Chaosometer: Measuring market efficiency via nonparametric mutual information

Omid M. Ardakani [†]

Abstract

Chaosometer maps financial return series to a scalar efficiency score on a zero-to-hundred scale measuring departure from the martingale hypothesis. The score is a monotone transform of the mutual information between consecutive returns, estimated nonparametrically via the Kraskov–Stögbauer–Grassberger algorithm and parametrically via Gaussian determinant ratios. A channel decomposition separates return-level dependence from volatility clustering. The Hill tail index and Ljung–Box portmanteau statistic complete the diagnostic panel. This document states every formula precisely and verifies each against the production codebase.

Keywords: Mutual information, market efficiency, KSG estimator, tail index, nonparametric dependence, volatility clustering

1 Framework

Let $\{r_t\}_{t=1}^{T}$ denote log-returns $r_t = \log P_t - \log P_{t-1}$. The weak-form efficient market hypothesis asserts $r_t \perp\!\!\!\perp (r_{t-1}, \ldots, r_{t-L})$, i.e., past returns carry no information about future returns. Chaosometer quantifies departure from this condition through mutual information (MI), a measure that captures linear and nonlinear dependence (Cover and Thomas, 2006).

1.1 Mutual information

The MI between r_t and $\mathbf{r}_{t-L}^{t-1} = (r_{t-1}, \ldots, r_{t-L})$ is

$$I(r_t;\, \mathbf{r}_{t-L}^{t-1}) = \mathcal{D}_{\mathrm{KL}}(p_{r_t, \mathbf{r}_{t-L}^{t-1}};\, p_{r_t} \otimes p_{\mathbf{r}_{t-L}^{t-1}}) = \int p(r_t, \mathbf{r}_{t-L}^{t-1}) \log \frac{p(r_t, \mathbf{r}_{t-L}^{t-1})}{p(r_t)\, p(\mathbf{r}_{t-L}^{t-1})}\, dr_t\, d\mathbf{r}_{t-L}^{t-1}, \qquad (1)$$

the Kullback–Leibler divergence of the joint density from the product of marginals (Kullback and Leibler, 1951). This quantity is non-negative, equals zero if and only if $r_t \perp\!\!\!\perp \mathbf{r}_{t-L}^{t-1}$, and is invariant under smooth invertible marginal transformations.

[†]Professor of Economics, Parker College of Business, Georgia Southern University; oardakani@georgiasouthern.edu.

MI has been applied to nonlinear lag identification (Granger and Lin, 1994), multivariate dependence (Joe, 1989), forecast evaluation (Ardakani et al., 2018, 2025b), portfolio construction (Ardakani, 2024c), and efficiency testing (Ardakani, 2025a). Information-theoretic measures of tail risk (Ardakani, 2023a,b, 2025c,b) and information loss from dependence misspecification (Ardakani et al., 2025a,c) motivate the present application.

1.2 Efficiency score

Definition 1. *The efficiency score* $\mathcal{E} : [0, \infty) \to [0, 100)$ *is*

$$\mathcal{E} = 100 \left(1 - \exp\left(-\kappa \hat{I}^*\right)\right), \tag{2}$$

where $\hat{I}^*$ *is the estimated MI (Section 2.4) and* $\kappa > 0$ *is a calibration constant.*

Proposition 1. *The map* $f(x) = 100(1 - e^{-\kappa x})$ *satisfies: (i)* $f(0) = 0$*; (ii)* $f'(x) = 100\kappa e^{-\kappa x} > 0$*; (iii)* $f''(x) = -100\kappa^2 e^{-\kappa x} < 0$*; (iv)* $\lim_{x\to\infty} f(x) = 100$.

Strict concavity compresses large MI values, preventing extreme readings from dominating portfolio-level summaries. The production system uses $\kappa = 10$, calibrated so that the S&P 500 yields $\mathcal{E} \approx 33$ and the VIX yields $\mathcal{E} \approx 63$.

2 Estimation

Two estimators are computed for each MI quantity: a parametric Gaussian estimator and a nonparametric KSG estimator. The maximum is retained, ensuring that nonlinear dependence invisible to the Gaussian method is captured.

2.1 Gaussian MI via Toeplitz determinant ratio

For a stationary process with autocorrelations $\rho_0 = 1, \rho_1, \ldots, \rho_L$, let R_m denote the $m \times m$ Toeplitz matrix with (i, j)-entry $\rho_{|i-j|}$.

Proposition 2. *Under joint Gaussianity,*

$$I_G(r_t;\, r_{t-1}, \ldots, r_{t-L}) = -\tfrac{1}{2} \log \frac{\det R_{L+1}}{\det R_L}. \tag{3}$$

For $L = 1$*, this reduces to* $I_G = -\frac{1}{2}\log(1 - \rho_1^2)$.

Proof. *Write $X = r_t$, $Y = (r_{t-1}, \ldots, r_{t-L})$, and let $\sigma^2 = \mathrm{Var}(r_t)$. The Gaussian entropies are*

$$\begin{aligned} H(X) &= \tfrac{1}{2}\log(2\pi e\,\sigma^2), \\ H(Y) &= \tfrac{1}{2}\log\det(2\pi e\,\sigma^2 R_L) = \tfrac{L}{2}\log(2\pi e\,\sigma^2) + \tfrac{1}{2}\log\det R_L, \\ H(X,Y) &= \tfrac{1}{2}\log\det(2\pi e\,\sigma^2 R_{L+1}) = \tfrac{L+1}{2}\log(2\pi e\,\sigma^2) + \tfrac{1}{2}\log\det R_{L+1}. \end{aligned}$$

Since $I = H(X) + H(Y) - H(X,Y)$, all $(2\pi e\,\sigma^2)$ terms cancel: $I = \frac{1}{2}\log\det R_L - \frac{1}{2}\log\det R_{L+1}$.

The sample autocorrelation

$$\hat{\rho}_\ell = \frac{\sum_{t=\ell+1}^{n}(r_t - \bar{r})(r_{t-\ell} - \bar{r})}{\sum_{t=1}^{n}(r_t - \bar{r})^2} \tag{4}$$

is substituted into R_{L+1}. If $\det R_{L+1} \le 0$ or $\det R_L \le 0$, the code falls back to $L = 1$.

2.2 Joint Gaussian MI

The joint channel conditions r_t on both lagged returns and lagged absolute returns. Let $X = r_t$ and $Y = (r_{t-1}, \ldots, r_{t-L'}, |r_{t-1}|, \ldots, |r_{t-L'}|) \in \mathbb{R}^{2L'}$ with $L' = \min(L, 3)$. From the sample covariance $\hat{\Sigma}$ of (X, Y),

$$I_{G,\text{joint}} = \tfrac{1}{2}\Big(\log\hat{\sigma}_X^2 + \log\det\hat{\Sigma}_Y - \log\det\hat{\Sigma}_{XY}\Big), \tag{5}$$

where $\hat{\sigma}_X^2 = \hat{\Sigma}_{11}$, $\hat{\Sigma}_Y$ is the $2L' \times 2L'$ submatrix, and $\hat{\Sigma}_{XY}$ the $(2L'+1) \times (2L'+1)$ matrix.

2.3 KSG nonparametric estimator

The Kraskov–Stögbauer–Grassberger Algorithm 1 (Kraskov et al., 2004) estimates MI via k-nearest-neighbor distances under the Chebyshev norm, without density estimation.

Definition 2. *Given n samples $(x_i, \mathbf{y}_i)$ with $x_i \in \mathbb{R}$, $\mathbf{y}_i \in \mathbb{R}^d$, let $\varepsilon(i)$ be the Chebyshev distance to the k-th nearest neighbor of point i in the joint $(1+d)$-dimensional space. Define*

$$n_x(i) = \#\{j \neq i : |x_j - x_i| < \varepsilon(i)\}, \qquad n_y(i) = \#\{j \neq i : \|\mathbf{y}_j - \mathbf{y}_i\|_\infty < \varepsilon(i)\}. \tag{6}$$

The KSG estimator is

$$\hat{I}_{\text{KSG}} = \psi(k) - \frac{1}{n}\sum_{i=1}^{n}\big[\psi(n_x(i)+1) + \psi(n_y(i)+1)\big] + \psi(n), \tag{7}$$

where ψ is the digamma function.

Each dimension is standardized to zero mean and unit variance before computing distances. Samples exceeding $n = 600$ are subsampled; $k = 3$ throughout. All KSG computations use lag depth $L = 1$ (bivariate or trivariate MI) for robustness in the high-dimensional Chebyshev norm.

2.4 Channel decomposition and scoring

Three channels capture distinct dependence structures:

$$I_{\text{level}} = I(r_t;\, r_{t-1}, \ldots, r_{t-L}), \tag{8}$$

$$I_{\text{vol}} = I(|r_t|;\, |r_{t-1}|, \ldots, |r_{t-L}|), \tag{9}$$

$$I_{\text{joint}} = I(r_t;\, r_{t-1}, \ldots, r_{t-L'},\, |r_{t-1}|, \ldots, |r_{t-L'}|), \quad L' = \min(L, 3). \tag{10}$$

The level channel detects momentum and mean-reversion. The volatility channel detects clustering in return magnitude. The joint channel conditions on both.

For each channel c, both estimators are computed and the maximum retained:

$$\hat{I}_c = \max\Big(\hat{I}_{c,\text{KSG}},\, \hat{I}_{c,G}\Big). \tag{11}$$

The score uses the single most informative channel:

$$\hat{I}^* = \max\Big(\hat{I}_{\text{level}},\, \hat{I}_{\text{vol}},\, \hat{I}_{\text{joint}}\Big). \tag{12}$$

The maximum prevents double-counting: the joint channel subsumes information from both the level and volatility channels, so whichever channel yields the highest MI estimate drives the score. The remaining channels are reported as diagnostics.

Remark 1. *$\hat{I}_{vol} \gg \hat{I}_{level}$ indicates volatility clustering; volatility-targeting strategies have edge. $\hat{I}_{level} \gg \hat{I}_{vol}$ indicates momentum or mean-reversion (Ardakani, 2024c,b).*

2.5 Lag selection

The Gaussian estimator uses lag depth

$$L = \begin{cases} 5 & n \geq 200, \\ 3 & 80 \leq n < 200, \\ 1 & n < 80. \end{cases} \tag{13}$$

The KSG estimator uses $L = 1$ for the level and volatility channels (bivariate MI) and $L = 1$ for the joint channel (trivariate: r_t against r_{t-1} and $|r_{t-1}|$).

3 Diagnostics

3.1 Hill tail index

The tail index α governs power-law decay $\mathbb{P}(|r_t| > u) \sim u^{-\alpha}$ (Nolan, 2020). For $\alpha < 2$, variance is infinite; this is common in financial returns (Cont, 2001; Ardakani, 2024a) and relevant to risk measurement (Ardakani, 2023b, 2025b) and option pricing (Ardakani, 2022).

Definition 3. *Let $|r|_{(1)} \geq \cdots \geq |r|_{(n)}$ be the order statistics of $|r_1|, \ldots, |r_n|$. Set $k = \max(5, \lfloor\sqrt{n}\rfloor)$. The Hill estimator (Hill, 1975) is*

$$\hat{\gamma} = \frac{1}{k}\sum_{i=1}^{k} \log \frac{|r|_{(i)}}{|r|_{(k+1)}}, \qquad \hat{\alpha} = \frac{1}{\hat{\gamma}}, \qquad \mathrm{SE}(\hat{\alpha}) = \frac{\hat{\alpha}}{\sqrt{k}}. \tag{14}$$

The implementation clips $\hat{\alpha}$ to $[1.01, 4.0]$.

3.2 Ljung–Box test

Definition 4. *The Ljung–Box statistic (Ljung and Box, 1978) is*

$$Q_L = n(n+2)\sum_{\ell=1}^{L} \frac{\hat{\rho}_\ell^2}{n-\ell}, \qquad Q_L \sim \chi^2_L \text{ under } H_0 : \rho_1 = \cdots = \rho_L = 0. \tag{15}$$

The implementation uses $L = 10$.

3.3 Interpretation

Table 1: Efficiency score interpretation.

$\hat{\mathcal{E}}$	Interpretation	Implication
$[0, 20)$	Efficient	Passive strategies appropriate
$[20, 50)$	Mild predictability	Tactical allocation
$[50, 100]$	Inefficient	Exploitable patterns present

4 Algorithm

Algorithm 1 states the full scoring pipeline.

Algorithm 1 Chaosometer: full analysis (`compute_full`)

Require: Log-returns $r_1, \ldots, r_n$, $n \geq 30$

Ensure: $\hat{\mathcal{E}} \in [0, 100]$ and diagnostics

Set L via (13)

Compute $\hat{\rho}_0, \ldots, \hat{\rho}_{10}$ via (4)

$\hat{I}_{G,\text{level}} \leftarrow -\frac{1}{2}\log(\det R_{L+1} / \det R_L)$ ▷ (3), ACF of r_t

$\hat{I}_{G,\text{vol}} \leftarrow -\frac{1}{2}\log(\det R_{L+1} / \det R_L)$ ▷ (3), ACF of $|r_t|$

$\hat{I}_{G,\text{joint}} \leftarrow$ (5) with $L' = \min(L, 3)$

$r \leftarrow$ last $\min(n, 600)$ returns

$\hat{I}_{\text{KSG,level}} \leftarrow$ (7) on $(r_t;\, r_{t-1})$, $k = 3$

$\hat{I}_{\text{KSG,vol}} \leftarrow$ (7) on $(|r_t|;\, |r_{t-1}|)$, $k = 3$

$\hat{I}_{\text{KSG,joint}} \leftarrow$ (7) on $(r_t;\, r_{t-1}, |r_{t-1}|)$, $k = 3$

for $c \in \{\text{level}, \text{vol}, \text{joint}\}$ **do**

    $\hat{I}_c \leftarrow \max(\hat{I}_{\text{KSG},c},\, \hat{I}_{G,c})$

end for

$\hat{I}^* \leftarrow \max(\hat{I}_{\text{level}},\, \hat{I}_{\text{vol}},\, \hat{I}_{\text{joint}})$ ▷ (12)

$\hat{\mathcal{E}} \leftarrow \min\big(100,\ 100(1 - e^{-10\hat{I}^*})\big)$ ▷ (2), $\kappa = 10$

Compute $\hat{\alpha}$ via (14), Q_{10} via (15)

return $\hat{\mathcal{E}}$, $\hat{I}_{\text{level}}$, $\hat{I}_{\text{vol}}$, $\hat{I}_{\text{joint}}$, $\hat{\alpha}$, $\hat{\rho}_1$, Q_{10}

Algorithm 2 Chaosometer: lightweight analysis (`compute_light`)

Require: Log-returns $r_1, \ldots, r_n$, $n \geq 30$

Ensure: $\hat{\mathcal{E}} \in [0, 100]$ and diagnostics

1: Set L via (13); compute $\hat{\rho}_0, \ldots, \hat{\rho}_{10}$

2: $\hat{I}_{G,\text{level}} \leftarrow$ (3) on r_t; $\quad \hat{I}_{G,\text{vol}} \leftarrow$ (3) on $|r_t|$

3: $\hat{I}^* \leftarrow \max(\hat{I}_{G,\text{level}}, \hat{I}_{G,\text{vol}})$

4: $\hat{\mathcal{E}} \leftarrow \min\big(100,\ 100(1 - e^{-10\hat{I}^*})\big)$

5: Compute $\hat{\alpha}$, Q_{10}

6: **return** $\hat{\mathcal{E}}$, $\hat{I}_{G,\text{level}}$, $\hat{I}_{G,\text{vol}}$, $\hat{\alpha}$, $\hat{\rho}_1$, Q_{10}

Algorithm 2 is used for multi-asset comparison and regime scanning (rolling windows of 21, 63, 126, 252, and 504 trading days), where speed is prioritized over nonlinear sensitivity.

References

Omid M. Ardakani. Option pricing with maximum entropy densities: the inclusion of higher-order moments. *Journal of Futures Markets*, 42(10):1821–1836, 2022.

Omid M. Ardakani. Capturing information in extreme events. *Economics Letters*, 231: 111301, 2023a.

Omid M. Ardakani. Coherent measure of portfolio risk. *Finance Research Letters*, 57:104222, 2023b.

Omid M. Ardakani. Bayesian extreme value models: asymptotic behavior, hierarchical convergence, and predictive robustness. *Econometrics and Statistics*, 2024a.

Omid M. Ardakani. Information content of inflation expectations: a copula-based model. *Studies in Nonlinear Dynamics & Econometrics*, 29(1):71–93, 2024b.

Omid M. Ardakani. Portfolio optimization with transfer entropy constraint. *International Review of Financial Analysis*, 96(Part A):103644, 2024c.

Omid M. Ardakani. Informational efficiency and rational bubbles. *International Review of Economics & Finance*, 103:104486, 2025a.

Omid M. Ardakani. Robust learning of tail dependence. *Econometrics*, 13(4):47, 2025b.

Omid M. Ardakani. Strategic information asymmetry in tail-risk markets. *North American Journal of Economics and Finance*, 79:102460, 2025c.

Omid M. Ardakani, Nader Ebrahimi, and Ehsan S. Soofi. Ranking forecasts by stochastic error distance, information and reliability measures. *International Statistical Review*, 86 (3):442–468, 2018.

Omid M. Ardakani, M. Asadi, and E. S. Soofi. Information about the research parameter or model parameter: a chicken and egg problem. *Applied Stochastic Models in Business and Industry*, 41(1):e2931, 2025a.

Omid M. Ardakani, R. F. Bordley, and E. S. Soofi. Expected information of noisy attribute forecasts for probabilistic forecasts. *European Journal of Operational Research*, 323(3): 1013–1023, 2025b.

Omid M. Ardakani, V. Dalko, and H. Shim. Information loss from perception alignment. *International Review of Economics and Finance*, 97:103830, 2025c.

Rama Cont. Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 1:223–236, 2001.

Thomas M. Cover and Joy A. Thomas. *Elements of information theory*. Wiley, 2nd edition, 2006.

Clive W. J. Granger and Jin-Lung Lin. Using the mutual information coefficient to identify lags in nonlinear models. *Journal of Time Series Analysis*, 15(4):371–384, 1994.

Bruce M. Hill. A simple general approach to inference about the tail of a distribution. *Annals of Statistics*, 3(5):1163–1174, 1975.

Harry Joe. Relative entropy measures of multivariate dependence. *Journal of the American Statistical Association*, 84(405):157–164, 1989.

Alexander Kraskov, Harald Stögbauer, and Peter Grassberger. Estimating mutual information. *Physical Review E*, 69(6):066138, 2004.

Solomon Kullback and Richard A. Leibler. On information and sufficiency. *Annals of Mathematical Statistics*, 22(1):79–86, 1951.

Greta M. Ljung and George E. P. Box. On a measure of lack of fit in time series models. *Biometrika*, 65(2):297–303, 1978.

John P. Nolan. *Stable distributions: models for heavy-tailed data*. Birkhäuser, 2020.