Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Forecastability Profiles: Horizon-Specific Limits to Prediction Before Model Selection

Peter M. Catt

peter.catt@virtualblue.co.nz

Independent Researcher

Research Article

Keywords: forecastability profile, horizon-specific prediction, mutual information, logarithmic loss, information-theoretic forecasting, time series forecasting

Posted Date: June 19th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9882043/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Forecastability Profiles: Horizon-Specific Limits to Prediction Before Model Selection

Peter M. Catt [1,*]

[1] Independent Researcher, Auckland, New Zealand

[*] Correspondence: peter.catt@virtualblue.co.nz; ORCID: 0009-0000-1324-0861

Abstract

Forecasting is often treated as a question of choosing the best model. This paper asks an earlier question: how much predictive information is available at each forecast horizon, given a declared information set? We define a horizon-indexed quantity, the forecastability profile, as the mutual information between a future observation and the available information. Under logarithmic loss, this profile has a precise interpretation: it equals the maximum possible reduction in expected log loss that any predictor using that information set can achieve. The profile therefore provides a horizon-specific ceiling on prediction, and its shape reveals whether a process is predictable at short lags, seasonal horizons, or not at all. We show that reducing the information set can only lower forecastability, that limiting prediction to a finite lag window imposes an exact information cost, and that periodic dependence produces corresponding structure in the profile. We also define an exploitability ratio that separates losses caused by limited information from losses caused by imperfect modelling. Five finite-sample illustrations show monotone decay, damped recurrence, nonlinear collapse, quasi-periodic dependence, and near-zero predictability. Together, these results offer a way to diagnose prediction limits before model selection, helping distinguish problems that call for better information from those that call for better algorithms.

Keywords: forecastability profile; horizon-specific prediction; mutual information; logarithmic loss; information-theoretic forecasting; time series forecasting

1. Introduction

Prediction is usually framed as a problem of model selection. A practitioner assembles candidate forecasting algorithms, evaluates them on held-out data, and selects the method with the lowest error, a paradigm that dominates the forecasting literature [1-3]. This framing takes the information structure for granted and treats forecasting quality as a property of the method.

Applied forecasting has advanced rapidly under this framing, from financial time series [4] to Gaussian-process models [5], yet current research evaluates algorithms, studies asymptotic consistency, and develops benchmark accuracy comparisons. It does not generally characterise predictive information as a horizon-specific, information-set explicit decision-theoretic object independent of any fitted forecasting

model. This paper addresses that prior question. Before asking which model best exploits available information, one should characterise what information is available, at which horizons, and how much of it any method could in principle capture. The exact ceiling result is derived under log loss and is specific to that criterion; the profile remains a well-defined dependence diagnostic under other criteria, but the decision-theoretic log-loss identity does not carry over.

The central object of this paper is the *forecastability profile*: the mutual information between a future observation and a declared information set, evaluated as a function of forecast horizon. The profile is a property of the information structure defined by the declared information set, not of any forecasting method, and its decision-theoretic interpretation is specific to log loss. Under log loss, strict properness of the scoring rule establishes that the profile equals the maximal achievable reduction in expected log loss for predictors measurable with respect to the declared information set. The information set matters: two studies using different lag embeddings of the same series obtain different profiles, and the profiles are not directly comparable without accounting for the information-set difference.

The profile is best understood as a horizon-specific structural object rather than as a single horizon-specific number. The forecastability profile preserves the organisation of predictive information across horizon, allowing dependence regimes with similar aggregate uncertainty to exhibit qualitatively distinct horizon distributions. A short-memory process and a seasonal process may share the same total predictive information yet distribute it very differently across horizon: the first concentrates it at short lead times, the second relocates it to seasonal multiples. Scalar entropy measures, which integrate over horizon, cannot separate these regimes; the horizon-specific profile can.

This paper positions the forecastability profile against the broader landscape of information-theoretic quantities used to characterise predictive structure in time series. These quantities fall into three broad groups. Scalar complexity measures, including entropy rate [6], approximate entropy [7], sample entropy, and permutation entropy [8], compress predictive structure to one number and so cannot resolve how predictive information is distributed across horizon. Directional dependence measures, including transfer entropy [9], Granger causality [10], and modern causal-discovery methods for nonlinear time series [11], quantify directed information flow between distinct processes rather than a horizon-specific predictive ceiling for a single target. Asymptotic predictive quantities, including excess entropy [12] and predictive information [13], characterise total stored or block-scaling complexity rather than per-horizon prediction limits. The forecastability profile differs structurally from all three: it is horizon-specific rather than scalar, conditioned on an explicitly declared information set rather than on a second process, grounded in a strictly proper scoring rule, and decision-theoretic in that its value is the achievable log-loss reduction under logarithmic scoring. To be explicit about what is inherited and what is new: the use of mutual information as a

dependence measure, proper scoring rules as a loss criterion, and lag embeddings as an information-set representation are all established. What is new is (i) the combination of these ingredients into a single horizon-specific object with an exact decision-theoretic interpretation under log loss, (ii) the three structural results (monotonicity, truncation budget, periodicity inheritance) that follow from that combination, and (iii) the exploitability decomposition that separates information-limited from approximation-limited prediction. Section 6 sharpens this comparison against the nearest existing quantities.

The contributions of this paper are as follows.

1. **Forecastability profile as a horizon-specific predictive object.** Predictive information is defined as the mutual information between a future observation and a declared information set, evaluated as a function of forecast horizon. The resulting profile is a horizon-specific function whose shape, not merely its level, encodes the dependence regime, retaining structure that scalar indices discard.

2. **Decision-theoretic interpretation.** Under logarithmic scoring, the profile is an intrinsic upper bound on the expected log-loss reduction attainable by any predictor measurable with respect to the declared information set. This is a pre-modelling ceiling that does not depend on any fitted model.

3. **Finite-window truncation budget.** Restricting the declared information set to a finite lag window of length p incurs an exact information cost equal to the conditional mutual information between the target and the remote past given the retained window. This is an identity, not an approximation.

4. **Periodicity inheritance and damped harmonic recurrence.** Exact joint-law periodicity propagates to the forecastability profile exactly. In stable seasonal systems, exact periodicity is replaced by damped harmonic recurrence: peaks at seasonal horizon multiples whose amplitudes decline across successive cycle transitions.

5. **Exploitability ratio (corollary).** As a corollary of the log-loss decomposition, the exploitability ratio separates information-limited from approximation-limited prediction, determining whether poor forecasting reflects an information problem or a modelling problem. It is subordinate to the profile and developed only briefly here.

In summary, the paper establishes horizon-specific mutual information relative to a declared information set as an exact decision-theoretic predictive ceiling under logarithmic scoring, whose geometry reveals dependence structure obscured by scalar entropy measures.

The objective of this paper is not to propose a forecasting algorithm, evaluate predictive methods competitively, or benchmark model performance. The forecastability profile is a structural quantity defined prior to any model selection, whose value depends on the process and the declared information set rather than on any estimation or modelling choice. The empirical illustrations in Section 8 demonstrate profile

geometry under controlled conditions; they are not validation studies, accuracy comparisons, or claims about practical forecasting performance.

Section 2 provides related-work context. Section 3 establishes notation. Section 4 develops the log-loss identity and the forecastability profile, including a treatment of why log loss occupies a privileged position and an introduction of the exploitability ratio. Section 5 develops the three structural properties. Section 6 relates the profile to existing information-theoretic quantities. Section 7 describes estimation and the surrogate test. Section 8 presents five empirical illustrations. Section 9 provides discussion, limitations, and conclusions.

2. Related Work and Positioning

2.1. Proper Scoring Rules and Logarithmic Loss

The decision-theoretic foundation of probabilistic forecasting rests on proper scoring rules [14, 15]. A scoring rule is strictly proper when the expected score is minimised by the true conditional distribution and by no other. Logarithmic loss is strictly proper: the expected excess loss of any predictive distribution relative to the truth equals the Kullback–Leibler divergence [16], providing the exact bridge between probabilistic forecasting and information theory that underpins the forecastability profile.

2.2. Mutual Information and Time Series

Lagged mutual information is established in embedding-dimension selection and state-space reconstruction [17, 18], horizon-aware dependence measures for coupled systems [19], and block-scaling characterisations of stored complexity [12]. This framework adds a decision-theoretic interpretation relative to a declared information set: the mutual information between the future and the information set is the forecastability bound on expected log-loss reduction, not merely a dependence measure.

2.3. Scalar Predictability Indices

Scalar complexity measures (entropy rate [6], approximate entropy [7], permutation entropy [8, 20], minimum description length [21]) compress the dependence structure of a process to a single number and are necessarily horizon-agnostic. The forecastability profile is a horizon-specific generalisation that retains information-theoretic rigour while exposing the variation across horizons that any scalar index must discard. Applied forecasting summarises series-level forecastability through related feature constructions, such as the spectral-entropy instance spaces used to visualise where particular methods succeed [22]; these too are horizon-agnostic scalar summaries.

2.4. Granger Causality and Information-Set Formalisms

Granger causality [10] and its information-theoretic extension, transfer entropy [9, 23], quantify predictive usefulness through inclusion of additional variables, with the

two criteria equivalent for Gaussian processes [24]; modern causal-discovery methods extend these to nonlinear high-dimensional settings via conditional-independence testing [11]. The forecastability profile differs in scope: it treats a single-target, self-lag information set, and the declared information set plays the role of the conditioning sigma-algebra with a log-loss ceiling as the criterion rather than variance reduction.

2.5. Positioning within the Information-Theoretic Literature

The framework here is closest in spirit to DelSole [25, 26], who develops information-theoretic predictability measures for atmospheric models building on the foundational bounds of Lorenz [27], and to the predictive information formalism of Bialek and colleagues [13]. Barnett and Bossomaier [28] study information dynamics and the predictive structure of coupled systems. The present author has previously synthesised physics-based, graphical, and information-theoretic perspectives on forecastability in an applied setting [29]. This paper differs in treating forecastability as a horizon-specific log-loss ceiling relative to an explicitly declared information set. The structural contribution is not the mutual-information identity alone, but the consequences of varying or truncating that information set: compression monotonicity, an exact finite-window information cost, profile recurrence under joint-law periodicity, and the exploitability decomposition. These results make the information set itself part of the object being measured.

3. Setting and Notation

Unadorned quantities such as $F(h; I_t)$ denote population objects; hatted quantities denote empirical estimates. Let $\{Y_t\}$ be a stochastic process indexed by the integers, assumed stationary unless stated otherwise. At time t, a forecaster declares an information set I_t, represented as a sub-sigma-algebra of the underlying probability space. The prediction problem is to specify a probabilistic forecast for Y_{t+h} at horizon $h \geq 1$.

The true predictive conditional distribution is $p^*_h(\cdot \mid I_t)$. A predictive method supplies a distribution $q_h(\cdot \mid I_t)$ intended to approximate it. Prediction quality is measured by log loss,

$$\ell(q, y) = -\log q(y), \tag{1}$$

where $q(y)$ denotes the density or probability mass assigned to y. The p-lag truncated information set is

$$I_t^{(p)} = \sigma(Y_t, Y_{t-1}, \ldots, Y_{t-p+1}), \quad p \geq 1, \tag{2}$$

and the full (idealised) information set I_t may be taken as the sigma-algebra generated by the entire past. Throughout, $H(\cdot)$, $H(\cdot \mid \cdot)$, and $I(\cdot\,;\,\cdot)$ denote entropy, conditional entropy, and mutual information respectively [30, 31]. The Kullback–Leibler divergence between distributions p and q is written $D_{KL}(p \,\|\, q)$.

The results that follow rest on four explicit assumptions, stated here once and assumed throughout unless noted otherwise. (A1) *Stationarity*: the process $\{Y_t\}$ is strictly stationary, so all joint laws are shift-invariant; this is required for the profile to be a function of horizon rather than of calendar time. (A2) *Measurability*: the target Y_{t+h} and the information set I_t are defined on standard Borel spaces, and the regular conditional distribution of Y_{t+h} given I_t exists; this guarantees the conditional entropy is well defined. (A3) *Finite entropy*: the relevant marginal and conditional entropies are finite, so the mutual information $I(Y_{t+h}; I_t)$ is well defined and finite; without this the log-loss decomposition is not integrable. (A4) *Forecasting target*: the prediction target is the full conditional distribution of the scalar Y_{t+h} at lead time $h \geq 1$, scored by log loss; the exact mutual-information identity of Theorem 1 is specific to this loss. Under (A1)–(A4) the forecastability profile and the structural results of Section 5 are population-level identities.

4. The Log-Loss Identity and the Forecastability Profile

4.1. Why Logarithmic Loss?

Logarithmic loss occupies a privileged position among loss functions in the context of the forecastability profile. Four distinct properties justify this position.

Strict properness. The expected log loss is minimised by the true conditional distribution and by no other, establishing an unambiguous performance target.

Coding interpretation. The log loss of a predictive distribution q evaluated at outcome y equals the code length assigned to y under q.

Cross-entropy equivalence. The excess expected log loss of any predictive distribution relative to the truth equals the Kullback–Leibler divergence $D_{KL}(p^*_h \| q_h)$.

Exact mutual-information identity. Under log loss, the maximum achievable reduction in expected loss over the marginal predictor equals exactly the mutual information $I(Y_{t+h}; I_t)$. Under mean-squared-error the analogous quantity is the conditional variance reduction, which coincides with mutual information only in the Gaussian case.

4.2. The Log-Loss Decomposition

Theorem 1. *(Log-loss decomposition). For any predictive distribution q_h with finite expected log loss,*

$$E[-\log q_h(Y_{t+h} \mid I_t)] = H(Y_{t+h} \mid I_t) + E[D_{KL}(p^*_h \| q_h)] \quad (3)$$

The expected loss is minimised if and only if $q_h = p^*_h$ almost surely.

Proof. Expand the cross-entropy inside the expectation using the definition of the Kullback–Leibler divergence, then apply the tower property of conditional expectations.

Non-negativity of the divergence establishes the minimum, with equality if and only if the two conditional distributions coincide. □

4.3. Forecastability as a Log-Loss Ceiling

The unconditional (marginal) predictor achieves expected log loss $H(Y_{t+h})$. The maximum improvement over this baseline obtainable from I_t is

$$H(Y_{t+h}) - H(Y_{t+h} \mid I_t) = I(Y_{t+h}; I_t) \tag{4}$$

Definition 1. *(Forecastability profile). The forecastability at horizon h relative to information set I_t is*

$$F(h; I_t) := I(Y_{t+h}; I_t) \tag{5}$$

The mapping $h \mapsto F(h; I_t)$ is the forecastability profile. Its decay rate reflects memory, its recurrence reflects periodic forcing, and its extinction marks a predictability horizon. The shape is a compact description of the dependence regime, recoverable from the horizon-specific profile without fitting any model. The distribution of predictive information across forecast horizon itself carries structural information about the underlying dependence regime: two processes with equal total predictive information but different horizon distributions are distinct predictive objects.

Remark 1. *(Gaussian special case). When the pair is jointly Gaussian, the forecastability reduces to $F(h; I_t^{(p)}) = -\tfrac{1}{2} \log(1 - R^2_h)$, where R^2_h is the population squared multiple correlation coefficient of Y_{t+h} on the lag vector.*

4.4. Corollary: The Exploitability Ratio

The forecastability profile is the maximal achievable reduction in expected log loss under a declared information set. A fitted probabilistic predictor may attain only a fraction of this ceiling. To make the gap between achievable and attained performance precise, define, for a fitted predictor q_h with finite expected log loss, the realised loss reduction over the marginal baseline as $G_q(h) = H(Y_{t+h}) - E[-\log q_h(Y_{t+h} \mid I_t)]$. The exploitability ratio at horizon h is then the ratio of realised reduction to the forecastability ceiling,

$$\chi_q(h) = G_q(h) / F(h; I_t) \tag{6}$$

whenever $F(h; I_t) > 0$. The decomposition follows directly from Theorem 1. Writing the realised loss as the sum of the conditional entropy and the expected divergence $E[D_{KL}(p^*_h \,\|\, q_h)]$ yields the exact identity $G_q(h) = F(h; I_t) - E[D_{KL}(p^*_h \,\|\, q_h)]$, so that

$$\chi_q(h) = 1 - E[D_{KL}(p^*_h \,\|\, q_h)] / F(h; I_t). \tag{7}$$

Corollary 1. *(Bounds on the exploitability ratio). Under (A1)–(A4), for any predictor q_h and any horizon h with $F(h; I_t) > 0$, the realised reduction satisfies $G_q(h) \le F(h; I_t)$, with*

equality if and only if $q_h = p^*_h$ almost surely. Consequently $\chi_q(h) \leq 1$, and $\chi_q(h) = 1$ if and only if the predictor recovers the true conditional law. If, in addition, q_h improves weakly on the marginal baseline, then $\chi_q(h) \in [0, 1]$.

Proof. The upper bound and its equality condition follow from the non-negativity and definiteness of the expected Kullback–Leibler divergence in Theorem 1. The lower bound follows because a predictor that improves weakly on the marginal baseline has $G_q(h) \geq 0$; dividing by the positive quantity $F(h; I_t)$ preserves the inequalities. □

The ratio is undefined where the ceiling is zero, which is the correct behaviour: where no predictive information exists, the notion of exploiting it is vacuous. In practice, horizons where the estimated profile is zero (after clipping) are excluded from exploitability calculations and marked as information-limited in any tabular or graphical presentation; the S&P 500 illustration in Section 8.5 exemplifies this regime, with all displayed horizons treated as undefined. A predictor that underperforms the marginal baseline, which can occur for a misspecified or poorly estimated model, yields a negative realised reduction and hence a negative ratio; this is diagnostically informative rather than pathological, flagging a model that performs worse than the unconditional forecast. Reporting $\chi_q(h)$ across horizon therefore separates two questions that aggregate error measures conflate: how much predictive information the information set contains, and how much of it the fitted model captures. This decomposition is developed here as a corollary; empirical estimation of $\chi(h)$ using fitted probabilistic models is left to future work.

The joint behaviour of the forecastability ceiling $F(h; I_t)$ and the exploitability ratio $\chi_q(h)$ defines four qualitative forecasting regimes:

- High forecastability, low exploitability: substantial predictive information exists, but the fitted model fails to extract it efficiently.
- Low forecastability, low exploitability: the declared information set contains little predictive information, so model improvements alone cannot materially improve prediction.
- High forecastability, high exploitability: the information set is rich and the predictor operates near the information-theoretic frontier.
- Low forecastability, high exploitability: the predictor extracts most of the available predictive information despite a low information-theoretic limit.

This taxonomy clarifies where forecasting effort should be directed. Poor predictive performance is often attributed to model inadequacy, motivating increasing model complexity. The decomposition shows that such effort is warranted only in approximation-limited regimes; in information-limited regimes, changing the model class alone is unlikely to improve prediction unless the declared information set itself changes. Because the forecastability ceiling is fixed before any forecasting model is selected, the regime can in principle be assessed prior to model development.

For example, chaotic systems may exhibit substantial short-horizon forecastability followed by rapid predictability collapse. A model applied to such a system might operate in an approximation-limited regime at short horizons, where a better architecture could extract more of the rich available structure, before transitioning into a strictly information-limited regime at longer horizons where the ceiling itself vanishes. Conversely, processes whose forecastability profile remains uniformly close to zero are intrinsically information-limited relative to the declared information set, so increasing model complexity alone is unlikely to materially improve prediction.

The exploitability ratio is subordinate to the profile and developed here only as a corollary. Its numerator is a difference of expected log losses, estimable from held-out scoring of the fitted predictor and the marginal baseline; its denominator is the estimated profile, which inherits the finite-sample bias of the mutual-information estimator discussed in Section 7 and is ill-conditioned near the estimation floor. The empirical study in this paper reports profiles rather than fitted exploitability ratios to keep the illustrations independent of any particular forecasting architecture; full empirical treatment is left to future work.

5. Structural Properties of the Forecastability Profile

5.1. Information-Set Monotonicity

Any compression of the information set can only reduce the forecastability ceiling. This follows from the data processing inequality and has a direct operational reading: richer architectures and feature-engineering pipelines can reduce approximation error but cannot create predictive information the information set does not contain.

Proposition 1. *(Data processing inequality for forecastability). Let T be any measurable transformation of I_t. Then for all $h \geq 1$,*

$$F(h; T(I_t)) \leq F(h; I_t) \tag{8}$$

Equality holds when Y_{t+h} and I_t are conditionally independent given $T(I_t)$, that is, when $T(I_t)$ is sufficient for Y_{t+h} relative to I_t.

Proof. Since $T(I_t)$ is a measurable function of I_t, the Markov structure $Y_{t+h} \to I_t \to T(I_t)$ holds. The data processing inequality [31] gives the result. □

5.2. Finite-Window Truncation Budget

Finite lag windows are universal in practice. The following result quantifies the exact information cost of this truncation as a conditional mutual information identity; Figure 1 gives the schematic.

Proposition 2. *(Truncation budget). For all $p \geq 1$ and $h \geq 1$,*

$$F(h; I_t) - F(h; I_t^{(p)}) = I(Y_{t+h}; \text{remote past} \mid Y_t, \ldots, Y_{t-p+1}) \quad (9)$$

The right-hand side is non-negative and equals zero if and only if Y_{t+h} is conditionally independent of the remote past given the retained window.

Proof. Write the full information set as the join of the recent window and the remote past. Applying the chain rule for mutual information [31] and rearranging yields the identity; non-negativity of conditional mutual information completes the proof. □

The truncation budget is an exact information-accounting consequence of the chain rule. In short-memory processes the conditional mutual information is negligible for moderate p; in long-memory processes it may remain substantial across horizons.

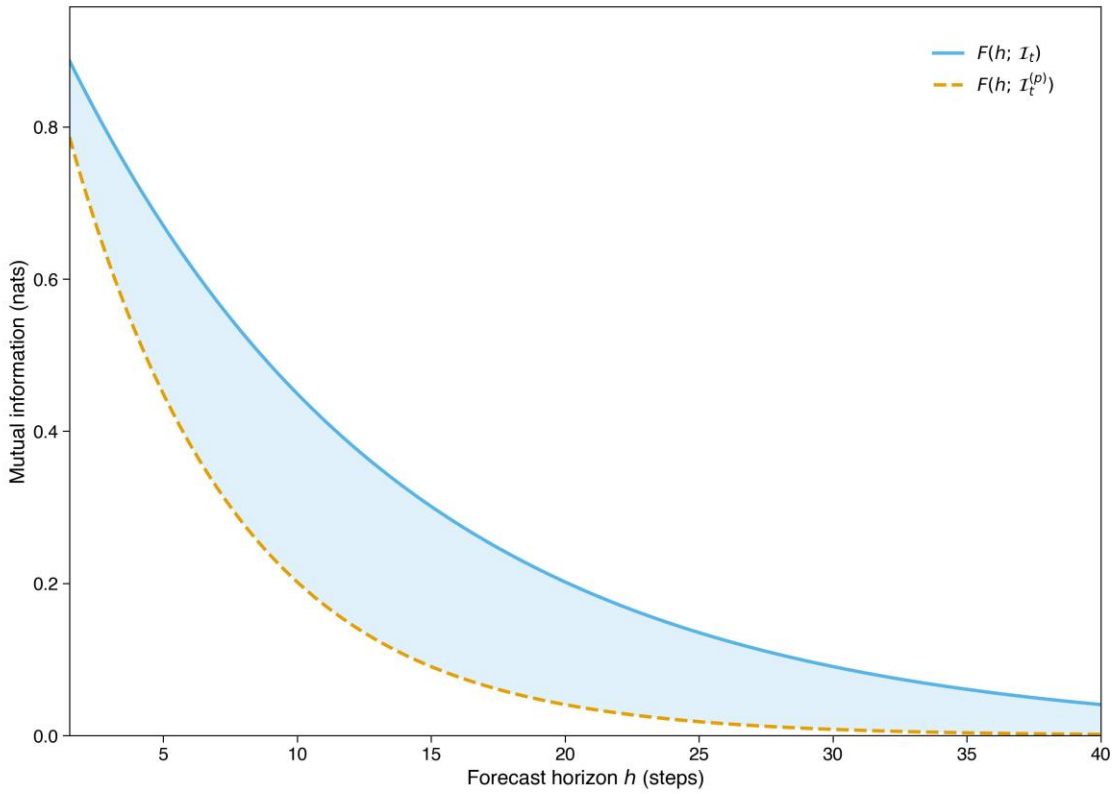


Fig. 1 Information-set truncation. The full profile $F(h; I_t)$ (solid blue) lies weakly above the truncated profile $F(h; I_t^{(p)})$ (dashed orange) for a representative AR process; the shaded gap is the truncation budget of Proposition 2 (Equation 9). A finite lag window necessarily discards the remote past, incurring an exact information cost equal to the conditional mutual information between the target and the omitted history.

5.3. Periodicity Inheritance

Many processes of practical importance have periodic conditional structure. Periodicity in the conditional law propagates directly to the forecastability profile.

Proposition 3. *(Periodicity inheritance). Let V_t be a random vector generating I_t. Suppose there exists s in the natural numbers such that for all h ≥ 1, the pairs (Y_{t+h}, V_t) and (Y_{t+h+s}, V_t) are equal in distribution. Then $F(h + s; I_t) = F(h; I_t)$ for all h ≥ 1.*

Proof. Joint-law periodicity implies equal marginal distributions, so the marginal entropies coincide; since conditional entropy is a functional of the joint law, the conditional entropies also coincide. Subtracting completes the proof. □

Remark 2. *(Damped harmonic recurrence). In stable seasonal systems, approximate joint-law periodicity yields damped harmonic recurrence: peaks at seasonal horizon multiples h = s, 2s, 3s, … whose amplitudes decline across successive cycle transitions. This pattern is distinct from both monotone decay (no seasonal structure) and exact periodicity (constant amplitudes). The seasonal autoregressive illustration in Section 8 shows peaks at h = 12, 24, 36 with estimated heights 0.858, 0.556, and 0.378 nats, confirming amplitude decline.*

6. Relation to Existing Information-Theoretic Quantities

Entropy rate. The Kolmogorov–Sinai entropy rate [6] characterises asymptotic per-step uncertainty as a scalar with no horizon structure. The profile at a fixed horizon recovers the per-step information available at that horizon, but is information-set relative and carries a direct log-loss interpretation.

Approximate and sample entropy. Approximate entropy [7] and sample entropy summarise short-sequence regularity as heuristic scalar indices with no decision-theoretic interpretation. The profile replaces these heuristics with an exact mutual-information identity grounded in proper scoring rules.

Excess entropy and predictive information. Excess entropy [12] is the mutual information between the semi-infinite past and future; predictive information [13] studies how past-future mutual information scales with block length. Neither is horizon-specific or carries a per-horizon log-loss interpretation.

Transfer entropy. Transfer entropy [9] measures directional information flow from a source process to a target and does not carry a log-loss ceiling interpretation in the forecasting sense. When source and target coincide it reduces to a form of auto-mutual information, the special case the profile builds on.

Granger causality and linear R^2. In jointly Gaussian settings the profile reduces to $-\frac{1}{2}\log(1 - R^2_h)$, aligning with mean-squared-error predictability in the Gaussian case. The profile generalises beyond linearity and Gaussianity.

Summary. The framework combines horizon-specific evaluation, information-set explicit conditioning, grounding in a strictly proper scoring rule, and a decision-theoretic interpretation as the maximum achievable reduction in expected log loss. Among the quantities reviewed here: entropy rate provides horizon-agnostic per-step uncertainty with no log-loss ceiling interpretation; transfer entropy quantifies cross-process

directional flow rather than a single-target self-lag ceiling; predictive information characterises block-scaling complexity without per-horizon resolution or a proper-scoring grounding. The profile differs from each of these on at least two of the four dimensions, as Table 1 summarises.

Table 1. Structural comparison of the forecastability profile against established information-theoretic quantities. ✓ = property holds; – = property does not hold or is not the primary purpose of the measure; (✓) = holds in special cases only.

Property	Entropy rate	Predictive information	Transfer entropy	Granger causality	Forecastability profile
Horizon indexed	–	–	–	–	✓
Information-set explicit	–	–	(✓)	(✓)	✓
Single-target (self-lag)	✓	✓	–	–	✓
Scoring-rule grounded	–	–	–	–	✓
Decision-theoretic ceiling	–	–	–	–	✓
Pre-modelling (no model fit)	✓	✓	✓	(✓)	✓
Directional / inter-process	–	–	✓	✓	–

7. Estimating Forecastability Profiles

7.1. The KSG Estimator

For the p-lag truncated information set the forecastability reduces to a mutual information between a scalar future and a p-dimensional lag vector,

$$F(h; I_t^{(p)}) = I(Y_{t+h}; Y_t, \ldots, Y_{t-p+1}) \tag{10}$$

estimated using the Kraskov–Stögbauer–Grassberger (KSG) k-nearest-neighbour estimator [32] with joint-radius marginal counts, the −1/k digamma correction, and coordinate-wise standardisation. The structural results of Sections 4 and 5 are population-level identities; the estimated profiles in Section 8 are finite-sample approximations subject to the estimator limitations described in Section 9.2, particularly downward bias that grows with embedding dimension. Readers should treat profile shape and surrogate separation as the informative quantities, not absolute magnitudes. The KSG estimator is used here because it is nonparametrically consistent under mild regularity conditions, adapts to local density variation without requiring a fixed bandwidth or distributional assumption, and is the standard choice in nonlinear dependence estimation where the conditional structure of the target is unknown [32, 33], including in analytics applications of information-theoretic diagnostics [34]. We fix $k = 5$ throughout; this value lies in the range typically recommended for moderate sample sizes ($N = 2500$) based on simulation studies in [32, 33], and sensitivity checks with $k \in \{3, 7, 10\}$ produced qualitatively identical profile shapes and surrogate outcomes for the five illustrative series, though absolute magnitudes vary with k as expected from the bias–variance trade-off. Qualitative profile structure, including decay rate, recurrence peaks, and surrogate separation, was stable across these nearby k choices.

Clipping convention. Since mutual information is non-negative at the population level, negative finite-sample estimates are set to zero when constructing displayed profiles. The KSG estimator is consistent under regularity conditions but exhibits a systematic downward bias that increases with embedding dimension p and decreases with sample size [33]; in high-dimensional lag embeddings the neighbour distances concentrate and the bias grows, so absolute magnitudes at large p are interpreted with caution and inference rests on profile shape and surrogate separation rather than on point values.

7.2. Profile-Level Block-Surrogate Test

Estimating the profile does not by itself determine whether it contains more integrated predictive information than expected from a locally structured but misaligned surrogate. The null hypothesis of the surrogate test is that the observed integrated positive forecastability mass $T+$ is no greater than would arise from a process with the same short-range temporal structure but no alignment between future values and the lag embedding. Under the null, the observed profile contains no horizon-specific

predictive information beyond what local autocorrelation structure alone would produce. Block-permuted surrogates preserve local temporal structure within blocks of length L = max(10, p, s), with s = 12 for monthly seasonal series, while destroying the alignment between future values and the lag embedding. The test statistic is the integrated positive forecastability mass,

$$T^{+} = \Sigma_{h=1}^{H} \max\{\hat{F}_{raw}(h), 0\} \quad (11)$$

and the one-sided surrogate p-value is the proportion of surrogate statistics at least as large as the observed value. We use B = 999 surrogates. The test operates at the profile level: it does not test for horizon-wise significance, and a small p-value alone does not imply practically useful forecastability.

8. Empirical Illustrations of Predictive-Information Geometry

The five examples below demonstrate distinct profile structures established by the theory in Sections 4 and 5 (monotone decay, damped harmonic recurrence, nonlinear collapse, quasi-periodic long-memory dependence, and near-zero weak dependence). They are structural illustrations, not forecasting benchmarks or model-performance claims. All share one estimation workflow (KSG estimator, k = 5, block-permutation surrogate test, B = 999). Specifications are in Table 2, results in Table 4, and profiles in Figure 2.

Table 2. Series specifications for the empirical illustration panel. N: series length; p: lag embedding order; H: maximum forecast horizon; k: KSG neighbour count.

Series	Type	N	p	H	k	Expected Geometry
AR(1), φ = 0.95	Synthetic	2500	1	40	5	Monotone decay
Seasonal AR	Synthetic	2500	1	40	5	Damped harmonic peaks
Logistic map, r = 3.9	Synthetic	2500	1	40	5	Nonlinear collapse
Sunspots	Real	2500	12	120	5	Slow quasi-periodic decay
S&P 500 returns	Real	2500	3	40	5	Near-zero

8.1. AR(1)

The AR(1) process ($\varphi = 0.95$) provides a Gaussian closed-form reference against which the estimated profile can be read. The profile decays monotonically from 1.025 nats at h = 1 to zero by h = 13, lying below the analytical ceiling throughout; the gap widens from 11.9% at h = 1 to 84.6% at h = 12 as signal approaches the finite-sample floor (Table 3; $p_{mass} = 0.001$, $T_{+} = 3.609$ nats).

Table 3. AR(1) (φ = 0.95, p = 1, N = 2500, k = 5): Gaussian analytical reference F_ref(h) = −½ ln(1 − φ^(2h)) versus the KSG non-negative estimate at selected horizons.

h	F_ref(h) (nats)	Estimate (nats)	Absolute gap	Percentage gap
1	1.164	1.025	0.139	11.9%
2	0.842	0.675	0.167	19.8%
3	0.664	0.517	0.147	22.2%
4	0.544	0.386	0.159	29.1%
12	0.173	0.027	0.146	84.6%

8.2. Seasonal AR

With p = 1 the information set is the single lag Y_t, informative mainly at seasonal multiples. Peaks at h = 12, 24, 36 of 0.858, 0.556, and 0.378 nats with near-zero values between show a declining envelope consistent with damped harmonic recurrence rather than exact periodicity: predictive information is localised, not absent (p_{mass} = 0.001, T_+ = 1.792 nats).

8.3. Logistic map

The logistic map at r = 3.9 is chaotic, with a positive Lyapunov exponent. The profile opens at 2.924 nats and collapses to the floor by h = 7, the fastest decay in the panel and consistent with the Lyapunov timescale of information loss (p_{mass} = 0.001, T_+ = 7.443 nats).

8.4. Sunspots

The monthly sunspot number (p = 12, H = 120) opens at 0.686 nats and decays broadly, capturing the slow rise and fall of the ~132-month solar cycle rather than harmonic peaks; residual elevation at h = 120 (0.132 nats) marks cycle completion (p_{mass} = 0.001, T_+ = 13.730 nats).

8.5. S&P 500 returns

A daily log-return series (1 January 2010 to 31 December 2019, fixed before estimation) serves as a weak-dependence benchmark. With p = 3 the estimated profile is zero at every horizon and the positive-mass test is non-significant (p_{mass} = 1.000, T_+ = 0.000 nats): under this information set the profile is indistinguishable from the surrogate null, consistent with the long-standing view that daily financial returns closely resemble random walks [35]. This result cannot distinguish genuine weak dependence

from signal below the KSG estimation floor; see Section 9.4 for the bias mechanism and its consequences for inference.

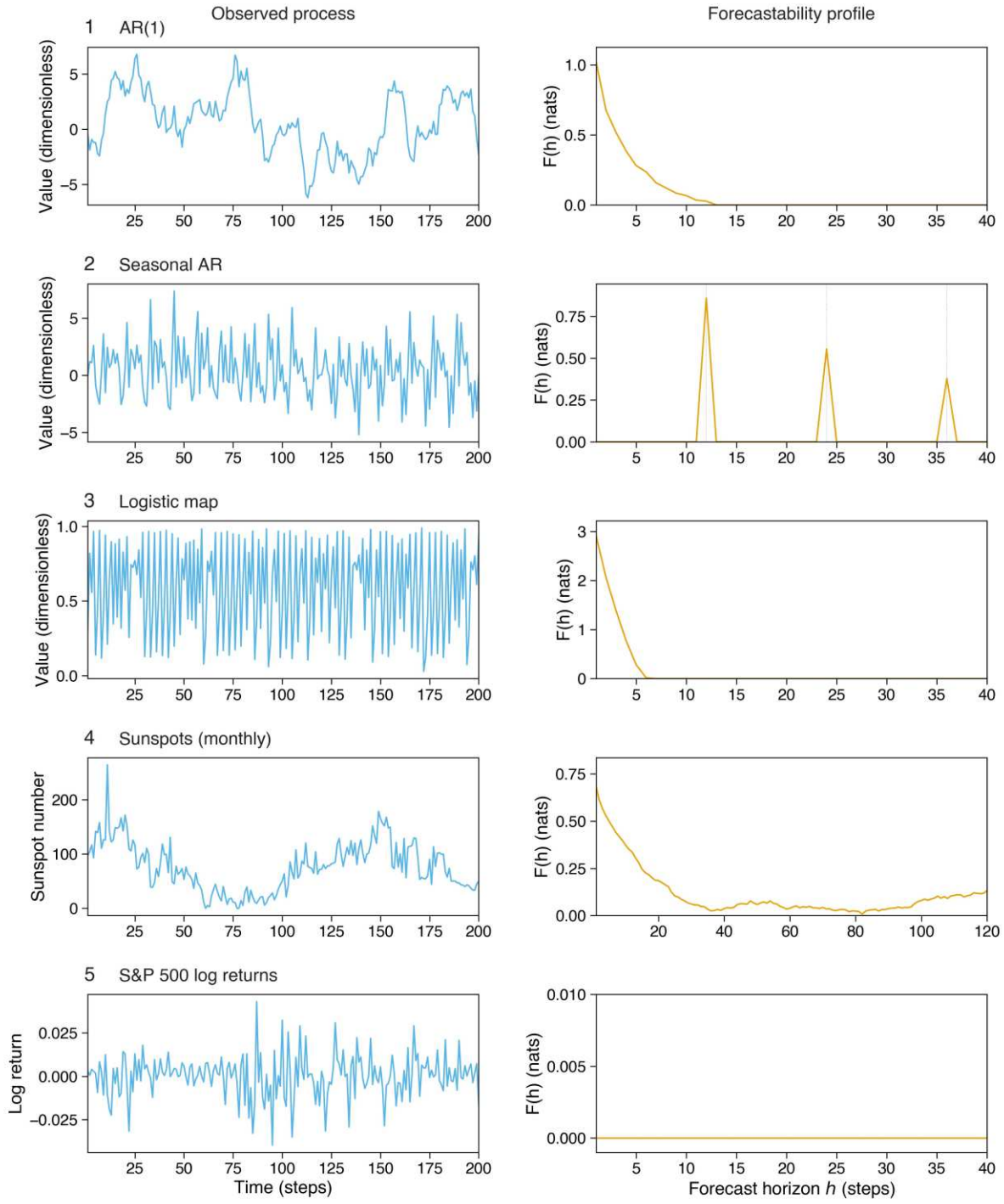


Fig. 2 Forecastability profiles for the five illustrative series. Left column: observed process (200 observations). Right column: estimated non-negative KSG profile F(h) in nats. Rows from top: AR(1) (φ = 0.95, p = 1, H = 40); seasonal AR (p = 1, H = 40); logistic map (r = 3.9, p = 1, H = 40); monthly sunspots (p = 12, H = 120); S&P 500 log returns (p = 3, H = 40). All series N = 2500, k = 5, B = 999. Each panel illustrates

a distinct horizon-specific dependence structure; the examples are not forecasting benchmarks.

Table 4. Profile-level block-surrogate positive-mass test. T+ is the integrated positive forecastability mass; mean F+ is the mean positive profile height. p_mass is the one-sided surrogate p-value (B = 999, k = 5); values ≤ 0.05 are marked with an asterisk.

Series	Type	Effect Class	T+	Mean F+	p_mass
AR(1), φ = 0.95	Synthetic	Monotone decay	3.609	0.090	0.001 *
Pure seasonal AR	Synthetic	Damped harmonic recurrence	1.792	0.045	0.001 *
Logistic map	Synthetic	Nonlinear collapse	7.443	0.186	0.001 *
Sunspots (monthly)	Real	Quasi-periodic dependence	13.730	0.114	0.001 *
S&P 500	Real	Weak dependence	0.000	0.000	1.000

9. Discussion and Conclusions

9.1. Theoretical Implications

Forecastability is a property of the information structure defined by the declared information set and the chosen loss criterion, not of any forecasting method. Under log loss, the profile is the predictive-information ceiling, relative to a declared information set. Any predictor using the declared information set operates within this ceiling; only enlarging the information set can raise it. The three structural properties govern how the profile behaves as a function of information-set specification. The central advance is that the profile preserves the organisation of predictive information across horizon: entropy rate, approximate entropy, and permutation entropy each return a single number and so cannot separate dependence regimes that differ in horizon structure but coincide in aggregate uncertainty. The horizon-specific object distinguishes short-memory decay, seasonal recurrence, and chaotic collapse without spectral analysis or model fitting.

9.2. Operational Implications

The forecastability ceiling and the exploitability ratio together place any prediction problem in one of the four regimes of Section 4.4, and the regime dictates where effort is worth spending: where the profile is high but only partly captured, richer model classes or better estimation can raise performance, whereas where the profile is low the

ceiling binds and only expanding or transforming the declared information set can help. The same process may move between regimes across horizon, as in the chaotic example, so the assessment is made per horizon rather than once for the whole series. Because the ceiling is fixed before any model is selected, this regime diagnosis can be carried out at the screening stage, redirecting effort away from model development where the information, not the model, is the limiting factor, a form of model-development triage based on the intrinsic prediction ceiling. The framework supports pre-modelling assessment of whether additional model complexity is likely to improve prediction under the declared information set.

9.3. Research Implications

Benchmark comparisons that evaluate models on series with near-zero profiles may largely reflect estimation noise or information limitation rather than modelling ability [36], so competition-based evaluations would benefit from pre-screening series by forecastability profile. The profile also complements feature-based model-selection approaches [22, 37] by supporting principled algorithm selection before any hold-out evaluation: monotone decay suggests short-memory specifications; damped harmonic recurrence identifies seasonal structure; near-zero profiles indicate that parsimonious baselines are appropriate.

9.4. Limitations

Scope of the log-loss identity. Theorem 1 holds specifically for log loss. The profile is an exact ceiling under log loss and a principled dependence diagnostic under other criteria.

Information-set specification. The profile is a property of the pair (series, information set), not of the series alone. Different lag lengths, exogenous variables, or transformations produce different profiles, and comparisons across studies require care.

Stationarity. The theoretical results assume stationarity. Trends, regime shifts, and seasonal nonstationarity can inflate apparent dependence. Both real-world series pass standard stationarity checks within the analysed window (augmented Dickey–Fuller and KPSS tests [38, 39]). Extending the profile to locally stationary processes [40], where it would be estimated as a function of both horizon and calendar time, is a natural relaxation of this assumption.

Estimator sensitivity. The empirical profiles inherit the known limitations of nearest-neighbour mutual-information estimation: the KSG estimator is consistent but downward-biased in finite samples, and the bias grows with embedding dimension p as neighbour distances concentrate [32, 33]. Genuine long-range or high-order dependence can therefore be pushed below the estimation floor and read as absent, so the near-zero S&P 500 profile in particular cannot distinguish weak dependence from sub-floor signal. Inference accordingly rests on profile shape and surrogate separation

rather than on absolute magnitudes. Adaptive-k and bias-corrected estimators would sharpen the profiles but lie beyond the scope of this primarily theoretical contribution.

9.5. Future Directions

1. Multivariate information sets: adding exogenous predictors alongside lag embeddings is handled without modification to the truncation budget or monotonicity results, but the practical question of how the profile changes as predictors are added or removed, an information-set selection problem, remains open.

2. Finite-sample estimation theory for KSG-estimated profiles, including calibrated pointwise confidence bands.

3. Estimating and interpreting the exploitability ratio in practice using fitted probabilistic models.

9.6. Conclusion

This paper develops the forecastability profile: the horizon-indexed mutual information between a future observation and a declared information set, which under log loss is the exact ceiling on achievable predictive improvement. Three structural results, compression monotonicity, an exact finite-window truncation budget, and periodicity inheritance, describe how that ceiling responds to the specification of the information set, and the exploitability ratio separates information-limited from approximation-limited prediction. Read as an analytics tool, the profile supports pre-modelling triage: before any model is fitted, it identifies the horizons at which predictive structure exists and the regimes in which further modelling effort can and cannot help. It is offered as a diagnostic that precedes model development rather than as a competing forecasting method, and the empirical illustrations demonstrate its geometry rather than benchmark performance.

Author contributions P.M.C. is the sole author and conceived the study, developed the theoretical results, implemented the analysis and figures, and wrote the manuscript.

Funding The author received no specific funding for this work.

Data availability All datasets used in the empirical illustrations are publicly available. Replication code and configuration files are archived at Zenodo (https://doi.org/10.5281/zenodo.19931875). Running the archived script regenerates the outputs reported in this manuscript.

Declarations

Competing interests The author declares no competing interests.

Ethics approval and consent to participate Not applicable. This study uses only publicly available and simulated data and involves no human participants or animals.

Consent for publication Not applicable.

Clinical trial number: not applicable.

References

1. Petropoulos F, Apiletti D, Assimakopoulos V, et al. Forecasting: theory and practice. Int J Forecast. 2022;38:705–871. https://doi.org/10.1016/j.ijforecast.2021.11.001.
2. Makridakis S, Spiliotis E, Assimakopoulos V. The M4 Competition: 100,000 time series and 61 forecasting methods. Int J Forecast. 2020;36:54–74. https://doi.org/10.1016/j.ijforecast.2019.04.014.
3. Makridakis S, Spiliotis E, Assimakopoulos V. Statistical and machine learning forecasting methods: concerns and ways forward. PLoS One. 2018;13:e0194889. https://doi.org/10.1371/journal.pone.0194889.
4. Chatterjee I, Chakraborti S, Tono T. A comparative study of bitcoin price prediction during pre-Covid19 and whilst-Covid19 period using time series and machine learning models. Discov Anal. 2024;2:19. https://doi.org/10.1007/s44257-024-00024-z.
5. Khurram H. Auxiliary-lag dependent Gaussian process model for forecasting using proposed kernels and multi-start optimization method. Discov Anal. 2026;4:14. https://doi.org/10.1007/s44257-026-00067-4.
6. Kolmogorov AN. A new metric invariant of transitive dynamical systems and automorphisms in Lebesgue spaces. Dokl Akad Nauk SSSR. 1958;119:861–4.
7. Richman JS, Moorman JR. Physiological time-series analysis using approximate entropy and sample entropy. Am J Physiol Heart Circ Physiol. 2000;278:H2039–49.
8. Bandt C, Pompe B. Permutation entropy: a natural complexity measure for time series. Phys Rev Lett. 2002;88:174102.
9. Schreiber T. Measuring information transfer. Phys Rev Lett. 2000;85:461–4.
10. Granger CWJ. Investigating causal relations by econometric models and cross-spectral methods. Econometrica. 1969;37:424–38.
11. Runge J, Nowack P, Kretschmer M, Flaxman S, Sejdinovic D. Detecting and quantifying causal associations in large nonlinear time series datasets. Sci Adv. 2019;5:eaau4996. https://doi.org/10.1126/sciadv.aau4996.
12. Crutchfield JP, Feldman DP. Regularities unseen, randomness observed: levels of entropy convergence. Chaos. 2003;13:25–54.
13. Bialek W, Nemenman I, Tishby N. Predictability, complexity, and learning. Neural Comput. 2001;13:2409–63.
14. Gneiting T, Raftery AE. Strictly proper scoring rules, prediction, and estimation. J Am Stat Assoc. 2007;102:359–78.

15. Gneiting T, Katzfuss M. Probabilistic forecasting. Annu Rev Stat Appl. 2014;1:125–51.
16. Kullback S. Information theory and statistics. New York: Wiley; 1959.
17. Fraser AM, Swinney HL. Independent coordinates for strange attractors from mutual information. Phys Rev A. 1986;33:1134–40.
18. Kantz H, Schreiber T. Nonlinear time series analysis. 2nd ed. Cambridge: Cambridge University Press; 2004.
19. Paluš M, Vejmelka M. Directionality of coupling from bivariate time series. Phys Rev E. 2007;75:056211.
20. Amigó JM, Dale R, Tempesta P. A generalized permutation entropy for noisy dynamics and random processes. Chaos. 2021;31:013115.
21. Rissanen J. Modeling by shortest data description. Automatica. 1978;14:465–71.
22. Kang Y, Hyndman RJ, Smith-Miles K. Visualising forecasting algorithm performance using time series instance spaces. Int J Forecast. 2017;33:345–58. https://doi.org/10.1016/j.ijforecast.2016.09.004.
23. Pompe B, Runge J. Momentary information transfer as a coupling measure of time series. Phys Rev E. 2011;83:051122.
24. Barnett L, Barrett AB, Seth AK. Granger causality and transfer entropy are equivalent for Gaussian variables. Phys Rev Lett. 2009;103:238701.
25. DelSole T. Predictability and information theory. Part I: measures of predictability. J Atmos Sci. 2004;61:2425–40.
26. DelSole T, Tippett MK. Predictability: recent insights from information theory. Rev Geophys. 2007;45:RG4002.
27. Lorenz EN. The predictability of a flow which possesses many scales of motion. Tellus. 1969;21:289–307.
28. Barnett L, Bossomaier T. Transfer entropy as a log-likelihood ratio. Phys Rev Lett. 2012;109:138105.
29. Catt PM. Forecastability: insights from physics, graphical decomposition, and information theory. Foresight Int J Appl Forecast. 2009;13:24–33.
30. Shannon CE. A mathematical theory of communication. Bell Syst Tech J. 1948;27:379–423.
31. Cover TM, Thomas JA. Elements of information theory. 2nd ed. Hoboken: Wiley; 2006.
32. Kraskov A, Stögbauer H, Grassberger P. Estimating mutual information. Phys Rev E. 2004;69:066138. https://doi.org/10.1103/PhysRevE.69.066138.
33. Gao S, Ver Steeg G, Galstyan A. Efficient estimation of mutual information for strongly dependent variables. In: Proceedings of the 18th International Conference on Artificial Intelligence and Statistics (AISTATS); 2015. p. 277–86.

34. Osman EGA. Entropy based solutions for detecting and treating multicollinearity in econometrics. Discov Anal. 2025;3:18. https://doi.org/10.1007/s44257-025-00047-0.
35. Tarani S, Kyriazi F, Thomakos D. Random walks, random chains and the contributions of Holbrook Working. Discov Anal. 2024;2:20. https://doi.org/10.1007/s44257-024-00027-w.
36. Hewamalage H, Ackermann K, Bergmeir C. Forecast evaluation for data scientists: common pitfalls and best practices. Data Min Knowl Discov. 2023;37:788–832. https://doi.org/10.1007/s10618-022-00894-5.
37. Montero-Manso P, Athanasopoulos G, Hyndman RJ, Talagala TS. FFORMA: feature-based forecast model averaging. Int J Forecast. 2020;36:86–92. https://doi.org/10.1016/j.ijforecast.2019.02.011.
38. Dickey DA, Fuller WA. Distribution of the estimators for autoregressive time series with a unit root. J Am Stat Assoc. 1979;74:427–31.
39. Kwiatkowski D, Phillips PCB, Schmidt P, Shin Y. Testing the null hypothesis of stationarity against the alternative of a unit root. J Econom. 1992;54:159–78.
40. Dahlhaus R. Fitting time series models to nonstationary processes. Ann Stat. 1997;25:1–37.