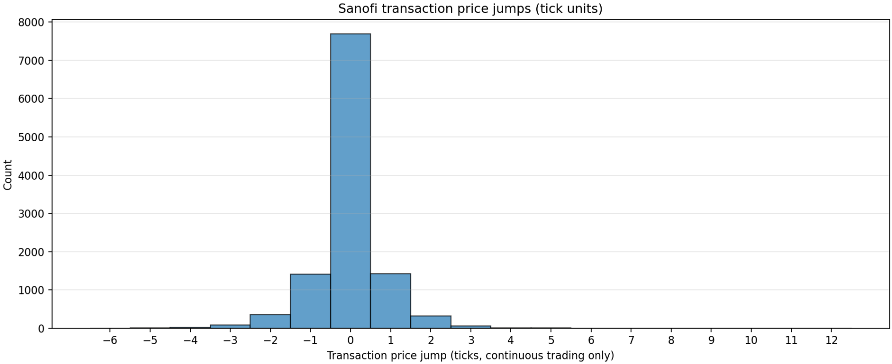
Causal inference and LOB

Wentao Wang, Yuhang Guo

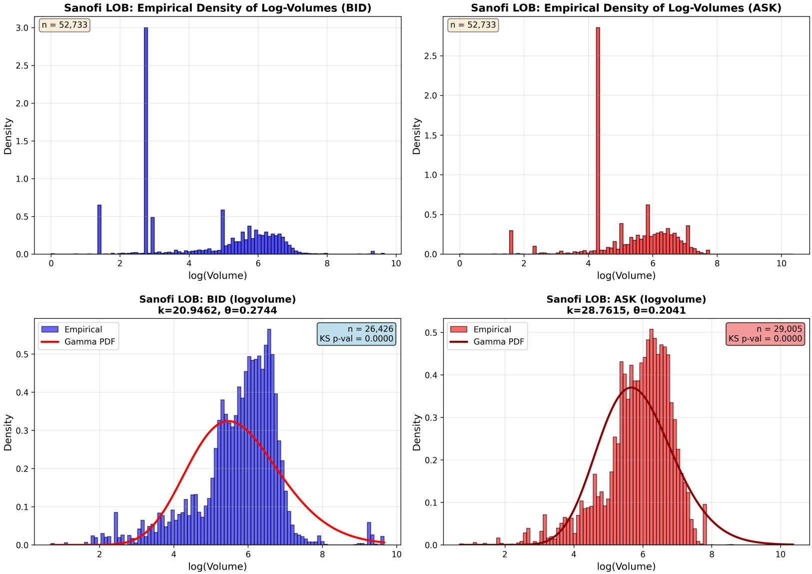
Présentation du cours APM_52112_EP

10 Feb 2026

transaction price jump distribution



Gamma distribution Sanofi



Introduction

We study the weak-form EMH, in which market efficiency is equivalent to whether prices can be represented as a martingale.

- Hurst index exhibits strong model dependency, not an effective indicator under real-world financial models.
- Shannon entropy gives no statistical significance explanation.

Key idea is to construct statistical test with null hypothesis that of EMH holds, rather than a continuous efficiency index.

Definitions

Risso's method (2008): binary classification of price rises and falls, for price serie $P_0, P_1, ..., P_n$, we define:

$$X_i = \mathbf{1}_{\{P_i - P_{i-1} > 0\}}, \quad i = 1, ..., n.$$

For $L \geq 1$, there are 2^L distinct sequences of length L, noting:

$$G_1^L, \ldots, G_{2^L}^L \in \{0,1\}^L.$$

We define for each G_i^L:

$$p_i^L = \mathbb{P}\big((X_1, \ldots, X_L) = G_i^L\big),$$

$$\pi_i^L = \mathbb{P}\big(X_{L+1} = 1 \,\big|\, (X_1, \ldots, X_L) = G_i^L\big).$$

Shannon Entropy

Recall the expression of Shannon Entropy:

$$H(Y,Z) = -\sum_{y,z} p(y,z)\, \log_2 p(y,z)$$

For $Y = (X_1,\ldots,X_L),\ Z = X_{L+1}$:

$$\begin{aligned} H_{L+1} &= H(Y,Z) = H\big((X_1,\ldots,X_L),\, X_{L+1}\big) \\ &= -\sum_{i=1}^{2^L} \Big[p_i^L \pi_i^L \log_2\big(p_i^L \pi_i^L\big) + p_i^L (1-\pi_i^L) \log_2\big(p_i^L (1-\pi_i^L)\big) \Big]. \end{aligned}$$

With weak-form EMH, we have:

$$\pi_i^L = \frac{1}{2}.$$

Plug in the former formula we have:

$$H_{L+1}^{\mathrm{EMH}} = 1 + H_L, \qquad H_L = -\sum_{i=1}^{2^L} p_i^L \log_2 p_i^L.$$

Statistic

We now define the statistic:

$$\boxed{I_{L+1} = H_{L+1}^{\mathrm{EMH}} - H_{L+1}}$$

The author prooved that:

Theorem (2.1)

For any $L \geq 1$, the market information satisfies

$$I_{L+1} \geq 0.$$

Moreover,

$$I_{L+1} = 0 \quad \Longleftrightarrow \quad \forall i \text{ such that } p_i^L > 0,\ \pi_i^L = \frac{1}{2}.$$

Estimator

We study the distuibution of $\widehat{I}_{L+1}$ under EMH, and decompose $\widehat{I}_{L+1}$ by its prefix, plugging in empirical frequency:

$$\widehat{I}_{L+1} = \sum_{j=1}^{2^L} \widehat{I}_j = \sum_{j=1}^{2^L} p_j^L \left[\widehat{\pi}_j^L \log_2\big(2\widehat{\pi}_j^L\big) + \big(1-\widehat{\pi}_j^L\big) \log_2\big(2(1-\widehat{\pi}_j^L)\big) \right].$$

With:

$$\widehat{\pi}_j^L = \frac{X_j}{n_j}, \qquad X_j \sim \mathrm{Binomial}\big(n_j, \tfrac{1}{2}\big)\,.$$

Let $x = \widehat{\pi}_j^L$, now we define:

$$g_j(t,x) = \exp(it\widehat{I}_{L+1}) = \exp\Bigg(it \Big[p_j^L\, x \log_2\big(p_j^L x\big) + p_j^L(1-x) \log_2\big(p_j^L(1-x)\big)\Big] - it\, p_j^L \log_2\Big(\frac{p_j^L}{2}\Big)\Bigg).$$

Asymptotic distribution

Under EMH, $\pi_i^L = \frac{1}{2}$, the author prooves that all impair terms are 0 at $x = \frac{1}{2}$, and the leading term is of second order. Thus:

$$\mathbb{E}\left[g_j\left(t, \frac{X_j}{n_j}\right)\right] = 1 + \frac{it\, p_j^L}{2\ln(2)\, n_j} + \mathcal{O}\left(n_j^{-2}\right).$$

With quasi-independency of prefix, we have the caracteristic function:

$$\begin{aligned}\varphi_{\widehat{I}_{L+1}}(t) = \prod_{j=1}^{2^L} \mathbb{E}\left[g_j\left(t, \widehat{\pi}_j^L\right)\right] &\sim \prod_{j=1}^{2^L}\left(1 + \frac{it}{2\ln(2)\, n}\right) \\ &\xrightarrow[n\to+\infty]{} \left(1 - \frac{it}{\ln(2)\, n}\right)^{-2^{L-1}}.\end{aligned}$$

The function correspond with Gamma Distribution, thus:

$$\boxed{\widehat{I}_{L+1} \xrightarrow{d} \Gamma\left(k = 2^{L-1},\ \theta = \frac{1}{\ln(2)\, n}\right).}$$

Gamma distribution

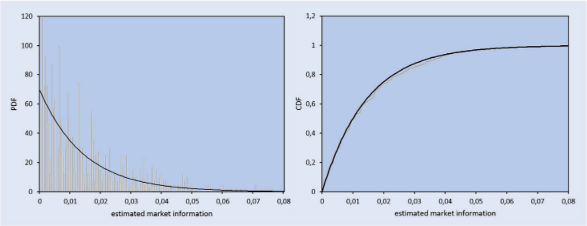


Figure 2. Asymptotic $\Gamma(2^{L-1}, 1/\ln(2)n)$ (black) and simulated (on 1000 trajectories, grey) distributions of the estimated market information $\widehat{I}^2$ for $n = 100$.

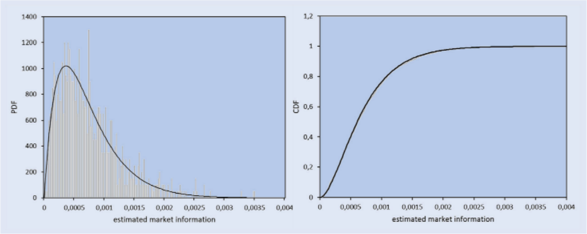


Figure 3. Asymptotic $\Gamma(2^{L-1}, 1/\ln(2)n)$ (black) and simulated (on 1000 trajectories, grey) distributions of the estimated market information $\widehat{I}^3$ for $n = 4000$.

Statistical test

We construct the following statistical test:

$$H_0 : \quad \pi_i^L = \tfrac{1}{2} \ \ \forall i; \qquad \text{(EMH)}$$

$$H_1 : \quad \exists i \text{ such that } \pi_i^L \neq \tfrac{1}{2}.$$

And the statistic:

$$T = \widehat{I}_{L+1}.$$

Under H_0:

$$T \sim \Gamma\left(2^{L-1}, \frac{1}{\ln(2)\, n}\right).$$

And the p-value:

$$\text{p-value} = \mathbb{P}\left(\Gamma\left(2^{L-1}, \frac{1}{\ln(2)\, n}\right) \geq \widehat{I}_{L+1}\right).$$

Given α:

$$\text{Reject } H_0 \quad \Longleftrightarrow \quad \widehat{I}_{L+1} > F^{-1}_{\Gamma(2^{L-1},\, 1/(\ln(2)\, n))}(1-\alpha).$$

Test on financial market

Considering $L = 1$ ($Markov(1)$), the author tested on the entire markets, single stocks and crypto currencies. Results are the follows:

- Significant inefficiencies mainly arise in small-cap indices, while large-cap indices exhibit at most temporary and weak departures from efficiency.
- At the stock level, weak-form inefficiency is primarily driven by market capitalization and liquidity rather than sectoral effects.
- Cryptocurrencies exhibit heterogeneous efficiency dynamics, with major assets showing time-varying inefficiencies and signs of market maturation, while others remain close to weak-form efficiency.

Overall, weak-form market efficiency is a dynamic and asset-dependent property, with robust inefficiencies concentrated in less liquid markets and diminishing as markets become larger and more mature.

Conclusion

Theoretical contributions:

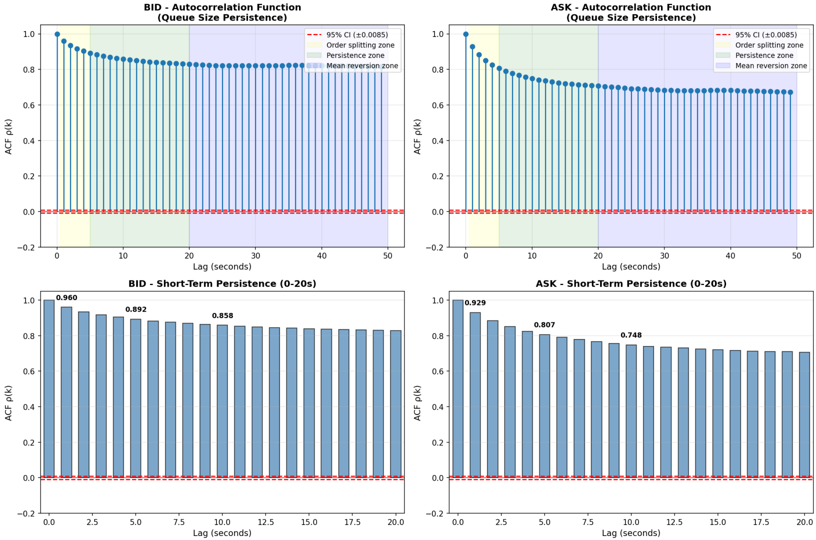
- Introduced model-free information theoretic test of weak-form EMH based on Shannon entropy;
- Derive the asymptotic Gamma distribution of the information estimator under EMH;
- Enable a closed-form statistical test with explicit p-values.

Empirical contributions:

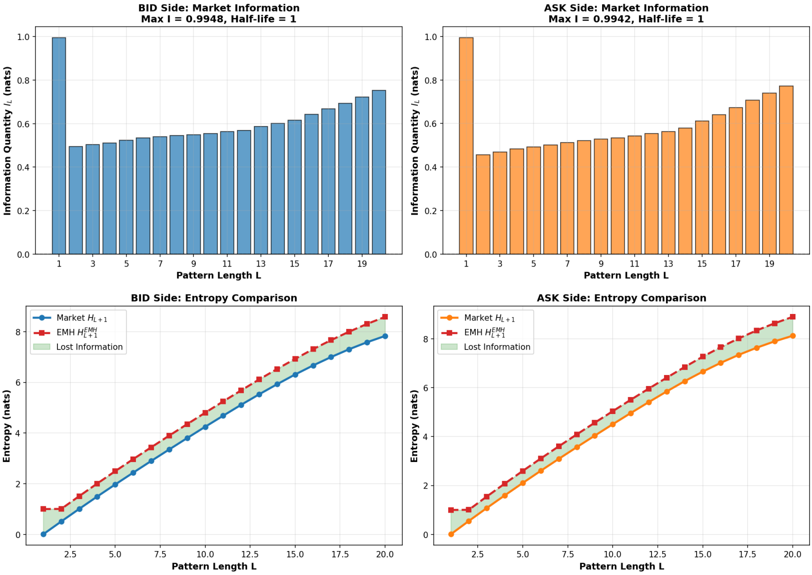
- The proposed test is implemented on stock indices, individual stocks, and cryptocurrencies using rolling windows.
- Market inefficiencies are shown to be time-varying and strongly dependent on asset class and liquidity.
- Robust and statistically significant inefficiencies mainly appear in small-cap, low-liquidity, and less mature markets.

$\Rightarrow$ This paper proposes an entropy-based statistical test of weak-form market efficiency and shows that robust inefficiencies mainly arise in less liquid and less mature markets.

ACF method


BID - Autocorrelation Function
(Queue Size Persistence)
ASK - Autocorrelation Function
(Queue Size Persistence)
95% CI (±0.0085)
Order splitting zone
Persistence zone
Mean reversion zone
ACF ρ(k)
Lag (seconds)
BID - Short-Term Persistence (0-20s)
ASK - Short-Term Persistence (0-20s)
0.960
0.892
0.858
0.929
0.807
0.748

Entropy method



Introduction

Core Question

- Are information-theoretic indicators and H equivalent, redundant, or complementary measures of market efficiency?
- The existing literature lacks a rigorous theoretical link between these two approaches.

Main Contributions of the Paper

1. The paper studies how market information varies with the size of the information set L, and not only with the time scale.
2. Within the fractional Brownian motion framework, derives explicit relationship between market information and H.
3. Introduces stationary fractal model, delampertized fBm, showing high level of market information may arise even when $H = 1/2$.

Model set-up

We define again, and to determine the explicit form:

$$I^m_{L+1} = 1 - H\left(G_{m,i+Lm} \,\middle|\, G_{m,i}, \ldots, G_{m,i+(L-1)m}\right).$$

A stochastic process $(X_t)_{t\geq 0}$ is said to be *H-self-similar* if, for any constant $c > 0$,

$$(X_{ct})_{t\geq 0} \overset{d}{=} (c^H X_t)_{t\geq 0},$$

where $\overset{d}{=}$ denotes equality in finite-dimensional distributions and $H \in (0,1)$ is called the *Hurst exponent*.

- $H = \frac{1}{2}$: standard Brownian motion;
- $H > \frac{1}{2}$: persistent, trend-following behaviour;
- $H < \frac{1}{2}$: characterized by mean-reverting or reversal effects.

fBm model

The fraction Brownian motion B_t^H is defined as follow:

- $B_0^H = 0$;
- $\mathbb{E}[B_t^H] = 0$;
- $\mathbb{E}\big[(B_t^H - B_s^H)^2\big] = \sigma^2|t-s|^{2H}$, for all $s, t \in \mathbb{R}$.

Thus the covariance:

$$\mathbb{E}[B_t^H B_s^H] = \frac{\sigma^2}{2}\Big(|t|^{2H} + |s|^{2H} - |t-s|^{2H}\Big).$$

- $H = \frac{1}{2}$: standard Brownian motion, Markov process with no memory.
- $H \neq \frac{1}{2}$: long-range temporal dependence.

Statistical study

Let:

$$Z = B_m^H - B_0^H, \qquad Y = B_{2m}^H - B_m^H.$$

Then:

$$G_{m,t} = 1 \iff Z > 0, \qquad G_{m,t+m} = 1 \iff Y > 0.$$

The random vector (Y, Z) is a bivariate Gaussian vector. Its correlation coefficient, entirely determined by the function $\rho(H)$, is given by:

$$\rho = \frac{\mathbb{E}[YZ]}{\sqrt{\mathbb{E}[Y^2]\mathbb{E}[Z^2]}} = 2^{2H-1} - 1.$$

We have for bivariate Gaussian vector:

$$\mathbb{P}(Y > 0 \mid Z > 0) = \frac{1}{2} + \frac{1}{\pi} \arctan\left(\frac{\rho}{\sqrt{1-\rho^2}}\right);$$

$$\mathbb{P}(Y > 0 \mid Z < 0) = \frac{1}{2} - \frac{1}{\pi} \arctan\left(\frac{\rho}{\sqrt{1-\rho^2}}\right).$$

fBm market information

Theorem

Let B_t^H be a fractional Brownian motion of Hurst exponent H and volatility parameter σ. Let the price P_t be equal to $\log(B_t^H)$ for all $t \in \mathbb{R}$, and let $m > 0$. Then, the market information I_2^m introduced in Eq. (6) is equal to

$$I_2^m = 1 + f\left(\frac{1}{2} - \frac{1}{\pi}\arctan\left(\frac{\rho}{\sqrt{1-\rho^2}}\right)\right) + f\left(\frac{1}{2} + \frac{1}{\pi}\arctan\left(\frac{\rho}{\sqrt{1-\rho^2}}\right)\right),$$

where $f : x > 0 \mapsto x\log_2(x)$ and $\rho = 2^{2H-1} - 1$.

Under fBm:

$$H = \tfrac{1}{2} \implies \rho = 0 \implies I_2^m = 0,$$
$$H \neq \tfrac{1}{2} \implies \rho \neq 0 \implies I_2^m > 0.$$

$$\Rightarrow \boxed{\text{EMH} \iff H = \frac{1}{2}.}$$

Delampertized fBm

We introduce delampertized fBm to tranform self-similar process to stable process:

$$(\mathcal{L}_H^{-1} X)_t = e^{-Ht} X_{e^t}.$$

Introducing a mean-reversion (stationarization) parameter $\theta > 0$, one defines the process:

$$\mathcal{X}_t^{H,\theta} = (\mathcal{L}_H^{-1} B^H)_{\theta t}.$$

If $H = \frac{1}{2}$, reduce to standard Ornstein–Uhlenbeck process and satisfies:

$$d\mathcal{X}_t^{1/2,\theta} = -\frac{\theta}{2} \mathcal{X}_t^{1/2,\theta} \, dt + \sqrt{\theta} \, dB_t.$$

Here, we have:

$$\rho = \rho(H, m\theta) = \frac{2 - h(2m\theta)}{4 - 2h(m\theta)} - 1,$$

where the function h is defined by:

$$h(x) = 2\cosh(Hx) - \left(2 \sinh \frac{x}{2}\right)^{2H}.$$

Delampertized fBm market information

Theorem

The market information I_2^m introduced previously is equal to

$$I_2^m = 1 + f\left(\frac{1}{2} - \frac{1}{\pi}\arctan\left(\frac{\rho}{\sqrt{1-\rho^2}}\right)\right) + f\left(\frac{1}{2} + \frac{1}{\pi}\arctan\left(\frac{\rho}{\sqrt{1-\rho^2}}\right)\right),$$

where $f : x > 0 \mapsto x\log_2(x)$.

Market information I_2^m can be significantly positive if θ is sufficiently large, even when $H = \frac{1}{2}$.

As a consequence,

$$\boxed{H = \tfrac{1}{2} \not\Rightarrow \text{EMH.}}$$

Market information and entropy

Theorem

Let $m > 0$ and $\mathcal{G}_i$ be the random stationary vector.
Then, the mapping: $L \mapsto H_L^m$ is increasing and concave in L.

Since the market information satisfies:

$$I_L^m = 1 + H_{L-1}^m - H_L^m,$$

The market information is equal to zero for all values of L if and only if $L \longmapsto H_L^m$ is a linear function of L with slope equal to 1.
An inefficient market must thus have a nonlinear entropy $L \longmapsto H_L^m$ or a slope lower than 1.

We introduce finally the partial market information:

$$\boxed{\mathcal{I}_L^m = I_L^m - I_{L-1}^m.}$$

Which measures the marginal contribution of the L-th lagged variable to the overall predictability.

Conclusion

Theoretical Contributions

- Within fBm, derives explicit analytical expression of market information, which equal to 0 iif $H = \frac{1}{2}$.
- Within delampertized fBm, shows that market information is also determined by mean-reversion and time-scale effects.
- Demonstrates that H characterizes scaling and dependence properties rather than predictability.

Practical Contributions

- Simulations validate the methodology and highlight the importance of marginal (partial) market information.
- Empirical applications show predictability is concentrated in a small number of recent lags and varies across time scales and markets.
- In several cases, significant market information is detected even when fractal indicators suggest near-random behavior.

$\Rightarrow$ Market efficiency is best understood as a multiscale information structure.

CIR model set-up

Let $(X_t)_{t\geq 0}$ follow a Cox–Ingersoll–Ross (CIR) diffusion:

$$dX_t = \kappa(\theta - X_t)\,dt + \sigma\sqrt{X_t}\,dW_t, \qquad \kappa > 0,\ \theta > 0,\ \sigma > 0.$$

Under the Feller condition $2\kappa\theta \geq \sigma^2$, the boundary 0 is unattainable and $X_t > 0$ a.s. In the stationary regime, X_t admits a Gamma invariant distribution:

$$X_t \sim \mathrm{Gamma}(\alpha,\ \beta)\,, \qquad \alpha = \frac{2\kappa\theta}{\sigma^2}, \quad \beta = \frac{\sigma^2}{2\kappa},$$

(shape α, scale β), hence:

$$\mathbb{E}[X_t] = \theta, \qquad \mathrm{Var}(X_t) = \alpha\beta^2 = \frac{\theta\sigma^2}{2\kappa}.$$

CIR autocorrelation

The CIR is affine, and its conditional mean is explicit:

$$\mathbb{E}[X_{t+s} \mid X_t] = \theta + (X_t - \theta)e^{-\kappa s}.$$

Therefore, for any $s \geq 0$,

$$\begin{aligned}
\mathrm{Cov}(X_t, X_{t+s}) &= \mathrm{Cov}\big(X_t, \mathbb{E}[X_{t+s} \mid X_t]\big) \\
&= \mathrm{Cov}\big(X_t, \theta + (X_t - \theta)e^{-\kappa s}\big) \\
&= e^{-\kappa s}\mathrm{Var}(X_t).
\end{aligned}$$

In the stationary regime this yields the **level autocorrelation**:

$$\rho_X(s) := \mathrm{Corr}(X_t, X_{t+s}) = e^{-\kappa s}.$$

Thus the serial dependence decays exponentially at rate κ.

Problem background and goal

- Conditional independence (CI) testing is central in graphical models and causal discovery:

$$H_0: X \perp\!\!\!\perp Y \mid Z \qquad \text{vs.} \qquad H_1: \text{otherwise.}$$

- Classical partial correlation works only under (approx.) Gaussian/linear assumptions; it may lose power under nonlinear dependence and mis-control Type I error when normality fails.
- Many omnibus CI tests rely on complicated asymptotic variance/bias estimation or expensive bootstraps.
- **Goal:** build a CI test that is
  - **distribution-free under** H_0 (easy calibration by simulation),
  - **powerful** for nonlinear alternatives,
  - **robust** (monotone-invariant; bounded transforms),
  - **computationally light** (quadratic time).

Key idea: CI $\Leftrightarrow$ mutual independence via Rosenblatt transform

Define (continuous case)

$$U := F_{X|Z}(X \mid Z), \qquad V := F_{Y|Z}(Y \mid Z), \qquad W := F_Z(Z).$$

Proposition 1 (equivalence)

Under mild continuity conditions on conditional CDFs,

$$X \perp\!\!\!\perp Y \mid Z \quad \Longleftrightarrow \quad (U, V, W) \text{ are mutually independent.}$$

- Under H_0, $U, V, W \sim \mathrm{Unif}(0, 1)$ and $U \perp\!\!\!\perp W$, $V \perp\!\!\!\perp W$.
- **Strategy:** convert CI into mutual-independence testing, while exploiting these intrinsic uniform/independence properties.

A new dependence index $\rho(X, Y \mid Z)$ with closed form

They measure mutual dependence in the frequency domain using characteristic functions, with a weight chosen as the density of i.i.d. standard Cauchy variables (yields a closed form and no moment conditions).

Definition (index)

A normalized index $\rho(X, Y \mid Z) \in [0, 1]$ is derived and can be written as

$$\rho(X, Y \mid Z) = c_0 \, \mathbb{E}\Big[A(U_1, U_2)\, A(V_1, V_2)\, e^{-|W_1 - W_2|}\Big],$$

where $A(\cdot, \cdot)$ is an explicit function (built from $e^{-|u-u'|}$ and marginal correction terms) and c_0 is a constant normalizer.

Theorem 2 (main properties)

- $0 \le \rho \le 1; \quad \rho = 0 \iff X \perp\!\!\!\perp Y \mid Z$.
- Symmetry: $\rho(X, Y \mid Z) = \rho(Y, X \mid Z)$.
- Invariance: $\rho(X, Y \mid Z) = \rho(m_1(X), m_2(Y) \mid m_3(Z))$ for strictly monotone m_1, m_2, m_3.

Estimator: nonparametric conditional CDFs + V-statistic

Given i.i.d. samples $\{(X_i, Y_i, Z_i)\}_{i=1}^n$:

- Estimate conditional CDFs by kernel smoothing (Li–Racine style):

$$\widehat{F}_{X|Z}(x \mid z) = \frac{\sum_{i=1}^n K_h(z - Z_i)\,\mathbf{1}(X_i \le x)}{\sum_{i=1}^n K_h(z - Z_i)}, \quad \widehat{F}_{Y|Z}(y \mid z) = \frac{\sum_{i=1}^n K_h(z - Z_i)\,\mathbf{1}(Y_i \le y)}{\sum_{i=1}^n K_h(z - Z_i)}.$$

- Transform data:

$$\hat{U}_i = \widehat{F}_{X|Z}(X_i \mid Z_i), \quad \hat{V}_i = \widehat{F}_{Y|Z}(Y_i \mid Z_i), \quad \hat{W}_i = \widehat{F}_Z(Z_i) = \frac{1}{n}\sum_{k=1}^n \mathbf{1}(Z_k \le Z_i).$$

- Plug-in V-statistic (quadratic time):

$$\hat{\rho} = c_0\,\frac{1}{n^2}\sum_{i,j=1}^n A(\hat{U}_i, \hat{U}_j)\,A(\hat{V}_i, \hat{V}_j)\,e^{-|\hat{W}_i - \hat{W}_j|}.$$

Key design choice: the statistic *uses* $U \perp\!\!\!\perp W$ and $V \perp\!\!\!\perp W$ to reduce the impact of nonparametric estimation bias.

Asymptotic theory: distribution-free null and power

Theorem 3 (null distribution)

Under H_0 and regularity conditions on the kernel, bandwidth, and smoothness in z,

$$n\hat{\rho} \xrightarrow{d} c_0 \sum_{j=1}^{\infty} \lambda_j \, \chi_j^2(1),$$

a weighted sum of i.i.d. $\chi^2(1)$, where $\{\lambda_j\}$ are eigenvalues of a kernel $h(\cdot)$ on $[0,1]^3$.

- **Distribution-free:** under H_0, (U, V, W) are i.i.d. independent $\mathrm{Unif}(0,1)$, so the limit depends only on this universal law.
- **Fixed alternative (Theorem 5):** if H_1 holds and $\rho > 0$,

  $$\sqrt{n}\,(\hat{\rho} - \rho) \Rightarrow \mathcal{N}(0, \sigma_0^2),$$

  hence the test is consistent for any fixed alternative.
- **Local alternatives (Theorem 5):** the test detects departures of order $O(n^{-1/2})$ with nontrivial limiting power.

Calibration without bootstrap: simulation-based critical values

Because the null limit is universal, critical values can be obtained by *simulating* (U, V, W) directly.

Simulation procedure (Theorem 4 justification)

1. Generate i.i.d. $\{(U_i^*, V_i^*, W_i^*)\}_{i=1}^n \sim \mathrm{Unif}(0,1)^3$ with mutual independence.
2. Compute $\hat{\rho}^*$ from $\{(U_i^*, V_i^*, W_i^*)\}_{i=1}^n$ using the same V-statistic formula.
3. Repeat for B times; set c_α as the $(1-\alpha)$ quantile of $T_b := n\hat{\rho}_b$.

Decision rule

Compute $T = n\hat{\rho}$ on data. Reject H_0 if $T > c_\alpha$ (or use Monte Carlo p-value).

- Only **one round** of simulation is needed (independent of the observed data), unlike resampling-based CI tests.

Full CI test workflow (end-to-end)

1. **Input:** data $\{(X_i, Y_i, Z_i)\}_{i=1}^{n}$, significance level α.
2. **Estimate transforms:**
   - choose kernel K and bandwidth h (rule-of-thumb is recommended in experiments),
   - compute $\hat{U}_i = \widehat{F}_{X|Z}(X_i \mid Z_i)$, $\hat{V}_i = \widehat{F}_{Y|Z}(Y_i \mid Z_i)$,
   - compute $\hat{W}_i = \widehat{F}_Z(Z_i)$ (empirical CDF).
3. **Compute statistic:** $T = n\hat{\rho}$ (a quadratic-time V-statistic).
4. **Calibrate null:** simulate B i.i.d. samples from $\mathrm{Unif}(0,1)^3$ to obtain c_α.
5. **Output:** reject if $T > c_\alpha$; otherwise do not reject.

Why this statistic (vs a naive mutual-independence plug-in)?

At the sample level, exploiting $U \perp W$ and $V \perp W$ reduces bias effects from estimating U, V, making the null behavior far less sensitive to bandwidth choice.

Numerical comparisons: size, power, robustness

They compare **CIT** (this paper) with three popular nonlinear CI tests:

- **CDC:** conditional distance correlation (energy statistics; higher-order U-statistics),
- **CMI:** conditional mutual information (kNN / permutation-based variants),
- **KCI:** kernel conditional independence test.

Main empirical findings (Tables 2–4)

- **Size control:** all methods are close to the nominal level in the reported settings.
- **Power:** CIT is typically **more powerful**, especially for small n and heavy-tailed designs.
- **Bandwidth sensitivity:** CIT is largely insensitive; a naive normalization-based variant is not.
- **Computation:** CIT is $O(n^2)$; CDC can be much heavier (often $O(n^3)$).

Application: causal discovery with the PC algorithm

- Under faithfulness, the **PC algorithm** recovers a Markov equivalence class by:
  1. removing edges using (conditional) independence tests,
  2. orienting edges using separation sets and logical rules.
- Performance depends critically on CI tests: early mistakes can change the recovered graph.

Real data experiment (n=392, 5 medical variables)

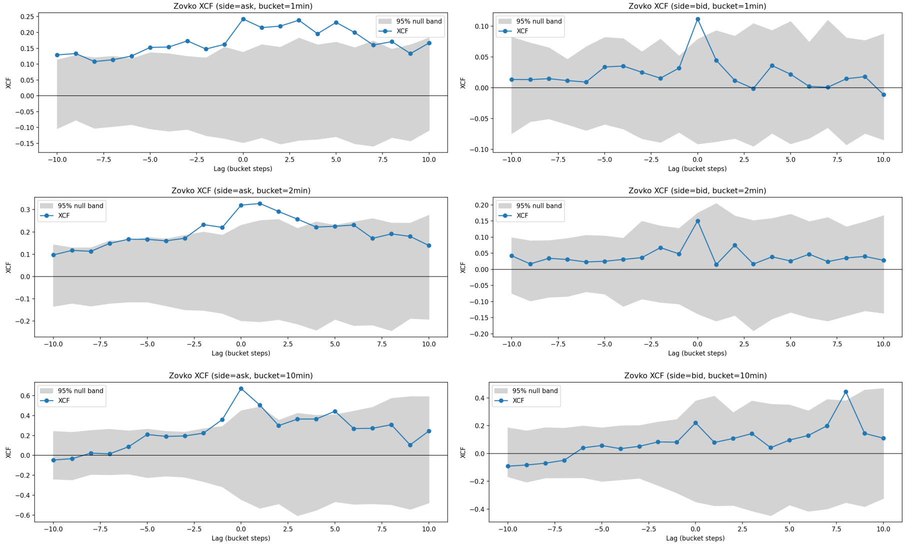
Variables: age, BMI, 2-hour serum insulin, plasma glucose, diastolic blood pressure.

- On the original (approximately normal) data: CIT yields the same estimated graph as partial correlation.
- After **log transformation**: CIT recovers the **same** structure (monotone-invariance), while partial correlation produces more false positives (normality violated).
- CMI may miss edges (higher false positives reported elsewhere); KCI tends to be conservative (misses edges).

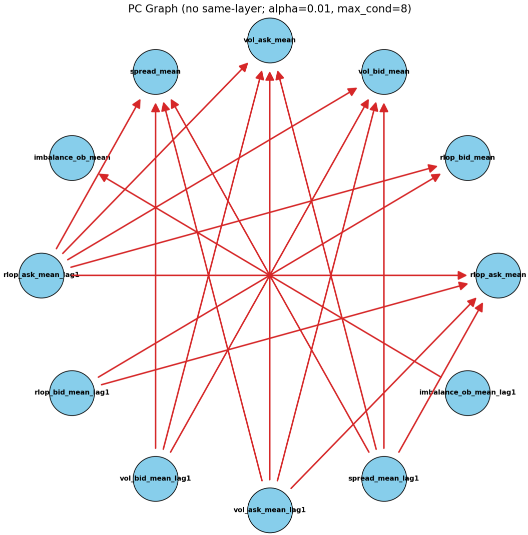
Takeaways

- **Conceptual bridge:** CI testing $\Rightarrow$ mutual-independence testing via Rosenblatt transforms.
- **New index ρ:** bounded in $[0, 1]$, equals 0 iff CI holds; monotone-invariant and robust.
- **Distribution-free null:** $n\hat{\rho}$ has a universal limiting law under H_0; critical values obtained by simulating i.i.d. $\mathrm{Unif}(0, 1)^3$ (no data bootstrap).
- **Practical strengths:** quadratic-time computation, low tuning sensitivity, strong finite-sample power.
- **Causal discovery:** plugging CIT into PC improves robustness under transformations and non-Gaussianity.

The cross autocorrelation of the time series of relative limit prices and volatilities Sanofi(Zovko)



Causal Inference: CIT + PC Algorithm



Limitation 1. The CIT assumes that the observations of (X, Y, Z) are i.i.d., which is generally violated for time-series data.

Limitation 2. Before running the PC algorithm, we remove all edges between variables within the same time layer and forbid edges pointing backward in time. These constraints may bias the inferred graph.