Estimating Procurement Risk with Prediction Markets, LLMs, and Bayesian Causal Models

Johannes Koponen

Executive Summary

Procurement risk analysts face increasing pressure to quantify the likelihood of disruptions in global supply networks. Traditional methods such as supplier self-reporting, static scorecards, or historical data analysis are insufficient for anticipating sudden geopolitical, logistical, or regulatory shocks. This white paper presents a generalizable methodology that integrates:

- **Prediction markets** to anchor real-time probabilities of external disruption events.
- **Large Language Models (LLMs)** to infer supplier-specific vulnerabilities from unstructured information.
- **Bayesian Belief Networks with Monte Carlo simulation** to propagate risks through supplier networks and estimate Risk Network Value at Risk (RNVaR).

We illustrate the approach with an automotive OEM case, but the framework applies across industries. Compared to self-reporting or standalone statistical models, the proposed approach offers:

- Quantified, transparent probabilities instead of qualitative risk scores.
- Continuous updating from external signals (markets, news, LLM inference).
- Explicit modeling of causal interdependencies and tail risk.
- Decision support aligned with risk appetite via expected utility metrics.

By adopting this approach, procurement teams can move from reactive crisis management to proactive, data-driven risk governance.

Introduction

Procurement risk is defined as the probability that ordered goods are not delivered within a specified lead-time threshold (e.g., within 7 days of the required date) [1]. IIn practice, this means quantifying the chance of late or incomplete deliveries that disrupt production or service. For instance, a 15% probability of a >7-day delay in a Tier-1 supplier may translate to €25M in expected losses for an automotive OEM. While procurement risk focuses narrowly on supplier delivery performance, it is embedded within broader *supply chain risk*, which also covers systemic factors such as logistics bottlenecks and demand variability. Recognizing this distinction is important: procurement risk estimation zeroes in on supplier-level delivery failures as the immediate operational concern. This risk is a critical concern for procurement and supply chain professionals because late or unfilled orders can halt production lines, incur financial penalties, or damage customer relationships. Recent global events (e.g. the COVID-19 pandemic, geopolitical conflicts, port closures) have underscored how vulnerable supply networks are to disruptions[2][3], making robust risk estimation methods more important than ever.

Traditional approaches to procurement risk management often rely on supplier self-assessments, static risk scorecards, or historical performance data. However, these methods have limitations. Self-reported supplier risk disclosures are often *biased or incomplete*[4] and provide only a snapshot of risk at a single point in time. Moreover, conventional risk registers and matrices tend to treat risks in isolation (e.g. evaluating each supplier or risk factor independently) and use point estimates that ignore compounding uncertainties[5][6]. Such simplistic methods struggle to capture the complex interdependencies in modern global supply chains or to anticipate low-probability/high-impact events. As a result, procurement risk analysts are seeking more dynamic, data-driven approaches that incorporate predictive intelligence and probabilistic modeling to foresee and quantify supply disruptions before they occur.

Objective: This white paper proposes a generalizable methodology for estimating procurement risk by combining three cutting-edge components: (1) Prediction Markets to anchor probabilities of disruptive events, (2) Large Language Models (LLMs) to derive calibrated, supplier-specific probability estimates from unstructured forecasts and geopolitical news, and (3) Causal Models (Bayesian Belief Networks with Monte Carlo simulation and expected utility considerations) to simulate how risks propagate through a supplier network and to quantify the severity of outcomes. We illustrate the approach with an automotive original equipment manufacturer (OEM) use case – an industry where complex, multi-tier supplier networks and just-in-time practices make timely deliveries critical – but the framework is generalizable across industries. The following sections define procurement risk more formally, review recent research in probabilistic supply chain risk modeling, present the integrated methodology, and compare it to conventional approaches including supplier self-reporting. A bibliography of scientific references is provided to ground our arguments in current research (primarily from the last 5–10 years).

Recent Advances in Procurement Risk Modeling

Traditional and Qualitative Methods: Historically, procurement and supply chain risks have been assessed via qualitative frameworks (e.g. risk matrices, FMEA/FMECA analyses, or supplier audits). These approaches catalog potential risk events (supplier financial failure, quality issues, delivery delays, etc.) and assign scores for likelihood and impact. While straightforward, they often fail to capture the dynamic nature of risk. The risk landscape is constantly evolving and unpredictable, as evidenced by events like sudden factory shutdowns or logistics disruptions turning reliable suppliers into high-risk overnight[3]. Manual risk assessments are also resource-intensive – large firms report spending 5–6 weeks to assess a single supplier[7], yet still struggle to keep information up to date in real time. Surveys indicate that ~73% of organizations have difficulty obtaining timely, transparent data from suppliers, and 77% admit to lacking sufficient data visibility for effective risk monitoring[8][9]. These challenges have driven interest in more automated, data-driven risk analytics.

Bayesian and Probabilistic Models: In the past decade, researchers have increasingly applied Bayesian probabilistic models to supply chain risk. Bayesian Belief Networks (BBNs) are a popular tool because they can capture causal interdependencies among risk factors in a directed graph[10]. Each node in a BBN represents a random variable (e.g. a specific risk event or performance outcome), and directed edges represent causal influence (e.g. "port closure" → "supplier delay" → "OEM production delay"). Unlike static risk matrices, BBNs naturally incorporate conditional probabilities and can update risk estimates as new evidence emerges. For example, Jindal and Sharma (2022) model supply risks with a BBN to incorporate factors like supplier reliability, geopolitical risk, and transportation disruptions, allowing a dynamic assessment of how each factor influences the probability of supply failure. In general, Bayesian network models have proven effective at capturing interdependencies and network-wide effects[11]. Qazi et al.(2018) highlight that two key gaps in earlier SCRM frameworks were the lack of interdependency modeling and the omission of decision-makers' risk appetite[12][13]. Their work introduced a Supply Chain Risk Network Management process that integrates BBNs for interdependencies and Expected Utility Theory (EUT) to account for how a decision-maker's risk tolerance affects mitigation strategies[10][14]. By using BBNs plus expected utility calculations, they showed one can prioritize risk mitigation options not just by likelihood of risk but also by the utility (or disutility) of outcomes to the firm[15]. Another advance in probabilistic modeling is the use of Monte Carlo simulation in conjunction with BBNs to capture the full distribution of possible outcomes, rather than a single expected value. For instance, Qazi and Simsekler (2022) propose a simulation-based process to compute the Risk Network Value at Risk (RNVaR) for supply chain performance measures[16]. RNVaR is defined analogously to financial VaR – it is "the maximum risk exposure expected at a given confidence level for a given timeframe" across the network[17]. By generating thousands of random scenarios of risk factor realizations via Monte Carlo methods, one can estimate the probability that, say, the percent of on-time deliveries falls below a critical threshold (the tail risk). These researchers found that point-estimate based approaches often underestimate tail risks, whereas a Bayesian network + Monte Carlo approach reveals the true exposure to extreme delays or disruptions[6][18]. They introduced new metrics like Risk Network Exposure (RNE) and Risk Network Resilience/Criticality to quantify each risk's contribution to the overall network VaR[19][20]. In sum, recent Bayesian approaches enable a richer, probabilistic understanding of procurement risk propagation (often called the "risk ripple effect" in supply networks) and support decision-making under uncertainty.

Machine Learning and AI Approaches: Alongside Bayesian methods, there is growing interest in applying machine learning to predict supply chain risks. Early efforts include classical predictive analytics using supplier historical data (delivery lead times, quality incidents, etc.) to predict late deliveries or failures. More recently, researchers have tried training machine learning models on broader datasets – for example, Rezki and Mansouri (2023) use artificial neural networks to forecast supply chain disruptions from patterns in operational metrics, and Nguyen et al. (2023) evaluate tree-based algorithms for predicting risk levels of suppliers based on financial and logistics indicators. These ML models can capture non-linear patterns and interactions in large data, but a common critique is that they function as "black boxes" and may not explicitly model causal relationships. This limits interpretability for risk analysts who need to explain why a supplier is high-risk and how to mitigate that risk. There is a move toward causal machine learning in supply chain risk (e.g. causal forests or structural models) to combine the strengths of ML with causal inference[21]. Another burgeoning area is the use of natural language processing (NLP) and generative AI to tap into unstructured data sources (news, social media, supplier reports) for risk signals. Large Language Models have recently demonstrated impressive ability to analyze text and extract forward-looking insights. For example, Fan et al. (2025) applied a generative LLM (a GPT-based model) to firms' earnings call transcripts and constructed a firm-level supply chain risk indicator[22]. Their LLM-derived risk measure had intuitive variation (spiking when known disruptions occurred) and importantly correlated with real risk outcomes – firms with higher LLM-estimated risk had higher capital costs, stock return volatility, and larger inventory buffers[23]. This suggests LLMs can read between the lines of managerial commentary to gauge supply continuity risk in ways traditional metrics miss. Likewise, Sun et al. (2024) explored using LLM-driven news analysis for supply chain risk prediction. They note that LLMs can process massive volumes of news and "discern the relevance and implications of news events in relation to specific supply chain contexts"[24]. For instance, an advanced LLM could infer that political unrest in Country X might threaten a raw material supplier's operations, even if an article doesn't explicitly mention that supplier[24]. In a 2024 workshop, Shahsavari et al. demonstrated an agent-based model where LLMs monitor live news feeds and flag "contributing events" that feed into a Bayesian network for proactive risk identification[25]. This hybrid AI approach was able to provide early warnings by linking seemingly disparate events (e.g. labor strikes, regulatory changes) to primary risk events like supply disruption[26][27]. Overall, the literature trend is toward integrating external data sources and AI/NLP to enhance risk situational awareness, while using probabilistic causal models to quantify how those signals impact the ultimate risk of not receiving goods on time.

Summary: Recent research underscores that no single method is a silver bullet; rather, combining techniques can yield a more comprehensive risk picture. Bayesian networks offer principled handling of uncertainties and dependencies, ML/LLM techniques bring in predictive signals from data and text, and crowd-based prediction markets (discussed next) add an external wisdom-of-crowds perspective. This sets the stage for our proposed methodology that fuses these elements.

Integrated Methodology for Probabilistic Procurement Risk Estimation

Our proposed methodology combines prediction markets, LLM-based probabilistic inference, and causal Bayesian networks into a structured approach for estimating procurement risk. Figure 1 provides a conceptual overview of how global events, supplier-specific factors, and network effects are linked in this approach. Below, we outline the methodology in three key steps, followed by the simulation and risk analysis outputs.

Step 1: Anchoring Event Probabilities with Prediction Markets

Prediction markets are exchange platforms where participants trade contracts tied to the outcome of future events – effectively crowdsourcing a probability forecast for those events. Decades of research have shown that prediction markets often produce highly accurate forecasts, frequently outperforming traditional methods like expert committees or surveys[28]. In fact, aggregated market prices (which can be interpreted as odds) have outshone professional forecasts in domains ranging from elections to economic indicators[28]. The power of prediction markets lies in incentivizing information sharing: traders "put their money where their forecast is," integrating diverse information and expert knowledge into a single probability. Some forward-looking organizations have even used internal prediction markets for project timelines and supply chain events[29]. For procurement risk, we leverage prediction markets to obtain baseline probabilities for disruptive events that could impact supply deliveries. These include macro events like geopolitical conflicts, natural disasters, or policy changes, as well as industry-specific events (e.g. "Will a major supplier in sector Y go bankrupt this quarter?"). For example, an automotive OEM might consult prediction market data on events such as "Chance of port closures in China this winter" or "Probability of a tariff on aluminum imports by year-end." If an open market is not directly available for a particular risk, analogous futures or crowd forecasts can be used (e.g. commodities futures implying probability of supply shortage, or crowdsourced forecasts from platforms like Good Judgment or Metaculus). The key is to obtain an unbiased, up-to-date probability for each salient external risk event. As an illustration, as of mid-2025 some regulated prediction exchanges are considering contracts on supply chain risks – "contracts on climate risk, supply chain disruption, and AI regulation could all become real trading vehicles", according to a recent report[30]. By anchoring our model inputs to market-driven probabilities, we ensure the risk estimates start from the collective best guess about future disruption events, grounded in real-time information. This mitigates the danger of relying on subjective or outdated odds.

Step 2: LLM-Based Probabilistic Inference for Supplier-Specific Risks

Having obtained baseline event probabilities, the next step is to translate those into supplier-specific risk estimates – essentially, how each event (or combination of events) impacts the probability that a given supplier fails to deliver on time. This is where Large Language Models come into play. LLMs like GPT-4 can ingest vast amounts of unstructured data (news, reports, analyst opinions) and produce calibrated guesstimates or conditional probability judgments by synthesizing that information. We use LLMs in two primary ways:

- Contextual Risk Assessment: An LLM can analyze textual information about a supplier's context and environment – for example, news about the supplier's country (political stability, labor unrest), financial reports on the supplier's health, or industry reports – and *infer the supplier's vulnerability* to certain disruptions. For instance, if Prediction Market A says "Conflict in Region X has a 30% chance in next 6 months," an LLM can evaluate how a given Tier-2 supplier in Region X might be affected. It might reason: *"Supplier ABC is located in Region X's capital; past conflicts in this region led to ~2-week port closures. If conflict erupts, there's a high likelihood (say 80%) that ABC's shipments will be delayed."* In doing so, the LLM essentially produces a **conditional probability** P(Supplier delay | event occurs). Combined with the event's prior probability, we can derive the unconditional risk contribution (0.3 * 0.8 = 24% in this example). The LLM's ability to connect subtle dots is valuable – as noted in one study, *LLMs can identify nuanced connections and interpret how political or logistical developments might impact specific suppliers, even if not explicitly stated*[24].

- **Bayesian Guesstimate Calibration:** Often we have only sparse or qualitative data on a supplier's risk. Here, an LLM can act as an *intelligent risk estimator* by drawing on analogous cases and expert knowledge embedded in its training data. We prompt the LLM with scenarios (e.g. *"Supplier XYZ is in a seismically active zone and has no secondary manufacturing site. Based on historical earthquake disruptions, what is a reasonable probability that XYZ cannot deliver within 2 weeks if a major earthquake (>7.0) hits their region?"*). The LLM might output a probability distribution or a range (perhaps "around 50% chance of significant delay given a major quake"). While this is a *guesstimate*, it is informed by the model's broad knowledge (it has effectively read thousands of articles about similar events). We treat this output as a **Bayesian prior** – a starting probability for that supplier's risk which can be later updated with any hard data. Importantly, we can calibrate the LLM's output by cross-checking with known reference points. For example, if industry data says historically 40% of suppliers without backups faced delays after earthquakes, we adjust the LLM's 50% accordingly. The LLM can also be used to generate *rich scenarios* ("if conflict happens during monsoon season, road closures could further delay shipping, raising risk to X%") that feed into our causal model structure.

By deploying LLMs in this manner, we effectively inject real-world knowledge and expert reasoning into our risk model on a supplier-by-supplier basis. The result of Step 2 is a set of probabilistic inputs tailored to each supplier: e.g. P(Supplier A late | Event 1), P(Supplier A late | no major event), P(Supplier B late | Event 1 ∧ Event 2), and so forth. These probabilities can be represented as conditional probability tables or parametric distributions (for continuous outcomes like length of delay) attached to the nodes of a Bayesian network (described next). It's worth noting that interpretability remains high – we can trace a high risk estimate back to specific narrative rationales provided by the LLM (e.g. "Supplier B is high-risk because an LLM reading news noted a recent strike at B's plant and financial troubles at their logistics provider"). This addresses the black-box concern of ML, giving analysts transparency into why a supplier is deemed risky.

Step 3: Causal Bayesian Network Simulation and Risk Propagation

At this stage, we construct a causal Bayesian Belief Network model of the supplier network, incorporating the probabilistic inputs from Steps 1 and 2. The Bayesian network serves as a unifying framework that links upstream events to downstream outcomes in a directed acyclic graph. Figure 1 illustrates a simplified example of such a network for an automotive OEM with two key suppliers and two potential disruptive events:

Figure 1: Simplified causal risk network for an OEM's supply chain. Global events (top nodes, e.g. a geopolitical conflict or a port strike) increase the risk of delays at specific suppliers (middle nodes). These supplier delay risks, in turn, propagate to the final outcome node – the probability that the OEM fails to

receive goods on time. The Bayesian network encodes these cause-effect relationships and enables probabilistic inference of the overall procurement risk given various scenarios.

In a real implementation, the network would include multiple layers and nodes, such as: nodes for macro risk events (economic downturns, natural disasters, regulatory changes), nodes for intermediate factors (e.g. "transportation capacity reduced" or "supplier production halt"), nodes for each critical supplier's status ("Supplier X on-time delivery" vs "delay"), and nodes for the OEM's key performance outcomes (e.g. "All critical parts received within 1 week of need date" – a success/failure node). Causal dependencies are defined by experts and data. For example, "Port Strike" → "Supplier B Delay" might be a link with a conditional probability that if a strike occurs, Supplier B has an 60% chance of delay (as informed by LLM/analyst input). Some suppliers might have multiple risk parent nodes (e.g. Supplier A could be impacted by both a raw material shortage event and a political event). The structure can be enriched with tiers – Tier-2 supplier risks feeding into Tier-1 supplier performance, etc., reflecting the multi-tier nature of modern supply chains[31].

Once the Bayesian network structure and conditional probabilities are in place, we use it to simulate risk propagation and outcomes. Two analytical techniques are particularly useful here:

- Inference and Scenario Analysis: We can input evidence or scenarios into the BBN and compute the posterior probabilities of outcomes. For instance, *conditional on a specific event happening* (say a major earthquake occurred), the model can update the probability that each supplier will delay and ultimately the probability the OEM misses its delivery target. This allows *stress-testing* of the supply chain: testing how resilient the network is to each event or combination of events (similar to scenario analysis in finance). It helps identify single points of failure (e.g. one supplier whose delay probability drives most of the risk) and highly correlated risk exposures.
- Monte Carlo Simulation for Distribution of Outcomes: To capture the full uncertainty, we perform Monte Carlo simulation by randomly sampling from the probability distributions in the BBN. Each simulation run might represent a possible "year" or "project" where certain events occur and certain suppliers fail or not. By thousands of such runs, we build up a probability distribution for outcomes like "percent of time the OEM experiences a significant delay" or "days of delay in worst-case 5% scenarios." This is where we compute risk metrics such as RNVaR (Risk Network Value at Risk) and expected impact. For example, using the methodology of Qazi *et al.* (2022), we determine the *RNVaR at 95% confidence* for the OEM's on-time delivery performance[16]. Suppose the result is that with 95% confidence, the OEM will lose at most 10 production days in the next year due to supplier delays – that 10-day loss is the RNVaR (the "value at risk" in terms of days lost). If we consider a *utility function* (say the utility = negative of cost or lost profit), we can also translate this into an expected utility for different risk mitigation decisions. For instance, if we have options to mitigate risk (like dual-sourcing a part or increasing safety stock), we can incorporate those as interventions in the network and evaluate the *expected utility gain* or *risk reduction* from each. This approach aligns with the expected utility framework introduced by Qazi *et al.* (2018) to integrate risk appetite into supply chain risk decisions[32][14]. A risk-averse decision-maker might prioritize actions that shave off the tail risk (reducing RNVaR), whereas a risk-neutral one might focus on improving the expected on-time rate – the model can accommodate either by applying the appropriate utility function to outcomes[33].

The outcome of Step 3 is a quantitative risk profile for the procurement in question. Instead of a single risk score, we obtain: (a) the probability that the procurement will not arrive within the desired timeframe (our definition of procurement risk), and (b) a distribution of possible delay severities or performance impacts, from which metrics like RNVaR, expected loss, and percentile outcomes are derived. We also gain insight into the risk propagation paths – for example, the simulation might reveal that "Supplier B's delay (often due to Port Strike) is the largest contributor to the 95th percentile worst-case outcome"[18]. This informs risk mitigation: the OEM could invest in contingency plans specifically for Supplier B or ports at risk. Additionally, by toggling different input scenarios (with or without certain events, or improved supplier resiliency), analysts can conduct sensitivity analysis to see how the overall procurement risk responds, guiding strategic decisions like diversification of suppliers or inventory buffering.

Justification and Scientific Basis

The rationale for this combined approach is supported by several strands of research. First, prediction markets provide well-calibrated probabilities for uncertain events, leveraging the "wisdom of crowds." Studies have repeatedly found that market-based forecasts outperform individual experts and even sophisticated models in many cases[28]. By using prediction market odds as inputs, we ground our risk assessment in the most current collective information about future events. This is especially valuable in today's volatile environment where geopolitical or environmental risks can change rapidly – markets update in real time, whereas periodic risk reports may lag. The rise of platforms like Kalshi and Polymarket indicates that ever more granular events (including supply chain disruptions) are becoming forecastable via markets[30]. As these mature, procurement teams can plug into these signals rather than relying solely on internal guesswork.

Second, the use of LLMs for probabilistic inference is justified by the dramatic progress in NLP and knowledge representation. Modern LLMs (GPT-4 and beyond) encapsulate a vast corpus of world knowledge including technical reports, news, and historical cases. When carefully prompted, they can output quantitative judgments that align with known reference data. Early academic evidence (e.g. Fan et al., 2025) shows that LLM-generated risk proxies have predictive validity in the real world[23]. Moreover, LLMs enable processing of unstructured data (continuous news feeds, social media, etc.) which traditional models typically ignore. This capability to harness weak signals and expert insights from text is a game-changer for supply risk assessment. Shahsavari et al. (2024) demonstrated that coupling LLM-based event detection with a Bayesian model improved early warning for supply chain disruptions[26]. The LLM essentially acts as an always-on analyst, scanning myriad sources to flag relevant risk events and estimate their implications, far beyond the capacity of human analysts alone[34]. By integrating LLM "knowledge" into a Bayesian framework, we retain a coherent probabilistic model while benefiting from the LLM's breadth of information.

Third, the Bayesian network and Monte Carlo simulation component is grounded in a rich body of operations research on supply chain risk. Bayesian networks have been advocated by many researchers for modeling risk interdependencies and cascading effects[11]. They force clarity in modeling assumptions (via causal structure) and can incorporate both data and expert input. The use of Monte Carlo simulation on top of BBNs, as in Qazi et al. (2021, 2022), is shown to unveil the tail-risk that deterministic or single-point methods miss[6]. Our approach explicitly computes RNVaR – ensuring that the severity of worst-case scenarios is quantified, not just the most likely scenario. This is crucial because procurement officers must plan not only for expected delays but also for rare but devastating disruptions (the COVID-19 lockdown of 2020 being a prime example that caught many off-guard). By simulating thousands of scenarios, we capture the distribution's tails and can answer questions like "what is the 99th percentile delay?" or "what probability do we have of more than 1 month delay?". The inclusion of expected utility considerations further aligns the model with decision-making – instead of just probabilities, we evaluate outcomes in terms of business impact (e.g. dollars of loss or utility). This allows integration of risk appetite: a risk-averse firm may effectively assign higher weight to tail outcomes, which our model can reflect by using a concave utility function[35].

Finally, it's important to note the generalizability of the approach. While our use case example is an automotive OEM, the methodology applies broadly. Any industry with a supplier network – electronics, pharmaceuticals, retail, aerospace, etc. – can utilize this framework by plugging in the relevant risk events and supplier nodes. The prediction market inputs would be chosen accordingly (e.g. a pharma company might track probability of a regulatory ban or a pandemic resurgence), and the LLM could be tuned to sector-specific news (say, mining reports for a tech company reliant on rare minerals). The Bayesian network topology would reflect that industry's supply chain structure and key risk propagation paths. This adaptability is a strength: as long as one can identify the major risk factors and has access to forecast information (market or expert) and data or knowledge about supplier vulnerabilities, the approach can be applied. Moreover, it is scalable – new risk factors or suppliers can be added as nodes without having to redesign the entire model, which is conducive to evolving risk landscapes.

Comparison to Conventional Approaches

Compared to other procurement risk assessment approaches, our integrated methodology offers distinct advantages:

- Versus Supplier Self-Reporting: Many organizations rely on suppliers to self-report risk via questionnaires or compliance forms (covering financial stability, disaster preparedness, etc.). While useful as a baseline, these self-assessments suffer from *staleness and bias* – suppliers have incentive to downplay risks and updates may occur only annually. In contrast, our approach is *externally driven and continuous*. It doesn't depend on a supplier's willingness to share information (which 73% of firms struggle with[8]). Instead, it picks up signals from market pricing and global news, which are harder to game. It also automates continuous monitoring: if a relevant event probability spikes (say risk of a logistics strike jumps from 5% to 30% in the market), our model immediately reflects that in higher procurement risk, whereas a questionnaire-based approach might not capture it until after the disruption has hit. Self-reports also lack network perspective – a supplier might report "we have backup facilities" giving a false sense of security, but a Bayesian network might reveal that their backup is in the same region (hence both primary and backup fail under the same event).
- Versus Standalone Statistical Models: Pure machine learning models that predict late deliveries from historical data might achieve decent accuracy in steady-state conditions, but they often fail under novel conditions (e.g. a pandemic with no precedent in the training data). Our approach is *forward-looking*, incorporating predictive market and LLM-generated foresight about events that have not occurred before. It is fundamentally scenario-based and causal, not just correlational. This means it can answer *"what if" questions* (what if an earthquake hits?) that a standard regression or ML model, trained on past data without such an event, cannot. Furthermore, by including expert knowledge and causality, our model is less likely to be misled by spurious correlations. For example, a black-box model might learn that "orders from supplier in country X are always late" (because historically that supplier had issues), but if that supplier has since improved and the main risk now is a possible flood season, the ML model might miss that shift.

Our approach would explicitly include "flood risk" as a factor and could capture the new reality even without prior examples of the improved performance.

- Versus Traditional Risk Matrices/Scoring: A common practice is to score each supplier on a 1-5 scale for likelihood of delay and impact, often averaging various sub-risk factors. This simplistic scoring loses nuance and tends to treat suppliers independently. Our Bayesian network approach, however, captures interdependencies: e.g. if two suppliers rely on the same sub-supplier or the same port, the model reflects that a port strike event will simultaneously affect both, possibly compounding the OEM's risk. Traditional matrices would treat them separately and might underestimate the combined effect (or fail to see that they are not independent risks). Additionally, by quantifying in probability terms (e.g. "20% chance of >1 week delay") and in units like days or dollars, our approach provides more actionable information than a generic "high risk" label on a heatmap.
- Transparency and Explainability: Compared to a purely AI/ML approach, the proposed method is highly explainable. Each component provides interpretable inputs – prediction market prices can be explained as consensus odds, LLM rationales can be summarized in natural language (the model could even output a narrative: "Supplier A is risky due to X, Y, Z factors"), and the Bayesian network structure is a visual map of cause and effect that stakeholders can understand. This is crucial when communicating with executives or making the case for mitigation investments. Rather than saying "our neural network predicts a 0.7 risk score", the analyst can say "there is a 25% probability we won't get these parts in time, primarily because of possible strikes in Supplier B's country (which have a 30% chance this year[30]). If that strike occurs, there's a 70% chance Supplier B is delayed, which cascades to a production stop at our plant." This kind of explanation builds confidence and urgency in a way black-box outputs cannot.
- Timeliness and Proactivity: Perhaps the biggest edge is the ability to be proactive. By monitoring prediction markets and news via LLM, the approach functions like an early-warning system. It can be updated continuously or on-demand. If a new risk emerges (e.g. a sudden coup in a country housing a key supplier), the framework can quickly ingest that (market odds of political instability, LLM reading news of unrest) and update the procurement risk. Traditional approaches might catch up only after delays start happening. As one industry article notes, continuous monitoring reduces reliance on periodic human-intensive assessments and *"minimizes the risk of costly errors"*[36]. Our approach embodies that principle.

Of course, integrating these components is not without challenges. It requires interdisciplinary expertise (market analysis, AI, Bayesian modeling) and quality data feeds. There are also potential limitations: prediction markets may not exist for all relevant events (we may need to rely on proxy markets or expert elicitation in some cases), LLM outputs need careful validation and can occasionally be inconsistent or require calibration, and Bayesian network models can grow complex if not constructed prudently (too many nodes can make parameter estimation difficult). However, these challenges are manageable. For missing markets, one can run internal prediction tournaments or use structured expert judgment to fill gaps. LLM calibration can be improved by fine-tuning on domain-specific text or by using ensemble of prompts. Complexity in BBNs can be tamed by focusing on the most critical risk factors (Pareto principle: a subset of risks likely drive the majority of procurement risk). Notably, even if certain probabilities are rough, the Monte Carlo simulation will reveal how sensitive outcomes are to them; if a particular probability is highly influential, that's where to invest in better data or expert estimates.

In summary, when compared to other approaches, the proposed methodology offers a richer, more responsive, and more holistic view of procurement risk. It brings together complementary strengths – the foresight of crowd intelligence, the analytical power of AI, and the rigor of probabilistic modeling – to support procurement risk analysts in making informed decisions to safeguard supply continuity.

Conclusion

Procurement risk – the likelihood of not receiving vital goods on time – is a complex function of many uncertain factors. In an era of globalized and just-in-time supply chains, traditional siloed risk assessment methods are proving inadequate for predicting and mitigating disruptions. This technical white paper introduced a novel, generalizable approach that fuses prediction market forecasts, LLM-based inference, and Bayesian causal networks to estimate procurement risk in a robust, quantifiable manner. By doing so, we harness collective intelligence, AI's text-processing prowess, and principled modeling of cause-effect relationships all in one framework.

The approach offers several key benefits: anticipation of emerging risks (through live market and news insights), quantification of uncertainty (through probabilistic simulation and RNVaR metrics), and decision support (via expected utility analysis and clear identification of critical risk drivers in the supplier network). The illustrative automotive OEM case shows how an organization could implement this: obtaining market odds for geopolitical or logistical events, using LLMs to translate those into supplier delay probabilities, and running a Bayesian network model to compute the probability of missing production targets along with the worst-case loss distribution. Armed with these results, procurement and supply chain managers can prioritize risk mitigation efforts (e.g. find alternate suppliers for the most risk-exposed part, increase inventory for components likely to be delayed, or hedge critical materials) and justify these decisions with data-driven forecasts. In essence, it shifts risk management from a reactive posture ("firefighting" after a supplier fails) to a proactive, predictive discipline.

Beyond the specific use case, the methodology can be generalized to any procurement context by customizing the risk factors and data inputs. It is flexible enough to evolve – as new risk information becomes available (say a new prediction market on shipping delays, or improved LLM models), those can be incorporated to refine the risk estimates. Over time, one can envision an automated system continuously ingesting risk signals and updating a dashboard of procurement risk metrics (much like a stock ticker for supply chain risk). This would enable near-real-time risk governance.

In closing, we emphasize that the multi-faceted nature of supply risk calls for a multi-disciplinary solution. The combination of market predictions, AI reasoning, and causal simulation presented here is at the forefront of procurement risk analytics. It aligns with the direction of recent academic and industry developments, as evidenced by the literature reviewed, and pushes the envelope by integrating these elements into a cohesive process. Organizations that adopt such advanced approaches will be better positioned to navigate uncertainties, ensure supply continuity, and gain competitive advantage by avoiding disruptions (or even capitalizing on them, for instance by securing supply when competitors cannot). As supply chain volatility persists in the global economy, a probabilistic, data-driven risk estimation approach will be an indispensable tool in the procurement risk analyst's toolkit.

Bibliography

1. Wolfers, J. & Zitzewitz, E. (2004). *Prediction Markets*. Journal of Economic Perspectives, 18(2), 107–126. (Demonstrates the accuracy of prediction markets in forecasting and their ability to aggregate information, often outperforming traditional prediction tools[28].)

2. Karimi, M. & Zaerpour, N. (2022). *Put your money where your forecast is: Supply chain collaborative forecasting with cost-function-based prediction markets*. European Journal of Operational Research, 300(3), 1035–1049. (Introduces a mechanism for supply chain partners to share forecasts via a prediction market, showing improved forecast accuracy and Pareto-improved profits in a supplier–retailer game[37][38].)
3. Masrani, A. (2025). *Kalshi Hits $2 Billion as Prediction Markets Go Mainstream*. TipRanks News, Aug 2025. (News article on the growth of Kalshi prediction market; notes the expansion of event contracts to domains like climate and supply chain disruptions[30] and validates prediction markets' forecasting prowess in various events[39].)
4. Shahsavari, M., Hussain, O. K., Saberi, M., & Sharma, P. (2024). *Empowering Supply Chains Resilience: LLMs-Powered BN for Proactive Supply Chain Risk Identification*. In Proc. of D2R2'24 Workshop (ESWC 2024). (Proposes an agent-based model using Large Language Models to monitor news and a Bayesian Network to identify and quantify supply chain risks; highlights how LLMs + BN can provide early warning of disruptions by linking contributing events to primary risks[25][27].)
5. Sun, J. (2024). *Application of News Analysis Based on Large Language Models in Supply Chain Risk Prediction*. Journal of Computer Technology and Applied Mathematics, 1(3), 56–68. (Investigates using LLMs to extract risk signals from news; finds that LLMs can analyze extensive news data to uncover subtle risk indicators and context-specific implications for supply chains[34][24].)
6. Fan, S., Wu, Y., & Yang, R. (2025). *Measuring firm-level supply chain risk using a generative large language model*. Finance Research Letters, 77, 107111. (Develops an LLM-based metric for supply chain risk from textual data (site visit transcripts); the LLM risk measure correlates with higher financing costs, stock volatility, and inventory levels, demonstrating the model's effectiveness in capturing firm-specific supply risk[22].)
7. Qazi, A., Dickson, A., Quigley, J., & Gaudenzi, B. (2018). *Supply chain risk network management: A Bayesian belief network and expected utility based approach for managing supply chain risks*. International Journal of Production Economics, 196, 24–42. (Presents a risk management process integrating Bayesian belief networks to model risk interdependencies and expected utility theory to account for decision-maker risk appetite; validated via case study[10][14], it introduces concepts like conditional expected utility matrices for risk propagation[40].)
8. Qazi, A., Simsekler, M.C.E., et al. (2022). *Supply chain risk network value at risk assessment using Bayesian belief networks and Monte Carlo simulation*. (Publication in International Journal of Production Research, forthcoming 2022). (Addresses the gap of assessing tail-risk (VaR) in supply networks; proposes a BBN+Monte Carlo method to compute Risk Network VaR at a given confidence for performance measures[16]. Demonstrated in a telecom supply chain case, introducing metrics like RNVaR, Risk Network Exposure, and comparing against point-estimate approaches[5][20].)
9. Qazi, A., *et al.* (2021). *Worst-Expected-Best method for probabilistic network VaR: Application in supply chain risk management*. International Journal of Quality & Reliability Management, 38(5), 1309–1336. (Proposes the "Worst-Expected-Best" method grounded in BBNs to prioritize risks by network-wide Value-at-Risk. Shows that point-estimate aggregation misjudges risk exposure, and the new method captures expected loss at confidence levels and incorporates decision-maker risk appetite into risk prioritization[41][18].)

10. Jindal, A. & Sharma, S. K. (2022). *Bayesian Belief Network Approach for Supply Risk Modelling*. International Journal of Information Systems and Supply Chain Management, 15(1), 1–22. (Uses a BBN to model supplier risk factors and their probabilistic relationships; highlights the ability of Bayesian models to update risk predictions with new information and improve over traditional supplier evaluation methods.)
11. Veridion (2023). *The Full Guide to Supplier Risk Assessment*. Veridion Blog, July 2023. (Industry guide discussing challenges in supplier risk assessment; notes dynamic nature of risks and difficulties with current practices. Emphasizes that supplier self-reported data can be unreliable[4] and advocates for continuous monitoring and third-party data to fill gaps.)
12. Ramp (2025). *7 Common Procurement Risks and How to Mitigate Them*. Ramp Procurement Blog, May 2025. (Defines procurement risk as the possibility of disruptions to obtaining goods/services[1]and overviews common risk types. Provides context for why proactive risk management is vital in light of recent global supply chain challenges[42].)