Open Access Research Journal of
Science and Technology
Journals home page: https://oarjst.com/
ISSN: 2782-9960 (Online)

OARJ OPEN ACCESS RESEARCH JOURNALS

(RESEARCH ARTICLE)

TEXT-DERIVED MINING SUPPLY-CHAIN RISK FOR RISK-AWARE DEMAND FORECASTING IN ZIMBABWE'S COAL AND MANUFACTURING VALUE CHAINS: A RESEARCH PROTOCOL

Antony G. Foya, Destine Mashava *, Takudzwa M. Muhla and Innocent Mapindu

Department of Industrial and Manufacturing Engineering, Faculty of Engineering, National University of Science and Technology (NUST), Bulawayo, Zimbabwe.
* Corresponding Author

ORCID Details
Antony G. Foya: https://orcid.org/0009-0003-5215-8504
Destine Mashava: https://orcid.org/0009-0002-9487-9038
Takudzwa M. Muhla: https://orcid.org/0000-0002-6212-8099
Innocent Mapindu: https://orcid.org/0009-0007-2024-1919

Open Access Research Journal of Science and Technology, 2026, 18(02), 001–016

Publication history: Received on 27 September 2026; revised on 03 November 2026; accepted on 05 November 2026

Article DOI: https://doi.org/10.53022/oarjst.2026.18.2.0102

Copyright © 2026 Author(s) retain the copyright of this article. This article is published under the terms of the Creative Commons Attribution License 4.0

ABSTRACT

Industrial value chains built around coal in Zimbabwe, spanning mining, power, metallurgical processing and manufacturing, face interacting disruptions: electricity interruptions, transport constraints, foreign-exchange shortages, regulatory change, climate stress, equipment failure and commodity-price volatility. These disruptions usually appear first as text, in news reports, utility notices, statutory instruments and corporate announcements, and only later in production or sales records. This paper specifies a reproducible research protocol that converts Zimbabwe-focused text into supply-chain risk indices and embeds them in demand and volatility forecasts. Latent Dirichlet Allocation (LDA) identifies persistent structural themes and BERTopic identifies emerging, context-dependent ones. Expert-validated topic probabilities are aggregated into lagged weekly or monthly risk indices that enter the mean and variance equations of ARIMAX-GARCH-X models. The protocol sets out an eight-family risk taxonomy; a multi-case portfolio described by process, product and sector, never by name; information-availability and leakage controls; a nested model ladder; six hypotheses stated in advance with matched statistical tests; human-in-the-loop decision rules; and minimum reporting standards. No empirical forecast improvement is claimed, because the firm-level and text-corpus data have not yet been assembled. The contribution is a testable and auditable design for data-constrained emerging-market settings, intended to support procurement, inventory, production scheduling, logistics, maintenance and industrial-policy planning.

Keywords: Supply-Chain Risk; Demand Forecasting; Topic Modelling; LDA; BERTopic; ARIMAX-GARCH-X; Coal Value Chain; Manufacturing; Zimbabwe; Industrial Resilience

1 INTRODUCTION

Demand forecasting is central to supply-chain coordination in mining and manufacturing. Mining operations use forecasts to plan extraction, blasting, stockpiles, maintenance, haulage, rail loading and customer commitments. Manufacturers use them to plan raw materials, electricity requirements, furnace or kiln schedules, labour, inventory and distribution. Forecast errors translate directly into stockouts, idle capacity, emergency procurement, missed export windows and excessive working-capital requirements.

Conventional forecasting methods assume that the past is a reliable guide to the future. That assumption is weak in Zimbabwe's industrial environment, where electricity availability, rail and road access, foreign-exchange conditions, policy changes, weather and regional demand can alter the relationship between past and future performance. Mining and quarrying contributed approximately 15.4% of Zimbabwe's real GDP in 2024, while gold, platinum-group metals and ferroalloys were major export earners [1].

Coal is particularly important because it links mining, electricity generation, industrial heat, coke production, ferroalloys, steel, cement and transport. Coal is designated a strategic mineral in Zimbabwe's mining framework. The principal national coal producer positions itself as an anchor of a broader coal-led industrial platform, and publicly documented linkages include the supply of coke from an independent coking-coal operation to ferrochrome smelting and to an integrated steel complex under development. In such a chain, forecasts that lean only on history fail precisely when the operating environment shifts, because one failed node (power, rail or foreign exchange) propagates to many others.

The central proposition of this paper is that historical demand and production data should be supplemented with structured signals extracted from unstructured text. Disruptions usually surface first as words, in a utility notice, a regulator's circular or a corporate announcement, and only later in the ledgers. Natural-language processing can identify themes that are difficult to encode in conventional spreadsheets, such as rail-wagon delays, power curtailment, foreign-exchange retention, coal-export restrictions, water shortages, kiln maintenance or community disputes. Text-based risk indices are established tools in macroeconomics [2, 3], but their use in operational demand forecasting for a single resource-based industrial system, with explicit treatment of forecast uncertainty, has not been specified as a complete and auditable protocol. This paper supplies that specification. It is a protocol and not a result: no forecast gains are asserted until the data have been gathered and the tests run.

Four research questions guide the study.

- **RQ1.** Which latent supply-chain risk themes can be identified from Zimbabwe-focused text covering coal mining, energy, logistics, metals and manufacturing?
- **RQ2.** Do text-derived risk indices improve point and density forecasts for coal dispatches, industrial energy demand and selected manufacturing outputs beyond historical-data baselines?
- **RQ3.** Do persistent risk themes and short-lived disruption themes contribute differently to forecast accuracy and operational decision-making?
- **RQ4.** How can firms and public institutions embed explainable text-based risk signals into procurement, inventory, production, logistics and policy processes without creating opaque automated decisions?

The contributions of the paper are threefold: (i) the adaptation of topic-modelling and volatility-forecasting methods to the Zimbabwean coal and manufacturing context; (ii) the extension of the unit of analysis from a single product-demand series to an interconnected industrial value-chain portfolio; and (iii) an implementation protocol comprising a risk taxonomy, a mathematical model specification, leakage controls, a validation design with pre-stated tests, governance rules and a decision-flow architecture.

The remainder of the paper is organised as follows. Section 2 reviews related work on supply-chain risk, text-based indicators, topic modelling and forecasting. Section 3 describes the industrial context and the risk taxonomy. Section 4 sets out the research design and data architecture. Section 5 presents the mathematical models and the inference strategy. Section 6 states the hypotheses and the validation design. Section 7 describes the decision-flow architecture and implementation roadmap. Section 8 discusses the implications and limitations, and Section 9 concludes.

2 LITERATURE REVIEW

2.1 Supply-chain risk and resilience

The resilience literature has long argued that disruption management is a design problem and not merely a crisis-response problem [4–6]. Later work stresses that unpredictable disruptions demand flexibility and early visibility [7]

and that interdependent networks must be assessed for survivability as systems, not as isolated links [8]. The present framework takes from this tradition the principle that visibility precedes response, and it treats the early-warning signal as a design component of the chain and not as an afterthought.

2.2 Text as data: risk indices and topic models

Text has been converted into economic indicators in several settings. Newspaper-based measures of policy uncertainty and geopolitical risk show that counts and shares of risk-related language carry information that conventional series reveal only later [2, 3]. These indices rely on keyword dictionaries, which are transparent but rigid. Topic models offer a data-driven alternative. LDA models each document as a mixture of latent topics and each topic as a probability distribution over words [9]; it suits stable, recurring themes in a large corpus. BERTopic combines transformer-based document embeddings, building on [10], dimensionality reduction [11], density-based clustering [12] and class-based TF-IDF representations [13], and is generally better suited to short, noisy or context-dependent documents. Structural topic models, which allow document-level covariates to shape topic prevalence, offer a further extension where metadata such as source type or province matter [14]. The two main methods are complementary: LDA is stable and interpretable at the level of the whole corpus, whereas BERTopic adapts more quickly to new vocabulary.

2.3 Separating the conditional mean from the conditional variance

The forecasting component separates the conditional mean from the conditional variance. The mean model estimates expected demand or production; the variance model estimates forecast uncertainty. The separation matters because a disruption may reduce expected output and simultaneously increase uncertainty. A risk signal that raises forecast variance can justify higher safety stock, more conservative scheduling or contingency transport arrangements even when expected demand is unchanged. The mean equation follows the ARIMA tradition [15, 16] and the variance equation the ARCH and GARCH tradition [17, 18].

2.4 Positioning of the present study

The three strands above (resilience, text-based indicators and mean-variance forecasting) have developed largely separately. Resilience research supplies the case for early visibility but rarely a quantitative early-warning signal; text-indicator research supplies the signal but typically targets aggregate macroeconomic variables; and volatility modelling supplies the uncertainty machinery but is seldom fed with operational text. The protocol proposed here joins them in a single chain:

- textual events;
- latent topics;
- validated risk categories;
- lagged risk indices;
- demand and volatility forecasts;
- operational response.

Each link is made testable: topics are validated by experts, indices are lagged and standardised without look-ahead, forecasts are scored against nested baselines, and responses are logged and audited.

3 INDUSTRIAL CONTEXT AND RISK TAXONOMY

3.1 Coal as an industrial-system node

The coal value chain comprises exploration, mine planning, extraction, beneficiation, stockpiling, road and rail transport, power generation, industrial boilers, coke production, metallurgical applications, exports and community-facing operations. A disruption at any node can affect the quantity, quality, timing and cost of delivered coal.

The chain is interconnected. Coal supports electricity and industrial heat; coke and ferroalloys support metals processing; steel, cement and other manufactured products support construction and infrastructure. A coal-supply disruption may therefore propagate through power generation, ferrochrome production, steel manufacturing, construction activity and transport demand, and demand shocks can travel in the opposite direction. Fig. 1 shows the coal-linked value chain and the types of text-detectable disruption signal associated with each node.

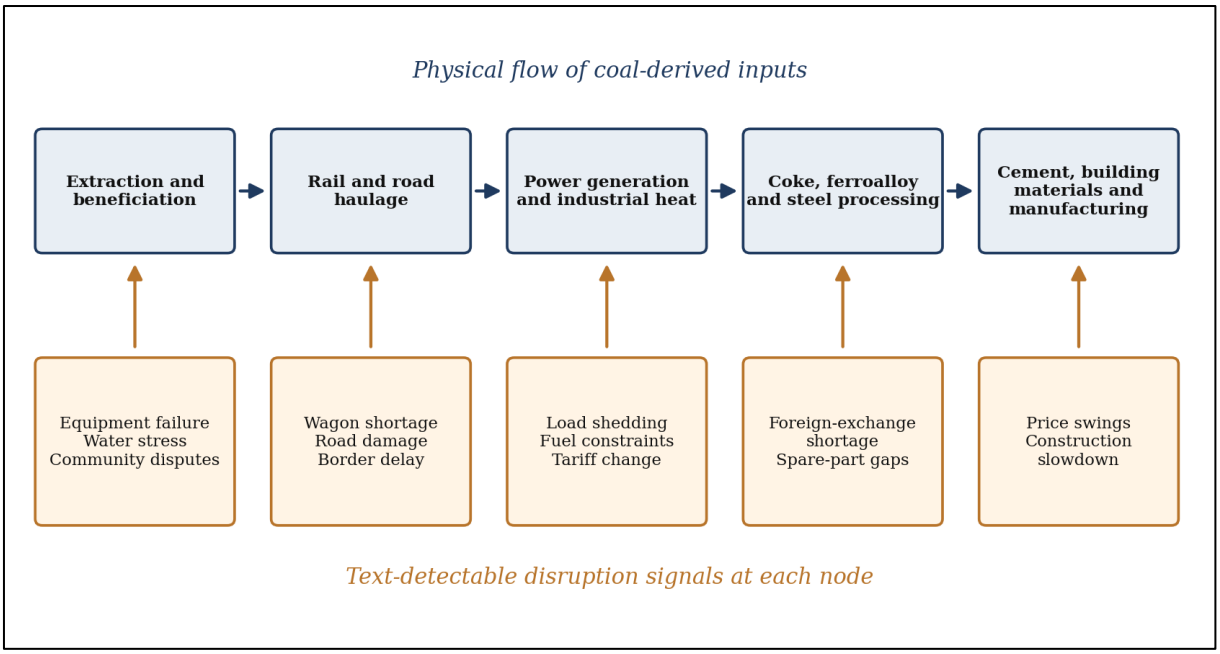


Figure 1: Schematic of the coal-linked value chain and the types of text-detectable disruption signal associated with each node.

3.2 Manufacturing and beneficiation linkages

Zimbabwe's policy direction emphasises domestic value addition and beneficiation. That direction is expected to require investment in mining, processing, heavy equipment, transport infrastructure and power solutions. As beneficiation expands, demand may shift from raw-mineral exports toward electricity, coke, steel inputs, chemicals, engineering services, transport capacity and skilled labour. Forecasting systems must therefore capture cross-sector dependencies rather than treat each firm or commodity in isolation.

3.3 Supply-chain risk taxonomy

The taxonomy in Table 1 organises disruptions into eight families. It is a starting vocabulary for expert annotation and not a closed list; new families should be admitted when the evidence demands it, and the admission rule (Section 4.5) should be recorded in the annotation protocol.

Table 1: Zimbabwe supply-chain risk taxonomy.

Risk family	Examples of events	Principal text sources	Likely operational effect
Energy risk	Load shedding, generation shortfall, fuel constraints, tariff changes	Utility notices, regulator statements, news	Reduced output, higher cost, production rescheduling
Logistics risk	Rail-wagon shortage, road damage, border delays, port congestion	Transport-operator and border-agency notices, news	Delayed deliveries, higher freight cost, stockouts
Financial and currency risk	Foreign-exchange shortage, payment delays, inflation, interest-rate changes	Central-bank and ministry statements, news	Input shortages, delayed procurement, working-capital pressure
Regulatory and political risk	Licensing changes, export restrictions, taxes, environmental rules	Statutory instruments, gazette notices, parliamentary records	Compliance cost, market-access uncertainty
Operational and technical risk	Equipment failure, spare-parts shortage, labour disruption, safety incident	Company releases, trade press	Downtime, quality loss, reduced throughput

Environmental and climate risk	Drought, flooding, heat, water scarcity, rehabilitation obligations	Weather alerts, water-authority notices, news	Production constraints, cost increases, social conflict
Supply-and-demand risk	Commodity-price movements, construction slowdown, customer shutdowns	Market reports, customer announcements	Demand volatility, inventory imbalance
Social and community risk	Labour disputes, community protests, resettlement issues	News, community and civil-society statements	Operational interruption, reputational and licence-to-operate risk

4 RESEARCH DESIGN AND DATA ARCHITECTURE

4.1 Research strategy

The proposed study uses a mixed-method, multi-case, longitudinal design. It combines computational text analysis, quantitative forecasting and qualitative validation by supply-chain, mining, engineering, finance, logistics and production experts.

The design has two layers: a *sector layer* using public and institutional data to construct Zimbabwe-wide or sector-wide indicators, and a *firm layer* using permissioned company data to estimate operational forecasts for participating firms. The layers should not be merged blindly. Public news may be timely but noisy, whereas internal enterprise-planning and production data may be accurate but available only after an event.

4.2 Units of analysis and candidate cases

The unit of analysis is a company–product–period observation: for example, tonnes of coal dispatched per week, megawatt-hours consumed per month, tonnes of coke produced per week, tonnes of ferrochrome or steel produced per month, cement dispatches per month, or selected manufacturing sales per month. Candidate cases are described by process and sector, not by name (Table 2). Participation should depend on data availability, value-chain relevance and organisational consent.

Table 2: Proposed multi-case portfolio, described by process, product and sector.

Case type	**Descriptor**	**Primary modelling target**
Thermal-coal producer	The principal national mining operation (surface and underground) and independent mid-scale producers supplying power, industry and export markets	Coal output, dispatches, sales
Coking-coal and coke producer	Independent coking-coal mining with coke ovens supplying metallurgical users	Coke output, dispatches
Ferroalloy and steel processor	Ferrochrome smelting, and integrated or secondary steel production, including a large integrated steel complex under development	Ferrochrome and steel output, electricity use
Energy-intensive materials manufacturer	Cement and other kiln- or furnace-based producers	Electricity demand, production volume
Downstream manufacturer	Construction-materials, food-and-beverage and other converting industries	Sales, dispatches, inventory

4.3 Data architecture

Four data streams feed the study (Table 3). Their different provenance and timing must be recorded, because information-availability rules are what separate genuine forecasting from hindsight.

Table 3: Data streams and variables.

Stream	Data type	Examples
A: Operational and demand data	Internal, permissioned	Production, dispatches, sales orders, stock levels, downtime, power use, transport movements, lead times
B: External text	Public or licensed	News, government notices, statutory instruments, utility communications, company releases, parliamentary records, weather alerts
C: Structured macro variables	Public or commercial	Exchange rates, inflation, interest rates, fuel prices, rainfall, temperature, commodity prices, public holidays
D: Expert annotations	Human-labelled	Risk category, affected node, direction, severity, lead time, confidence, prior visibility

4.4 Text preprocessing and temporal integrity

Documents should be deduplicated and dated by publication time rather than retrieval time; where a document is revised, each version carries its own timestamp and only the version available at the forecast origin may be used. Documents should be linked to an entity dictionary covering mines, processors, utilities, rail operators, border posts, provinces, products, regulators and spelling variants. The preprocessing pipeline comprises tokenisation; lower-casing where appropriate; removal of boilerplate and non-informative stop words; lemmatisation; phrase detection for terms such as “load shedding”, “foreign exchange” and “rail wagon”; named-entity recognition; preservation of negation; and translation or documented handling of Shona, Ndebele and other local-language content.

Negation deserves particular care. A report that “no power interruption is expected” must never be treated as equivalent to one that warns of an interruption. Preprocessing choices, the pretrained embedding model and all random seeds should be frozen and versioned before any outcome series is examined.

4.5 Expert annotation and validation

Topic output is a statistical object until experts have given it operational meaning. A stratified sample of documents and all topic top-word lists should be labelled independently by at least two experts from different functions (for example mining operations, logistics and finance) using the taxonomy in Table 1. The annotation protocol should record the risk category, affected node, expected direction, severity and lead time, with a confidence rating. Inter-annotator agreement should be reported with Krippendorff’s alpha [19], disagreements should be adjudicated by a third expert, and the minimum acceptable agreement should be fixed in advance. Topic quality should be judged jointly by automated coherence [20], diversity, stability across random seeds and the experts’ interpretability ratings. A candidate topic that fits no existing family and recurs across at least two independent sources and several periods should be proposed as a new family and reviewed by the expert panel.

5 MATHEMATICAL MODELS

Table s1 in Appendix B lists the notation. Equation labels follow the order of use.

5.1 LDA topic model

Let D denote the number of documents, K the number of topics and V the vocabulary size. In LDA each document d is a mixture over topics and each topic k a distribution over words [9]:

$$\boldsymbol{\theta}_d \sim \text{Dirichlet}(\boldsymbol{\alpha}), \qquad \boldsymbol{\phi}_k \sim \text{Dirichlet}(\boldsymbol{\eta}), \tag{1}$$

$$z_{dn} \sim \text{Categorical}(\boldsymbol{\theta}_d), \qquad w_{dn} \sim \text{Categorical}\left(\boldsymbol{\phi}_{z_{dn}}\right), \tag{2}$$

where $\boldsymbol{\theta}_d$ is the topic distribution of document d, $\boldsymbol{\phi}_k$ is the word distribution of topic k, z_{dn} is the topic assignment of word n in document d, w_{dn} is the observed word, and $\boldsymbol{\alpha}$ and $\boldsymbol{\eta}$ are Dirichlet hyperparameters. The number of topics K should be selected using topic coherence, topic diversity, stability across random seeds and expert interpretability, and not by likelihood alone.

5.2 BERTopic representation

BERTopic embeds each document with a transformer language model, reduces the dimensionality of the embeddings, clusters the documents and describes each cluster using class-based TF-IDF [13]:

$$\mathbf{e}_d = f_{\text{Transformer}}(x_d), \qquad \mathbf{u}_d = \text{UMAP}(\mathbf{e}_d), \qquad c_d = \text{HDBSCAN}(\mathbf{u}_d), \tag{3}$$

$$W_{k,v} = \text{tf}_{k,v} \log\left(1 + \frac{A}{f_v}\right), \tag{4}$$

where $\mathbf{e}_d$ is the document embedding, x_d the cleaned text, $\mathbf{u}_d$ the reduced embedding, c_d the cluster (topic) assignment, $\text{tf}_{k,v}$ the frequency of term v in topic k, f_v the frequency of term v across all topics, A the average number of words per topic, and $W_{k,v}$ the resulting class-based TF-IDF weight. Because HDBSCAN assigns some documents to no cluster, the share of unassigned documents in each period should be reported, and soft cluster-membership probabilities should be used in place of hard labels when constructing indices.

5.3 Topic-based risk indices

For topic k in period t, the basic risk index is

$$R_{k,t} = \frac{1}{N_t} \sum_{d \in t} w_d \ \theta_{d,k}, \tag{5}$$

where N_t is the number of documents published in period t, $\theta_{d,k}$ is the probability that document d belongs to topic k (for BERTopic, the soft membership probability), and w_d is a weight reflecting source reliability, relevance or document length. Three refinements are needed before an index can enter a regression.

First, periods with few documents give noisy ratios. A minimum document count per period should be set, thin periods should be pooled or shrunk towards the running mean, and a separate *attention* series (the document count itself) should be retained so that volume and content are not confounded.

Second, each index is standardised using training-window statistics only,

$$\tilde{R}_{k,t} = \frac{R_{k,t} - \bar{R}_k^{\text{train}}}{s_k^{\text{train}}}, \tag{6}$$

so that no information from the evaluation period leaks into the scale.

Third, topic shares are compositional: if every document has weight one, the topic indices sum to a constant in each period, and including all of them with an intercept produces exact collinearity. One reference topic must therefore be dropped, or indices must be aggregated to risk categories before estimation. A directional category-level index for risk category c in Table 1 is

$$R_{c,t}^{*} = \sum_{k \in c} R_{k,t} \ S_k \ D_k, \tag{7}$$

where S_k is the expert-validated severity score and D_k the expected direction of operational impact. Where the direction cannot be established reliably, a non-directional intensity index should be retained,

$$R_{c,t}^{\text{int}} = \sum_{k \in c} R_{k,t} \ S_k, \tag{8}$$

and the forecasting model should then estimate the sign of the effect empirically.

5.4 ARIMAX-GARCH-X forecasting model

Let Y_t denote a demand, production, dispatch or consumption series, transformed where necessary (logarithm or Box–Cox) to stabilise variance and to keep forecasts non-negative; series with structural zeros may use $\log(1 + Y_t)$. The conditional mean equation is

$$Y_t = \mu + \sum_{i=1}^{p} \phi_i \, Y_{t-i} + \sum_{j=1}^{q} \psi_j \, \varepsilon_{t-j} + \sum_{m=1}^{M} \beta_m \, X_{m,t-1} + \sum_{k=1}^{K_R} \gamma_k \, \tilde{R}_{k,t-1} + \varepsilon_t, \tag{9}$$

where Y_{t-i} are lagged values of the target, ε_t is the forecast error, $X_{m,t-1}$ are structured macroeconomic and operational controls, $\tilde{R}_{k,t-1}$ are lagged, standardised text-derived risk indices, and ϕ_i, ψ_j, β_m and γ_k are parameters to be estimated. (Moving-average coefficients are written ψ_j so as not to collide with the topic probabilities $\theta_{d,k}$.) For an h-step-ahead direct forecast, the single lag in Eq. (9) is replaced by a lag of at least h periods, so that every regressor is known at the forecast origin. The conditional variance equation is

$$\sigma_t^2 = \omega + \sum_{i=1}^{r} a_i \, \varepsilon_{t-i}^2 + \sum_{j=1}^{s} b_j \, \sigma_{t-j}^2 + \sum_{h=1}^{H} \delta_h \, Z_{h,t-1}, \tag{10}$$

where σ_t^2 is the conditional variance, $Z_{h,t-1}$ are lagged risk variables used to explain volatility, a_i capture the effect of recent shocks, b_j capture volatility persistence and δ_h measure the effect of risk signals on uncertainty. A positive and statistically significant δ_h would imply that the risk signal increases forecast uncertainty, supporting larger safety stocks or contingency planning, as illustrated schematically in Fig. 2.

Two conditions are needed for Eq. (10) to define a valid variance. Positivity requires $\omega > 0$, $a_i \geq 0$, $b_j \geq 0$, and $\delta_h Z_{h,t-1} \geq 0$; the last condition holds when $Z_{h,t-1}$ is a non-negative intensity index (Eq. (8), rescaled to be non-negative) and $\delta_h \geq 0$. Standardised indices that can be negative should not enter Eq. (10) directly. Where a signed effect on variance is wanted, the exponential form of Nelson [21] removes the constraint:

$$\ln\sigma_t^2 = \omega + \sum_{i=1}^{r} a_i \left(|z_{t-i}| - \mathrm{E}|z_{t-i}|\right) + \sum_{j=1}^{s} b_j \ln\sigma_{t-j}^2 + \sum_{h=1}^{H} \delta_h \, Z_{h,t-1}, \tag{11}$$

with $z_t = \varepsilon_t / \sigma_t$. Covariance stationarity requires $\sum a_i + \sum b_j < 1$ in Eq. (10); estimates near the boundary should be treated as a sign of structural breaks and not as genuine persistence.

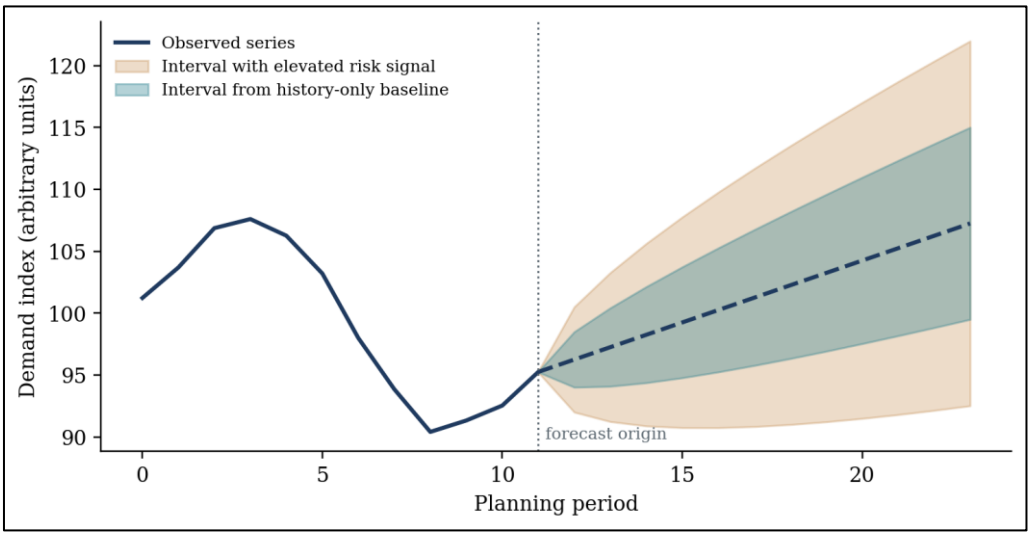


Figure 2: Schematic only: how a risk signal is expected to widen the prediction interval around an unchanged expected path. The series is illustrative and is not empirical data.

5.5 Estimation, dimensionality and inference

Mean and variance equations are estimated jointly by quasi-maximum likelihood, with a heavy-tailed innovation distribution (Student's t) as a robustness alternative. Weekly series over a few years yield only a few hundred observations, so the number of free parameters must be kept small. Where the number of candidate indices is large relative to the sample, indices should be aggregated to the eight risk categories, and penalised selection by the lasso [22] or the elastic net [23] should be applied inside each rolling estimation window, never on the full sample, with the selected specification re-estimated before inference. Where a GARCH-type variance model is poorly identified in short samples, a transparent alternative is to regress absolute or squared forecast errors, or quantiles of the forecast distribution, on the lagged risk indices.

Risk indices are themselves estimated, so they are *generated regressors*: ordinary standard errors ignore the sampling error of the first stage and are generally too small [24]. Inference on γ_k and δ_h should therefore use a bootstrap that resamples documents or time blocks and repeats topic modelling and index construction, and the topic models must be refitted, or at minimum their vocabulary and clusters frozen, inside each rolling window so that later documents do not shape earlier indices.

5.6 Forecast-evaluation metrics

Let n be the length of the estimation sample and T the number of out-of-sample forecasts. Point forecasts are scored using

$$\mathrm{MAE} = \frac{1}{T}\sum_{t=1}^{T}\left|Y_t - \hat{Y}_t\right|, \qquad \mathrm{RMSE} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(Y_t - \hat{Y}_t\right)^2}, \tag{12}$$

$$\mathrm{MASE} = \frac{\frac{1}{T}\sum_{t=1}^{T}\left|Y_t - \hat{Y}_t\right|}{\frac{1}{n-m}\sum_{t=m+1}^{n}\left|Y_t - Y_{t-m}\right|}, \tag{13}$$

where m is the seasonal period ($m = 1$ for the non-seasonal naïve forecast) and the denominator is the in-sample mean absolute error of the naïve forecast, so that values below one indicate an improvement on that benchmark [25].

Density forecasts should be evaluated using prediction-interval coverage [26], interval width, the weighted interval score and the logarithmic score, all of which reward honest uncertainty [27]. For a central $(1-\alpha)$ prediction interval $[l, u]$ and outcome y, the interval score is

$$\mathrm{IS}_\alpha(l, u; y) = (u - l) + \frac{2}{\alpha}(l - y)\,\mathbf{1}\{y < l\} + \frac{2}{\alpha}(y - u)\,\mathbf{1}\{y > u\}, \tag{14}$$

and the weighted interval score combines the median forecast m_{med} with intervals at levels $\alpha_1, \ldots, \alpha_K$ [28]:

$$\mathrm{WIS} = \frac{1}{K + 1/2}\left(1/2\,|y - m_{\mathrm{med}}| + \sum_{k=1}^{K}\frac{\alpha_k}{2}\,\mathrm{IS}_{\alpha_k}(l_k, u_k; y)\right). \tag{15}$$

Competing models are compared on the loss differential $d_t = L\left(e_t^{\mathrm{base}}\right) - L\left(e_t^{\mathrm{risk}}\right)$, where L is squared or absolute error and a positive mean indicates that the risk-aware model is better. The Diebold–Mariano statistic [29] is

$$\mathrm{DM} = \frac{\bar{d}}{\sqrt{\hat{V}\left(\bar{d}\right)}}, \tag{16}$$

where $\hat{V}\left(\bar{d}\right)$ is a heteroskedasticity- and autocorrelation-consistent variance estimate with a small-sample correction [30]. Two cautions apply. The risk-aware model nests its baseline, and for nested models the Diebold–Mariano statistic is not asymptotically normal under the null, so the Clark–West adjustment [31] is the primary test for nested pairs; the Diebold–Mariano test is retained for non-nested comparisons such as those against machine-learning models. Where

the question is whether the gain is concentrated in disruption periods, conditional predictive ability tests [32] are more informative than unconditional averages.

5.7 Model comparison framework

Table 4 sets out the ladder of models to be estimated. Each rung must justify itself against the one below it.

Table 4: Forecast-model comparison.

Model	Inputs	Purpose	Expected role
Seasonal naïve	Historical target series	Minimum benchmark	Tests whether sophisticated models add value
SARIMA	Historical target series	Captures trend and seasonality	Statistical baseline
ARIMAX	Target series plus macro and operational controls	Adds structured explanatory variables	Tests value of structured data
ARIMAX-GARCH-X	Target series, controls, lagged topic-risk indices	Models mean and volatility	Tests value of text-derived risk signals
Machine-learning robustness models	Engineered features, lagged variables	Nonlinear robustness check	Gradient boosting, random forest, LSTM or temporal convolutional networks

Machine-learning models should be treated as robustness checks and not as replacements for interpretable baselines. Operational planners need to understand why a forecast has changed, and a model that cannot explain itself will be overruled the first time it is inconvenient.

6 HYPOTHESES AND VALIDATION DESIGN

6.1 Research hypotheses

Six hypotheses are advanced and, together with their tests, fixed before any outcome series is examined.

- **H1** (point accuracy). Adding lagged text-derived risk indices improves out-of-sample point-forecast accuracy relative to historical-data-only baselines, that is, $\mathrm{E}[d_t] > 0$ for the loss differential defined in Section 5.6.
- **H2** (density calibration). Adding risk indices improves density-forecast calibration and weighted interval score, especially during disruption periods.
- **H3** (horizon). BERTopic-derived indices contribute more to short-horizon forecasts during abrupt disruptions, while LDA-derived indices contribute more to medium-horizon forecasts dominated by persistent conditions.
- **H4** (interconnection). Forecast improvements are larger for interconnected value-chain targets, such as coke, ferrochrome, steel, power-linked production and cement.
- **H5** (validation and diversity). Operational value increases when topic labels are validated by domain experts and sources are geographically and institutionally diverse.
- **H6** (human review). A human-in-the-loop decision process reduces false-positive operational interventions compared with unreviewed automated alerts.

Table 5 links each hypothesis to the evidence that would support or refute it. It states the test in advance, so that the conclusion cannot be chosen after the data are seen. Because six hypotheses are examined across several targets and horizons, the family-wise error rate should be controlled within each hypothesis family, for example by the Holm procedure [33], and every departure from the analysis plan should be reported.

Table 5: Hypotheses, required evidence and proposed tests.

Hypothesis	Research question	Primary evidence	Proposed test
H1	RQ2	MAE, RMSE and MASE of risk-aware versus baseline models on rolling origins	Clark–West test for nested pairs; Diebold–Mariano test with small-sample correction for non-nested pairs
H2	RQ2	Interval coverage, width, weighted interval score and logarithmic score	Coverage tests and score comparison, with disruption periods stratified
H3	RQ3	Horizon-specific accuracy by index family (BERTopic versus LDA)	Horizon-by-horizon model comparison
H4	RQ2, RQ3	Gain over baseline by target type	Comparison of gains across interconnected and isolated targets
H5	RQ1, RQ4	Forecast and decision outcomes with and without expert-validated labels; source-diversity audits	Ablation of validation and of source groups
H6	RQ4	Rate of unwarranted interventions with and without human review	Paired comparison of decision logs

6.2 Validation, bias and robustness

The study should use rolling-origin or sliding-window validation aligned with the planning horizon. One-week-ahead forecasts may support dispatch and logistics decisions; one-month-ahead forecasts may support production and procurement; quarterly forecasts may support strategic capacity planning. A minimum initial estimation window and the refitting frequency should be fixed in advance. Six safeguards are proposed: (i) recording publication timestamps and information-availability rules; (ii) lagging all topic indices relative to the forecast target; (iii) comparing company-originated text with independent sources; (iv) conducting placebo and permutation tests using randomly timed topics; (v) auditing source coverage, language, geography and political framing; and (vi) testing topic stability across rolling windows.

Statistical accuracy is necessary but not sufficient. A decision-based evaluation should therefore replay historical planning decisions, applying the safety-stock rule of Section 7.1 with and without the risk signal and comparing the resulting service levels and holding costs. Vocabulary drift should also be anticipated: terms such as “load shedding”, “curtailment” and “power rationing” may describe similar events in different periods, and a robust system should monitor whether topic meanings remain stable over time. Every one of these checks is cheaper than the cost of acting on a spurious signal.

7 DECISION-FLOW ARCHITECTURE AND IMPLEMENTATION ROADMAP

7.1 Text-to-decision pipeline

A forecast that never reaches a decision is an academic exercise. Fig. 3 and Table 6 set out how text becomes action, with human review built into the flow. Where the forecast distribution has mean $\hat{\mu}_{t+h}$ and standard deviation $\hat{\sigma}_{t+h}$, a simple and transparent buffer rule for a target service level τ is

$$s_{t+h} = \hat{\mu}_{t+h} + z_\tau \, \hat{\sigma}_{t+h}, \tag{17}$$

where z_τ is the standard normal quantile. A risk signal that raises $\hat{\sigma}_{t+h}$ through Eq. (10) raises the buffer even if $\hat{\mu}_{t+h}$ is unchanged, which is precisely the operational meaning of the volatility channel. For markedly skewed series the empirical forecast quantile should replace the normal approximation.

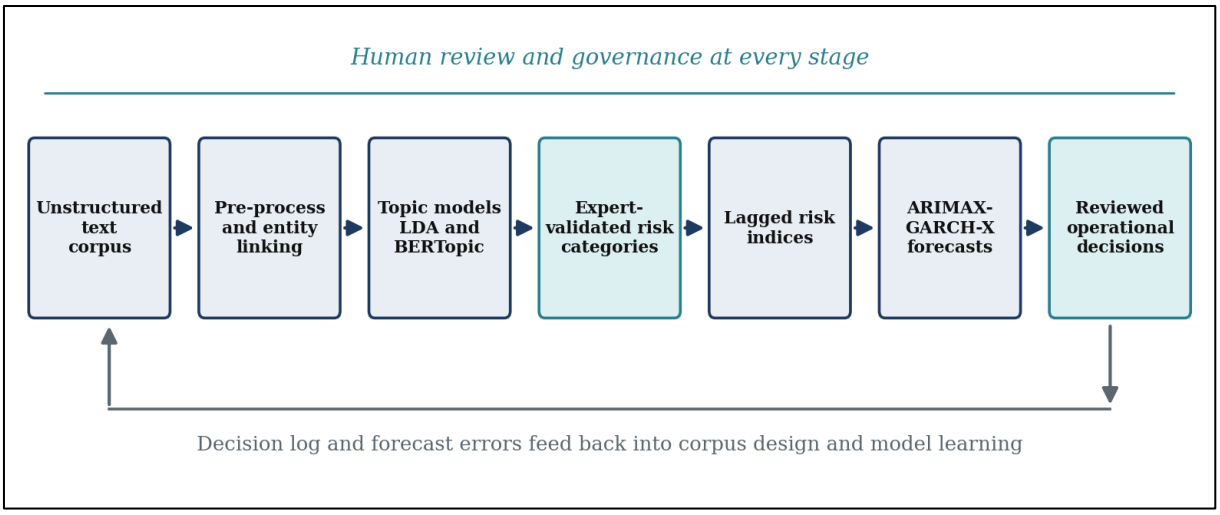


Figure 3: Text-to-decision pipeline, with human review at each stage and a feedback loop from decisions to model learning

Table 6: Text-to-decision workflow.

Stage	Activity	Output
1. Monitor	Ingest text and structured indicators daily or weekly	Time-stamped data repository
2. Classify	Assign documents to topics and risk categories	Topic probabilities and confidence scores
3. Aggregate	Construct risk indices by category, geography, entity and horizon	Risk-index dashboard
4. Forecast	Generate baseline and risk-aware forecasts	Point forecasts and prediction intervals
5. Review	Cross-functional review when thresholds are crossed	Documented decision recommendation
6. Act and learn	Adjust procurement, inventory, production, logistics or maintenance	Decision log and model-learning feedback

An operational alert should show the affected risk category, the change from the previous period, representative source excerpts, the entities and locations involved, the forecast impact, the uncertainty range and the recommended decision owner.

7.2 Implementation roadmap

Table 7 sets out a phased plan. The phases are sequential by design: baselines come before text models, and text models before any claim about operational value.

Table 7: Phased implementation plan.

Phase	Duration	Key activities
1. Feasibility and governance	0–3 months	Select target series, confirm data owners, establish lawful text access, define taxonomy
2. Corpus and baselines	3–6 months	Build text corpus, clean operational data, estimate naïve and SARIMA baselines
3. Topic modelling and validation	6–9 months	Fit LDA and BERTopic, test stability, label topics with experts

4. Forecast integration	9–12 months	Estimate ARIMAX and ARIMAX-GARCH-X models, compare against baselines
5. Decision pilot	12–18 months	Integrate selected alerts into planning meetings and track decisions
6. Scale and audit	18+ months	Extend to additional firms or products after governance and performance review

8 DISCUSSION

8.1 Company-level implications

For coal producers, the framework can support dispatch planning, stockpile management, rail and truck allocation, spare-parts procurement, customer prioritisation and maintenance scheduling. For smelters, steel plants, cement producers and other manufacturers, it can support energy procurement, raw-material ordering, production sequencing, inventory buffers and customer-service planning.

Risk management should be integrated with demand planning rather than treated as a separate annual compliance exercise. A text signal is valuable only when it changes a decision, or improves confidence in not changing one. Firms are therefore advised to begin with two or three high-value series, to build plain baselines first, and to add text signals only where they demonstrably earn their keep.

8.2 Public-sector and policy implications

Public institutions can use aggregated, non-confidential risk indices to identify recurring bottlenecks across coal, power, rail, roads, borders and manufacturing. Such a system would complement, not replace, official statistics and sector consultation. The approach also supports beneficiation policy by helping policymakers judge whether infrastructure, energy supply, transport capacity and supplier networks are expanding in line with domestic-processing ambitions.

8.3 Sustainability and social implications

The framework should include environmental and social signals, and should not define resilience merely as uninterrupted output. Safety events, water stress, emissions, community impacts and rehabilitation obligations should be linked to escalation rules. Artificial intelligence must never be used to justify unsafe production acceleration or to suppress legitimate worker and community concerns. Human review, source transparency and a clear distinction between early-warning signals and verified facts are essential: every alert should be reviewed by people who know the plant, the line or the haul road.

8.4 Limitations and threats to validity

This paper presents a research framework and empirical study protocol. It does not report forecast metrics, company performance or causal effects, because none have yet been produced. The principal threats to validity are as follows.

- **Construct validity.** Public text may overrepresent crises and politically salient events, and English-language models may underrepresent local-language information. Topic labels may be unstable when the corpus is small.
- **Internal validity.** Improved forecasts do not prove causality: a risk index may proxy for an unobserved driver of both text and demand. Generated-regressor error and look-ahead leakage are controlled by the designs in Sections 5.5 and 6.2 but cannot be eliminated.
- **Statistical-conclusion validity.** Company data may be incomplete or inconsistently recorded, small firm-level samples may favour parsimonious models over deep-learning alternatives, and structural breaks in currency, energy, policy and trade conditions may complicate historical comparison and the estimation of volatility persistence.
- **External validity.** The candidate portfolio is specific to Zimbabwe's coal-linked chain, and findings should not be assumed to transfer to other commodities or countries without re-estimation.

These limitations make data provenance, expert validation, leakage control and out-of-sample testing essential rather than optional.

8.5 Future work

Future work follows the phased roadmap in Table 7: assembling the corpus and operational data, estimating baselines, fitting and validating the topic models, estimating the risk-aware forecasting models, piloting selected alerts in planning meetings and, after governance and performance review, extending the approach to additional firms and products.

9 CONCLUSION

Zimbabwe's coal, mining and manufacturing value chains are strategically important to energy security, export earnings, beneficiation and industrialisation. Their operating conditions are shaped by risks that appear in both structured and unstructured data. This paper has proposed a practical and testable framework for converting text into risk-aware forecasts. LDA can identify persistent structural themes; BERTopic can surface emerging contextual disruptions; expert validation can translate topics into meaningful supply-chain categories; and ARIMAX-GARCH-X models can connect those signals to demand and volatility.

The future of the approach depends less on selecting the most advanced model than on building a reliable information-to-decision system. Four recommendations follow for the firms and institutions that take it up. First, begin with a small number of high-value forecast targets. Second, insist on transparent baselines, and make every new model beat them in out-of-sample tests. Third, draw on diverse, lawful sources, and record what each one can and cannot see. Fourth, keep experienced people in the loop, because a signal is only as good as the judgement that weighs it.

STATEMENTS ON COMPLIANCE WITH ETHICAL STANDARDS

Acknowledgments

The authors gratefully acknowledge the National University of Science and Technology for supporting this research.

Disclosure of conflict of interest

The authors declare that there are no financial or personal relationships that could have inappropriately influenced the research, analysis, or interpretation presented in this paper.

Data Availability and Ethical Approval

This manuscript specifies a future empirical design and does not claim access to confidential company data. Organisations are deliberately described by process, product and sector, and no organisation is identified or implied to have agreed to participate. Any full empirical implementation should obtain informed organisational consent, respect data-protection and contractual restrictions, document source licences, protect commercially sensitive information and report model uncertainty. Inclusion of a sector or process in the candidate portfolio does not imply participation; this should be confirmed separately with each organisation.

REFERENCES

[1] U.S. Geological Survey. Zimbabwe: Minerals Yearbook country summary. U.S. Department of the Interior; 2025.

[2] Baker SR, Bloom N, Davis SJ. Measuring economic policy uncertainty. *Q J Econ*. 2016;131(4):1593–1636.

[3] Caldara D, Iacoviello M. Measuring geopolitical risk. *Am Econ Rev*. 2022;112(4):1194–1225.

[4] Christopher M, Peck H. Building the resilient supply chain. *Int J Logist Manag*. 2004;15(2):1–14.

[5] Tang CS. Robust strategies for mitigating supply chain disruptions. *Int J Logist Res Appl*. 2006;9(1):33–45.

[6] Pettit TJ, Fiksel J, Croxton KL. Ensuring supply-chain resilience: Development of a conceptual framework. *J Bus Logist*. 2010;31(1):1–21.

[7] Simchi-Levi D, Schmidt W, Wei Y. From superstorms to factory fires: Managing unpredictable supply-chain disruptions. *Harv Bus Rev*. 2014;92(1/2):96–101.

[8] Ivanov D, Dolgui A. Viability of intertwined supply networks: Extending the supply-chain resilience angles towards survivability. *Int J Prod Res*. 2020;58(10):2904–2915.

[9] Blei DM, Ng AY, Jordan MI. Latent Dirichlet allocation. *J Mach Learn Res*. 2003;3:993–1022.

[10] Devlin J, Chang M-W, Lee K, Toutanova K. BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of NAACL-HLT*; 2019. p. 4171–4186.

[11] McInnes L, Healy J, Melville J. UMAP: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint* arXiv:1802.03426. 2018.

[12] Campello RJGB, Moulavi D, Sander J. Density-based clustering based on hierarchical density estimates. In: *Advances in Knowledge Discovery and Data Mining*. 2013. p. 160–172.

[13] Grootendorst M. BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv preprint* arXiv:2203.05794. 2022.

[14] Roberts ME, Stewart BM, Tingley D, et al. Structural topic models for open-ended survey responses. *Am J Polit Sci*. 2014;58(4):1064–1082.

[15] Box GEP, Jenkins GM, Reinsel GC, Ljung GM. *Time series analysis: Forecasting and control*. 5th ed. Hoboken (NJ): John Wiley & Sons; 2015.

[16] Hyndman RJ, Athanasopoulos G. *Forecasting: Principles and practice*. 3rd ed. Melbourne: OTexts; 2021.

[17] Engle RF. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*. 1982;50(4):987–1007.

[18] Bollerslev T. Generalized autoregressive conditional heteroskedasticity. *J Econom*. 1986;31(3):307–327.

[19] Krippendorff K. *Content analysis: An introduction to its methodology*. 4th ed. Thousand Oaks (CA): SAGE; 2018.

[20] Röder M, Both A, Hinneburg A. Exploring the space of topic coherence measures. *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*; 2015. p. 399–408.

[21] Nelson DB. Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*. 1991;59(2):347–370.

[22] Tibshirani R. Regression shrinkage and selection via the Lasso. *J R Stat Soc Series B Stat Methodol*. 1996;58(1):267–288.

[23] Zou H, Hastie T. Regularization and variable selection via the elastic net. *J R Stat Soc Series B Stat Methodol*. 2005;67(2):301–320.

[24] Pagan A. Econometric issues in the analysis of regressions with generated regressors. *Int Econ Rev*. 1984;25(1):221–247.

[25] Hyndman RJ, Koehler AB. Another look at measures of forecast accuracy. *Int J Forecast*. 2006;22(4):679–688.

[26] Christoffersen PF. Evaluating interval forecasts. *Int Econ Rev*. 1998;39(4):841–862.

[27] Gneiting T, Raftery AE. Strictly proper scoring rules, prediction, and estimation. *J Am Stat Assoc*. 2007;102(477):359–378.

[28] Bracher J, Ray EL, Gneiting T, Reich NG. Evaluating epidemic forecasts in an interval format. *PLoS Comput Biol*. 2021;17(2):e1008618.

[29] Diebold FX, Mariano RS. Comparing predictive accuracy. *J Bus Econ Stat*. 1995;13(3):253–263.

[30] Harvey D, Leybourne S, Newbold P. Testing the equality of prediction mean squared errors. *Int J Forecast*. 1997;13(2):281–291.

[31] Clark TE, West KD. Approximately normal tests for equal predictive accuracy in nested models. *J Econom*. 2007;138(1):291–311.

[32] Giacomini R, White H. Tests of conditional predictive ability. *Econometrica*. 2006;74(6):1545–1578.

[33] Holm S. A simple sequentially rejective multiple test procedure. *Scand J Stat*. 1979;6(2):65–70.

APPENDIX A. MINIMUM REPORTING STANDARDS FOR THE EMPIRICAL STUDY

Table S1: Items to be reported in any empirical implementation.

Item	What should be reported
Corpus	Sources, licences, date range, languages, deduplication rules, document counts per period
Information availability	Publication-time rules, lags applied, and the cut-off for every forecast origin
Topic models	Selected K, coherence, diversity, seed stability, hyperparameters and embedding model
Expert validation	Number and background of experts, labelling protocol, agreement statistics
Forecast design	Target series, horizons, rolling-origin scheme, estimation windows and benchmarks
Evaluation	Point and density metrics, Clark–West and Diebold–Mariano results, and stratification by disruption period
Inference	Bootstrap scheme for generated regressors, multiple-testing adjustment, deviations from the analysis plan
Robustness	Placebo and permutation tests, topic-stability results, machine-learning comparison
Governance	Decision log, alert thresholds, rate of unwarranted interventions, ethical approvals

APPENDIX B. NOTATION

Table S2: Principal notation.

Symbol	Meaning
D, K, V	Number of documents, topics and vocabulary terms
θ_d, ϕ_k	Topic distribution of document d; word distribution of topic k
α, η	Dirichlet hyperparameters
$R_{k,t}$, $\tilde{R}_{k,t}$	Raw and standardised risk index of topic k in period t
$R^{*}_{c,t}$, $R^{\mathrm{int}}_{c,t}$	Directional and intensity indices of risk category c
S_k, D_k	Expert-validated severity and expected direction of topic k
Y_t	Target series (demand, production, dispatch or consumption)
$X_{m,t}$	Structured macroeconomic and operational controls
ϕ_i, ψ_j, β_m, γ_k	Autoregressive, moving-average, control and risk-index coefficients
σ_t^2	Conditional variance of the forecast error
a_i, b_j, δ_h	Shock, persistence and risk-signal coefficients in the variance equation
$Z_{h,t}$	Lagged risk variables entering the variance equation
d_t	Loss differential between baseline and risk-aware forecasts
τ, z_τ	Target service level and its standard normal quantile