{t=L}^{N-1} b_t(x_{t+1} | h_t). \quad (8)$$

This construction keeps the physical initial block and replaces subsequent kernels. It is a statistical comparison law, not automatically an iid path law or a pricing measure.

THEOREM 2.3 (Pathwise likelihood ratio) *Suppose $\mathbb{P}_N^M \ll \mathbb{P}_N^B$. Then, $\mathbb{P}_N^M$ -almost surely,*

$$\frac{d\mathbb{P}_N^M}{d\mathbb{P}_N^B}(X_{1:N}) = \prod_{t=L}^{N-1} \frac{m_t(X_{t+1} | H_t)}{b_t(X_{t+1} | H_t)}, \quad (9)$$

and

$$D_{\text{KL}}(\mathbb{P}_N^M \| \mathbb{P}_N^B) = \sum_{t=L}^{N-1} \mathbb{E}^M[D_{\text{KL}}\{m_t(\cdot | H_t) \| b_t(\cdot | H_t)\}]. \quad (10)$$

If the benchmark kernels are normalized traded state prices, $b_t = q_t$, write $\mathbb{P}_N^Q$ for the corresponding path law. If W_t denotes complete-market Kelly wealth measured in units of the cash account, then

$$\frac{W_N}{W_L} = \frac{d\mathbb{P}_N^M}{d\mathbb{P}_N^Q}(X_{1:N}), \quad \mathbb{E}^M \left[\log \frac{W_N}{W_L} \right] = D_{\text{KL}}(\mathbb{P}_N^M \| \mathbb{P}_N^Q).$$

Proof. The chain rule for path probabilities gives (9). Taking logarithms and physical expectations gives the path divergence. Conditioning each summand on H_t yields (10). The final statement follows by applying Theorem 2.1 at each date. $\square$

3. The Brouty–Garcin statistic and deterministic coarsening

3.1. The original entropy construction

For a finite probability vector $r = (r_1, \dots, r_m)$, define Shannon entropy in bits, following Shannon (1948), by

$$H_2(r) := - \sum_{j=1}^m r_j \log_2 r_j, \quad 0 \log_2 0 := 0. \quad (11)$$

Let $X_t \in \{0, 1\}$ be the sign process, fix a history length L , and enumerate the 2^L prefixes as $G_1^L, \dots, G_{2^L}^L$. Prefixes with zero probability are omitted from conditional expressions; entropy sums retain them under $0 \log 0 = 0$. Write

$$p_i := \mathbb{P}^M(H = G_i^L), \quad \pi_i := \mathbb{P}^M(X = 1 | H = G_i^L) \quad (p_i > 0).$$

When $p_i = 0$, assign π_i arbitrarily; every occurrence is multiplied by p_i . The prefix entropy is

$$H_L := - \sum_{i=1}^{2^L} p_i \log_2 p_i,$$

and the binary entropy function is

$$h_2(u) := -u \log_2 u - (1 - u) \log_2(1 - u).$$

Under the physical law, the entropy of the length- $(L + 1)$ sequence is

$$\begin{aligned} H_{L+1}^M &= - \sum_{i=1}^{2^L} [p_i \pi_i \log_2(p_i \pi_i) + p_i(1 - \pi_i) \log_2\{p_i(1 - \pi_i)\}] \\ &= H_L + \sum_{i=1}^{2^L} p_i h_2(\pi_i). \end{aligned} \quad (12)$$

The conditionally fair one-step benchmark retains p_i and assigns probability $1/2$ to each next sign. Its length- $(L + 1)$ entropy is

$$H_{L+1}^B = - \sum_{i=1}^{2^L} \left[\frac{p_i}{2} \log_2 \frac{p_i}{2} + \frac{p_i}{2} \log_2 \frac{p_i}{2} \right] = H_L + 1. \quad (13)$$

Brouty and Garcin define market information as

$$I_{L+1} := H_{L+1}^B - H_{L+1}^M.$$

Equations (12)–(13) make the cancellation of the prefix entropy explicit.

PROPOSITION 3.1 (Exact relative-entropy representation) *Define one-step joint laws by*

$$\mathbb{P}^M(H = G_i^L, X = 1) = p_i \pi_i, \quad \mathbb{P}^M(H = G_i^L, X = 0) = p_i(1 - \pi_i),$$

and

$$\mathbb{P}^B(H = G_i^L, X = 1) = \mathbb{P}^B(H = G_i^L, X = 0) = \frac{p_i}{2}.$$

Then

$$\begin{aligned} I_{L+1} &= 1 - \sum_{i=1}^{2^L} p_i h_2(\pi_i) \\ &= \sum_{i=1}^{2^L} p_i \left[\pi_i \log_2 \frac{\pi_i}{1/2} + (1 - \pi_i) \log_2 \frac{1 - \pi_i}{1/2} \right] \\ &= D_2(\mathbb{P}_{H,X}^M \| \mathbb{P}_{H,X}^B). \end{aligned} \quad (14)$$

Proof. The first two lines follow directly from (12)–(13). Expanding the joint divergence gives

$$D_2(\mathbb{P}_{H,X}^M \| \mathbb{P}_{H,X}^B) = \sum_{i:p_i>0} p_i \pi_i \log_2 \frac{p_i \pi_i}{p_i/2} + \sum_{i:p_i>0} p_i(1 - \pi_i) \log_2 \frac{p_i(1 - \pi_i)}{p_i/2}.$$

Zero-mass prefixes contribute zero and are omitted from these sums. The positive prefix probabilities cancel inside each logarithm, leaving the second line of (14). $\square$

The law $\mathbb{P}_{H,X}^B$ in Proposition 3.1 is a one-step conditional benchmark: it preserves the physical prefix probabilities and changes only the next-sign kernel. It agrees with the corresponding block law of an iid fair-sign process exactly when the prefix law is also iid fair. A pricing interpretation requires traded fair odds.

COROLLARY 3.2 (Hypothetical fair-odds interpretation) *If, after history G_i^L , up and down Arrow–Debreu claims are each offered at normalized price $1/2$, the Kelly payoff is*

$$\ell_i^*(1) = 2\pi_i, \quad \ell_i^*(0) = 2(1 - \pi_i),$$

and averaging its conditional expected log return over prefixes with weights p_i gives I_{L+1} bits per transition. Zero payoffs are admissible on physical-null branches.

Proof. Apply Theorem 2.1 with physical probabilities $(\pi_i, 1 - \pi_i)$ and normalized state prices $(1/2, 1/2)$, then average the conditional values over prefixes with weights p_i and divide by $\log 2$. $\square$

The non-negativity result in Brouty and Garcin (2023) is therefore an instance of Gibbs’ inequality. Equality holds exactly when $\pi_i = 1/2$ for every prefix of positive mass. The entropy-gap representation works because the fair Bernoulli law is the unique maximum-entropy binary law; for a general benchmark, equality of entropy does not identify equality of distributions.

3.2. Fine and coarse alphabets

The finite-alphabet extension resolves the entropy-level ambiguity noted by Brouty and Garcin (2023): relative entropy compares complete probability vectors once a benchmark is specified. It does not, however, imply that a finer test is uniformly more powerful.

THEOREM 3.3 (Information under deterministic coarsening) *Let $\mathbb{P}_1 \ll \mathbb{P}_0$ be probability laws on a finite set Ω , let $T : \Omega \rightarrow \mathcal{Z}$ be deterministic, and write $\mathbb{P}_{j,T} := \mathbb{P}_j \circ T^{-1}$ for $j \in \{0, 1\}$. Then*

$$\begin{aligned} D_{\text{KL}}(\mathbb{P}_1 \| \mathbb{P}_0) &= D_{\text{KL}}(\mathbb{P}_{1,T} \| \mathbb{P}_{0,T}) + \sum_{z: \mathbb{P}_{1,T}(z) > 0} \mathbb{P}_{1,T}(z) D_{\text{KL}}\{\mathbb{P}_1(\cdot | T = z) \| \mathbb{P}_0(\cdot | T = z)\} \\ &\geq D_{\text{KL}}(\mathbb{P}_{1,T} \| \mathbb{P}_{0,T}). \end{aligned} \tag{15}$$

Equality holds if and only if $\mathbb{P}_1(\cdot | T = z) = \mathbb{P}_0(\cdot | T = z)$ for every z of positive $\mathbb{P}_{1,T}$ -mass; equivalently, $d\mathbb{P}_1/d\mathbb{P}_0$ is constant on every relevant fibre of T . The inequality is strict whenever this condition fails on a fibre of positive mass.

Proof. Factor the likelihood ratio as

$$\frac{\mathbb{P}_1(\omega)}{\mathbb{P}_0(\omega)} = \frac{\mathbb{P}_{1,T}(T\omega)}{\mathbb{P}_{0,T}(T\omega)} \frac{\mathbb{P}_1(\omega | T\omega)}{\mathbb{P}_0(\omega | T\omega)},$$

take $\mathbb{P}_1$ -expectations, and apply Gibbs’ inequality to the conditional divergences. $\square$

For symbolic histories, let $\phi : \mathcal{A}_f \rightarrow \mathcal{A}_c$ aggregate fine states and take

$$T(h, a) = (\phi^L(h), \phi(a))$$

on the joint history–outcome space. Applying Theorem 3.3 to the physical and benchmark joint laws compares the correct fine and coarse conditional divergences. The coarse benchmark is the conditional law induced by pushing forward the fine benchmark; it is not chosen independently after aggregation.

Remark 2 (Information monotonicity does not imply power monotonicity) Coarsening changes both the alternative signal and the model dimension, and hence the critical value, so Theorem 3.3 alone gives no power ordering. With an unknown history-free marginal, the fine K -state test has $(K^L - 1)(K - 1)$ degrees of freedom, while a J -state coarsening has $(J^L - 1)(J - 1)$. For $K = 4$, $J = 2$, and $L = 1$, these are 9 and 1. Along a sign-only alternative whose likelihood ratio is constant within the two sign fibres, fine and binary population information – and hence local noncentrality – are equal, but the binary test has the smaller critical value. Along a within-sign magnitude alternative, the coarse divergence can be zero while the fine divergence is positive. Thus neither test uniformly dominates. Section 5 gives a controlled finite-sample comparison.

4. Transition-likelihood inference with overlapping histories

4.1. Counts, plug-in information, and deviance

The maintained sampling model in this section is an order- L Markov process: the conditional law given the full past depends on it only through the displayed L -symbol history. Within that model, a history-free categorical kernel makes subsequent symbols iid conditional on the initial block; with iid initialization, as imposed for the full-sequence permutation null, the entire path is iid categorical. Longer-memory or nonstationary data-generating processes lie outside the null calibration; the statistic still measures only dependence visible through the selected history and alphabet.

Let $X_1, \dots, X_N$ take values in a fixed K -symbol alphabet, fix a memory length L , and put

$$H_t = (X_{t-L+1}, \dots, X_t), \quad t = L, \dots, N-1.$$

The number of usable transitions is

$$n := N - L. \quad (16)$$

Define

$$N_{h,a} := \sum_{t=L}^{N-1} \mathbf{1}_{\{H_t=h, X_{t+1}=a\}}, \quad N_h := \sum_a N_{h,a}, \quad \hat{m}_h(a) := \frac{N_{h,a}}{N_h}$$

when $N_h > 0$. Consecutive pairs overlap, but

$$\ell_n(m) = \sum_{t=L}^{N-1} \log m_{H_t}(X_{t+1}) = \sum_{h,a} N_{h,a} \log m_h(a) \quad (17)$$

is the ordinary conditional likelihood of an order- L Markov kernel.

Let $b_h(a; \theta)$ be a regular benchmark family and let $\hat{\theta}$ be its conditional maximum-likelihood estimator. The plug-in benchmark divergence is

$$\hat{\mathcal{I}}_n := \sum_{h: N_h > 0} \frac{N_h}{n} \sum_a \hat{m}_h(a) \log \frac{\hat{m}_h(a)}{b_h(a; \hat{\theta})}, \quad (18)$$

with $0 \log(0/y) = 0$. The same quantity in bits is denoted by $\hat{I}_2 := \hat{\mathcal{I}}_n / \log 2$.

PROPOSITION 4.1 (Transition-deviance identity) *Let $\hat{m}$ maximize (17) over the closed product of*

transition simplices. Then

$$\Lambda_n := 2\{\ell_n(\hat{m}) - \ell_n(b(\hat{\theta}))\} = 2 \sum_{h,a} N_{h,a} \log \frac{\hat{m}_h(a)}{b_h(a; \hat{\theta})} = 2n\hat{\mathcal{I}}_n. \quad (19)$$

Proof. At the unrestricted maximum, $\hat{m}_h(a) = N_{h,a}/N_h$. Substituting $N_{h,a} = N_h \hat{m}_h(a)$ into the conditional log-likelihood difference gives the middle expression in (19); factoring out n gives the final equality. $\square$

4.2. Wilks theory for overlapping transitions

THEOREM 4.2 (Conditional Markov Wilks theorem) *Let K and L be fixed as $N \rightarrow \infty$. Suppose the true kernel belongs to a null family $\{b(\cdot; \theta) : \theta \in \Theta_0 \subset \mathbb{R}^s\}$, where θ_0 is interior, all $b_h(a; \theta_0) > 0$, the kernel is three times continuously differentiable near θ_0 , and its derivative has rank s . Assume the induced chain of histories is irreducible and aperiodic. The initial block may have an arbitrary distribution. Then*

$$\Lambda_n \Rightarrow \chi_{r-s}^2, \quad r = K^L(K-1). \quad (20)$$

Proof. Represent the process as a first-order chain on the K^L histories. An allowed transition appends the new symbol and drops the oldest one. In smooth coordinates, the conditional score increment at time t is a martingale difference because

$$\mathbb{E}_{\theta_0}[\partial_\theta \log b_{H_t}(X_{t+1}; \theta_0) \mid \mathcal{F}_t] = \sum_a b_{H_t}(a; \theta_0) \partial_\theta \log b_{H_t}(a; \theta_0) = 0.$$

The martingale central limit theorem and ergodic theorem yield asymptotic normality of the score and convergence of observed information. Positivity and the rank condition make the information nonsingular on the r -dimensional full tangent space. The quadratic expansion of the conditional likelihood, maximized over the full tangent space and its s -dimensional null subspace, gives the chi-square law with codimension $r - s$. The initial-block likelihood is $O(1)$. This is the transition-likelihood analogue of the classical likelihood-ratio result of Wilks (1938); the finite-order Markov argument follows Anderson and Goodman (1957) and Billingsley (1961). See also the information-theoretic formulation in Kontoyiannis and Skoularidou (2016). $\square$

Finite samples may contain empty cells, so the unrestricted MLE can lie on the boundary. Under the positive fixed-dimensional null, every expected cell count diverges and the probability of an empty cell tends to zero. Histories of zero population probability are not identifiable and are excluded from r .

COROLLARY 4.3 (History-free categorical benchmarks) *For the history-free null $b_h(a; \eta) = \eta_a$, the nuisance rank is $s = K - 1$, the conditional MLE is*

$$\hat{\eta}_a = \frac{1}{n} \sum_h N_{h,a},$$

and

$$\Lambda_n \Rightarrow \chi_{(K^L-1)(K-1)}^2. \quad (21)$$

For a fixed benchmark, $s = 0$ and the degrees of freedom are $K^L(K - 1)$.

Proof. The unrestricted order- L kernel has $K^L(K-1)$ free parameters. A common categorical row has $K-1$ parameters, whose conditional MLE is the pooled next-symbol frequency shown above; subtracting these dimensions gives (21). $\square$

For binary signs, a fixed fair benchmark has 2^L degrees of freedom; an estimated history-free Bernoulli benchmark has $2^L - 1$. For $K = 4, L = 1$, the estimated history-free categorical benchmark has 9 degrees of freedom.

4.3. The Brouty–Garcin limit

Under the fixed binary fair-sign null,

$$\Lambda_n = 2n \log 2 \hat{I}_{L+1}.$$

COROLLARY 4.4 (Brouty–Garcin asymptotics) *Under the assumptions of Theorem 4.2,*

$$2n \log 2 \hat{I}_{L+1} \Rightarrow \chi_{2^L}^2. \quad (22)$$

Equivalently, with the gamma law parameterized by shape and scale,

$$n \hat{I}_{L+1} \Rightarrow \Gamma\left(2^{L-1}, \frac{1}{\log 2}\right). \quad (23)$$

Proof. Corollary 4.3 with a fixed binary fair-sign benchmark gives (22). Since $\chi_{2^L}^2$ is $\Gamma(2^{L-1}, 2)$ in shape–scale notation, division by $2 \log 2$ yields (23). $\square$

This recovers the asymptotic component of Brouty and Garcin (2023). Their exact moment-generating function and error bounds give finite-sample information for the binary fair-sign model that the limit theorem does not.

4.4. Contiguous alternatives and local power

Let $b^0 = b(\cdot; \theta_0)$ satisfy Theorem 4.2, and let $\pi_0(h)$ be the null stationary probability of history h . On the transition-kernel tangent space define

$$\langle u, v \rangle_0 := \sum_h \pi_0(h) \sum_a \frac{u_h(a) v_h(a)}{b_h^0(a)}, \quad \sum_a u_h(a) = \sum_a v_h(a) = 0.$$

Let $\mathcal{T}_0 = \{\dot{b}_{\theta_0} c : c \in \mathbb{R}^s\}$ be the null tangent space and Π_0 its orthogonal projection in this inner product.

PROPOSITION 4.5 (Local noncentral chi-square power) *Suppose the data are generated by kernels*

$$m_h^{(n)}(a) = b_h^0(a) + n^{-1/2} \delta_h(a) + o(n^{-1/2}), \quad \sum_a \delta_h(a) = 0, \quad (24)$$

uniformly over (h, a) . Then

$$\Lambda_n \Rightarrow \chi_{r-s}^2(\lambda), \quad \lambda = \|\delta - \Pi_0 \delta\|_0^2. \quad (25)$$

For a fixed benchmark,

$$\lambda = \sum_h \pi_0(h) \sum_a \frac{\delta_h(a)^2}{b_h^0(a)}.$$

For a history-free categorical benchmark $b_h^0(a) = \eta_a$, let $\bar{\delta}(a) = \sum_h \pi_0(h) \delta_h(a)$. Then

$$\lambda = \sum_h \pi_0(h) \sum_a \frac{\{\delta_h(a) - \bar{\delta}(a)\}^2}{\eta_a}, \quad \pi_0(h) = \prod_{j=1}^L \eta_{h_j}.$$

Proof. The martingale-score expansion makes the conditional Markov likelihood locally asymptotically normal. Le Cam's third lemma shifts the Gaussian central sequence under (24). The likelihood-ratio statistic is the squared Fisher norm of its component orthogonal to the nuisance tangent space, and therefore has the stated noncentral chi-square and not Gaussian limit; see van der Vaart (1998) for the general LAN framework. $\square$

Writing $F_{\nu, \lambda}$ for the cdf of a noncentral chi-square variable with ν degrees of freedom and noncentrality parameter λ , the local asymptotic power of a level- α test is

$$1 - F_{r-s, \lambda}(\chi_{r-s, 1-\alpha}^2).$$

4.5. Variability under fixed alternatives

PROPOSITION 4.6 (Fixed-alternative central limit theorem) *Assume a fixed positive, irreducible, and aperiodic order- L physical kernel m , and let π_M denote the stationary distribution of its induced history chain. Let θ^* be the unique interior maximizer of $\mathbb{E}^M \log b_{H_t}(X_{t+1}; \theta)$ and suppose $b_h(a; \theta^*) > 0$ for all (h, a) . Define*

$$\mathcal{I}^* := \sum_h \pi_M(h) \sum_a m_h(a) \log \frac{m_h(a)}{b_h(a; \theta^*)}.$$

Then

$$\sqrt{n}\{\widehat{\mathcal{I}}_n - \mathcal{I}^*\} \Rightarrow N(0, \sigma^2), \tag{26}$$

where, for the stationary chain,

$$g_t := \log \frac{m_{H_t}(X_{t+1})}{b_{H_t}(X_{t+1}; \theta^*)} - \mathcal{I}^*, \quad \sigma^2 := \text{Var}(g_0) + 2 \sum_{j \geq 1} \text{Cov}(g_0, g_j).$$

The covariance series is absolutely convergent for a positive finite-state kernel.

Proof. Apply the finite-state Markov-chain central limit theorem to the empirical transition measure and then the delta method. The derivative of conditional relative entropy with respect to a transition cell is the displayed log likelihood ratio. Derivatives through θ^* vanish by the envelope theorem. $\square$

At the null, the first derivative degenerates and Theorem 4.2 supplies the n -rate limit. If the corresponding first two moments are uniformly integrable, then, in nats and with $d = r - s$,

$$\mathbb{E}_0(\widehat{\mathcal{I}}_n) = \frac{d}{2n} + o(n^{-1}), \quad \text{Var}_0(\widehat{\mathcal{I}}_n) = \frac{d}{2n^2} + o(n^{-2}). \tag{27}$$

In bits, divide the mean by $\log 2$ and the variance by $(\log 2)^2$.

4.6. Finite-sample calibration and sparse tables

For fixed K, L , an interior consistently estimated null parameter, and a positive ergodic kernel, a fitted parametric bootstrap is valid: keep the observed initial L -symbol history fixed, simulate the next n symbols under $b(\hat{\theta})$, reconstruct the overlapping histories, refit both null and alternative, and recompute Λ_n ; see Efron and Tibshirani (1993) for general bootstrap principles. If R_{boot} paths are simulated, we use

$$\hat{p}_{\text{boot}} = \frac{1 + \#\{\Lambda_r^* \geq \Lambda_{\text{obs}}\}}{R_{\text{boot}} + 1}. \quad (28)$$

For the unknown history-free categorical null there is a stronger finite-sample calibration.

PROPOSITION 4.7 (Conditional permutation test) *Suppose $X_1, \dots, X_N$ are iid categorical under the history-free null. Conditional on the total category counts, all distinct orderings are equiprobable. Consequently, permuting the full symbolic sequence, reconstructing all overlapping histories, re-estimating the history-free marginal from the $n = N - L$ next symbols, and recomputing any transition statistic gives its exact conditional null distribution. Full enumeration is exact; the Monte Carlo p -value obtained from (28) is finite-sample valid.*

Proof. Under the iid categorical null, every sequence with the observed category counts has the same probability, so the conditional law is uniform over its distinct permutations. Reconstructing histories and re-estimating the nuisance marginal after each permutation applies the same statistic to every element of this conditional sample space; enumeration is therefore exact, and the add-one Monte Carlo rule is finite-sample valid. $\square$

The permutation test automatically preserves marginal imbalance, sparse category counts, and the overlap structure. If one wishes to allow an arbitrary initial block, the same statement holds conditionally on that block after permuting the remaining symbols conditional on their counts.

All asymptotic statements in this paper keep K and L fixed. Growing alphabets or memories would require a triangular-array analysis, at minimum $np_{\min, n} \rightarrow \infty$ for the smallest relevant transition-cell probability and additional restrictions on model dimension. Adjusting the degrees of freedom to the number of observed histories does not account for empty outcome cells; in sparse designs we rely on resampling.

Equation (27) motivates the first-order null-centering amount

$$c_{d,n} := \frac{d}{2n \log 2} \quad \text{bits}. \quad (29)$$

Classical first-order bias corrections for plug-in entropy estimators originate with Miller (1955); Verdú (2019) surveys the broader information-measure estimation literature. Here $c_{d,n}$ only centers the statistic under the null; it is not a bias correction under fixed or local alternatives or in sparse tables. We report $\max\{0, \hat{I}_2 - c_{d,n}\}$ as a sensitivity measure, and inference uses the raw deviance.

5. Finite-sample comparison of binary and four-state tests

The simulation uses common four-state paths and coarsens those same paths to binary signs. The fine states are ordered (D_2, D_1, U_1, U_2) and indexed $j = 0, \dots, 3$: large down, small down, small up, and large up. The deterministic coarsening maps the first two states to down and the last two to

up. We consider balanced marginal probabilities $(0.25, 0.25, 0.25, 0.25)$ and asymmetric probabilities $(0.10, 0.20, 0.30, 0.40)$.

There are three stationary first-order data-generating mechanisms. Under the null, every transition row equals the marginal p . For the sign-only alternative, if $g(i) \in \{0, 1\}$ is the sign group, $s_g = \sum_{j:g(j)=g} p_j$, and $\rho \in (0, 1)$,

$$T_{ij}^{\text{sign}} = \left[(1 - \rho)s_{g(j)} + \rho \mathbf{1}_{\{g(i)=g(j)\}} \right] \frac{p_j}{s_{g(j)}}. \quad (30)$$

The likelihood ratio is constant within each sign fibre, so binary coarsening retains all population information. For the within-sign magnitude alternative, set

$$v = \left(1, -\frac{p_0}{p_1}, 1, -\frac{p_2}{p_3} \right), \quad T_{ij}^{\text{mag}} = p_j(1 + \theta v_i v_j).$$

Each sign total remains history free, so the binary population divergence is zero. In each marginal design, ρ and θ are calibrated so that the fine conditional information is exactly 0.02 nats per transition. The common value gives one effect-size comparison; the information contrast itself does not depend on it: for every admissible ρ , the sign-only likelihood ratio is constant within sign fibres and the binary and fine divergences coincide; for every admissible θ , the within-sign construction leaves the binary divergence exactly zero. All transition probabilities remain positive and $pT = p$.

We vary the number of usable overlapping transitions $n \in \{50, 100, 250, 500\}$, alphabet size $K \in \{2, 4\}$, and tested memory $L \in \{1, 2\}$. Each path has $n + 2$ symbols, and all four tests use the same n target symbols; the $L = 1$ tests therefore omit the first usable transition. Every test estimates its history-free categorical marginal. For each design point, 2,000 outer replications are run; each paired fitted-null bootstrap uses 999 paths, holds the observed initial two symbols fixed (only the last L enter a given statistic), and re-estimates the nuisance marginal. The nominal level is 5%, and the pseudorandom seed is 20260715. The Monte Carlo standard error of a size estimate near 5% is about 0.5 percentage points.

Table 1. Empirical null rejection frequencies. Each cell reports asymptotic/parametric-bootstrap size at nominal 5%. The columns identify alphabet size K and memory L .

Marginal	n	$K = 2, L = 1$	$K = 2, L = 2$	$K = 4, L = 1$	$K = 4, L = 2$
Balanced	50	.060/.048	.059/.048	.104/.055	.073/.090
	100	.050/.049	.060/.055	.061/.043	.234/.065
	250	.053/.055	.051/.049	.051/.043	.115/.042
	500	.046/.044	.057/.053	.050/.048	.068/.045
Asymmetric	50	.068/.057	.066/.061	.074/.051	.006/.033
	100	.054/.051	.060/.052	.083/.062	.053/.046
	250	.056/.054	.057/.052	.070/.057	.102/.050
	500	.053/.051	.050/.049	.051/.044	.095/.053

Table 1 shows that asymptotic calibration can be badly distorted in sparse four-state, two-lag tables. At $K = 4, L = 2$ there are 45 degrees of freedom. In the balanced design, the proportions of null samples containing at least one empty transition cell at $n = 50, 100, 250, 500$ are 1.000, 1.000, .714, .024; under the asymmetric marginal they are 1.000, 1.000, 1.000, .971. The non-monotone size pattern reflects competing boundary effects: under extreme sparsity, unvisited histories and cells can reduce the realized dimension below the nominal 45 and make the chi-square reference conservative, whereas partially filled rows can place excess mass in the deviance's right tail and produce over-rejection before the fixed-dimensional approximation stabilizes. The fitted bootstrap is much closer to 5% in most cases, but its 9.0% size in the most sparse balanced $n = 50$ design

shows that the fitted bootstrap itself can be inaccurate when the table is nearly uninformative. Such a specification should be avoided.

Table 2. Fitted-bootstrap rejection frequencies under the two alternatives for $L = 1$. Each cell reports binary/four-state power; both alternatives contain 0.02 nats of fine-state conditional information.

Marginal	Alternative	$n = 50$	$n = 100$	$n = 250$	$n = 500$
Balanced	Sign only	.212/.089	.459/.186	.864/.530	.992/.898
Balanced	Within sign	.049/.097	.041/.174	.052/.529	.057/.904
Asymmetric	Sign only	.232/.139	.449/.183	.873/.535	.991/.878
Asymmetric	Within sign	.052/.129	.052/.197	.049/.519	.052/.892

Table 3. Fitted-bootstrap rejection frequencies under the two alternatives for $L = 2$. Each cell reports binary/four-state power; the design otherwise matches Table 2. The $n = 50$ four-state entries should be read with the severe-sparsity warning from Table 1.

Marginal	Alternative	$n = 50$	$n = 100$	$n = 250$	$n = 500$
Balanced	Sign only	.141/.089	.315/.120	.721/.237	.965/.549
Balanced	Within sign	.046/.086	.046/.116	.047/.227	.056/.561
Asymmetric	Sign only	.179/.077	.323/.143	.731/.285	.964/.560
Asymmetric	Within sign	.061/.054	.047/.113	.049/.276	.055/.557

Because Table 1 shows material size distortion for several asymptotic tests, Tables 2–3 compare power using fitted-bootstrap calibration.

The power reversal in Tables 2–3 is the finite-sample counterpart of Remark 2. Sign-only dependence survives aggregation with no information loss; the binary test is generally more powerful because it pays for fewer degrees of freedom. Within-sign magnitude dependence is eliminated by aggregation, so binary rejection remains near size while four-state power rises with n . Increasing the tested memory to $L = 2$ adds no population information for these first-order alternatives but raises the degrees of freedom from 1 to 3 and from 9 to 45, reducing power in the informative designs. Alphabet and memory should therefore be chosen for the alternative of interest and kept within what the sample size supports.

6. Attainable log growth under a pricing law

The statistical benchmark $\mathbb{P}^B$ defines a null; a trading interpretation requires prices. Let $\mathbb{P}^Q$ be a normalized pricing law for the available payoffs on a finite state space Ω , assume $\mathbb{P}^M \ll \mathbb{P}^Q$, and let $\mathcal{V}$ be a family of strictly positive attainable cash-relative payoff densities satisfying

$$\mathbb{E}^Q[v] = 1, \quad 1 \in \mathcal{V}. \quad (31)$$

Each $v \in \mathcal{V}$ induces an attainable implied law $\mathbb{P}^V$ by

$$\frac{d\mathbb{P}^V}{d\mathbb{P}^Q} = v. \quad (32)$$

PROPOSITION 6.1 (Expected log return and divergence decomposition) *For every $v \in \mathcal{V}$,*

$$\mathbb{E}^M[\log v] = D_{\text{KL}}(\mathbb{P}^M \parallel \mathbb{P}^Q) - D_{\text{KL}}(\mathbb{P}^M \parallel \mathbb{P}^V). \quad (33)$$

Proof. Use $d\mathbb{P}^V/d\mathbb{P}^Q = v$ and subtract the two log likelihood ratios. Strict positivity of v ensures $\mathbb{P}^M \ll \mathbb{P}^V$. $\square$

THEOREM 6.2 (Projected attainable log growth) *Define*

$$\mathcal{G}_V(\mathbb{P}^M \parallel \mathbb{P}^Q) := \sup_{v \in \mathcal{V}} \mathbb{E}^M[\log v].$$

Then

$$\mathcal{G}_V(\mathbb{P}^M \parallel \mathbb{P}^Q) = D_{\text{KL}}(\mathbb{P}^M \parallel \mathbb{P}^Q) - \inf_{v \in \mathcal{V}} D_{\text{KL}}(\mathbb{P}^M \parallel \mathbb{P}^V), \quad 0 \leq \mathcal{G}_V(\mathbb{P}^M \parallel \mathbb{P}^Q) \leq D_{\text{KL}}(\mathbb{P}^M \parallel \mathbb{P}^Q). \quad (34)$$

If the induced attainable laws form a compact set, the supremum is attained. If that set is also convex, $\mathbb{P}^M$ has full support, and the payoff parameterization is identifiable, strict convexity gives uniqueness of the minimizing implied law and payoff.

Proof. Take the supremum in (33). The no-trade density $v = 1$ gives the lower bound, and non-negativity of relative entropy gives the upper bound. $\square$

Remark 3 (Why two states are special) With two states and cash available, a projection gap requires trading constraints or the absence of a nondegenerate risky payoff. Let states be d, u , let $q_u = q$, and let one risky payoff satisfy $s_u \neq s_d$. Put $z(a) = s(a) - \mathbb{E}^Q s$. Every positive normalized density v can be written as $v(a) = 1 + \alpha z(a)$, with

$$\alpha = \frac{v_u - v_d}{s_u - s_d}.$$

Thus cash and one nondegenerate risky payoff span the normalized two-state payoff space when unrestricted long/short exposure is allowed. For $K > 2$, normalized payoff space has dimension $K - 1$, so cash and one risky direction are generically incomplete.

EXAMPLE 6.3 (A three-state incomplete market) *Let $\mathbb{P}^Q = (0.3, 0.4, 0.3)$ and $\mathbb{P}^M = (0.45, 0.35, 0.20)$. The complete-market edge is $D_{\text{KL}}(\mathbb{P}^M \parallel \mathbb{P}^Q) \approx 0.0546$. If only*

$$v_\alpha = (1 + \alpha, 1, 1 - \alpha), \quad -1 < \alpha < 1,$$

is attainable, then

$$\mathbb{E}^M[\log v_\alpha] = 0.45 \log(1 + \alpha) + 0.20 \log(1 - \alpha).$$

Its first-order condition gives $\alpha^* = 0.25/0.65 \approx 0.3846$, and $\mathcal{G}_V(\mathbb{P}^M \parallel \mathbb{P}^Q) \approx 0.0493$. The gap 0.0053 is the divergence from $\mathbb{P}^M$ to the closest attainable implied law.

Because the empirical categorical benchmark of Section 7 is not derived from Arrow–Debreu prices, Section 7.5 studies a stylized frictionless cash/index payoff and compares history-dependent with history-free allocations.

7. Empirical illustration: S&P 500 around COVID-19

7.1. Data and exact empirical nulls

We use daily closing values of the S&P 500 price index from the FRED series SP500 (Federal Reserve Bank of St. Louis 2026). The price sample runs from 2 January 2019 to 30 December 2022. Daily log

returns define the symbols; simple returns define the cash/index payoff. The five return windows are Pre-COVID (3 January 2019–19 February 2020), crash (20 February–30 April 2020), crash-to-June (20 February–30 June 2020), post-June 2020 (1 July 2020–30 December 2022), and the full return sample (3 January 2019–30 December 2022). The illustration uses one index and one stressed episode to show how benchmark, alphabet, horizon, memory, and attainable payoff affect the conclusion.

For the binary sign $X_t = \mathbf{1}_{\{r_t > 0\}}$, the fixed null is

$$H_0^{(1/2)} : \quad \mathbb{P}^M(X_{t+1} = 1 \mid X_{1:t}) = \frac{1}{2} \quad \text{for every } t.$$

The fitted history-free null is

$$H_0^{(\eta)} : \quad \mathbb{P}^M(X_{t+1} = 1 \mid X_{1:t}) = \eta \quad \text{for every } t,$$

with η estimated from the $n = N - L$ next signs entering the transition likelihood. Within the maintained order- L Markov model, we impose iid Bernoulli initialization for the full-sequence permutation null, so the entire null path is exchangeable. The null absorbs unconditional sign imbalance, while the saturated alternative allows dependence on the specified history. The test does not address longer memories or dependence hidden within the chosen symbols.

7.2. Binary evidence

Table 4 reports $L = 1$. For the fixed fair-sign null, resampling means 100,000 parametric paths from Bernoulli(1/2). For the fitted history-free null, the reported resampling p-value is based on 200,000 conditional permutations of the observed signs. Every replication reconstructs overlapping histories and recomputes the statistic.

Table 4. Daily binary sign tests, $L = 1$. N is the number of returns and $n = N - 1$ the number of usable transitions. Information is in bits per transition.

Window	N	n	$\hat{\eta}$	Fixed 50–50 benchmark			Fitted history-free benchmark		
				$\hat{I}_2$	p_{χ^2}	p_{boot}	$\hat{I}_2$	p_{χ^2}	p_{perm}
Pre-COVID	284	283	.601	.0295064	.0031	.0031	.0000421	.8977	.9024
Crash window	50	49	.429	.0976446	.0363	.0384	.0828728	.0177	.0220
Crash to June	92	91	.527	.0345692	.1130	.1162	.0323904	.0432	.0499
Post-June 2020	631	630	.517	.0013966	.5434	.5422	.0005168	.5017	.5258
Full sample	1007	1006	.542	.0058234	.0172	.0174	.0007882	.2944	.3012

The fixed fair-sign test rejects in the pre-COVID bull market, where up days are substantially more frequent than down days. The fitted history-free test removes this marginal imbalance and does not reject. Proposition 2.2 explains the difference: the fixed statistic includes the marginal term $D_2(\bar{m}||1/2)$, while the fitted statistic is the plug-in mutual information between the selected history and the next sign. Because $\hat{\eta}$ is estimated from the $n = N - L$ usable next symbols, it is their empirical marginal and the benchmark-mismatch term vanishes exactly.

During the crash window the fitted history-free test rejects. After a down day, the next day is up about 57.1% of the time, while after an up day it is up about 23.8% of the time. The evidence weakens when the window is extended to June and is absent outside the crisis. These are statements about one-day sign ordering, not about the full weak-form hypothesis.

7.3. Four return states

Define four states by splitting negative and positive log returns separately at their window medians:

$$X_t \in \{D_2, D_1, U_1, U_2\},$$

where

$$D_2 = \{r_t \leq c_-\}, \quad D_1 = \{c_- < r_t < 0\}, \quad U_1 = \{0 \leq r_t \leq c_+\}, \quad U_2 = \{r_t > c_+\}.$$

The labels mean large down, small down, small up, and large up. No return in the sample is exactly zero. For $K = 4, L = 1$, the fitted history-free categorical null has $(4^1 - 1)(4 - 1) = 9$ degrees of freedom.

Table 5. Four-state tests for $L = 1$. Cutoffs are log returns in percent. Information is in bits per transition. The null-centered sensitivity is $\widehat{I}_2^{\text{nc}} = \max\{0, \widehat{I}_2 - 9/(2n \log 2)\}$. Parametric-bootstrap and full-symbol Monte Carlo permutation p-values use 200,000 replications. Both symbolic resampling procedures condition on the window-specific cutoffs c_- and c_+ and do not propagate threshold-estimation uncertainty.

Window	N	n	c_-	c_+	$\widehat{I}_2$	$c_{9,n}$	$\widehat{I}_2^{\text{nc}}$	Λ_n	p_{χ^2}	p_{boot}	p_{perm}
Crash window	50	49	-2.851	3.011	.2950	.1325	.1625	20.04	.0177	.0384	.0396
Crash to June	92	91	-1.783	1.448	.1420	.0713	.0707	17.91	.0362	.0523	.0498

The observed transition matrices, in the order (D_2, D_1, U_1, U_2) , are

$$\text{Crash window:} \quad \begin{pmatrix} 3 & 3 & 3 & 6 \\ 5 & 1 & 4 & 3 \\ 1 & 7 & 2 & 1 \\ 6 & 2 & 2 & 0 \end{pmatrix},$$

and

$$\text{Crash to June:} \quad \begin{pmatrix} 4 & 4 & 5 & 9 \\ 6 & 2 & 7 & 7 \\ 4 & 5 & 8 & 7 \\ 8 & 10 & 4 & 1 \end{pmatrix}.$$

The empty cell in the crash-window table, together with only 49 transitions for 9 degrees of freedom, is why resampling is treated as primary. The short-window conclusion survives both the fitted bootstrap and Monte Carlo conditional-permutation calibration, while the longer window is borderline and its classification depends on calibration at the third decimal.

7.4. Return horizon and history length

The return horizon and memory length define different nulls rather than innocuous tuning choices. We therefore repeat the fitted history-free binary test at $L = 2$ and for signs of disjoint two-day log returns. For two-day returns, phase 0 starts with the first return in the window and phase 1 drops that return before forming blocks; reporting both avoids selecting a block origin. Table 6 uses 200,000 conditional permutations.

Table 6. Binary robustness to memory and non-overlapping return horizon.

Window	Specification	Symbols N	Transitions n	p_{perm}
Crash window	1-day, $L = 1$	50	49	.0220
	1-day, $L = 2$	50	48	.0961
	2-day, phase 0, $L = 1$	25	24	.8477
	2-day, phase 1, $L = 1$	24	23	.0953
Crash to June	1-day, $L = 1$	92	91	.0499
	1-day, $L = 2$	92	90	.1255
	2-day, phase 0, $L = 1$	46	45	.8865
	2-day, phase 1, $L = 1$	45	44	.2073
Full sample	1-day, $L = 1$	1007	1006	.3012
	1-day, $L = 2$	1007	1005	.7925
	2-day, phase 0, $L = 1$	503	502	.2319
	2-day, phase 1, $L = 1$	503	502	.1594

In both crisis windows, only the daily $L = 1$ tests in Table 6 reject at the 5% level. Moving to $L = 2$ triples the degrees of freedom, which can cost power even if first-order dependence persists, while the disjoint two-day series roughly halve the number of transitions and give p-values that depend strongly on the block origin (.848 and .095 in the crash window). The full-sample benchmark-relative test also remains non-significant under all specifications reported in Table 6. A four-state $L = 2$ test would use 45 degrees of freedom against only 48 or 90 usable crisis-window transitions; the simulation shows that such a table is uninformative.

7.5. *A stylized cash/index diagnostic*

To ask whether conditioning on the symbolic history improves log growth under bounded index exposure, we approximate the one-day cash return by zero and let R_{t+1} denote the FRED price-index simple return, treated as a frictionless index-exposure proxy. This omits dividends, transaction costs, financing spreads, and the tracking instrument, so it is a diagnostic rather than a realized total-return backtest. After observing four-state history h , invest fraction $\alpha_h \in [-1, 1]$ in the index proxy and the remainder in cash. Within this stylized two-asset model, the one-period self-financing wealth factor is

$$1 + \alpha_h R_{t+1}.$$

The interval permits bounded index exposure from fully short to fully long, involves no cash borrowing, and keeps wealth positive in the sample. To remove unconditional index exposure, compare the four history-specific allocations with the best single history-free allocation:

$$\Delta \hat{\mathcal{G}}_{\text{id},2} := \frac{1}{n \log 2} \left[\max_{\alpha_h \in [-1,1]} \sum_{t=1}^n \log\{1 + \alpha_{H_t} R_{t+1}\} - \max_{\alpha \in [-1,1]} \sum_{t=1}^n \log(1 + \alpha R_{t+1}) \right]. \quad (35)$$

Because zero exposure is feasible in both optimization problems, each maximized in-sample log-growth level is nonnegative; the conditional family nests the history-free strategy, so their increment is also nonnegative by construction. The sign of the increment is therefore uninformative; a precommitted or out-of-sample strategy, by contrast, can lose money. Significance is assessed by permuting the raw returns, recomputing the states and re-solving both optimizations. This uses the stronger null that the observed raw returns are exchangeable, so it diagnoses temporal ordering in realized index returns rather than recalibrating the symbolic transition null. Table 7 uses 50,000 permutations.

Table 7. History-free and history-dependent index/cash-proxy log growth, in bits per transition. The p-value tests the incremental value in (35) under exchangeable return order.

Window	N	n	History-free	History-dependent	Increment / p_{perm}
Crash window	50	49	.00162	.01575	.01414 / .181
Crash to June	92	91	.00008	.01075	.01067 / .073

Neither increment is significant at the 5% level ($p = .181$ and $.073$). The statistical transition tests and the trading diagnostic ask different questions: the former uses the saturated conditional distribution over symbols, while the latter permits only a bounded one-dimensional index exposure after each state and controls for the best unconditional exposure.

8. Conclusion

The Brouty–Garcin statistic measures conditional relative entropy, so its interpretation depends on the benchmark. A fair-sign benchmark captures both an unconditional imbalance between up and down days and dependence on past signs. Fitting the history-free marginal accounts for the imbalance and leaves the dependence between the chosen history and the next symbol. The same interpretation applies to larger finite alphabets. Overlapping histories are handled through the transition likelihood, which gives chi-square limits under the stated Markov assumptions when alphabet and memory are fixed. Under the null that the symbols are independent and identically distributed, conditional permutation gives exact calibration.

The simulations show why a more detailed alphabet may not need to produce a more powerful test. When dependence operates through signs alone, binary coarsening loses no information, and the binary test benefits from fewer degrees of freedom. Four states help when dependence lies within the sign categories and disappears on aggregation. For the first-order alternatives considered here, longer histories add parameters without adding information. With four states and two lags, the resulting tables can be so sparse that even the fitted bootstrap gives inaccurate rejection rates.

Whether this dependence can be exploited depends on the available payoffs and their prices. If the reference probabilities are normalized traded state prices, relative entropy gives the maximal expected log growth relative to cash in a complete market. With a restricted payoff family, the projection result determines how much growth is attainable. In the S&P 500 application, empirical category frequencies provide the statistical benchmark. Around the COVID crash, daily signs show significant dependence at the 5% level with one lag, but significance disappears with two lags or disjoint two-day returns. Using the four-state history to choose a bounded cash/index exposure produces no statistically significant improvement in in-sample log growth in either crisis window.

Funding

This work was partially supported by the Grant Agency of the Czech Republic under grant 24-11146S.

Disclosure statement

No potential conflict of interest was reported by the author.

Data availability statement

The empirical illustration uses publicly available daily S&P 500 index data from the FRED series SP500, available at <https://fred.stlouisfed.org/series/SP500>. The direct CSV download is

available from <https://fred.stlouisfed.org/graph/fredgraph.csv?id=SP500>.

Declaration of generative AI use

I used ChatGPT 6 Astra (OpenAI) and Claude 5.5 Opus (Anthropic) to edit language, check mathematical arguments, and assist with numerical code. I reviewed and revised their output and take full responsibility for the article's originality, accuracy, and integrity, including its computations and references.

References

- Ammy-Driss, A. and Garcin, M. (2023). Efficiency of the financial markets during the COVID-19 crisis: Time-varying parameters of fractional stable dynamics. *Physica A*, 609, 128335.
- Anderson, T. W. and Goodman, L. A. (1957). Statistical inference about Markov chains. *Ann. Math. Statist.*, 28(1), 89–110.
- Baker, S. R., Bloom, N., Davis, S. J., Kost, K., Sammon, M. and Viratyosin, T. (2020). The unprecedented stock market reaction to COVID-19. *Rev. Asset Pricing Stud.*, 10(4), 742–758.
- Barnett, L. and Bossomaier, T. (2012). Transfer entropy as a log-likelihood ratio. *Phys. Rev. Lett.*, 109(13), 138105.
- Billingsley, P. (1961). Statistical methods in Markov chains. *Ann. Math. Statist.*, 32(1), 12–40.
- Brouty, X. and Garcin, M. (2023). A statistical test of market efficiency based on information theory. *Quant. Financ.*, 23(6), 1003–1018.
- Brouty, X. and Garcin, M. (2024). Fractal properties, information theory, and market efficiency. *Chaos Solitons Fractals*, 180, 114543.
- Choi, S.-Y. (2021). Analysis of stock market efficiency during crisis periods in the US stock market: Differences between the global financial crisis and COVID-19 pandemic. *Physica A*, 574, 125988.
- Cover, T. M. and Thomas, J. A. (2006). *Elements of Information Theory*, 2nd ed. Wiley, Hoboken.
- Dima, B., Dima, S. M. and Ioan, R. (2021). Remarks on the behaviour of financial market efficiency during the COVID-19 pandemic: The case of VIX. *Financ. Res. Lett.*, 43, 101967.
- Efron, B. and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. Chapman & Hall, New York.
- Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *J. Finance*, 25(2), 383–417.
- Federal Reserve Bank of St. Louis (2026). S&P 500 [SP500]. FRED, Federal Reserve Bank of St. Louis.
- Frezza, M., Bianchi, S. and Pianese, A. (2021). Fractal analysis of market (in)efficiency during the COVID-19. *Financ. Res. Lett.*, 38, 101851.
- Kelly, J. L. (1956). A new interpretation of information rate. *Bell Syst. Tech. J.*, 35(4), 917–926.
- Kontoyiannis, I. and Skoularidou, M. (2016). Estimating the directed information and testing for causality. *IEEE Trans. Inform. Theory*, 62(11), 6053–6067.
- Kullback, S. and Leibler, R. A. (1951). On information and sufficiency. *Ann. Math. Statist.*, 22(1), 79–86.
- Miller, G. A. (1955). Note on the bias of information estimates. In H. Quastler (ed.), *Information Theory in Psychology: Problems and Methods*, pp. 95–100. Free Press, Glencoe, IL.
- Navratil, R., Taylor, S. and Vecer, J. (2021). On equity market inefficiency during the COVID-19 pandemic. *Int. Rev. Financ. Anal.*, 77, 101820.
- Okorie, D. I. and Lin, B. (2021). Adaptive market hypothesis: The story of the stock markets and COVID-19 pandemic. *North Am. J. Econ. Finance*, 57, 101397.
- Richard, M. and Vecer, J. (2021). Efficiency testing of prediction markets: Martingale approach, likelihood ratio and Bayes factor analysis. *Risks*, 9(2), 31.
- Risso, W. A. (2008). The informational efficiency and the financial crashes. *Res. Int. Bus. Finance*, 22(3), 396–408.
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell Syst. Tech. J.*, 27(3), 379–423, 623–656.
- van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press, Cambridge.

- Vecer, J. (2026). Analytical solution for Kelly's criterion for multiple outcomes. *Quantitative Finance*, 26(5), 671–684. <https://doi.org/10.1080/14697688.2026.2629952>.
- Verdú, S. (2019). Empirical estimation of information measures: A literature guide. *Entropy*, 21(8), 720.
- Wilks, S. S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Ann. Math. Statist.*, 9(1), 60–62.