arXiv:2504.19612v2 [physics.data-an] 21 Sep 2026

When Should Team KPIs Be Absolute or Relative for Match-Outcome Prediction?

Rowan Brown*
Swansea University, Swansea, UK

Liam P. Kilduff
Swansea University, Swansea, UK

G. A. Scott
Swansea University, Swansea, UK

Abstract

In rugby union and association football, team key performance indicators (KPIs) can be represented in absolute terms or relative to the opponent. Relativisation sometimes improves match-outcome prediction and sometimes harms it. There has been no general account of when each occurs.

The answer depends on how much the two teams differ in variability and how strongly their KPI values rise and fall together. We combine these properties into the Paired Efficiency Factor (PEF), which generalises Fisher's paired-efficiency result to the unequal-variance conditions typical of competitive sport. The PEF also connects a KPI's statistical efficiency to how much information its relative form carries about the outcome. Combined KPIs can interact in complex ways, so we analyse each indicator on its own.

Across 86 team KPIs from professional rugby union and association football, the PEF places every metric in one of four regimes. Anti-correlation is common, especially for high-volume competitive counts, and absolute measures are then usually preferable. Relativisation can still improve prediction when a noisier difference carries more outcome information. An idealised simulation and one representative KPI from each regime confirm both signs under team-blocked cross-validation. The same pairing geometry appears in healthcare, genomics, finance, and manufacturing.

The PEF turns an ad hoc feature-engineering choice into a transparent, data-informed diagnostic for when to relativise performance metrics and when not to.

1 Introduction

1.1 Why Relativisation? From Cross-Domain Practice to Sport

Across quantitative disciplines, a recurring analysis decision is whether to represent a measurement in absolute terms or relative to a comparator. In many cases, the relative quantity is preferred because it cancels a shared baseline and isolates the signal of interest. For example, in finance, asset returns are routinely benchmarked against the market [1, 2]; in clinical medicine, pulse pressure, the difference between systolic and diastolic blood pressure, is an established relative biomarker [3, 4]; and in industrial process control, statistical control charts track deviations from a reference rather than raw values [5]. Management research more broadly emphasises that competitive benchmarking shifts attention from absolute totals to relative performance [6].

The same choice arises especially sharply in competitive sport: whether to represent a team's performance in absolute terms or relative to the opponent. In rugby union and association football, several studies contrast absolute and relative team-level formulations of key performance indicators (KPIs) for match-outcome prediction [7, 8, 9, 10, 11]. These studies typically assess

*Corresponding author. m.r.brown@swansea.ac.uk.

KPIs by building match-outcome classifiers, most commonly logistic regression, from team-level features, evaluated by cross-validated predictive accuracy [12, 13]. Quantitative international sports ratings illustrate the same long-standing reliance on structured comparison under common rules [14]. The rugby KPI literature is large and often inconsistent across competitions [15, 16]. The reported results run in several directions. Relative formulations improve accuracy in the United Rugby Championship [10]. They show no significant improvement over absolute features in international women's rugby [11]. The same pattern holds in Australian sub-elite rugby [9]. A recent six-league football preprint again finds that team–opponent differences outperform isolated team statistics [17]. Those studies treat the choice as help versus no help. They do not give a regime in which relativisation should be avoided. Observational win-probability estimators are themselves high-variance objects [18]. The central question of this paper is therefore simple to state: *why does relativisation of some KPIs improve outcome prediction, while for others it is predictively harmful?*

Underlying this question is an apparent paradox. The baseline-cancellation rationale above presumes that differencing reduces noise. Yet when two competitors are negatively associated, so that one performs well precisely when the other performs poorly, the difference can have *larger* variance than either raw measurement. Classical reasoning treats that larger variance as harmful to inference. Practitioners nonetheless observe that relative features can predict better in exactly these competitive settings. How can a feature with higher variance carry more predictive information? Answering this requires returning to the statistical principle that underlies relativisation.

1.2 Fisher's Paired Efficiency

The principle that correlated measurements yield more efficient inference than independent samples has been recognised since Fisher's work on experimental design [19]. For two correlated measurements X_{A} and X_{B} with equal variances σ^2 and correlation ρ, the variance of their difference is

$$\mathrm{Var}(X_{\mathrm{A}} - X_{\mathrm{B}}) = 2\sigma^2(1-\rho).$$

Compared to independent measurements (variance $2\sigma^2$), the efficiency gain is $1/(1-\rho)$. This result, which is purely algebraic and requires no distributional assumptions beyond finite second moments, underpins the paired t-test [20] and decades of experimental design across medicine, psychology, and agriculture [19].

This classical formula, however, rests on a single structural assumption: equal variances. The competitive data that motivate our question routinely violate it [21]. Examining KPIs from professional rugby union and football, we observe two features Fisher's equal-variance relation cannot represent. First, for each KPI the per-match paired-difference series $X_{\mathrm{A},i} - X_{\mathrm{B},i}$ is a sampled distribution across fixtures. Those differences are approximately normal even where the individual team series are mildly right-skewed, the behaviour expected of paired differences (Supplementary section 6). Second, the two teams' per-match series rarely share a common variance, and their cross-fixture correlation takes both positive values (from shared environmental conditions) and negative values (from competitive dynamics) depending on the KPI (table 4). Variance asymmetry and signed correlation are therefore the norm rather than the exception, and Fisher's equal-variance formula cannot by itself answer when relativisation helps.

1.3 The Paired Efficiency Factor

These observations, approximately normal differences with non-zero means and unequal second moments, motivate a formula parameterised jointly by the variance ratio and the correlation. We allow unequal variances σ_{A}^2 and σ_{B}^2 with variance ratio $\kappa = \sigma_{\mathrm{B}}^2/\sigma_{\mathrm{A}}^2$. The variance of the difference

in this case then becomes

$$\mathrm{Var}(X_\mathrm{A} - X_\mathrm{B}) = \sigma_\mathrm{A}^2\big(1 + \kappa - 2\sqrt{\kappa}\,\rho\big).$$

Defining the Paired Efficiency Factor as the ratio of unpaired to paired variance, similar to the Fisher efficiency gain relation above, we have:

$$\eta = \frac{1+\kappa}{1+\kappa-2\sqrt{\kappa}\,\rho}. \tag{1}$$

where $\eta > 1$ indicates that pairing reduces variance. Across sports KPIs, variance ratios and signed correlations occupy a wide range, with a broader spread in the supporting domains (table 4). Specifically, $\rho = \mathrm{Corr}(X_\mathrm{A}, X_\mathrm{B})$ is the Pearson correlation between the home and away KPI values *across fixtures*: for each match i, $X_{\mathrm{A},i}$ and $X_{\mathrm{B},i}$ are the two teams' counts from the same game, and $\hat{\rho}$ is computed from the n such paired observations (section 3).

Like the classical result, this formula is purely algebraic, and it recovers Fisher's equal-variance ($\kappa = 1$) as a special case. Thus, η depends jointly on correlation strength and variance asymmetry, a dependence the classical formula cannot express. Under bivariate normality and equal group sizes, the PEF equals the Pitman asymptotic relative efficiency of the paired t-test versus the independent two-sample t-test, confirming the variance-ratio formula (see Supplementary section 6 and [22] for the general framework).

1.4 Resolving the Tension: Efficiency versus Information

From (1) it is clear that negative correlation yields $\eta < 1$: pairing increases the variance of the difference relative to independent measurement, which classical reasoning treats as harmful. Yet match-outcome models routinely find that relative features ($X_\mathrm{A} - X_\mathrm{B}$) outperform absolute features (X_A, X_B) in precisely these competitive settings [7, 10]. The resolution requires distinguishing between **statistical efficiency**, how precisely parameters can be estimated, and **information content**, how much a feature reduces uncertainty about outcomes. We show these quantities are functionally related but can diverge, and we derive the conditions under which they do.

Resolving this tension carries a concrete practical payoff. Feature engineering, the choice between absolute and relative features, is currently governed largely by trial-and-error cross-validation. When κ, ρ, and the signal strength δ/σ_A can be estimated reliably, they provide a principled first-pass diagnostic for that choice, supplementing rather than replacing cross-validated checks. For the contested question of how predictable competitive outcomes really are, the framework supplies a per-feature account of both directions. It explains why a relative measure can carry more decision-relevant information than its absolute counterpart. It also identifies when the difference is the worse classifier. That account is a potential foundation on which multi-feature predictability studies can build.

1.5 Scope and Contributions

The primary empirical analysis draws on KPI data from professional rugby union and football, where head-to-head matchups produce natural pairing across a range of (κ, ρ) values, including negative within-match correlation. Performance-indicator construction in invasion games is well established [23], and modern team-sports analytics emphasises spatiotemporal and match-context structure [24]. A complementary programme evaluates the statistical properties of sports metrics before they enter prediction models [25, 26]. We supplement this with validation data from healthcare (NHANES oscillometric systolic and diastolic readings on the same visit), transcriptomics (GEO GSE47462), finance, and manufacturing (real CNC process telemetry) to assess generalisability. Supporting-domain summaries appear in table 4. The Supplementary

Information records the mathematical derivations, the sports KPI landscape, and the simulation diagnostics.

The empirical argument is organised around three questions. Why does relativisation improve outcome prediction for some KPIs and harm it for others? How can a relative feature with $\eta < 1$ still carry more information about the outcome? Does the same (κ, ρ) geometry appear outside invasion-game sport? Section 4 addresses each. Section 5 interprets the answers in that order. The theoretical framework and the methods underpin all three questions rather than answering them in sequence.

This paper makes the following contributions:

1. **PEF generalisation.** We extend Fisher's paired efficiency to unequal variances through a distribution-free formula (equation (1)).

2. **Information content relationship.** Under bivariate normality, we derive a closed-form relationship between η and the mutual information $I(X;Y)$ [27, 28], clarifying the efficiency–power tension.

3. **Quadrant taxonomy.** We partition the (κ, ρ) parameter space into four regimes with distinct statistical and predictive properties, each with corresponding feature engineering guidance.

4. **Mechanistic confirmation.** We confirm the predicted mechanism in two stages. An idealised probit simulation under assumptions (A1)–(A2) reproduces the η–$I(X;Y)$ mapping and the efficiency–power tension in a controlled setting (section 4; Supplementary section 6). Four single-KPI exemplars, one per quadrant, then test the sign of ΔML in real data when signal strength (δ/σ_A) is held comparable (figure 3 and table 3). The exemplars confirm help and harm at matched signal. They are not a claim of large Quadrant 4 predictive gain. Multi-KPI feature interactions are identified as a direct extension (section 5).

5. **Cross-domain applicability.** We show that market-adjusted returns, paired clinical designs, manufacturing control charts, and relative sports metrics are all instances of correlation-based variance reduction, differing only in their (κ, ρ) values. Supporting validation across four additional domains is consistent with the framework's generalisability.

1.6 Organisation

Section 2 develops the PEF formula, the information-content relationship, and the quadrant taxonomy. Section 3 specifies the three-tier validation. Those sections supply the shared apparatus. Section 4 addresses the three questions. Section 5 interprets the answers. Mathematical derivations with full intermediate steps appear in Supplementary section 6.

2 Theoretical Framework

This section develops the theoretical core sketched in the Introduction. We first define how pairing efficiency depends on the variance ratio κ and the correlation ρ, using only second-moment algebra that requires no distributional assumptions. We then connect efficiency to predictive information under bivariate normality, and translate the resulting (κ, ρ) geometry into quadrant-level guidance for relativisation. Readers focused on application may skim the formal derivations and proceed to the quadrant taxonomy (table 5) and the signal-strength discussion (section 2.5).

2.1 The Paired Efficiency Factor

2.1.1 Derivation

We begin from the variance of the difference between two measurements. When entity A and entity B are observed on the same occasion, their values may covary; the variance of $X_A - X_B$ captures how much noise remains after differencing. For random variables X_A and X_B with variances σ_A^2, σ_B^2 and correlation ρ,

$$\mathrm{Var}(X_A - X_B) = \sigma_A^2 + \sigma_B^2 - 2\rho\sigma_A\sigma_B. \tag{2}$$

The cross term $-2\rho\sigma_A\sigma_B$ is where correlation enters: positive ρ shrinks the variance of the difference because shared fluctuations partially cancel, whilst negative ρ amplifies it. Introducing $\kappa = \sigma_B^2/\sigma_A^2$ and factoring,

$$\mathrm{Var}(X_A - X_B) = \sigma_A^2\big(1 + \kappa - 2\sqrt{\kappa}\,\rho\big). \tag{3}$$

The variance ratio κ separates *how unequal* the two entity variances are from *how correlated* their measurements are; holding κ fixed, ρ alone determines whether differencing helps or harms efficiency.

The efficiency question is how this paired variance compares with treating the two measurements as independent. For unpaired measurements ($\rho = 0$), the baseline variance is $\sigma_A^2(1+\kappa)$. The Paired Efficiency Factor is their ratio, i.e. the expression in equation (1):

$$\eta = \frac{1+\kappa}{1+\kappa-2\sqrt{\kappa}\,\rho}.$$

with η, κ, and ρ as defined for equation (1) in section 1.3.

Remark. No distributional assumptions are required: equations (1) and (3) hold for any random variables with finite second moments.

2.1.2 Properties

Intuitively, stronger positive correlation means that shared fluctuations cancel more completely in the difference, so efficiency gain rises as ρ increases. Formally, PEF is strictly increasing in ρ for fixed $\kappa > 0$,

$$\frac{\partial\eta}{\partial\rho} = \frac{2\sqrt{\kappa}(1+\kappa)}{\big(1+\kappa-2\sqrt{\kappa}\,\rho\big)^2} > 0, \tag{4}$$

and convex in ρ,

$$\frac{\partial^2\eta}{\partial\rho^2} = \frac{8\kappa(1+\kappa)}{\big(1+\kappa-2\sqrt{\kappa}\,\rho\big)^3} > 0. \tag{5}$$

The sign of η relative to unity partitions pairing into three interpretable regimes:

- $\eta > 1$: pairing reduces variance (positive correlation dominates);
- $\eta = 1$: pairing provides no efficiency change (Fisher baseline);
- $\eta < 1$: pairing increases variance (negative correlation dominates).

Figure 1 visualises the PEF landscape across the (ρ, κ) parameter space.

2.1.3 Reduction to Classical Case

When the two competitors have equal variability ($\kappa = 1$, so $\sigma_A = \sigma_B$), the general formula simplifies to Fisher's classical result:

$$\eta = \frac{2}{2 - 2\rho} = \frac{1}{1 - \rho},$$

recovering the paired-efficiency formula from the Introduction [19]. That is, positive correlation shrinks the variance of the difference below the independent-measurement baseline ($\eta > 1$), whilst negative correlation amplifies it ($\eta < 1$). This is exactly the efficiency logic of the paired t-test [20]. The PEF formula therefore *contains* Fisher's result as a special case; unequal variances ($\kappa \neq 1$) are the ingredient the classical framework cannot represent on its own.

2.1.4 Provenance and Scope of Contribution

The variance-of-difference identity for unequal variances (equations (2) and (3)) is a standard second-moment result, and the efficiency-of-pairing argument under equal variances is due to Fisher [19]; the same paired-margin machinery underlies the Pitman–Morgan test for equality of correlated variances [29, 30]. Our contribution is not this algebra but its reframing: we collect it into a single closed-form efficiency ratio (equation (1)), parameterised by (κ, ρ), and repurpose it from an inferential setting toward feature construction and predictive information in competitive measurement.

2.2 The PEF–Information Content Relationship

Statistical efficiency (η) alone cannot explain why relative features sometimes improve prediction when $\eta < 1$. To address that paradox, we need a second quantity: how much the KPI difference reduces uncertainty about the match outcome.

2.2.1 Setup and Assumptions

We now connect PEF with predictive power. Let $Y \in \{0, 1\}$ denote the **binary match outcome** (whether entity A outperforms entity B on the measure of interest), and let $X = X_A - X_B$ be the corresponding relative feature. The predictive setting introduces a quantity beyond the second-moment pair (κ, ρ) that governs PEF: the **signal**, or mean separation $\delta = \mu_A - \mu_B$ between the two entities. Whereas (κ, ρ) fix the variance geometry of the relative feature, it is the magnitude of δ relative to σ_A that determines how much of that geometry translates into predictive power; this distinction is the crux of the efficiency–power tension and its consequences are developed in section 2.5. To express that link in closed form, two distributional assumptions are needed: one on the paired measurements themselves, and one on how the relative feature relates to the match outcome.

(A1) (X_A, X_B) follows a bivariate normal distribution with means μ_A, μ_B, variances σ_A^2, σ_B^2, and correlation ρ. Under (A1), $X = X_A - X_B$ is normally distributed with mean $\delta = \mu_A - \mu_B$ and variance given by equation (3).

(A2) The class-conditional distributions of X given Y follow a symmetric Gaussian discriminant model:

$$X \mid Y = 1 \;\sim\; \mathcal{N}\big(+\tfrac{\delta}{2},\, \mathrm{Var}(X)\big)\,, \quad X \mid Y = 0 \;\sim\; \mathcal{N}\big(-\tfrac{\delta}{2},\, \mathrm{Var}(X)\big)\,,$$

with equal class priors $\mathbb{P}(Y = 1) = \mathbb{P}(Y = 0) = \frac{1}{2}$.

Under (A2), Y is *not* a deterministic function of X: the outcome depends on many factors beyond the single observed KPI difference, so $H(Y \mid X) > 0$. The Gaussian discriminant model

is the natural linear classifier for normally distributed features and is consistent with (A1) when the within-class variance equals the marginal $\mathrm{Var}(X)$.

Remark. Assumption (A1) is empirically reasonable for the present data: the per-match paired difference series $X = X_{\mathrm{A}} - X_{\mathrm{B}}$ in professional rugby and football are approximately normal, consistent with the central-limit behaviour of paired differences even where the individual team series are mildly skewed (section 3.7; Supplementary section 6). Note that (A1) is required only for the information content derivation; the PEF formula itself (equation (1)) is distribution-free. Assumption (A2) is an additional approximation for the predictive setting and is likewise not required for the PEF formula.

2.2.2 Derivation

Under (A1) and (A2), predictive information is quantified as the mutual information between the relative feature X and the outcome Y. The mutual information [31, 27] between X and Y is

$$I(X;Y) = H(Y) - H(Y \mid X). \tag{6}$$

Here $I(X;Y)$ is the mutual information (in bits) between the relative feature X and the binary outcome Y; $H(Y)$ is the Shannon entropy of Y; and $H(Y \mid X) = \mathbb{E}_X[H(Y \mid X = x)]$ is the conditional entropy. For equiprobable outcomes, $H(Y) = 1$ bit. Under (A2), the posterior probability of winning given the observed KPI difference is

$$\mathbb{P}(Y = 1 \mid X = x) = \Phi\left(\frac{x}{\sqrt{\mathrm{Var}(X)}}\right),$$

where Φ is the standard normal CDF. The Bayes error rate is accordingly $\Phi\big(-\delta/(2\sqrt{\mathrm{Var}(X)})\big)$, and the expected conditional entropy evaluates to

$$H(Y \mid X) = H\left(\Phi\left(\frac{\delta}{2\sqrt{\mathrm{Var}(X)}}\right)\right), \tag{7}$$

where $H(p) = -p\log_2 p - (1-p)\log_2(1-p)$ is the binary entropy function. (Full derivation in Supplementary section 6.) The PEF formula rearranges to $\mathrm{Var}(X) = \sigma_{\mathrm{A}}^2(1+\kappa)/\eta$, linking variance geometry directly to η. Substituting into equation (7) yields

$$I(X;Y) = 1 - H\left(\Phi\left(\frac{\delta}{2\sigma_{\mathrm{A}}\sqrt{(1+\kappa)/\eta}}\right)\right). \tag{8}$$

Here δ, σ_{A}, Φ, $H(p)$, κ, and η retain their earlier definitions (section 2.2.1; equations (1) and (7)). This is the **PEF–information content relationship**. η and $I(X;Y)$ are functionally linked through the signal-to-noise ratio $\delta/\big(2\sigma_{\mathrm{A}}\sqrt{(1+\kappa)/\eta}\big)$. In plain terms, $I(X;Y)$ measures how much knowing the relative KPI value X reduces uncertainty about who wins: higher mutual information means the feature carries more decision-relevant signal about the outcome. Figure 2 visualises the information content surface $I(X;Y)$ across the (κ, ρ) parameter space for the nominal case $\delta/\sigma_{\mathrm{A}} = 1$; sensitivity of the surface to this parameter is examined in Supplementary figure S1.

2.2.3 The Efficiency–Power Tension

From equation (8), η enters the information content formula through the denominator $\sqrt{(1+\kappa)/\eta}$. Consider two effects of negative correlation ($\rho < 0$):

1. **Variance increases:** $\eta < 1$, meaning $\mathrm{Var}(X) > \mathrm{Var}_{\mathrm{unpaired}}(X)$. This harms statistical efficiency.

2. **Signal may increase relative to noise:** if the mean difference δ is large relative to the amplified standard deviation, $I(X;Y)$ can still be substantial.

Predictive power depends on $I(X;Y)$, not η alone. When δ is sufficiently large relative to the amplified standard deviation, the information content remains high despite increased variance. This accounts for cases where machine learning models benefit from relative features even when $\eta < 1$.

Statistical harm ($\eta < 1$) and predictive harm ($\Delta\text{ML} < 0$; section 3.6) are not the same test. The first means the difference has larger variance than unpaired measurement. The second means the difference is a worse classifier than the absolute KPI. The two can disagree (section 4.4).

2.3 Sources of Correlation in Paired Comparisons

Having linked η to predictive information, we turn to the origins of ρ itself. Paired measurements acquire correlation through three generic mechanisms, each of which appears across competitive and non-competitive domains:

- **Shared exogenous factors.** Conditions common to both entities move their measurements together, inducing positive correlation. Examples include market-wide movements in finance, batch, instrument, or site effects in manufacturing and clinical measurement, and weather, venue, or officiating in sport.
- **Zero-sum interaction.** When one entity's gain is the other's loss, their measurements move oppositely, inducing negative correlation. Examples include rivals competing for a fixed market share, and head-to-head contests in which possession or territory ceded by one competitor accrues directly to the other.
- **Non-zero-sum (common-mode) constraints.** When both entities can succeed or fail together, positive correlation can arise even under competition. Examples include firms lifted or depressed by a common sector trend, and matches in which both teams perform well in open, high-tempo play.

The sign and magnitude of ρ therefore reflect which mechanism dominates in a given setting. The PEF formula accommodates all correlation signs within a single expression, covering shared-factor regimes (typically $\rho > 0$) and zero-sum competitive regimes (potentially $\rho < 0$). The competitive sports KPIs analysed in section 4 provide concrete instances spanning this range, including Quadrant 4 cases in which the efficiency–power tension (section 2.2.3) is most pronounced. In the primary sports application, each observation is a same-fixture home–away pair and each match contributes one point to the bivariate sample of size n; the estimated $\hat{\rho}$ therefore summarises cross-fixture co-movement between competitors (high home counts tending to co-occur with high or low away counts in the same fixture) rather than serial dependence of a single team's KPI across time. Estimation scope and the associated independence considerations are detailed in section 3.

2.4 Quadrant Taxonomy

Each KPI occupies a position in the (κ, ρ) plane shaped by these mechanisms. Dividing that plane at $\kappa = 1$ and $\rho = 0$ yields four quadrants with distinct statistical and predictive properties (table 5). The same partition appears in figures 1 and 2, which map η and $I(X;Y)$ respectively at nominal signal strength; the four confirmatory exemplars of table 3 (one per quadrant, with season-to-season drift segments) are annotated on both figures. The regimes are:

Quadrant 1 ($\kappa > 1$, $\rho > 0$): High efficiency, high information. Positive correlation combined with variance asymmetry produces $\eta > 1$ with high information content. Relative

Table 1: Cross-domain mapping of familiar methods to typical PEF regimes.

Field	Method	Correlation source	Typical PEF regime
Statistics	Paired t-test	Repeated measures	$\eta = 1/(1-\rho)$ when $\kappa = 1$
Finance	Market-adjusted returns	Market factors	Q1: $\eta > 1$, high ρ
Healthcare	Paired clinical trials	Temporal/biological	Q1–Q2: $\eta \gg 1$
Manufacturing	Control charts	Process dynamics	Q2: $\eta > 1$, moderate ρ
Sports	Relative metrics	Environmental/competitive	Q3-dominant; Q4 when $\kappa > 1$

features outperform absolute features in this regime. Exemplified by market-adjusted returns and paired clinical trials.

Quadrant 2 ($\kappa < 1$, $\rho > 0$): High efficiency, moderate information. Positive correlation provides variance reduction even with low variance ratio, though information content is lower than Quadrant 1. Exemplified by manufacturing control charts.

Quadrant 3 ($\kappa < 1$, $\rho < 0$): Low efficiency, low information. Negative correlation with low variance asymmetry amplifies variance without commensurate information gain. Relativisation is then predictively harmful ($\Delta\text{ML} < 0$), so absolute features are preferred. Exemplified by high-volume invasion-game KPIs (passes, pressures, duels) and by weakly competitive or near-independent measurements.

Quadrant 4 ($\kappa > 1$, $\rho < 0$): Low efficiency, variable information. This is where the efficiency–power tension is most pronounced. $\eta < 1$ indicates statistical harm, yet information content may remain substantial when δ is large. The decision to relativise requires explicit analysis of both η and $I(X;Y)$. Exemplified by asymmetric head-to-head pairings ($\kappa > 1$ with $\rho < 0$), including some sports KPIs and competitive business comparisons.

2.5 Signal Strength and the Limits of η as a Predictor

Quadrant position fixes the directional regime; signal strength fixes the magnitude of predictive gain within that regime. The PEF–information content relationship (equation (8)) shows that η is not the sole determinant of predictive gain. The mutual information $I(X;Y)$ depends on η through the denominator $\sqrt{(1+\kappa)/\eta}$, but also on the signal-to-noise ratio

$$\frac{\delta}{\sigma_A} = \frac{\mu_A - \mu_B}{\sigma_A}, \tag{9}$$

the standardised mean difference between the two entities. Two KPIs with identical (κ, ρ) positions, and therefore identical η, will yield different information content if their signal strengths differ.

This has a practical consequence: η alone does not determine ML gain. η encodes the (κ, ρ) position and identifies the quadrant regime, but two KPIs in the same quadrant will yield different ML outcomes if their signal strengths differ. The appropriate diagnostic is therefore the $(\kappa, \rho, \delta/\sigma_A)$ triple: quadrant for the directional prediction, δ/σ_A estimated from data for the magnitude. Supplementary figure S3 shows how iso-η and iso-$I(X;Y)$ contours diverge once δ/σ_A varies.

The practical implication is that the four-quadrant taxonomy and the $I(X;Y)$ formula together constitute the appropriate predictive tool, with δ/σ_A estimated from data. Section 4.4 illustrates this with four quadrant exemplars spanning the empirical range of $(\kappa, \rho, \delta/\sigma_A)$ values.

2.6 Cross-Domain Connections

The quadrant logic applies uniformly across fields. Table 1 situates familiar methods on the (κ, ρ) plane.

When two domains share similar (κ, ρ) values, techniques from one may transfer to the other, for example, methods from healthcare (high-ρ settings) to manufacturing contexts with similar correlation structures.

3 Methodology

3.1 Study Design

The validation proceeds in three tiers of increasing empirical complexity. The first tier is a **theoretical/numerical foundation**: an idealised probit simulation (section 4) that generates bivariate-normal (X_A, X_B) on a $(\kappa, \rho, \delta/\sigma_A)$ grid under assumptions (A1)–(A2) and confirms the closed-form mapping from PEF to information content (equation (8)) in a controlled setting where those assumptions hold by construction (full model specification and validation in Supplementary section 6). Because the sports KPIs analysed here are well approximated by bivariate normality (normality testing, section 3.7), this idealisation provides an informative reference for the empirical data rather than a disconnected model. The second tier is the **primary empirical analysis** on professional sports KPIs. Here the full KPI inventory (86 studies: 22 rugby, 64 football) provides a *landscape characterisation* of where real metrics sit on the (κ, ρ) plane (Supplementary figures S4 and S5), whilst *mechanistic confirmation* rests on four single-KPI exemplars, one per quadrant, with comparable signal strength (figure 3, section 4.4, and table 3). The third tier is **supporting validation** on published data from four additional domains, assessing generalisability. Across these tiers the objectives were to: (1) validate the PEF–information content relationship (equation (8)), (2) confirm quadrant-specific predictions and the role of signal strength δ/σ_A as a moderator, (3) characterise the efficiency–power tension in Quadrant 4, and (4) test cross-domain applicability. Those four objectives are the operational form of the three questions in the Introduction. The estimation and design choices below serve more than one objective.

3.2 Primary Data: Sports KPIs

The primary analysis uses match-level KPI data from two professional leagues:

- **Rugby union.** United Rugby Championship (URC): $n = 283$ games pooled across seasons 23/24 and 24/25 ($m = 16$ teams, double round-robin). KPIs include rucks won, kick metres, kicks from hand, carries, penalties conceded, turnovers won, and others (Supplementary figure S4).
- **Football.** English Championship: $n = 1114$ games pooled across the same two seasons ($m = 24$ teams, double round-robin). KPIs include duels, pressures, interceptions, fouls, and shot volume (Supplementary figure S4). Association-football score and outcome modelling has a long econometric lineage [12, 32], with contemporary machine-learning approaches incorporating domain knowledge [13].

Head-to-head matchups provide natural pairing through shared environmental conditions (weather, venue, officiating), and KPI variability across matches is well documented in professional sport [33]. Those rugby union lines of work motivate treating head-to-head KPIs as a primary testbed for the present framework [7, 8, 9, 10, 11]. The sports data are well suited to testing the framework because the KPIs span a range of correlation values (including negative correlation from competitive dynamics) and variance ratios, populating all four quadrants. To assess temporal stability, we compute $(\hat{\kappa}, \hat{\rho})$ for each KPI in both seasons separately; figures 1 and 2 show the four confirmatory exemplars of table 3 with season 23/24–24/25 drift segments, whilst Supplementary figure S4 visualises the full KPI cloud.

3.3 Supporting Data: Cross-Domain Validation

To assess generalisability beyond sports, we applied the framework to four additional domains. Domain-level summaries are reported in table 4; mean (κ, ρ) positions for the supporting tier appear as overlay markers in figure 1.

Healthcare. Paired systolic (BPXOSY1) and diastolic (BPXODI1) blood pressure readings from the same oscillometric measurement event, drawn from the NHANES 2017 public-use examination file `P_BPXO` [34]. Each participant contributes one (systolic, diastolic) pair; pulse pressure PP = SBP − DBP is the relative feature, an established cardiovascular risk biomarker [3, 4]. Observations with SBP ≤ DBP were excluded as physiologically implausible (N = 10,352 retained).

Clinical Genomics. Normalised RNA-seq gene-level expression from GEO series GSE47462: patient-matched normal breast tissue vs early neoplasia [35, 36]. Pairs were defined by shared subject identifiers in the series sample annotations; per-gene PEF summaries aggregate correlation and variance-ratio structure across those paired specimens.

Finance. Daily log returns for 18 S&P 100 constituents relativised to the S&P 500 over the pinned window 2020–2023, obtained from Yahoo Finance. Market-adjusted returns provide the pairing [1, 37, 2, 38, 39].

Manufacturing. Real industrial process telemetry from the Bosch production-line CNC dataset on OpenML (dataset id 752) [40]. For each production cycle (one row), paired vectors are formed axis-wise: angular position θ_i versus angular velocity $\dot{\theta}_i$ $(i = 1, \ldots, 6)$ and torque τ_j versus displacement metric dm_j $(j = 1, \ldots, 5)$, yielding two PEF estimates per cycle. Summaries aggregate N = 16,384 such pairings (8,192 cycles × two pairing types). The NASA C-MAPSS turbofan simulation [41] was used in earlier drafts but is not the default supporting manufacturing source here because it is simulated rather than observational plant data.

3.4 Parameter Estimation

For each dataset, the variance ratio and correlation were estimated using standard sample estimators

$$\hat{\kappa} = \frac{s_{\mathrm{B}}^2}{s_{\mathrm{A}}^2}, \quad \hat{\rho} = \frac{\sum_{i=1}^{n}(X_{\mathrm{A},i} - \bar{X}_{\mathrm{A}})(X_{\mathrm{B},i} - \bar{X}_{\mathrm{B}})}{\sqrt{\sum_{i=1}^{n}(X_{\mathrm{A},i} - \bar{X}_{\mathrm{A}})^2 \sum_{i=1}^{n}(X_{\mathrm{B},i} - \bar{X}_{\mathrm{B}})^2}}. \tag{10}$$

Under bivariate normality (Assumption (A1)), these are maximum likelihood estimators. PEF was computed by substituting into equation (1). For sports KPIs, index i runs over fixtures; $X_{\mathrm{A},i}$ and $X_{\mathrm{B},i}$ denote the home and away KPI counts from game i, so $(\hat{\kappa}, \hat{\rho})$ characterise the cross-fixture distribution of those paired observations rather than correlation within a single match or serial correlation of one team's time series.

3.5 Empirical Units and Outcome Definitions

Each sports KPI constitutes one *study*: match-level home and away measurements pooled across seasons 23/24 and 24/25, yielding a single $(\hat{\kappa}, \hat{\rho}, \hat{\eta})$ estimate per KPI. The inventory comprises 86 studies (22 rugby union, 64 football). Score, scoring events, cards, expected-goals and on-ball-value models, and shots on target are omitted, because those quantities are circular with the binary match outcome. Action KPIs (passes, carries, shot volume, tackles) are retained. Per-KPI estimates, season-to-season positions, and machine-learning outcomes for this inventory are reported in Supplementary section 6 (Figures figures S4 and S5 and the accompanying table).

Two reporting roles are distinguished. **Landscape characterisation** uses the full inventory to map where invasion-game KPIs populate the (κ, ρ) plane, to summarise domain-level distributions, and to provide context for Figures 1–2. These aggregates are descriptive: the inventory is Q3-heavy (64.0% of KPIs), so pooled statistics primarily reflect the dominant negative-correlation regime rather than a balanced sample of the quadrant space. **Mechanistic confirmation** uses

the idealised probit simulation (Tier 1) together with four confirmatory exemplars, one per quadrant (figure 3 and table 3), at broadly comparable δ/σ_A (range 0.16–0.32). The lines $\kappa = 1$ and $\rho = 0$ define a coarse partition only: η and $I(X;Y)$ are nonlinear in $(\kappa, \rho, \delta/\sigma_A)$, so KPIs within a quadrant can differ in ΔML. The exemplars are fixed anchor points at distinct $(\hat{\kappa}, \hat{\rho})$ positions that test directional predictions locally, not linear representatives of whole-quadrant behaviour. One action KPI was taken from each quadrant, omitting quantities circular with the match outcome, with δ/σ_A held in a common band (0.16–0.32). The Q4 row sits on the $\kappa = 1$ boundary and is therefore a limited-signal boundary case, not a strong-κ tension demonstration. Supplementary table S1 summarises the remaining inventory by quadrant.

Where domain summaries report a **landscape efficiency rate**, this is the proportion of KPI studies with $\hat{\eta} > 1$ (pairing reduces variance relative to independent measurement). This is a descriptive summary of the (κ, ρ) inventory, not a hypothesis-test outcome; quadrant-level aggregates appear in Supplementary table S1.

3.6 Machine Learning Validation

To assess the relationship between quadrant position and ML performance, logistic regression was fitted with five-fold cross-validation [42], implemented in MATLAB (`glmfit`, Statistics and Machine Learning Toolbox). The signed change $\Delta\text{ML} = 100(\text{acc}_{\text{rel}} - \text{acc}_{\text{abs}})/\text{acc}_{\text{abs}}$ compares the relative feature $X_A - X_B$ with the absolute feature X_A only. A two-feature absolute model (X_A, X_B) would be a multivariate comparator and is outside the univariate scope of this paper. Positive ΔML means relativisation improves classification. Negative ΔML means it is predictively harmful: the difference is a worse univariate predictor than the raw KPI. Signal strength was estimated as $\hat{\delta}/\hat{\sigma}_A = (\bar{X}_A - \bar{X}_B)/s_A$ (equation (9)).

Team-blocked cross-validation. Each match is one observation, but teams recur across fixtures, so random match-level folds can place the same side in both training and test sets and inflate accuracy. We therefore partition teams (not matches) into five disjoint sets with a fixed random seed. A match is assigned to the fold of its home team; for fold f, the test set comprises all matches whose home team belongs to set f, and the training set comprises the remaining matches. No home team therefore appears in both partitions within a fold. Away opponents may appear in training when facing a held-out home side, a mild, standard form of leakage in league settings that still removes the dominant source of dependence (repeated home-side observations). The same fold assignment is used for absolute and relative features so that ΔML compares like with like. Landscape ΔML values characterise the inventory; mechanistic claims rest on the quadrant exemplars at comparable δ/σ_A (section 4.4).

3.7 Inference and Quality Control

The 86-KPI landscape is descriptive. It is not a family-wise hypothesis battery, so Bonferroni, FDR, and Holm corrections are not applied. Uncertainty for $(\hat{\kappa}, \hat{\rho}, \hat{\eta})$ uses a team-cluster bootstrap: teams are resampled with replacement and all home-side matches for each drawn team are retained [43]. Supplementary section 6 reports the Quadrant 4 check ($B = 300$). Predictive comparisons use team-blocked ΔML (section 3.6). The companion effect size is the standardised mean difference δ/σ_A (equation (9)), not Cohen's d or partial η^2.

Paired-difference series were checked with the Shapiro–Wilk test. Mean W was 0.977 in rugby and 0.986 in football, with mean skewness 0.028 and 0.069 (Supplementary section 6). Analysis code was version-controlled. The canonical run is `run_paper_pipeline.m` in MATLAB R2025b with the Statistics and Machine Learning Toolbox (section 6).

3.8 Sports Data: Independence Considerations

The sports datasets present a hierarchical structure requiring careful treatment. Each team appears in multiple games, inducing correlation across observations involving the same team through latent team ability θ_j:

$$X_{j,i} = \theta_j + \epsilon_{j,i}. \tag{11}$$

We model game-level observations using a bivariate normal approximation

$$\left(X_{\mathrm{A}(i)}, X_{\mathrm{B}(i)}\right) \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \quad i = 1, \ldots, n, \tag{12}$$

treating games as conditionally independent. Three considerations justify this approximation for PEF analysis:

1. **Target of inference.** PEF focuses on cross-fixture (same-occasion) correlation $\rho = \mathrm{Corr}(X_{\mathrm{A},i}, X_{\mathrm{B},i})$ summarised over fixtures $i = 1, \ldots, n$, capturing shared environmental factors and competitive dynamics on each game occasion rather than serial dependence of a single team across games. Team quality primarily affects mean structure rather than the correlation structure driving PEF.

2. **Variance estimation.** The total variance $\sigma_{\mathrm{A}}^2 = \mathrm{Var}(\theta_j) + \mathrm{Var}(\epsilon_{j,i})$ correctly reflects the observable variance relevant to feature engineering decisions.

3. **Robustness checks.** For $(\hat{\kappa}, \hat{\rho}, \hat{\eta})$, bootstrap confidence intervals resample *teams* with replacement and retain all home-side matches for each drawn team (cluster bootstrap on the home-team index; Supplementary section 6). For predictive validation, team-blocked cross-validation (section 3.6) prevents home teams from appearing in both training and test folds. Together, these checks address the main dependence induced by recurring teams without altering the distribution-free PEF formula itself.

This parallels established practice in sports analytics [44, 45] and panel data econometrics [46, 47].

4 Results

The results address the three questions posed in the Introduction. The controlled mapping and the efficiency–power analyses show how a higher-variance relative feature can still inform. The sports landscape and the quadrant exemplars show why relativisation helps for some KPIs and harms others. The supporting domains show whether the same (κ, ρ) geometry appears outside sport.

4.1 Theoretical Foundation: PEF–Information Content

The theoretical relationship between η and information content (equation (8)) was examined in four named scenarios from the theory-aligned idealised probit simulation (figure S2; Supplementary section 6) under assumptions (A1)–(A2), with $n = 5{,}000$ per cell, 50 Monte Carlo trials, and five-fold logistic cross-validation matching the empirical protocol. The simulation's data-generating process, estimands, and parameter-recovery checks are specified in full in Supplementary section 6.

On the admissible (κ, ρ) grid at fixed $\delta/\sigma_{\mathrm{A}} = 1$, analytic η and $I(X;Y)$ from equation (8) are strongly correlated ($r \approx 0.87$ across 16 grid cells in the idealised simulation). The four scenarios in table 2 span the quadrant space under controlled (A1)–(A2) conditions; the first two illustrate the efficiency–power tension ($\eta < 1$ with positive mean ML improvement), consistent with section 2.2.3.

Table 2: Scenario-level confirmation from the idealised probit simulation (A1)–(A2).

Scenario	κ	ρ	η	$I(X;Y)$ (bits)	ML impr.[†]
High competitive dynamics	2.0	−0.3	0.780	0.029	9.4%
Moderate competitive dynamics	1.2	−0.15	0.870	0.044	8.2%
Environmental dynamics	1.1	0.4	1.665	0.087	8.6%
Balanced (Fisher baseline)	1.0	0.0	1.000	0.056	7.9%

[†]Mean five-fold cross-validated accuracy improvement (relative minus absolute features) across 50 Monte Carlo trials per scenario cell ($\delta/\sigma_A = 1$). Standard errors are ≈0.8–1.0 percentage points (Supplementary section 6).

4.2 Sports KPI Landscape

This inventory is the empirical setting of the help-versus-harm question. The full sports inventory spans all four quadrants (Supplementary figure S4; 86 KPI studies, section 3.5). Rugby union (URC) contributes 22 KPIs and football (English Championship) 64; per-KPI $(\hat{\kappa}, \hat{\rho}, \hat{\eta})$ estimates and season-to-season drift appear in Supplementary figures S4 and S5. Domain-level landscape summaries are collected in table 4: rugby mean $\hat{\eta} = 1.342$ (95% CI: $[0.909, 1.776]$; 54.5% of KPIs with $\hat{\eta} > 1$), football mean $\hat{\eta} = 0.859$ (95% CI: $[0.809, 0.909]$; 18.8% with $\hat{\eta} > 1$). These figures characterise the empirical (κ, ρ) distribution, notably the Q3-heavy structure of invasion-game KPIs (55 of 86 studies; 64.0%). They do not serve as the primary confirmatory test.

KPIs with positive correlation (e.g., kicks from hand, carries) concentrate in Quadrants 1–2 and typically yield $\hat{\eta} > 1$; KPIs with negative or near-zero correlation (e.g., turnovers won, penalties conceded) fall predominantly in Quadrants 3–4 with $\hat{\eta} \leq 1$, yet several still show positive ML improvement in the full inventory (Supplementary figure S5), consistent with the efficiency–power tension (section 2.2.3). Mechanistic confirmation of the taxonomy's directional predictions is assessed through the four quadrant exemplars below (figure 3 and section 4.4), not through pooled success rates over the Q3-dominated inventory.

4.3 Efficiency–Power Alignment Across Quadrants

This section addresses how a less efficient relative feature can still inform. The four quadrants differ not only in mean η but in the relationship between efficiency and predictive gain. Three supplementary analyses characterise this structure.

Idealised simulation under (A1)–(A2). On the theory-aligned probit grid (Supplementary figure S2; $n = 5{,}000$, 50 trials, five-fold CV), all 16 out of 16 Quadrant 4 cells with $\eta < 1$ produced positive mean ML improvement. At fixed δ/σ_A, η and $I(X;Y)$ are strongly correlated (within-slice $r \approx 0.87$ at the empirical median $\delta/\sigma_A = 0.2920$). However, the pooled-across-δ/σ_A correlation is negative ($r = -0.70$), because signal strength shifts the $I(X;Y)$ level independently of the (κ, ρ) position. The practical consequence is that iso-η and iso-$I(X;Y)$ contours diverge once δ/σ_A varies across KPIs (Supplementary figure S3).

Quadrant distribution. Of the 86 sports KPI studies, 55 (64.0%) lie in Q3: the negative-correlation regime where competitive dynamics dominate and absolute features are typically preferred. This reflects a structural feature of professional invasion-game sport: high-volume counts (passes, pressures, duels) are anti-correlated across opponents. Because the dataset is Q3-heavy, any aggregate statistic pooled across all KPIs primarily characterises this dominant regime rather than the mechanism across the full quadrant space (Supplementary table S1). The exemplar approach (table 3) directly addresses this by selecting one KPI per quadrant, allowing the taxonomy's directional predictions to be assessed without the distributional confound.

Quadrant 4 and the Bayes bound. Team-stratified bootstrap intervals for tackles (rugby, Q4) lie entirely below $\eta = 1$ (Supplementary section 6). Under (A1)–(A2) with equal priors,

Table 3: Quadrant exemplars: four KPIs spanning the (κ, ρ) parameter space.

Q	KPI (sport)	κ	ρ	η	δ/σ_A	ΔML (%)
Q1	Kick metres (rugby)	1.06	+0.65	2.84	0.19	+4.3
Q2	Long balls (football)	0.91	+0.23	1.30	0.16	+5.5
Q3	Passes (football)	0.84	−0.65	0.61	0.32	−0.8
Q4	Goalkeeper long balls (football)	1.00	−0.24	0.81	0.26	−0.3

Two-season pooled estimates (κ, ρ, η, δ/σ_A) and team-blocked five-fold cross-validated logistic change (ΔML; section 3.6). Signal strengths ($\delta/\sigma_A \approx 0.16$–$0.32$) are broadly comparable across exemplars. The table tests sign at matched signal: Q1 and Q2, $\Delta\text{ML} > 0$ with $\eta > 1$; Q3, predictive harm ($\Delta\text{ML} < 0$) with $\eta < 1$; Q4, $\eta < 1$ with near-zero ΔML. Relative gain despite $\eta < 1$ is shown in the idealised grid and the full inventory (section 4.3 and figure S5), not by a large Q4 exemplar gain.

the Gaussian Bayes accuracy of the absolute feature is the theoretical comparator. For the Q4 confirmatory KPI, team-blocked CV accuracies for absolute and relative features are close to each other and sit below that bound. The gap is expected: real match outcomes are not generated from a single KPI by the probit link (A2). Relativisation does not lift accuracy here, matching the Q4 row of table 3. This is the limited-signal Quadrant 4 case, not the efficiency–power tension. Relative gain despite $\eta < 1$ appears in the simulation grid and in parts of the landscape (figures S2 and S5). Statistical efficiency and information content remain related but distinct: the value of relativisation cannot be read from η alone (section 2.2.3).

4.4 Quadrant Exemplars and Signal-Strength Context

The confirmatory test of help versus harm is the sign of ΔML at matched signal. Figure 3 shows the four confirmatory exemplars on the $\hat{\eta}$–ΔML plane, with rucks won as a high-η foil at negligible δ/σ_A (table 3). The full inventory remains a landscape characterisation (Supplementary figures S4 and S5).

Each ΔML in table 3 comes from *univariate* logistic regression (one KPI feature per model: X_A or $X_A - X_B$). Percentage changes are modest in absolute terms yet substantively informative. The confirmatory question is the *sign*. $\Delta\text{ML} > 0$ means the relative feature classifies match outcome more accurately than X_A alone. $\Delta\text{ML} < 0$ means relativisation is predictively harmful: the difference is a worse univariate classifier than the absolute KPI. That harm is the empirical content of the Quadrant 3 recommendation, not a failed gain. Statistical harm ($\eta < 1$) and predictive harm ($\Delta\text{ML} < 0$) can disagree. Q3 combines both. Rucks won has $\eta > 1$ with $\Delta\text{ML} < 0$. Q4 has $\eta < 1$ with near-zero ΔML (figure 3 and table 3). The present design does not construct an optimal multi-feature match predictor. Prior rugby and football work that motivates this study combined many KPIs in a single classifier, where per-feature changes can compound, cancel, or interact (section 5). The single-KPI protocol is conservative by construction. Team-blocked cross-validation (section 3.6) avoids the optimistic bias of random match folds when the same side recurs across fixtures. It also reduces leakage of a team's early and late home matches across train and test partitions; season-blocked validation is noted as a further tightening step in section 5.

The four confirmatory KPIs occupy distinct regions of the (κ, ρ) plane (figure 1 and table 3) at comparable signal strength; they are not interchangeable summaries of their quadrants. Because η and $I(X;Y)$ respond nonlinearly to (κ, ρ) and to δ/σ_A (equation (8)), each exemplar illustrates *local* regime behaviour. Interpreting a quadrant label as a uniform prescription would therefore overstate what the taxonomy claims.

The Q4 exemplar (goalkeeper long balls, $\eta = 0.81$) has $\Delta\text{ML} = -0.3\%$ at team-blocked cross-validation. The change is near zero at this δ/σ_A, as expected when pairing inflates variance

and signal is limited. This row is not a demonstration of relative benefit despite $\eta < 1$. Cases with $\eta < 1$ and $\Delta\text{ML} > 0$ appear in the full inventory (figure S5 and section 4.3). Here $\kappa \approx 1.00$, so $\eta < 1$ is driven by negative correlation ($\rho = -0.24$) rather than variance asymmetry. The KPI therefore sits on the boundary between Quadrants 3 and 4.

The Q3 exemplar (passes, football; $\rho = -0.65$, $\eta = 0.61$) is the complementary result. Strong negative correlation places this KPI deep in the low-efficiency regime, and $\Delta\text{ML} = -0.8\%$ shows that relativisation is predictively harmful. The absolute feature is the better univariate classifier, which is the Quadrant 3 recommendation in table 5. Under (A1)–(A2) the idealised probit grid cannot produce $\Delta\text{ML} < 0$ (Supplementary section 6). With a Bayes-sufficient relative feature, pairing does not reduce accuracy. Empirical harm is therefore a property of real match outcomes: the difference can discard useful absolute-level information, or it can feed the classifier a noisier feature. For high-volume counts such as passes, the difference can discard game volume or tempo.

Rucks won (rugby; Q2: $\kappa = 0.82$, $\rho = +0.82$, $\eta = 5.38$, $\delta/\sigma_A = 0.07$) shows that harm is not confined to $\eta < 1$. Despite the highest η in the dataset, $\Delta\text{ML} = -1.6\%$ because the two teams are almost equally matched on average. Pairing efficiency without signal still yields a worse predictor than the absolute KPI.

4.5 Supporting Validation: Cross-Domain Results

The third question is whether similar (κ, ρ) geometry appears beyond sport. The supporting domains are second-moment checks, not match-outcome replications. Table 4 summarises results across all six domains. They show how empirical (κ, ρ) structures propagate into $\hat{\eta}$ and landscape efficiency rates (proportion with $\hat{\eta} > 1$; section 3.5):

- **Healthcare** [34]: pooled $\eta = 2.533$ (bootstrap 95% CI: $[2.448, 2.622]$) across $N = 10{,}352$ NHANES participants pairing systolic and diastolic blood pressure from the same oscillometric measurement (Quadrant 2). Pulse pressure (SBP − DBP) is the clinically established relative feature. The one healthcare row in table 4 is that single paired study, not a participant count.

- **Clinical genomics** [36, 35]: mean $\eta = 2.696$ (SD $= 1.601$) across 3500 genes ranked by paired-differential signal in GSE47462; 97.0% of genes yielded $\eta > 1$.

- **Finance**: mean $\eta = 3.129$ (SD $= 1.108$) across 18 S&P 100 stocks relativised to daily S&P 500 returns (2020–2023 pinned snapshot); 100.0% with $\hat{\eta} > 1$.

- **Manufacturing** [40]: mean $\eta = 1.382$ (SD $= 1.401$) on Bosch OpenML 752 CNC cycles ($N = 16384$ axis-wise pairings). Approximately 50.5% of pairings yielded $\eta > 1$.

Summary statistics for all six domains are collected in table 4; the sports landscape with supporting-domain overlay markers appears in figure 1. Detailed per-study breakdowns for the non-sports domains are not separately visualised given the cross-sectional nature of those datasets.

5 Discussion

The Introduction posed three questions. Section 4 addressed each. Here we interpret the answers in the same order.

5.1 Why Relativisation Helps and Harms

Relativisation helps or harms according to a KPI's (κ, ρ) regime. At matched signal the exemplars confirm both signs (table 3). The studies that motivated this work framed the choice as help

versus no help [7, 8, 9, 10, 11]. They did not classify a regime in which the difference is the worse classifier. The Quadrant 3 result is that missing regime.

Quadrant membership remains a coarse guide. The PEF surface is nonlinear in (κ, ρ) and δ/σ_A, so two KPIs in the same quadrant need not share the same ΔML. The confirmatory table tests local agreement with the taxonomy, not a linear quadrant effect.

The analysis is univariate by design. Those same motivating studies used multi-KPI classifiers, where inter-feature correlations and joint decision boundaries lie outside univariate PEF. Reported ΔML values here are therefore conservative: they benchmark one paired feature at a time (section 3.6), and the exemplars test directional agreement rather than match-outcome accuracy.

5.2 Efficiency versus Information

Statistical efficiency (η) and predictive power ($I(X;Y)$) are functionally related through equation (8), yet they measure different things. Classical variance-reduction reasoning advises against relativisation when $\eta < 1$. That rule addresses only one dimension of the feature choice. It does not account for information content.

The PEF formula (equation (1)) requires only finite second moments. The information-content relationship requires bivariate normality (A1). The efficiency characterisation is therefore the more general result. Relative gain when $\eta < 1$ is a property of information content, not a failure of η.

5.3 Practical Guidance

The first two answers yield a prior for the feature choice:

1. Estimate κ and ρ from available data.
2. Compute η using equation (1).
3. Identify the quadrant from table 5.
4. For Quadrants 1–2, use relative features (they improve both statistical efficiency and information content).
5. For Quadrant 3, prefer absolute features: relativisation is predictively harmful.
6. For Quadrant 4, estimate δ/σ_A from the same sample and confirm the sign of ΔML under team-blocked cross-validation. The (A1)–(A2) formula for $I(X;Y)$ is a prior for whether relativisation could help, not a substitute for that check.

Possession-normalised rates are a different construction from opponent differences [48]. They adjust for game state rather than pairing two teams' counts. The signed-harm recommendation in step 5 applies to opponent pairing, not to that rate construction. An open implementation is provided (Supplementary section 6; section 6).

5.4 Cross-Domain Unification

The third question is whether the same geometry appears outside sport. Market-adjusted returns, paired clinical designs, control charts, and relative sports metrics are instances of correlation-based variance reduction with different (κ, ρ) values. A further analogue is value-added benchmarking in education, which compares teacher performance with that of comparable cohorts [49].

Healthcare pairing sits in Quadrant 2 (table 4), the high-ρ regime in which the taxonomy recommends relative features. Invasion-game KPIs occupy all four quadrants and are Q3-heavy (section 4). Relativisation helps for some sports measurements and is harmful for others because (κ, ρ) differs by KPI, not because sport is a single regime.

5.5 Limitations

Distributional assumptions. The PEF formula is distribution-free, but the information content relationship (equation (8)) assumes bivariate normality. Paired differences are often closer to normal than the raw series (Supplementary section 6). Strongly skewed or multimodal distributions may still need a non-parametric mapping. For count-based KPIs, $Z = \log(X + 1)$ can improve normality whilst preserving the (κ, ρ) structure. Across 22 rugby KPIs the rank ordering by ρ was largely preserved (Spearman $r_s = 0.967$), and the aggregate PEF change was not significant ($\bar{\eta}_{\log}/\bar{\eta} = 1.085$, $p = 0.357$; Supplementary section 6). Individual count-based KPIs can still move substantially (rucks won).

Stationarity. The framework treats κ and ρ as fixed. In the sports inventory, most KPIs drift modestly between seasons 23/24 and 24/25, though a minority of rugby KPIs cross quadrant boundaries (figures 1 and 2; Supplementary figure S4). Pooled two-season estimates (as in table 3) therefore summarise a moving target.

Binary outcomes. Both equation (8) and the sports protocol target binary classification: which team wins a head-to-head match. That choice matches the probit link under (A1)–(A2) and keeps the KPI comparison on a common scale. Multi-class outcomes and continuous targets would need a corresponding information-theoretic mapping. The per-KPI (κ, ρ) diagnostic still applies wherever paired absolute and relative features can be formed.

Strategic interactions. In invasion-game sport, teams respond to opponents' tactics and game state. The (κ, ρ) structure of a KPI is therefore partly an equilibrium property of competition, not a fixed physical law. Parameters here are estimated from observed match data without a model of within-game adaptation. Where such dynamics are material, (κ, ρ) should be read as descriptive summaries at the chosen time scale.

Primary evidence base. The full KPI inventory, machine-learning validation, and quadrant exemplars rest on rugby union and football. Healthcare, clinical genomics, finance, and manufacturing show that the same (κ, ρ) geometry recurs (table 4), but without the same per-unit ML scrutiny or temporal replication.

5.6 Future Directions

A priority is a distribution-free analogue of $I(X; Y)$, for example a rank-based or permutation measure. Quadrant 4 decisions would then not rely on (A1)–(A2) when skewness or heavy tails are material (Supplementary section 6).

Multivariate PEF-style summaries would show how per-KPI effects combine when several paired features enter one classifier. A complementary test is within-quadrant landscape design: stratify high and low κ and high and low $|\rho|$ inside each quadrant, with $\delta/\sigma_{\mathrm{A}}$ held comparable. That design would ask whether ΔML tracks local iso-η or iso-$I(X; Y)$ structure rather than quadrant labels alone.

Rolling-window or state-space estimates of (κ, ρ) would track parameter drift. Season-blocked or chronological cross-validation would tighten temporal leakage beyond the team-blocked folds used here (section 3.6).

Relativisation is often justified as baseline cancellation rather than variance reduction. Connecting PEF geometry to causal estimands would show when differencing preserves a treatment or exposure contrast and when it introduces collider or selection bias.

Deeper per-unit machine-learning checks in social-science panels, environmental paired measurements, and engineering telemetry would test whether the taxonomy transfers before the practical guidance is used outside invasion-game sport.

6 Conclusion

The paper asked why relativisation helps for some KPIs and harms others, how a less efficient relative feature can still inform, and whether the same geometry appears beyond sport.

The Paired Efficiency Factor $\eta = (1+\kappa)/(1+\kappa-2\sqrt{\kappa}\,\rho)$ extends Fisher's paired efficiency to unequal variances. The formula is distribution-free. Under bivariate normality, η and information content are related but distinct, which is why a relative feature can improve prediction when $\eta < 1$.

The sports inventory occupies all four (κ, ρ) quadrants and is Q3-heavy. Four exemplars at comparable signal strength confirm both help and harm (table 3). The same geometry appears, with different mean $\hat{\eta}$, in healthcare, clinical genomics, finance, and manufacturing (table 4). Read with δ/σ_A, the taxonomy is a prior for when relative features should be used and when they should not.

Data and Code Availability

Sports KPI data were analysed under agreement with the United Rugby Championship and Swansea City AFC and are not redistributed. Supporting-domain provenance is described in Methods (section 3).

All PEF estimates and ΔML values were produced by `run_paper_pipeline.m` in MATLAB R2025b with the Statistics and Machine Learning Toolbox. Team-blocked fold construction is described in Methods (section 3.6). Interactive SI2 and SI3 tools (analytic probit surface and practitioner diagnostic) are available at `https://github.com/AroundInteger/pef-tools`. Input schemas and run instructions are given in the Supplementary Information and the tools repository README.

References

[1] W. F. Sharpe. 'Mutual Fund Performance'. In: *The Journal of Business* 39.1 (1966), pp. 119–138. DOI: 10.1086/294846.

[2] E. F. Fama. 'Efficient Capital Markets: A Review of Theory and Empirical Work'. In: *The Journal of Finance* 25.2 (1970), pp. 383–417. DOI: 10.1111/j.1540-6261.1970.tb00518.x.

[3] S. S. Franklin, S. A. Khan, N. D. Wong, M. G. Larson and D. Levy. 'Is Pulse Pressure Useful in Predicting Risk for Coronary Heart Disease? The Framingham Heart Study'. In: *Circulation* 100.4 (1999), pp. 354–360. DOI: 10.1161/01.CIR.100.4.354.

[4] J. Blacher, A. P. Guerin, B. Pannier, S. J. Marchais, M. E. Safar and G. M. London. 'Impact of Aortic Stiffness on Survival in End-Stage Renal Disease'. In: *Circulation* 99.18 (1999), pp. 2434–2439. DOI: 10.1161/01.CIR.99.18.2434.

[5] D. C. Montgomery. *Introduction to Statistical Quality Control*. 7th ed. John Wiley & Sons, 2012. ISBN: 978-1-118-14681-1.

[6] T. L. Keiningham, A. Buoye and J. Ball. 'Competitive Context Is Everything: Moving from Absolute to Relative Metrics'. In: *Global Economics and Management Review* 20.1 (2015), pp. 18–25. DOI: 10.1016/j.gemrev.2015.07.001.

[7] M. Bennett, N. Bezodis, D. A. Shearer, D. Locke and L. P. Kilduff. 'Descriptive Conversion of Performance Indicators in Rugby Union'. In: *Journal of Science and Medicine in Sport* 22.3 (2019), pp. 330–334. DOI: 10.1016/j.jsams.2018.08.008.

[8] M. Bennett, N. E. Bezodis, D. A. Shearer and L. P. Kilduff. 'Predicting Performance at the Group-Phase and Knockout-Phase of the 2015 Rugby World Cup'. In: *European Journal of Sport Science* 21.3 (2021), pp. 312–320. DOI: 10.1080/17461391.2020.1743764.

Table 4: Validation results across domains.

Domain	**Tier**	n **studies**	**Mean** η	**SD**	**% with** $\eta > 1$	**Data source**
Sports (Rugby)	Primary	22	1.342	1.037	54.5%	URC
Sports (Football)	Primary	64	0.859	0.204	18.8%	English Championship
Healthcare	Supporting	1[†]	2.533	0.04[†]	100.0%	[34]
Clinical Genomics	Supporting	3500[‡]	2.696	1.601	97.0%	[36, 35]
Finance	Supporting	18	3.129	1.108	100.0%	Yahoo Finance
Manufacturing	Supporting	16384[§]	1.382	1.401	50.5%	[40]

Note: Landscape efficiency rate: proportion of units with $\hat{\eta} > 1$ (section 3.5). The n studies column is the number of PEF units, not a participant count: healthcare is one paired study of $N = 10{,}352$ participants. [†]Paired NHANES oscillometric records (systolic vs diastolic, first reading); SD column gives a bootstrap standard error for the pooled η (not a cross-unit sample SD). [‡]Genes in the ranked subset of the GSE47462 paired design (see Methods). [§]Axis-wise paired process features per Bosch production cycle (OpenML 752); SD is the sample SD of η across pairings. Primary empirical tier: landscape characterisation in Figures 1–2 and Supplementary figures S4 and S5; mechanistic confirmation via figure 3 and table 3. Supporting tier: parameter estimation and PEF calculation summarised here and in section 4; mean (κ, ρ) overlay in figure 1.

Table 5: Quadrant taxonomy and practical guidance.

Q	**Parameters**	η	$I(X;Y)$	**Recommendation**
Q1	$\kappa > 1$, $\rho > 0$	> 1	High	Use relative features
Q2	$\kappa < 1$, $\rho > 0$	> 1	Moderate	Use relative features
Q3	$\kappa < 1$, $\rho < 0$	< 1	Low	Prefer absolute features
Q4	$\kappa > 1$, $\rho < 0$	< 1	Variable	Analyse both η and $I(X;Y)$

Examples: Q1, market-adjusted returns, paired clinical trials; Q2, control charts, quality metrics; Q3, high-volume anti-correlated sports KPIs; Q4, asymmetric head-to-head comparisons (some sports KPIs; competitive business).

Table 6: Boundary behaviour (illustrative).

Condition	η	$I(X;Y)$	**Interpretation**
$\kappa = 1$, $\rho = 0$ (Fisher)	1.000	0.056	Baseline: independent, equal variances
$\kappa \to 0$	1.000	0.108	Entity B negligible
$\kappa \to \infty$	1.000	$\to 0$	Signal drowned in noise
$\rho \to -1$ $(\kappa = 1)$	0.500	0.028	Maximum variance amplification
$\rho \to +1$ $(\kappa = 1)$	$\to \infty$	$\to 1$	Perfect cancellation of noise

Note: $I(X;Y)$ computed with $\delta/\sigma_A = 1$.

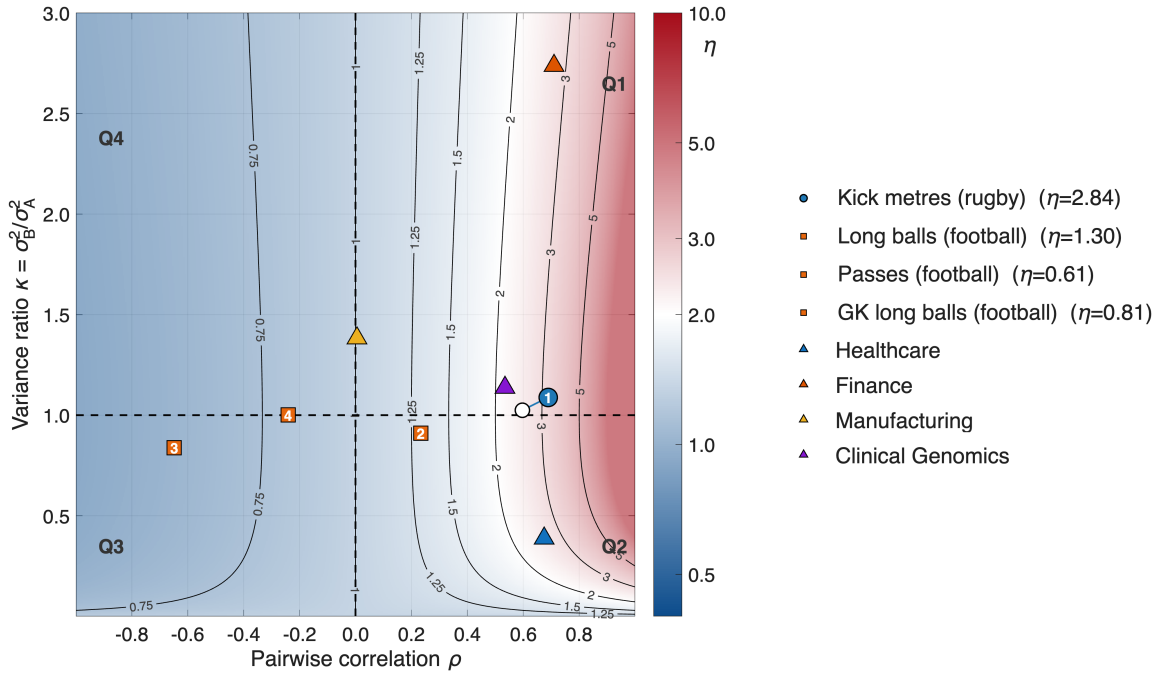


Figure 1: PEF landscape: η as a function of ρ and κ, with four quadrants labelled. Background shading uses a $\log_{10}$ colour mapping of η (tick labels show η) so that the admissibility blow-up as $\rho \to +1$ (Supplementary section 6) remains visible; contour lines are in raw η. The dotted curve marks the admissibility boundary $1 + \kappa - 2\sqrt{\kappa}\,\rho = 0$, beyond which η is undefined. **Sports exemplars** (rugby union URC, circle; football Championship, squares): the four confirmatory KPIs of table 3, one per quadrant, with open markers at season 23/24, filled markers at season 24/25, and segments showing year-on-year drift (Q1–Q4 labels on filled markers; pooled two-season $\hat{\eta}$ in the legend). **Supporting domains** (triangles): mean (κ, ρ) for healthcare, clinical genomics, finance, and manufacturing. The full action-KPI cloud is in Supplementary figure S4.

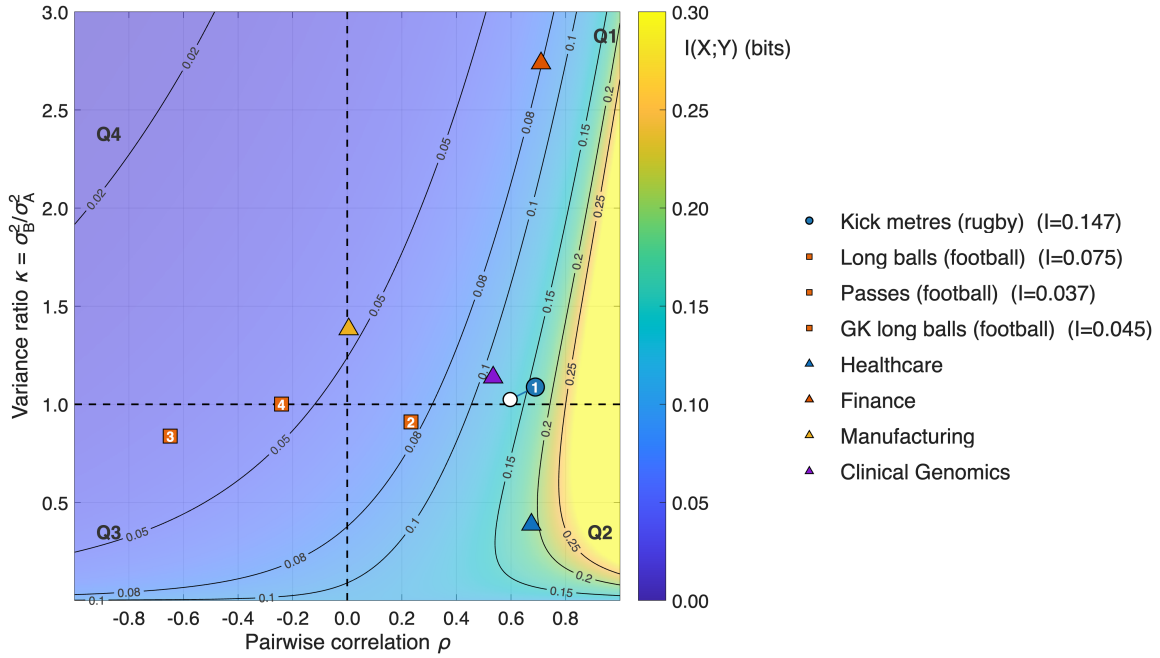


Figure 2: Information-theoretic surface $I(X;Y)$ as a function of ρ and κ (equation (8)), computed with nominal signal strength $\delta/\sigma_A = 1$. Layout matches figure 1: the same four confirmatory exemplars (table 3) are plotted at their estimated (κ, ρ) positions on this $\delta/\sigma_A = 1$ surface, with season 23/24–24/25 drift segments, supporting-domain triangles, quadrant boundaries, and the admissibility boundary (dotted curve). Background shading uses a *sequential* colour scale over $I(X;Y) \in [0, 0.3]$ bits (low to high information); figure 1 uses a *diverging* $\log_{10}$ scale of η centred on the efficiency boundary $\eta = 1$. Sensitivity to $\delta/\sigma_A \in \{0.5, 1.0, 2.0\}$ appears in Supplementary figure S1.

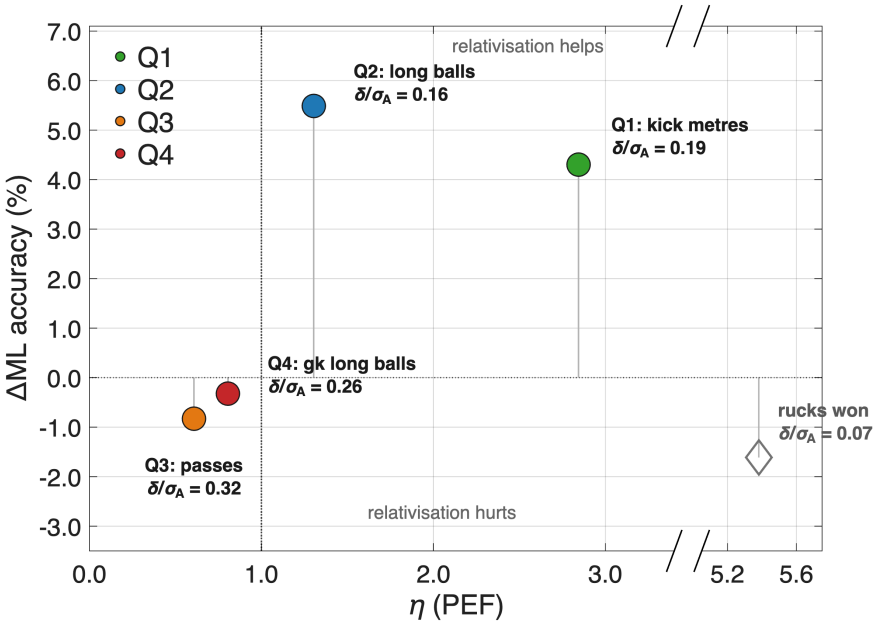


Figure 3: Confirmatory machine-learning check: ΔML versus $\hat{\eta}$ for four quadrant exemplars at comparable δ/σ_A (table 3). $\Delta\text{ML} = 100(\text{acc}_{\text{rel}} - \text{acc}_{\text{abs}})/\text{acc}_{\text{abs}}$ from team-blocked five-fold CV logistic regression (section 3.6). The zero line is the decision: above it, use the relative feature; below it, keep the absolute feature. Direction at matched signal: Q1 and Q2, $\eta > 1$ with $\Delta\text{ML} > 0$; Q3, $\eta < 1$ with $\Delta\text{ML} < 0$; Q4, $\eta < 1$ with near-zero ΔML. Rucks won (diamond) is a high-η foil with negligible δ/σ_A: pairing efficiency without signal still yields $\Delta\text{ML} < 0$. The x-axis is broken between 3.4 and 5.05 to keep the foil on scale. Colours match figures 1 and 2. The full inventory, including cases with $\eta < 1$ and $\Delta\text{ML} > 0$, is in Supplementary figures S4 and S5.

[9] T. J. Mosey and L. J. G. Mitchell. 'Key Performance Indicators in Australian Sub-Elite Rugby Union'. In: *Journal of Science and Medicine in Sport* 23.1 (2020), pp. 35–40. DOI: 10.1016/j.jsams.2019.08.014.

[10] G. A. Scott, N. Bezodis, M. Waldron, M. Bennett, S. Church, L. P. Kilduff and M. R. Brown. 'Performance Indicators Associated with Match Outcome Within the United Rugby Championship'. In: *Journal of Science and Medicine in Sport* 26.1 (2023), pp. 63–68. DOI: 10.1016/j.jsams.2022.09.006.

[11] G. A. Scott, O. Edwards, N. E. Bezodis, M. Waldron, E. Roberts, D. B. Pyne, J. Mara, C. Cook, L. Mason, M. R. Brown and L. P. Kilduff. 'Classifying Winning Performances in International Women's Rugby Union'. In: *International Journal of Sports Physiology and Performance* 18.9 (2023), pp. 1072–1078. DOI: 10.1123/ijspp.2023-0086.

[12] M. J. Dixon and S. G. Coles. 'Modelling Association Football Scores and Inefficiencies in the Football Betting Market'. In: *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 46.2 (1997), pp. 265–280. DOI: 10.1111/1467-9876.00065.

[13] D. Berrar, P. Lopes and W. Dubitzky. 'Incorporating Domain Knowledge in Machine Learning for Soccer Outcome Prediction'. In: *Machine Learning* 108.1 (2019), pp. 97–126. DOI: 10.1007/s10994-018-5747-8.

[14] R. Stefani and R. Pollard. 'The Methodology of Officially Recognized International Sports Rating Systems'. In: *Journal of Quantitative Analysis in Sports* 7.1 (2011). DOI: 10.2202/1559-0410.1347.

[15] C. M. E. Colomer, D. B. Pyne, M. Mooney, A. McKune and B. G. Serpell. 'Performance Analysis in Rugby Union: A Critical Systematic Review'. In: *Sports Medicine – Open* 6.1 (2020), p. 4. DOI: 10.1186/s40798-019-0232-x.

[16] N. Watson, I. Durbach, S. Hendricks and T. Stewart. 'On the Validity of Team Performance Indicators in Rugby Union'. In: *International Journal of Performance Analysis in Sport* 17.4 (2017), pp. 609–621. DOI: 10.1080/24748668.2017.1376998.

[17] Â. Brito. *Absolute and Relative Performance Indicators Associated with Match Success Across Six European Professional Football Leagues*. Research Square preprint. Posted 5 August 2026; not peer reviewed. 2026. DOI: 10.21203/rs.3.rs-10445439/v1.

[18] R. S. Brill, R. Yurko and A. J. Wyner. 'Exploring the Difficulty of Estimating Win Probability: A Simulation Study'. In: *Journal of Quantitative Analysis in Sports* (2025). DOI: 10.1515/jqas-2024-0130.

[19] R. A. Fisher. *The Design of Experiments*. Edinburgh: Oliver and Boyd, 1935.

[20] Student. 'The Probable Error of a Mean'. In: *Biometrika* 6.1 (1908), pp. 1–25. DOI: 10.2307/2331554.

[21] B. L. Welch. 'The Generalization of 'Student's' Problem When Several Different Population Variances Are Involved'. In: *Biometrika* 34.1–2 (1947), pp. 28–35. DOI: 10.1093/biomet/34.1-2.28.

[22] E. L. Lehmann. *Elements of Large-Sample Theory*. New York: Springer, 1999. DOI: 10.1007/b98855.

[23] M. D. Hughes and R. M. Bartlett. 'The Use of Performance Indicators in Performance Analysis'. In: *Journal of Sports Sciences* 20.10 (2002), pp. 739–754. DOI: 10.1080/026404102320675602.

[24] L. Bornn, D. Cervone and J. Fernández. 'Spatiotemporal Analysis of Team Sports'. In: *Annals of Applied Statistics* 15.2 (2021), pp. 773–794. DOI: 10.1214/20-AOAS1363.

[25] A. M. Franks, A. D'Amour, D. Cervone and L. Bornn. 'Meta-Analytics: Tools for Understanding the Statistical Properties of Sports Metrics'. In: *Journal of Quantitative Analysis in Sports* 12.4 (2016), pp. 151–165. DOI: 10.1515/jqas-2016-0098.

[26] D. Matteo, M. C. Varley, J. D. Ruddy and D. L. Carey. 'Understanding the Statistical Properties of Performance Indicators in Australian Football'. In: *Journal of Sports Sciences* 44.14 (2026). DOI: 10.1080/02640414.2026.2703355.

[27] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. 2nd ed. Hoboken, NJ: John Wiley & Sons, 2006. ISBN: 978-0-471-24195-9.

[28] I. Guyon and A. Elisseeff. 'An Introduction to Variable and Feature Selection'. In: *Journal of Machine Learning Research* 3 (2003), pp. 1157–1182. URL: https://www.jmlr.org/papers/v3/guyon03a.html.

[29] E. J. G. Pitman. 'A Note on Normal Correlation'. In: *Biometrika* 31.1/2 (1939). Pitman–Morgan test for equal paired margins, pp. 9–12.

[30] W. A. Morgan. 'A Test for the Significance of the Difference between the Two Variances of a Pair of Variables'. In: *Biometrika* 31.1/2 (1939). Pitman–Morgan test for equal paired margins, pp. 13–19.

[31] C. E. Shannon. 'A Mathematical Theory of Communication'. In: *The Bell System Technical Journal* 27.3 (1948), pp. 379–423. DOI: 10.1002/j.1538-7305.1948.tb01338.x.

[32] B. L. Boulier and H. O. Stekler. 'Predicting the Outcomes of National Football League Games'. In: *International Journal of Forecasting* 19.2 (2003), pp. 257–270. DOI: 10.1016/S0169-2070(01)00144-3.

[33] R. M. Malcata and W. G. Hopkins. 'Variability of Competitive Performance of Elite Athletes: A Systematic Review'. In: *Sports Medicine* 44.12 (2014), pp. 1763–1774. DOI: 10.1007/s40279-014-0239-x.

[34] National Center for Health Statistics. *NHANES 2017 Examination Data: Blood Pressure, Oscillometric Measurement (P_BPXO)*. Centers for Disease Control and Prevention, U.S. Department of Health and Human Services. Public-use SAS transport file `P_BPXO.xpt`. 2021. URL: https://wwwn.cdc.gov/nchs/nhanes/.

[35] NCBI Gene Expression Omnibus. *GEO Series GSE47462*. NCBI, National Library of Medicine. Paired normal vs early neoplasia expression; normalized RefSeq matrix in series supplementary files. 2014. URL: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE47462.

[36] A. L. Brunner, J. Li, X. Guo, R. T. Sweeney, S. Varma, S. X. Zhu, R. Li, R. Tibshirani and R. B. West. 'A Shared Transcriptional Program in Early Breast Neoplasias Despite Genetic and Clinical Distinctions'. In: *Genome Biology* 15.5 (2014). PubMed: 24887547, R71. DOI: 10.1186/gb-2014-15-5-r71.

[37] W. F. Sharpe. 'The Sharpe Ratio'. In: *The Journal of Portfolio Management* 21.1 (1994), pp. 49–58. DOI: 10.3905/jpm.1994.409501.

[38] M. M. Carhart. 'On Persistence in Mutual Fund Performance'. In: *The Journal of Finance* 52.1 (1997), pp. 57–82. DOI: 10.1111/j.1540-6261.1997.tb03808.x.

[39] E. F. Fama and K. R. French. 'Common Risk Factors in the Returns on Stocks and Bonds'. In: *Journal of Financial Economics* 33.1 (1993), pp. 3–56. DOI: 10.1016/0304-405X(93)90023-5.

[40] OpenML. *Dataset 752: Bosch Production Line Performance*. OpenML. Real CNC process telemetry; accessed via `sklearn.datasets.fetch_openml`. 2016. URL: https://www.openml.org/d/752.

[41] A. Saxena and K. Goebel. *Turbofan Engine Degradation Simulation Data Set*. Tech. rep. Publicly available via the NASA Prognostics Center of Excellence Data Set Repository. Moffett Field, CA: NASA Ames Prognostics Center of Excellence, NASA Ames Research Center, 2008. URL: https://www.nasa.gov/intelligent-systems-division/discovery-and-systems-health/pcoe/pcoe-data-set-repository/.

[42] R. Kohavi. 'A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection'. In: *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence (IJCAI)*. Vol. 14. Morgan Kaufmann, 1995, pp. 1137–1143.

[43] B. Efron. 'Bootstrap Methods: Another Look at the Jackknife'. In: *The Annals of Statistics* 7.1 (1979), pp. 1–26. DOI: 10.1214/aos/1176344552.

[44] E. Ben-Naim, F. Vazquez and S. Redner. 'Parity and Predictability of Competitions'. In: *Journal of Quantitative Analysis in Sports* 2.4 (2006). DOI: 10.2202/1559-0410.1034.

[45] P. D. Owen. 'Measurement of Competitive Balance and Uncertainty of Outcome'. In: *Handbook on the Economics of Professional Football*. Ed. by J. Goddard and P. Sloane. Edward Elgar, 2014, pp. 41–59.

[46] A. C. Cameron and D. L. Miller. 'A Practitioner's Guide to Cluster-Robust Inference'. In: *Journal of Human Resources* 50.2 (2015), pp. 317–372. DOI: 10.3368/jhr.50.2.317.

[47] J. M. Wooldridge. *Econometric Analysis of Cross Section and Panel Data*. 2nd ed. MIT Press, 2010. ISBN: 978-0-262-23258-6.

[48] O. Sever, Ç. Öncü, Y. Öztaşyonar, Ö. Aksoy and S. Bayrakdaroğlu. 'Match Success in Football: Insights from Raw and Possession-Normalised Key Performance Indicators in the Turkish Professional League'. In: *BMC Sports Science, Medicine and Rehabilitation* 18 (2026), p. 340. DOI: 10.1186/s13102-026-01768-9.

[49] E. A. Hanushek and S. G. Rivkin. 'Generalizations About Using Value-Added Measures of Teacher Quality'. In: *American Economic Review* 100.2 (2010), pp. 267–271. DOI: 10.1257/aer.100.2.267.

Supplementary Information

This supplement follows the main paper: theoretical derivations, then the idealised probit check, then the sports KPI analysis. Figures are numbered in order of appearance (S1–S5). File names on disk retain their generator tags and need not match the printed S-number.

SI block	**Main paper**	**Contents**
§1 Theory	Theory; Fig. 2; Intro. (Pitman)	Algebra; $\tau = \frac{1}{2}\log\kappa$; Pitman ARE; Fig. S1
§2 Probit validation	Methods Tier 1; Results	Spec.; Figs S2–S3
§3 Empirical analysis	Methods; Results; Discussion	Figs S4–S5; Table S1; QC

SI Section 1: Theoretical Derivations

This section supports the theoretical framework (section 2). The main text states the closed form (equation (8)) and the four-quadrant partition. This section records the omitted algebra and extends main-text Figure 2 (figure 2).
The algebraic derivation of the PEF formula (equation (1)) from the variance of correlated differences, together with its reduction to Fisher's classical case at $\kappa = 1$, is given in full in section 2 (§2.1). This section records a log-ratio remapping of κ, the omitted information-content algebra, and the Pitman asymptotic relative efficiency proof.

Log-ratio coordinate for the variance ratio

The variance ratio $\kappa = \sigma_B^2/\sigma_A^2$ is multiplicative. An additive remapping that places equal variances at the origin is the log standard-deviation ratio

$$\tau \;=\; \log \frac{\sigma_B}{\sigma_A} \;=\; \tfrac{1}{2}\log \kappa. \tag{13}$$

The map $\kappa \mapsto \tau$ is strictly increasing on $(0, \infty)$, and $\tau = 0$ if and only if $\kappa = 1$. Interchanging the labels of A and B sends $\kappa \mapsto 1/\kappa$ and therefore $\tau \mapsto -\tau$. The four quadrants of the (κ, ρ) plane then occupy the four open quadrants of the (ρ, τ) plane. Figure S2A uses this coordinate so that the design points sit symmetrically about the Fisher line. All arguments in the main text remain in (κ, ρ).

Information Content Derivation

Setup Under Assumption (A1), $X = X_A - X_B \sim \mathcal{N}\big(\delta, \sigma_A^2(1 + \kappa - 2\sqrt{\kappa}\,\rho)\big)$ with $\delta = \mu_A - \mu_B$.

The binary outcome $Y \in \{0, 1\}$ is the match outcome; it is *not* a deterministic function of X. Under Assumption (A2) (Gaussian discriminant model, section 2.2.1), the class-conditional distributions are

$$X \mid Y = 1 \sim \mathcal{N}(+\delta/2,\, \mathrm{Var}(X)), \quad X \mid Y = 0 \sim \mathcal{N}(-\delta/2,\, \mathrm{Var}(X)),$$

with equal priors. Mutual information:

$$I(X; Y) = H(Y) - H(Y \mid X).$$

Unconditional Entropy For equiprobable outcomes ($\mathbb{P}(Y = 1) = \mathbb{P}(Y = 0) = 0.5$),

$$H(Y) = 1 \text{ bit}.$$

Conditional Entropy Under (A2), the posterior is $\mathbb{P}(Y = 1 \mid X = x) = \Phi(x/\sqrt{\mathrm{Var}(X)})$, so the expected conditional entropy is

$$H(Y \mid X) = \mathbb{E}_X\big[H\big(\mathbb{P}(Y = 1 \mid X)\big)\big].$$

Evaluating this expectation under the marginal distribution of X (which has mean $\delta/2$ under the equal-prior mixture) yields the closed-form approximation

$$H(Y \mid X) \approx H\left(\Phi\left(\frac{\delta}{2\sqrt{\mathrm{Var}(X)}}\right)\right),$$

where $H(p) = -p\log_2 p - (1-p)\log_2(1-p)$ is the binary entropy function and the right-hand side equals the binary entropy of the Bayes error rate $\Phi(-\delta/(2\sqrt{\mathrm{Var}(X)}))$.

Substitution From equation (1): $1+\kappa-2\sqrt{\kappa}\,\rho = (1+\kappa)/\eta$. Therefore $\mathrm{Var}(X) = \sigma_{\mathrm{A}}^2(1+\kappa)/\eta$, giving

$$I(X;Y) = 1 - H\left(\Phi\left(\frac{\delta}{2\sigma_{\mathrm{A}}\sqrt{(1+\kappa)/\eta}}\right)\right).$$

Connection to Pitman Asymptotic Relative Efficiency

This section establishes that, under bivariate normality and equal group sizes, the PEF equals the Pitman asymptotic relative efficiency (ARE) of the paired t-test relative to the independent two-sample t-test. We state this as a proposition, give a self-contained proof, verify the classical reduction, and record the qualifications that bound the result.

Proposition **Proposition (PEF as Pitman ARE).** *Under Assumption (A1) and equal allocation (n observations per group), the Pitman ARE of the paired t-test relative to the independent two-sample t-test equals*

$$\mathrm{ARE}_{\mathrm{paired/indep}} = \frac{1+\kappa}{1+\kappa-2\sqrt{\kappa}\,\rho} = \eta. \tag{14}$$

Proof Let $N = 2n$ be the total number of observations available. Fix $\mu_{\mathrm{A}} - \mu_{\mathrm{B}} = \delta > 0$ and let $n \to \infty$. We compare the noncentrality parameters of the two tests at the same total sample size.

Step 1: Paired test. With n matched pairs, differences $D_i = X_{\mathrm{A},i} - X_{\mathrm{B},i} \sim \mathcal{N}(\delta, \sigma_{\mathrm{A}}^2(1+\kappa-2\sqrt{\kappa}\,\rho))$ are i.i.d. under (A1). The noncentrality parameter of $T_p = \bar{D}/\sqrt{s_D^2/n}$ is

$$\lambda_p(n) = \frac{\delta\sqrt{n}}{\sigma_{\mathrm{A}}\sqrt{1+\kappa-2\sqrt{\kappa}\,\rho}}. \tag{15}$$

Step 2: Independent two-sample test. With n independent observations per group (total $N = 2n$, equal allocation), the noncentrality parameter of the Welch t-test $T_u = (\bar{X}_{\mathrm{A}} - \bar{X}_{\mathrm{B}})/\sqrt{s_{\mathrm{A}}^2/n + s_{\mathrm{B}}^2/n}$ is

$$\lambda_u(n) = \frac{\delta\sqrt{n}}{\sigma_{\mathrm{A}}\sqrt{1+\kappa}}. \tag{16}$$

Step 3: ARE. Both tests use the same total sample $N = 2n$. The Pitman ARE is the limiting ratio of sample sizes required for equal power at fixed significance and effect size, which equals the ratio of squared noncentrality parameters per observation [22]:

$$\mathrm{ARE}_{\mathrm{paired/indep}} = \lim_{n\to\infty}\frac{n_u^*}{n_p^*} = \frac{\lambda_p^2(n)/n}{\lambda_u^2(n)/n} = \frac{\delta^2/\left(\sigma_{\mathrm{A}}^2(1+\kappa-2\sqrt{\kappa}\,\rho)\right)}{\delta^2/\left(\sigma_{\mathrm{A}}^2(1+\kappa)\right)} = \frac{1+\kappa}{1+\kappa-2\sqrt{\kappa}\,\rho} = \eta \quad \square. \tag{17}$$

Classical Verification Setting $\kappa = 1$:

$$\mathrm{ARE}_{\mathrm{paired/indep}} = \frac{2}{2-2\rho} = \frac{1}{1-\rho},$$

recovering Fisher's classical paired-design efficiency [19]. The PEF is therefore the direct generalisation of Fisher's result to unequal variances, expressed in the language of Pitman efficiency.

Qualifications Three conditions bound equation (14):

1. **Normality.** Steps 1–2 use the normal noncentrality parameter. Under non-normality, the central limit theorem ensures the limiting noncentrality is still given by equations (15) and (16) under finite second moments, so the ARE result is asymptotically robust; however, the small-sample comparison may differ.

2. **Equal group allocation.** Step 2 assumes $n_A = n_B = n$. Under Neyman-optimal allocation $n_A/n_B = \sigma_A/\sigma_B = 1/\sqrt{\kappa}$, the noncentrality of the independent test becomes $\delta\sqrt{N}/(2\sigma_A\sqrt{\kappa}/(1+\sqrt{\kappa})\cdot\sqrt{1+\kappa})$, which modifies the ARE. Equal allocation is the natural comparison when paired and unpaired designs collect data in the same proportions.

3. **Fixed alternative.** The result above uses a fixed δ and $n \to \infty$. Under local (Pitman) contiguous alternatives $\delta_n = c/\sqrt{n}$, the noncentrality parameters remain $O(1)$ and the same ratio η is obtained in the limit, so the result is unchanged [22].

The PEF formula is distribution-free (equation (1)); only the ARE *interpretation* in equation (14) requires normality. The distribution-free efficiency characterisation (the variance ratio) is the primary result; the Pitman ARE connection provides a classical testing-theory corroboration.

Information surface

Figure S1 examines sensitivity of the $I(X;Y)$ surface to δ/σ_A on a controlled grid; empirical KPI positions appear in section 6.

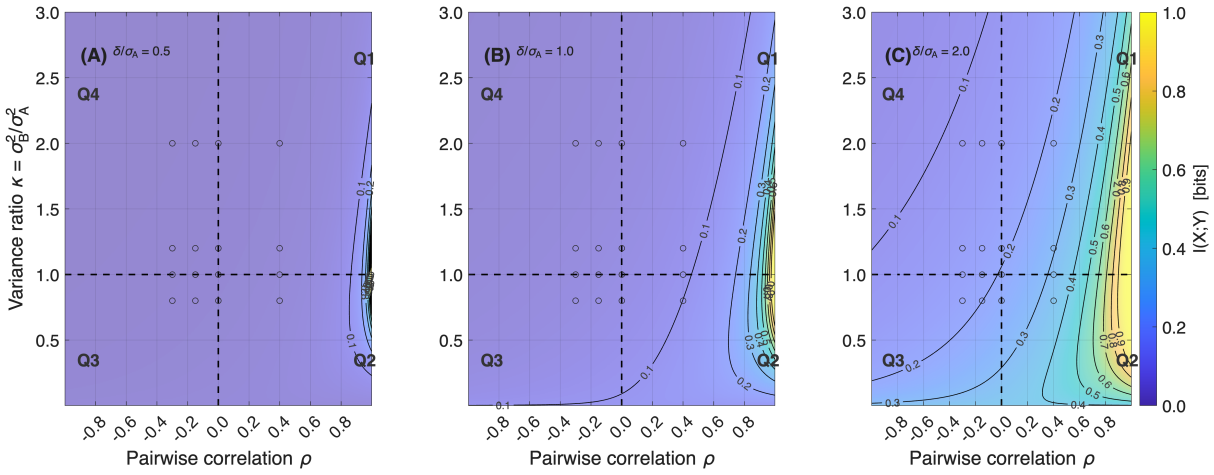


Figure S1: Sensitivity of the information-content surface $I(X;Y)$ to δ/σ_A (panels at 0.5, 1.0, 2.0; centre panel matches figure 2). Markers are idealised factorial (κ, ρ) design points; empirical KPI maps are in figure S4.

SI Section 2: Idealised Probit Validation

This section supports the first validation tier (section 3): an idealised probit simulation under (A1)–(A2) that confirms the closed-form PEF–information mapping (equation (8)) and the efficiency–power tension (section 2.2.3) before turning to empirical KPIs. Main-text scenario results appear in section 4 and table 2; the specification below is the full model; figures S2 and S3 visualise grid-level and contour-level behaviour cited in sections 2.5 and 4.

Idealised Probit Simulation: Model Specification and Validation

Purpose. The first validation tier (section 3) isolates the PEF–information content relationship (equation (8)) from the finite-sample noise, mild non-normality, and team-level clustering present in the empirical KPIs. Data are generated to satisfy (A1)–(A2) exactly. The simulation then checks two claims. First, the closed-form mapping is recovered by five-fold logistic regression, the same classifier family as the empirical protocol. Pairs are independent by construction, so the simulation does not use team-blocked folds. Second, the efficiency–power tension (section 2.2.3) arises as predicted. This is a check of internal consistency under the model's own assumptions, not an independent empirical test. The empirical tiers (section 4) provide the latter.

Data-generating process (A1). For each grid cell $(\kappa, \rho, \delta/\sigma_A)$ we draw n independent pairs from the bivariate normal

$$\begin{pmatrix} X_A \\ X_B \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \delta/2 \\ -\delta/2 \end{pmatrix}, \sigma_A^2 \begin{pmatrix} 1 & \rho\sqrt{\kappa} \\ \rho\sqrt{\kappa} & \kappa \end{pmatrix} \right), \tag{18}$$

with $\sigma_A = 1$. This parameterisation fixes $\mathrm{Var}(X_A) = \sigma_A^2$, $\mathrm{Var}(X_B) = \kappa\sigma_A^2$, and $\mathrm{Corr}(X_A, X_B) = \rho$, so the variance ratio and correlation match their targets by construction. The relative feature $X = X_A - X_B$ then has mean δ and variance $\mathrm{Var}(X) = \sigma_A^2(1+\kappa-2\sqrt{\kappa}\,\rho)$, exactly the denominator of equation (1).

Outcome model (A2). Binary outcomes are drawn from the probit link consistent with equation (7):

$$Y_i \sim \mathrm{Bernoulli}\left(\Phi\left(\frac{X_i}{\sqrt{\mathrm{Var}(X)}} \right) \right), \tag{19}$$

which makes X the Bayes-sufficient statistic for Y and yields analytic information content $I(X;Y)$ as in equation (8) and Bayes error $\Phi\big(-\delta/(2\sqrt{\mathrm{Var}(X)})\big)$.

Grid and Monte Carlo design. The factorial grid is $\kappa \in \{0.8, 1.0, 1.2, 2.0\}$, $\rho \in \{-0.3, -0.15, 0, 0.4\}$, and $\delta/\sigma_A \in \{0.3, 0.5, 1.0, 2.0\}$ (64 admissible cells; the admissibility constraint $1 + \kappa - 2\sqrt{\kappa}\,\rho > 0$ holds throughout). Each cell uses $n = 5{,}000$ pairs and is repeated over 50 Monte Carlo trials with deterministic per-cell seeds (base seed 20,260,520). The four named scenarios of table 2 are simulated identically.

Estimands. For each cell we record (a) the analytic quantities η, $\mathrm{Var}(X)$, $I(X;Y)$, and the Bayes accuracy bound, computed directly from $(\kappa, \rho, \delta, \sigma_A)$; and (b) Monte Carlo estimates from the machine-learning protocol used throughout the paper, namely five-fold cross-validated logistic-regression accuracy for absolute features (X_A only), relative features ($X = X_A - X_B$), and the full pair (X_A, X_B). The headline machine-learning improvement is the relative-minus-absolute accuracy difference, matching section 3.

Validation 2 (PEF–information mapping). At fixed signal strength, analytic η and $I(X;Y)$ are strongly positively correlated ($r \approx 0.87$ within the δ/σ_A slice nearest the empirical median; section 4). Pooled across δ/σ_A the correlation reverses sign ($r \approx -0.70$), because signal strength shifts $I(X;Y)$ independently of the (κ, ρ) position: the iso-η versus iso-I divergence shown in figures S2 and S3. This confirms the manuscript's central caveat that η alone does not determine predictive gain.

Validation 3 (efficiency–power tension). All 16 Quadrant 4 grid cells with $\eta < 1$ produced positive mean machine-learning improvement (figure S2), reproducing the predicted tension (section 2.2.3) in a setting free of empirical confounds.

Validation 4 (a master surface and its envelope). Under (A1)–(A2) the relativisation gain is a function of two coordinates, with closed-form bounds. The outcome depends on the pair only through X (equation (19)). The relative feature is therefore Bayes-sufficient, and its optimal rule is the sign of X. Write the standardised relative signal

$$d_{\mathrm{rel}} \;=\; \frac{|\delta|}{\sqrt{\mathrm{Var}(X)}} \;=\; \frac{|\delta|}{\sigma_{\mathrm{A}}}\sqrt{\frac{\eta}{1+\kappa}}, \qquad u \equiv \frac{X}{\sqrt{\mathrm{Var}(X)}} \sim \mathcal{N}(d_{\mathrm{rel}}, 1). \tag{20}$$

The Bayes accuracy of the relative feature depends on d_{rel} alone,

$$\mathrm{acc}_R(d_{\mathrm{rel}}) \;=\; \mathbb{E}\big[\Phi(|u|)\big] \;=\; g(d_{\mathrm{rel}}), \qquad g(d) \equiv \mathbb{E}_{u\sim\mathcal{N}(0,1)}\Phi(|d+u|). \tag{21}$$

The absolute feature $A = X_{\mathrm{A}}$ is jointly Gaussian with X. Their correlation is

$$r \;\equiv\; \mathrm{Corr}(A, X) \;=\; \frac{1-\rho\sqrt{\kappa}}{\sqrt{1+\kappa-2\sqrt{\kappa}\,\rho}} \;=\; (1-\rho\sqrt{\kappa})\sqrt{\frac{\eta}{1+\kappa}}. \tag{22}$$

Set $w = (A - \delta/2)/\sigma_{\mathrm{A}} \sim \mathcal{N}(0,1)$. The posterior is $\mathbb{E}[Y \mid A] = \Phi\big((d_{\mathrm{rel}} + rw)/\sqrt{2-r^2}\big)$, which is monotone in A. The Bayes rule therefore thresholds A, and

$$\mathrm{acc}_A(d_{\mathrm{rel}}, r) \;=\; \mathbb{E}_w\left[\Phi\left(\frac{|d_{\mathrm{rel}} + rw|}{\sqrt{2-r^2}}\right)\right]. \tag{23}$$

The relativisation gain is the two-parameter surface

$$\Delta\mathrm{ML}(d_{\mathrm{rel}}, r) \;=\; 100\,\frac{\mathrm{acc}_R(d_{\mathrm{rel}}) - \mathrm{acc}_A(d_{\mathrm{rel}}, r)}{\mathrm{acc}_A(d_{\mathrm{rel}}, r)}. \tag{24}$$

Across the 64 grid cells this analytic surface reproduces the simulated ΔML almost exactly ($r = 0.9998$; equation (24) versus five-fold logistic CV). The machine-learning pipeline therefore attains the Bayes-optimal gain. No third coordinate is needed. The accuracy acc_A increases with r from the majority-class baseline to acc_R. At each d_{rel} the surface is bracketed by two closed-form envelopes:

$$\underbrace{0}_{\text{floor } (r\to 1)} \;\le\; \Delta\mathrm{ML} \;\le\; \underbrace{100\left(\frac{g(d_{\mathrm{rel}})}{\Phi(d_{\mathrm{rel}}/\sqrt{2})} - 1\right)}_{\text{ceiling } (r\to 0)}. \tag{25}$$

The floor is exact. With X Bayes-sufficient, relativisation cannot reduce accuracy in this idealised setting, so $\Delta\mathrm{ML} \ge 0$. The gain vanishes as $r \to 1$, where A coincides with X, and under signal saturation, where both accuracies approach one. The ceiling is attained as $r \to 0$, where A carries no outcome information and collapses to the marginal rate $\Phi(d_{\mathrm{rel}}/\sqrt{2}) = \Pr(Y = 1)$. From equation (20), large η raises d_{rel} rather than lowering it. The percentage gain is largest when d_{rel} is modest, so accuracies have not saturated, and r is small, so A is poorly aligned with X. Figure S2C shows the design points between these curves, coloured by r. The grid samples $r \in [0.32, 0.83]$, so the visible upper edge is the $r = 0.32$ contour, not the $r \to 0$ ceiling.

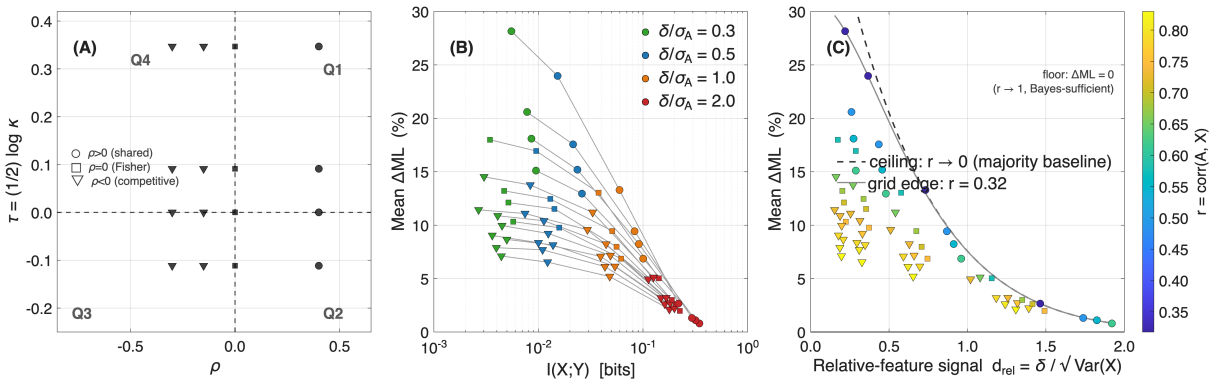


Figure S2: Idealised probit simulation (SI §2) under (A1)–(A2). Grid: $\kappa \in \{0.8, 1.0, 1.2, 2.0\}$, $\rho \in \{-0.3, -0.15, 0, 0.4\}$, $\delta/\sigma_{\mathrm{A}} \in \{0.3, 0.5, 1.0, 2.0\}$ (64 cells; $n = 5{,}000$, 50 trials, five-fold logistic CV). (A) The 16 (κ, ρ) design points plotted against $\tau = \frac{1}{2}\log\kappa$ (equation (13)), so that $\kappa = 1$ is the horizontal axis. This paper's arguments use (κ, ρ). Dashed lines mark $\rho = 0$ and $\tau = 0$ (equivalently $\kappa = 1$). Marker shape encodes sign(ρ) (circle $\rho > 0$, square $\rho = 0$, triangle $\rho < 0$). (B) $I(X;Y)$ versus mean ΔML (logarithmic x-axis). Colour is signal strength. Grey trajectories join the same (κ, ρ) across the four $\delta/\sigma_{\mathrm{A}}$ values. The colour bands are signal strata, not PEF regimes. (C) Analytic gain surface $\Delta\mathrm{ML} = F(d_{\mathrm{rel}}, r)$ from SI §2 ($r = 0.9998$ versus simulated CV). Points are coloured by $r = \mathrm{Corr}(X_{\mathrm{A}}, X)$. The dashed curve is the $r \to 0$ ceiling; the grey curve is the grid edge at $r = 0.32$.

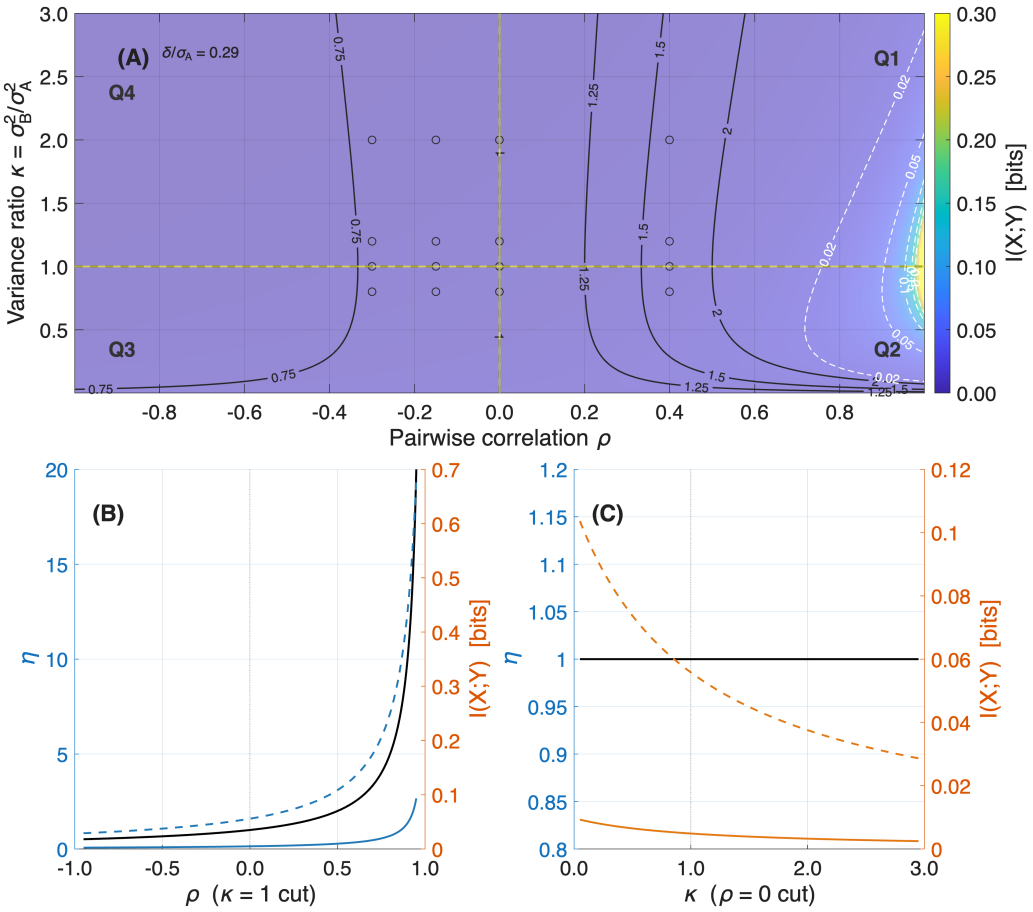


Figure S3: Idealised probit simulation (SI §2): iso-η vs. iso-$I(X;Y)$. (A) Surface at median empirical $\delta/\sigma_A = 0.28$ (solid: η; dashed: I); markers are factorial design points; yellow lines mark the cuts in (B)–(C). (B) Cut at $\kappa = 1$: $\eta(\rho)$ is signal-invariant while $I(\rho)$ rises from the empirical-median signal (solid) to the main-text Figure 2 nominal $\delta/\sigma_A = 1$ (dashed). (C) Cut at $\rho = 0$: $\eta = 1$ for all κ; again only I moves with signal strength.

SI Section 3: Empirical Analysis

This section supports landscape characterisation (sections 3.5 and 4): the full 86-KPI inventory on the (κ, ρ) plane and quadrant aggregates. Main-text Figures 1–2 map exemplars on the PEF and information surfaces; Figure 3 is the confirmatory η–ΔML check. This section gives per-KPI detail (figures S4 and S5), quadrant-level aggregates (Table S1), and quality-control checks. Throughout, $I_{\text{pred}}(X;Y)$ denotes the plug-in of equation (8) at the estimated $(\hat{\kappa}, \hat{\rho}, \hat{\delta}, \hat{\sigma}_{\text{A}})$. It is not a fitted mutual-information estimator, and it is not a global predictor of ΔML.

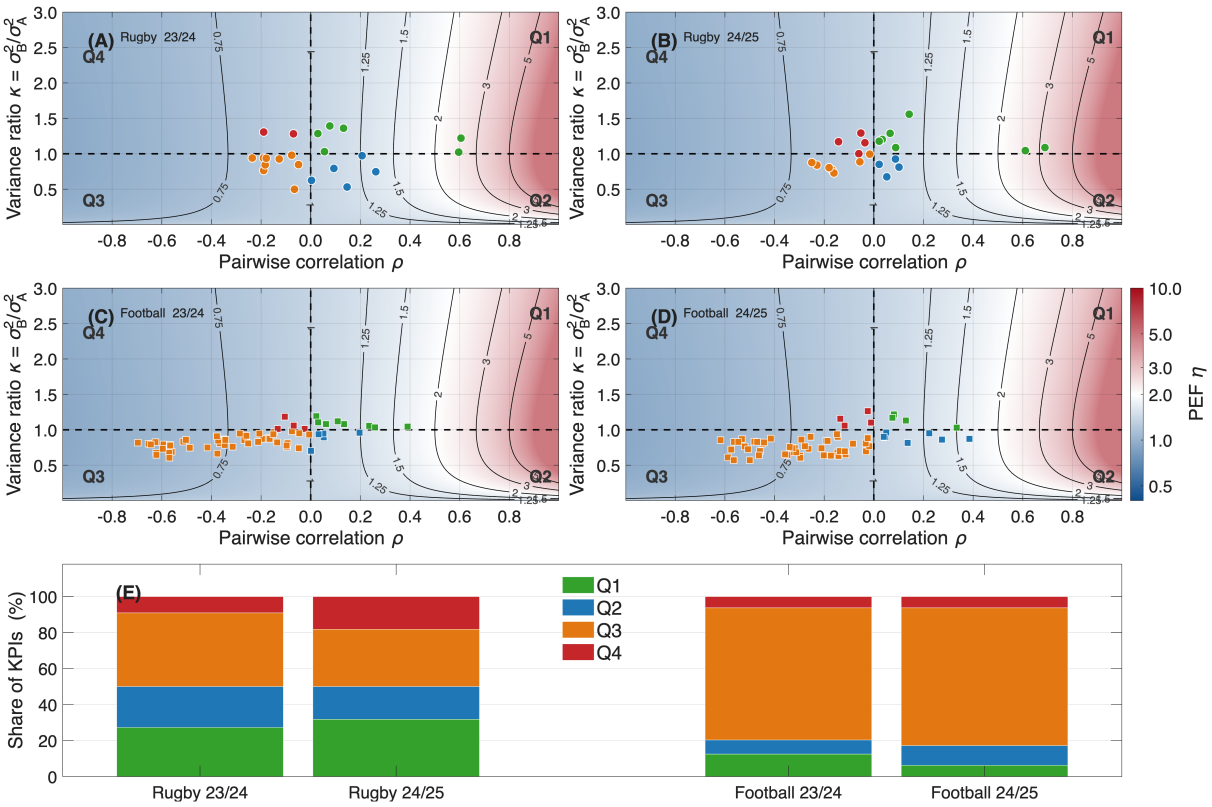


Figure S4: Season-specific KPI positions and quadrant occupancy (rugby URC; football Championship; seasons 23/24–24/25). (A, B) rugby; (C, D) football. Markers are coloured by quadrant (Q1 green, Q2 blue, Q3 orange, Q4 red). (E) share of KPIs in each quadrant, by season. Year-on-year identity is not traced: arrows overstate small moves across $\kappa = 1$ or $\rho = 0$. A minority of rugby KPIs cross a quadrant boundary between seasons. Football remains Q3-heavy. Main-text figures 1 and 2 show four confirmatory exemplars only (table 3).

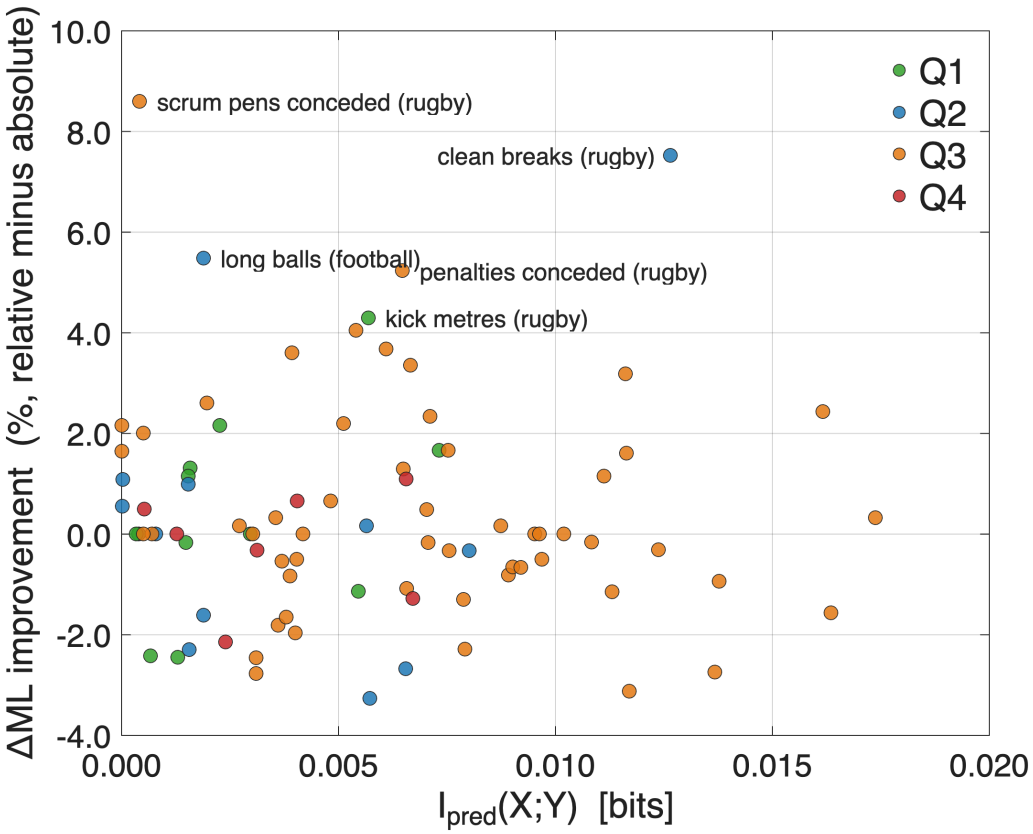


Figure S5: Plug-in $I_{\text{pred}}(X;Y)$ versus team-blocked five-fold ΔML across $N = 86$ sports KPIs (pooled seasons 23/24–24/25). Landscape characterisation only (section 3.5). The inventory is Q3-heavy. Linear correlation is weak ($r = -0.046$). Labelled points are the five largest $|\Delta\text{ML}|$ values among action KPIs. This panel does not support a global I or η mapping onto machine-learning gain.

Table S1: Descriptive quadrant-level statistics across the sports KPI inventory (seasons 23/24–24/25 pooled).

Q	n	**Mean** $\hat{\eta}$	**95% CI**	**%** $\hat{\eta} > 1$	**Mean** ΔML
Q1	12	1.346	$[0.986, 1.706]$	100	0.4%
Q2	12	1.547	$[0.857, 2.237]$	100	0.5%
Q3	55	0.788	$[0.759, 0.818]$	0	0.4%
Q4	7	0.913	$[0.867, 0.959]$	0	-0.2%

Mean ΔML: team-blocked five-fold cross-validated accuracy improvement (relative minus absolute features), pooled across KPIs in the quadrant (section 3.6). 95% CIs are ± 1.96 standard errors of the within-quadrant mean $\hat{\eta}$. Landscape characterisation only; mechanistic confirmation uses table 3. It is not a recommendation to relativise generic Q3 KPIs.

Quality control

This subsection supports normality and transformation checks in Methods (section 3.7) and distributional limitations in Discussion.

Paired-difference normality

The information-content mapping (equation (8)) uses assumption (A1) on the pair (X_A, X_B). The relevant check for that approximation is the per-match paired difference $X = X_A - X_B$, not the separate home or away series. On the two-season primary window, mean Shapiro–Wilk W for the difference series is 0.977 in rugby and 0.986 in football, with mean skewness 0.028 and 0.069 respectively (`normality_commentary.csv`). Those figures are the source of the 0.98 and ≈ 0.03–0.07 ranges quoted in the Introduction and in the Theory remark. Home and away series are more right-skewed, as expected for count KPIs. The PEF formula itself does not require normality.

Log-transform and Quadrant 4 bootstrap

A log-transform $Z = \log(X+1)$ on the 22 rugby action KPIs preserves rank order by ρ (Spearman $r_s = 0.967$; mean $|\rho_{\log} - \rho| = 0.032$). The mean PEF ratio is $\eta_{\log}/\eta = 1.085$ (median 1.002; paired t, $p = 0.357$). That mean is pulled by rucks won. The transform is a sensitivity check for right-skewed counts, not a default step.

Team-stratified bootstrap 95% intervals ($B = 300$; home-team resampling) for tackles (rugby, Q4) are $[0.773, 0.960]$ on $\hat{\eta}$, entirely below one. The confirmatory Q4 point estimate remains $\hat{\eta} = 0.81$ for goalkeeper long balls (table 3).

The open SI2 and SI3 tools for the six-step procedure in section 5.3 are documented at section 6.