# Literature foundations for Advuman's rule-based risk index framework

Advuman's three composite indices — RPI, LSI, and CPI — can be built on a robust body of established methodology spanning composite indicator science, statistical process control, financial early warning systems, and evidence fusion theory. **This review identifies 28 high-quality sources** across eight thematic clusters, each providing specific mathematical formulas, calibration methods, or production-system design patterns directly applicable to a transparent, rule-based trade lane intelligence service. The most critical finding is that Advuman's architecture does not require novel invention: the OECD composite indicator pipeline, the KLR signal approach from crisis economics, CUSUM/EWMA from statistical process control, and Dempster-Shafer evidence combination from information fusion collectively provide a complete, proven, interpretable mathematical toolkit. What distinguishes Advuman is the specific combination of these methods applied to UK-India trade lane monitoring for SMEs.

---

## 1. Composite index construction: the mathematical backbone

Seven sources provide the core methodology for building composite indices from heterogeneous signals.

**OECD/JRC Handbook on Constructing Composite Indicators: Methodology and User Guide** Michela Nardo, Michaela Saisana, Andrea Saltelli, Stefano Tarantola, Anders Hoffmann, Enrico Giovannini (2008). OECD Publishing. DOI: 10.1787/9789264043466-en. Full PDF: https://www.oecd.org/content/dam/oecd/en/publications/reports/2008/08/handbook-on-constructing-composite-indicators-methodology-and-user-guide_g1gh9301/9789264043466-en.pdf

This is the foundational reference for Advuman's entire index architecture. OECD It covers the complete **10-step pipeline**: theoretical framework → data selection → imputation → multivariate analysis → normalization → weighting → aggregation → robustness testing → decomposition → visualization. OECD It provides the core normalization formulas — **min-max** ($\tilde{x} = (x - x_{min})/(x_{max} - x_{min})$), **z-score** ($\tilde{x} = (x - \mu)/\sigma$), and distance-to-reference — alongside detailed comparison of aggregation methods: additive ($CI = \Sigma\, w_i \cdot \tilde{x}_i$), geometric ($CI = \Pi\, \tilde{x}_i{}^\wedge w_i$), and non-compensatory multi-criteria approaches. Its Table 3 maps compatibility between weighting methods (equal, PCA, AHP, budget allocation) and aggregation functions, OECD making it the single most important methodological checklist for Advuman.

**On the Methodological Framework of Composite Indices: A Review of the Issues of Weighting, Aggregation, and Robustness** Salvatore Greco, Alessio Ishizaka, Menelaos Tasiou, Gianpiero Torrisi (2019). Social Indicators Research, Vol. 141, pp. 61–94. DOI: 10.1007/s11205-017-1832-9

The most comprehensive modern review of composite index methodology. Springer It categorizes weighting methods into statistical (PCA, Data Envelopment Analysis), participatory (AHP, budget allocation), and hybrid approaches. Its critical contribution is formalizing how **the meaning of weights changes with aggregation method**: in additive aggregation, weights are trade-off rates (substitution rates between dimensions), while in non-compensatory aggregation they function as importance coefficients. Knowledge for policy This distinction directly determines whether Advuman should use additive or geometric aggregation when combining sub-indicators within RPI, LSI, and CPI.

**Constructing Composite Indicators with Shannon Entropy: The Case of Human Development Index** Grigoris Karagiannis, Stella Karagiannis (2020). Socio-Economic Planning Sciences, Vol. 70. DOI: 10.1016/j.seps.2019.03.006

Proposes an objective, data-driven weighting scheme based on Shannon entropy: indicators with higher dispersion receive higher weight because they carry more discriminating information. ScienceDirect The formula — $\mathbf{w_j = (1 - E_j) / \Sigma(1 - E_k)}$, where $E_j = -(1/\ln m)\ \Sigma\ p_{ij} \cdot \ln(p_{ij})$ — provides a transparent alternative to fixed expert weights for Advuman's sub-indicators. It is computationally lightweight, fully interpretable, and auto-calibrates when new data sources are added.

**Aggregating Composite Indicators through the Geometric Mean: A Penalization Approach** Matteo Ferretti, Mauro Ferrante (2022). Computation (MDPI), Vol. 10, No. 4, Article 64. DOI: 10.3390/computation10040064

Demonstrates that with indicators (21,1,1,1) versus (6,6,6,6), arithmetic aggregation gives equal scores (6.0) while geometric gives **2.14 vs. 6.00** Knowledge for policy — a large difference reflecting imbalance across dimensions. europa For Advuman's risk monitoring, where a single extreme dimension (e.g., severe regulatory pressure) should not be masked by normalcy in other dimensions, this paper validates using geometric or penalized geometric aggregation when combining RPI, LSI, and CPI into an overall risk state.

**JRC Competence Centre on Composite Indicators: 10-Step Guide and Toolkit** European Commission Joint Research Centre (2020). https://knowledge4policy.ec.europa.eu/composite-indicators/toolkit_en

The online companion toolkit to the OECD Handbook, providing worked numerical examples contrasting additive, geometric, and Condorcet aggregation with identical input data. europa It introduces the **COINr R package** for composite indicator development, implementing multiple normalization methods, weighting schemes, and aggregation functions Bluefoxr — serving as both a methodological checklist and software reference for Advuman's implementation pipeline.

**The GSCPI: A New Barometer of Global Supply Chain Pressures** Gianluca Benigno, Julian di Giovanni, Jan J.J. Groen, Adam I. Noble (2022). Federal Reserve Bank of New York Staff Report No. 1017.
https://www.newyorkfed.org/medialibrary/media/research/staff_reports/sr1017.pdf

The closest practical analogue to Advuman's index design. The GSCPI combines **27 heterogeneous variables** (Baltic Dry Index, Harpex container rates, BLS airfreight indices, PMI sub-components from 7 economies) TRADING ECONOMICS via a two-step process: first, demand-purging regressions isolate supply-side signals; then PCA extracts a single composite index expressed as **standard deviations from historical mean**. newyorkfed Liberty Street Economics This z-score-based output creates an intuitive scale (0 = normal, +2σ = elevated, +4σ = extreme) directly applicable to Advuman's LSI and CPI construction and state-transition thresholds.

**Connecting to Compete 2023: Trade Logistics in the Global Economy (World Bank LPI)** Jean-François Arvis, Lauri Ojala, Ben Shepherd et al. (2023). World Bank.
https://lpi.worldbank.org/about

The LPI scores logistics performance across six dimensions (customs efficiency, infrastructure, shipment ease, services quality, tracking, timeliness) using PCA-based aggregation of structured surveys from 652 freight forwarders. The 2023 edition added high-frequency container movement data. Wikipedia India's component-level scores provide baseline calibration data for Advuman's UK-India LSI, and the PCA methodology validates combining heterogeneous logistics signals into a single index.

---

## 2. Early warning systems and threshold calibration for state transitions

Four sources from financial crisis economics provide the theoretical and practical foundation for Advuman's Stable → Watch → Active threshold design.

**Leading Indicators of Currency Crises** Graciela Kaminsky, Saul Lizondo, Carmen M. Reinhart (1998). IMF Staff Papers, Vol. 45, No. 1, pp. 1–48. DOI: 10.2307/3867328

The foundational paper for threshold-based early warning. The KLR "signals approach" issues a warning when an indicator exceeds a percentile-based threshold, ScienceDirect optimized by minimizing the **noise-to-signal ratio** = (false alarms / total tranquil periods) ÷ (correct signals / total crisis periods). This framework maps directly to Advuman's state transitions: threshold calibration is fundamentally a tradeoff between Type I errors (missed risks) and Type II errors (false alarms), resolved via grid search over percentile cutoffs of the indicator's historical distribution.

**Assessing Early Warning Systems: How Have They Worked in Practice?** Andrew Berg, Eduardo Borensztein, Catherine Pattillo (2004). IMF Working Paper WP/04/52. https://www.imf.org/external/pubs/ft/wp/2004/wp0452.pdf

The most comprehensive real-time evaluation of multiple EWS models (KLR, DCSD logit, Goldman Sachs GS-WATCH, Credit Suisse EMRI). It formalizes the cutoff threshold selection problem using a **loss function**: $L(\theta) = \omega_1 \cdot P(\text{false alarm} \mid \theta) + \omega_2 \cdot P(\text{missed crisis} \mid \theta)$, where weights reflect the user's risk aversion. The finding that in-sample optimal thresholds transfer reasonably well out-of-sample suggests Advuman can set useful initial thresholds even with short baselines, then recalibrate as data accumulates.

**Towards a New Early Warning System of Financial Crises** Matthieu Bussière, Marcel Fratzscher (2006). Journal of International Money and Finance, Vol. 25, No. 6, pp. 953–973. DOI: 10.1016/j.jimonfin.2006.07.007. ECB Working Paper: https://www.ecb.europa.eu/pub/pdf/scpwps/ecbwp145.pdf

Exceptionally relevant because it introduces a **multinomial logit EWS with three distinct states** — tranquil, pre-crisis, and post-crisis/recovery — directly paralleling Advuman's Stable/Watch/Active framework. The authors prove that three-state classification outperforms binary models by eliminating "post-crisis bias," and derive optimal threshold placement between states as a function of an explicit loss function reflecting the relative costs of false alarms versus missed events. europa

**Financial Cycles — Early Warning Indicators of Banking Crises?** Sally Chen, Katsiaryna Svirydzenka (2021). IMF Working Paper WP/21/116. https://www.imf.org/-/media/files/publications/wp/2021/english/wpiea2021116-print-pdf.pdf

Uses **deviation-from-trend gap measures** derived from financial cycle variables — directly analogous to Advuman's deviation-from-baseline approach. Thresholds are calibrated by comparing CDFs of crisis versus non-crisis distributions and finding where the distance is maximized. International Monetary Fund The paper also constructs an **Overheating Index** that aggregates multiple threshold-breach signals into a composite risk state, closely paralleling how Advuman could aggregate simultaneous RPI, LSI, and CPI breaches.

---

## 3. Statistical process control adapted for risk monitoring

Three sources bridge industrial quality control methods with financial and risk monitoring, providing the CUSUM, EWMA, and Shewhart formulas Advuman needs for baseline deviation detection.

**Sequential Detection of US Business Cycle Turning Points: Performances of Shiryayev-Roberts, CUSUM and EWMA Procedures** Bakhodir A. Ergashev (2004). Econometrics Working Paper 0402001, University Library of Munich. https://ideas.repec.org/p/wpa/wuwpem/0402001.html

Applies CUSUM, EWMA, and Shiryayev-Roberts procedures to detect US business cycle turning points using leading economic indicators. The key finding for Advuman: **CUSUM performs best with volatile signals** (sharp regulatory shifts in RPI), while **EWMA's exponential weighting better suits smoother, trend-driven signals** (logistics stress in LSI, cost pressure in CPI). This suggests Advuman should match detection method to signal characteristics.

Core formulas:

- **CUSUM**: $S_t = \max(0, S_{t-1} + (x_t - \mu_0 - k))$, signal when $S_t > h$
- **EWMA**: $Z_t = \lambda \cdot x_t + (1-\lambda) \cdot Z_{t-1}$, signal when $|Z_t - \mu_0| > L \cdot \sigma_z$

**Introduction to Statistical Quality Control (8th Edition)** Douglas C. Montgomery (2019). Wiley. ISBN: 978-1-119-39930-8

The definitive SPC reference covering Shewhart charts (3σ control limits, in-control $ARL_0 \approx 370$), CUSUM (tabular formulation with parameters K and H), and EWMA (steady-state variance: $Var(Z_t) = \sigma^2 \cdot [\lambda/(2-\lambda)] \cdot [1-(1-\lambda)^{2t}]$). Critically, it covers **Phase I** (estimating in-control parameters from historical baseline) versus **Phase II** (online monitoring), mapping to Advuman's baseline establishment followed by real-time state transitions. The textbook explicitly addresses non-manufacturing applications of SPC. Google Books

**A Novel Application of Statistical Process Control Charts in Financial Market Surveillance** Kaveh Bastani, Mehdi Malakooti, Fatemeh Ghaseminejad (2023). PLOS ONE, 18(7), e0288627. DOI: 10.1371/journal.pone.0288627

Demonstrates that SPC control charts (Hotelling $T^2$, EWMA, CUSUM) work effectively in financial surveillance when applied to **model residuals rather than raw values** — addressing autocorrelation, non-normality, and time-varying volatility. For Advuman, this means control charts should monitor deviations from baseline trends (e.g., ARMA-fitted residuals) rather than absolute index levels, reducing false signals from predictable trend effects.

## 4. Signal fusion, confidence weighting, and evidence combination

Four sources address how to combine heterogeneous signals with different confidence levels — the core challenge of fusing news reports, government data, port statistics, and FX movements into unified indices.

**Combination of Evidence in Dempster-Shafer Theory** Kari Sentz, Scott Ferson (2002). Sandia National Laboratories, SAND2002-0835. https://www.stat.berkeley.edu/~aldous/Real_World/dempster_shafer.pdf

The definitive reference for evidence combination under uncertainty. It covers Dempster's combination rule ResearchGate — **$m_{12}(A) = [\Sigma\_\{B \cap C=A\}\ m_1(B) \cdot m_2(C)] / (1-K)$** — alongside Yager's, Zhang's, and Dubois & Prade's alternatives. ResearchGate Most critically for Advuman, it details **Shafer's discount-and-combine method** for source reliability weighting: each source gets a trust degree ($1-\alpha_i$), and its mass function is discounted before combination. This maps directly to Advuman's Low/Medium/High confidence levels — Low-confidence sources receive heavy discounting (mass redistributed toward uncertainty), High-confidence sources retain full evidential weight.

**A Multi-Source Information Fusion Approach Based on Improved Dempster-Shafer Evidence Theory** Bo Wu, Weixing Qiu, Wei Huang et al. (2022). Scientific Reports, 12, Article 3626. DOI: 10.1038/s41598-022-07171-x

Combines three heterogeneous data types — statistical data, physical sensor data, and expert judgment — into unified Basic Probability Assignments fused via improved D-S theory. nature Multi-source fusion achieved **98.1% accuracy vs. 78.8% for single-source models**, validating the value of combining weak/diverse signals. The improved D-S method addresses the classic limitation when evidence sources highly conflict — essential for trade intelligence where port statistics may indicate normalcy while news reports flag disruption.

**Dempster-Shafer Theory for Combining Evidence and Estimating Uncertainty in Risk Assessment** James F. Rathman, Chihae Yang et al. (2018). Computational Toxicology, 6, 16-31. DOI: 10.1016/j.comtox.2018.03.001

Provides the clearest articulation of combining heterogeneous evidence sources with explicit reliability weighting via D-S theory. Its key advantage over pure Bayesian methods: **D-S does not require prior probabilities**, which are often unknown for novel trade lane disruptions. The framework generates predictions with quantitative uncertainty estimates rather than point values, ScienceDirect supporting Advuman's need for confidence-level reporting alongside risk scores.

**J.P. Morgan/Reuters RiskMetrics Technical Document (4th Edition)** J.P. Morgan/Reuters (1996). https://www.msci.com/documents/10199/5915b101-4206-4ba0-aee2-3449d5c7e95a

Introduced the EWMA model as the industry standard for time-weighting observations in risk estimation. Yale Library Value-at-Risk The core formula — $\boldsymbol{\sigma^2_n = \lambda\sigma^2_{n-1} + (1-\lambda)r^2_{n-1}}$ — uses a single decay parameter λ (typically **0.94 for daily, 0.97 for monthly** data). Half-life calculation: $t_1/_2 = -\ln(2)/\ln(\lambda)$. With $\lambda = 0.94$, a 30-day-old signal retains only ~16% of its original weight. This provides Advuman's signal aging architecture with a proven, tunable, single-parameter decay mechanism.

---

## 5. Bayesian approaches for sparse signals and sequential updating

Two sources address how to maintain and update risk beliefs when historical disruption data is limited — a central challenge for UK-India trade lane monitoring.

**Bayesian Networks for Supply Chain Risk, Resilience and Ripple Effect Analysis: A Literature Review** Seyedmohsen Hosseini, Dmitry Ivanov (2020). Expert Systems with Applications, 161, 113649. DOI: 10.1016/j.eswa.2020.113649. https://pmc.ncbi.nlm.nih.gov/articles/PMC7305519/

The first comprehensive review of Bayesian Network applications in supply chain risk management (63 papers, 2007–2019). nih BNs enable both **forward propagation** (predicting disruption impacts from observed triggers) and **backward propagation** (diagnosing root causes from observed signals). Dynamic Bayesian Networks model how risk states evolve over time — applicable to Advuman's sequential updating of RPI, LSI, and CPI. The core updating formula: $\boldsymbol{P(H|E_1,E_2) = P(E_2|H) \cdot P(H|E_1) / P(E_2|E_1)}$, applied recursively as signals arrive.

**Modelling Operational Risk Using Bayesian Inference** Pavel V. Shevchenko (2011). arXiv: 0904.1067 / Springer monograph. https://arxiv.org/abs/0904.1067

Provides the mathematical framework for Bayesian inference when data is sparse. It demonstrates how to construct **informative prior distributions** from expert opinion using conjugate priors (Gamma-Poisson for event frequencies, LogNormal for severities), then perform recursive updating as observations arrive. For Poisson frequency with Gamma prior: **Posterior ~ Gamma(α + n_obs, β + T)**. The Bayesian approach naturally produces credible intervals rather than point estimates, making uncertainty transparent — essential for Advuman's confidence reporting.

---

## 6. Rolling baselines and time-decay functions

**Exponentially Weighted Moving Models** Eric Luxenberg, Stephen Boyd (2024). Stanford University. arXiv: 2404.08136. https://web.stanford.edu/~boyd/papers/pdf/ewmm.pdf

Generalizes EWMA into a broader class of Exponentially Weighted Moving Models parameterized by **half-life H**: the number of periods for weight to decay to one-half. Core formulas:

- EWMA update: $\hat{\mu}_t = (1-\alpha)\cdot\hat{\mu}_{t-1} + \alpha\cdot x_t$, where $\boldsymbol{\alpha = 1 - 2^{(-1/H)}}$
- Weight at lag k: $w_k = (1-\alpha)^k$
- Half-life to decay factor: $\lambda = 2^{(-1/H)}$

EWMMs avoid the "cliff effect" of fixed rolling windows (where observations abruptly exit the window). Corporate Finance Institute For Advuman, recommended starting half-lives: **30 days for operational signals** (port congestion, FX movements), **90 days for structural/regulatory signals** (policy changes, enforcement actions).

---

## 7. Rule-based production systems, operational risk, and cluster escalation

Seven sources provide practical design patterns for deployed rule-based risk systems and multi-factor compounding logic.

**Key Risk Indicators: Sound Practice Guidance** Institute of Operational Risk (2010). https://www.ior-institute.org/public/IORKRIGuidanceNov2010.pdf

The most directly relevant practitioner guide for Advuman's KRI program design. It covers composite/index indicator construction (Section 4.4, Appendix 8.6), threshold and escalation trigger design (cap/floor/collar thresholds, tiered escalation, **"repetitive touch" triggers** where boundaries must be breached across consecutive periods before alerts fire), and false positive management. ior-institute The guidance recommends collecting **6–12 months of baseline data** before setting initial thresholds and advocates periodic tightening as the system matures. ior-institute

**Scorecard Models for Operational Risk Management** Paolo Giudici (2007). SIS Conference, Venice. https://www.sis-statistica.org/old/htdocs/files/pdf/atti/SIS%202007%20Venezia%20intermedio_63-69.pdf

Presents a traffic-light rating system (A/green, B/yellow, C/red) derived from expert self-assessment of frequency, severity, and control effectiveness. sis-statistica Sis-statistica Ratings use the median class as location measure and the **normalized Gini index for consensus measurement**. sis-statistica An integrated Bayesian scorecard combines forward-looking self-assessment with backward-looking loss data, Sis-statistica producing composite VaR estimates more prudent than pure actuarial models. sis-statistica Adaptable to Advuman's hybrid automated-plus-analyst architecture.

**Operational Risk — Revisions to the Simpler Approaches** Basel Committee on Banking Supervision (2014). https://www.bis.org/publ/bcbs291.pdf

Contains the Business Indicator (BI) methodology — a composite from three macro-components with size-based marginal coefficients (**12%, 15%, or 18%** depending on thresholds at €1bn and €30bn). bis Annex 2 details OpCaR calibration against empirical loss distributions, and Annex 3 systematically evaluates **20 potential proxy indicators** ranked by predictive power. bis The principle of testing multiple candidates, selecting the best composite, and calibrating against empirical data transfers directly to Advuman's indicator selection process.

**Study on Diversification in Internal Models: Public Report** EIOPA Project Group (2024). https://www.eiopa.europa.eu/system/files/2024-01/EIOPA's%20Comparative%20Study%20on%20diversification%20in%20internal%20models.pdf

The authoritative source for **cluster escalation logic**. It provides: Joint Quantile Exceedance (JQE) for measuring whether RPI, LSI, and CPI simultaneously breach thresholds; variance-covariance aggregation using the formula **$SCR = \sqrt{(\Sigma\Sigma\, Corr_{ij} \times SCR_i \times SCR_j)}$**; and copula-based dependency structures (Gaussian, Clayton, t-copula) for capturing tail dependence. europa The study found that copula choice materially impacts risk estimates at extreme quantiles europa — directly relevant to amplifying overall risk when multiple indices enter red zones simultaneously.

**Challenges for Customs Risk Management Today: A Literature Review** Laszuk et al. (2024). Journal of Risk and Financial Management (MDPI), Vol. 17, No. 8, Article 321. https://www.mdpi.com/1911-8074/17/8/321

Comprehensive survey of customs risk management systems as actually deployed, covering the evolution from manual to automated risk profiling. MDPI Reviews how customs services balance **trade facilitation versus control** — the exact tension Advuman faces in setting thresholds sensitive enough to catch genuine risk without overwhelming SME users with false positives.

---

## 8. Trade and customs risk frameworks deployed at scale

Four sources demonstrate production-deployed rule-based risk systems in trade and customs contexts.

**Customs Risk Management Framework (CRMF)** European Commission, DG Taxation and Customs Union. Ongoing (CRMS2 deployed January 2022). https://taxation-customs.ec.europa.eu/customs/customs-risk-management/customs-risk-management-framework-crmf_en

A **production-deployed, rule-based system** operating across the EU's entire external border (670 customs offices, 2,900+ risk experts, 100+ items per second). Key design elements: Common Risk Criteria (CRC) as standardized rule sets for automated flagging; European Commission three-tier classification (high/medium/low) with explicit treatment rules; UNECE Priority Control Areas (PCAs) for time-limited intensified monitoring; European Commission and Risk Information Forms (RIFs) for real-time alert dissemination. European Commission This is the closest production analogue to Advuman's architecture.

**SAFE Framework of Standards to Secure and Facilitate Global Trade** World Customs Organization (2005; latest edition 2025). https://www.wcoomd.org/-/media/wco/public/global/pdf/topics/facilitation/instruments-and-tools/tools/safe-package/safe-framework-of-standards.pdf

Establishes the global standard for customs risk management, mandating automated risk management systems using advance electronic cargo information. Risk profiles are defined as "a combination of risk criteria and control areas which indicates the existence of risk and leads to a proposal to carry out a control measure." EUR-Lex The AEO trusted-trader concept directly informs how risk scores should differentiate between established and unknown operators.

**OECD Country Risk Classification** OECD Country Risk Experts Group. Ongoing (methodology ~1997; latest revision ~2023–2024). https://www.oecd.org/en/topics/country-risk-classification.html

Uses a two-step quantitative-qualitative methodology: the **CRAM model** produces scores based on four indicator groups (ECA payment experience, financial situation, economic situation, institutional situation), followed by expert adjustment on a **0–7 scale**. OECD India is currently classified at Category 3. Contract Directory The quantitative-model-plus-expert-overlay approach validates Advuman's hybrid architecture.

**Country and Sector Risks Handbook 2024** Coface Economic Research Department (2024).
https://www.coface.com/content/download/49421/file/COFACE_GUIDE_2024_EN_WEB.pdf

Combines structural analysis (5 long-term dimensions) with cyclical assessment, integrated via a **"structural corridor"** that prevents abrupt rating swings Coface — a smoothing mechanism directly applicable to preventing false signal oscillation in Advuman's state transitions. The separate sector risk methodology (proprietary payment data + financial forecasts + multifactorial criteria) offers a validated template for sector-specific cost pressure assessment.

---

## 9. OSINT and text mining for risk signal extraction

**A Systematic Review of Text Mining Analytics for Supply Chain Risk Management Using Online Data** Wichmann, Bose, Caldwell et al. (2025). Supply Chain Analytics (Elsevier). https://www.sciencedirect.com/science/article/pii/S2949863525000676

The most current systematic review of text mining for SCRM, covering 33 studies. It taxonomizes NLP techniques (LDA topic modeling, sentiment analysis, named entity recognition, transformer models, TF-IDF) and their tradeoffs between accuracy and computational cost. Directly validates Advuman's approach of extracting risk signals from unstructured OSINT sources and provides selection criteria for NLP methods suited to regulatory, logistics, and cost signal detection.

**A Global Supply Chain Risk Management Framework: Text-Mining to Identify Region-Specific Supply Chain Risks** Ni, Srinivasan, Sun et al. (2020). Advanced Engineering Informatics (Elsevier).
https://www.sciencedirect.com/science/article/abs/pii/S1474034620300227

Proposes a practical pipeline for converting unstructured text into region-tagged risk indicators using term frequency analysis, correlation analysis, and **LDA topic modeling** across multiple corpora. The multi-corpus validation approach (progressively larger datasets confirming stability of identified risk factors) offers a model for Advuman's signal validation methodology on the UK-India corridor.

---

## 10. Precedent-based risk assessment via case-based reasoning

**Case-Based Reasoning Approach to Construction Safety Hazard Identification** Y. M. Goh, D. K. H. Chua (2010). Journal of Construction Engineering and Management (ASCE), 136(2), 170-178. DOI: 10.1061/(ASCE)CO.1943-7862.0000116

Demonstrates the **4R cycle** (Retrieve, Reuse, Revise, Retain) for using historical analogues to inform current risk assessment. For Advuman, past trade disruptions (2021–22 container crisis, CBAM regulatory changes, past currency interventions) can be stored as structured cases and retrieved when similar patterns emerge, providing precedent-based context alongside statistical risk scores. The paper addresses how to compute similarity metrics between current situations and historical cases and how to adapt past solutions to new contexts.

---

## Mathematical formula reference for Advuman's implementation

The table below consolidates the most applicable formulas extracted across all sources, organized by function within Advuman's pipeline.

| Function | Formula | Recommended source |
|---|---|---|
| **Min-max normalization** | $\tilde{x} = (x - x_{min}) / (x_{max} - x_{min})$ | OECD Handbook |
| **Z-score normalization** | $\tilde{x} = (x - \mu) / \sigma$ | OECD Handbook; GSCPI |
| **Additive aggregation** | $CI = \Sigma\, w_i \cdot \tilde{x}_i$, where $\Sigma w_i = 1$ | OECD Handbook; Greco et al. |
| **Geometric aggregation** | $CI = \Pi\, \tilde{x}_i\text{^}w_i$ | Ferretti & Ferrante |
| **Shannon entropy weight** | $w_j = (1 - E_j) / \Sigma(1 - E_k)$ | Karagiannis & Karagiannis |
| **EWMA update** | $\hat{\mu}_t = (1-\alpha)\cdot\hat{\mu}_{t-1} + \alpha\cdot x_t$ | Luxenberg & Boyd; RiskMetrics |
| **Half-life decay** | $\alpha = 1 - 2\text{^}(-1/H)$; weight at lag k: $w_k = (1-\alpha)\text{^}k$ | Luxenberg & Boyd |
| **CUSUM** | $S_t = \max(0, S_{t-1} + x_t - \mu_0 - k)$; signal if $S_t > h$ | Montgomery; Ergashev |
| **EWMA control chart** | $Z_t = \lambda x_t + (1-\lambda)Z_{t-1}$; $UCL = \mu_0 + L\sigma\sqrt{[\lambda/(2-\lambda)]}$ | Montgomery |
| **KLR noise-to-signal** | $NSR = [B/(B+D)] / [A/(A+C)]$ | Kaminsky et al. |
| **Threshold loss** | $L(\theta) = \omega_1\cdot P(\text{false alarm}) + \omega_2\cdot P(\text{missed}$ | Berg et al.; Bussière & |

| | | |
|---|---|---|
| **function** | crisis) | Fratzscher |
| **Dempster's combination** | $m_{12}(A) = [\Sigma\_\{B \cap C=A\} m_1(B) \cdot m_2(C)] / (1-K)$ | Sentz & Ferson |
| **Shafer's confidence discount** | $m'(A) = \alpha \cdot m(A)$; $m'(\Theta) = 1 - \alpha(1 - m(\Theta))$ | Sentz & Ferson |
| **Bayesian sequential update** | P(H | E) = P(E |
| **Cluster escalation (VarCovar)** | $SCR = \sqrt{(\Sigma\Sigma\ Corr_{i\,j} \times SCR_i \times SCR_j)}$ | EIOPA |
| **RiskMetrics half-life** | $t_1/_2 = -\ln(2) / \ln(\lambda)$ | RiskMetrics |

## Architectural recommendations emerging from the literature

The literature converges on a clear, implementable pipeline for Advuman that remains fully transparent and rule-based.

**Normalization**: Z-score normalization for each sub-signal within RPI, LSI, and CPI, following the GSCPI precedent. This enables intuitive interpretation as "standard deviations from baseline" and natural threshold-setting (e.g., **+1.5σ for Watch, +2.5σ for Active**).

**Weighting**: Start with expert-determined fixed weights, validated against Shannon entropy weights as a sensitivity check. The OECD Handbook's compatibility matrix should guide the pairing of weighting and aggregation methods.

**Aggregation**: Additive (weighted sum) within each sub-index for simplicity; **geometric aggregation** when combining RPI, LSI, and CPI into an overall risk state, to prevent low-risk dimensions from masking high-risk ones.

**Baseline and decay**: EWMA with configurable half-lives — **30 days for operational signals**, **90 days for structural/regulatory signals** — rather than fixed rolling windows, avoiding cliff effects.

**Threshold calibration**: KLR noise-to-signal ratio minimization for initial threshold setting, with the three-state design validated by Bussière & Fratzscher. IOR's "repetitive touch" requirement (consecutive-period breach before alert fires) for false positive reduction.

**Cluster escalation**: EIOPA's variance-covariance formula for quantifying compounding risk when multiple indices simultaneously breach thresholds. When all three indices exceed

Watch thresholds, the combined risk score should exceed the sum of individual scores.

**Confidence weighting**: Shafer's discount function mapping Low/Medium/High confidence to numeric reliability weights (e.g., 0.3/0.6/0.9), applied before signal aggregation. This is transparent, tunable, and does not require Bayesian priors.

**Signal aging**: RiskMetrics exponential decay with $\lambda = 0.94$ (daily, ~11-day half-life) for fast-moving signals, $\lambda = 0.97$ (monthly, ~23-day half-life) for slower structural signals.

This architecture draws entirely from established, peer-reviewed methods. No component requires machine learning, neural networks, or any black-box approach. Every parameter is interpretable, every threshold is auditable, and every formula is traceable to a specific methodological source — aligning completely with Advuman's philosophy of transparent, explainable risk intelligence.